ClinGen Guidance and Recommendations for Monogenic Disease Nomenclature

1. Background

Due to the long history of discovery, characterization, and naming of different monogenic diseases, our understanding of these disease entities, their phenotypic features, and the names by which they are known may change over time. Historical practices in defining disease entities and the corresponding naming conventions used at different points in time or within different medical specialties have contributed to heterogeneity in disease names. At times, there have been consensus efforts to rename an entire group of monogenic diseases. This evolution of the naming of disease entities has resulted in a set of highly disparate and not always informative disease names being used in current practice.

The Clinical Genome Resource (ClinGen) views monogenic diseases as unitary and distinct entities defined by the causal relationship between a genotype and a resulting phenotype (in the individual) or spectrum of phenotypes observed (in a population with disease-causing genotypes). While each individual with a given monogenic disease has their own unique phenotype, the overall penetrance and expressivity of the disease entity can be described as the collective phenotypic spectrum observed in a population of individuals with disease-causing genotypes (Figure 1).

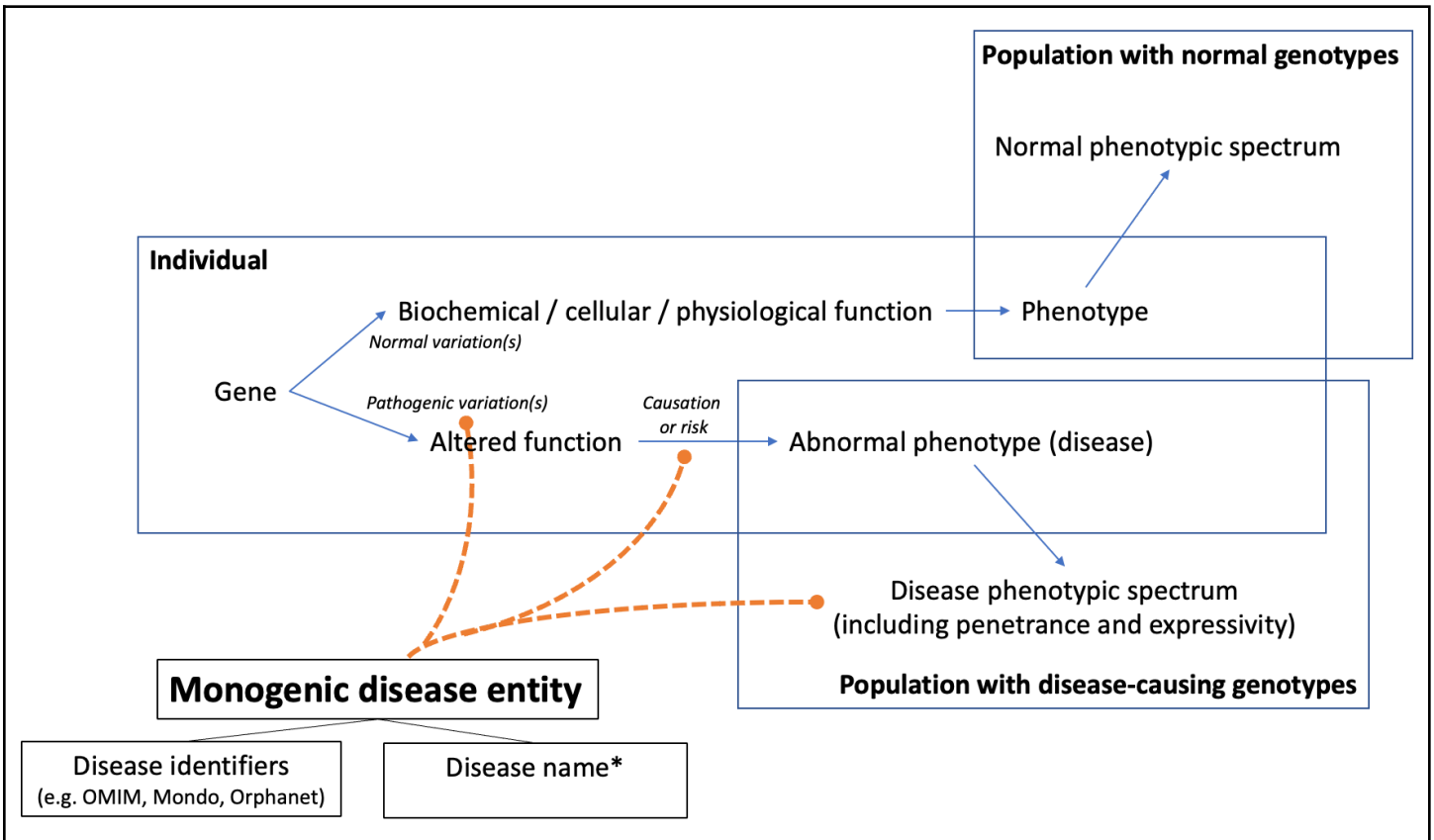

Figure 1 : Defining a monogenic disease entity versus normal genetic variation. At an individual level, genes serve specific functions that contribute to organismal phenotypes. In the general population there are genetic variations that result in

observable phenotypes or phenotypic spectrum that are considered “normal variation.” However, some genetic variations lead to a severely altered function that may result in an “abnormal” phenotype, termed disease. Within the collective population of individuals with disease-causing genotypes, a phenotypic spectrum can be described based on penetrance and expressivity. The combination of a single gene, the pathogenic variation(s) within that gene, and the causation or risk to develop a detectable phenotype are what are termed a “monogenic disease entity. ” These defined entities can be referred to either by an unambiguous numeric disease identifier that is easily computable or a name that is recognized by humans and used colloquially to refer to the disease. The disease ID should remain stable over time unless changes in evidence or understanding of genotype/phenotype relationships require “lumping” or “splitting” to redefine the disease entity. Similarly, the disease name may evolve and change (*) with knowledge gained over time. There may also be multiple different names that are synonymous for the unitary disease entity, and therefore using the unambiguous disease ID can help to recognize these different names that refer to the same disease entity. In a dyadic naming system, the disease name should clearly indicate both the gene and a phenotypic label that communicates information about the disease.

There is a distinction between the “entity” itself and the labels that are used to identify it. The “name” of any disease entity should be readily recognizable by humans so that it can be used in communication. Because of the causal relationship between altered function of a gene and the resulting abnormal phenotypes (or risk to develop those phenotypes), **we endorse the use of a dyadic strategy that explicitly recognizes the responsible gene in addition to a descriptive phenotypic label when naming curated disease entities**. Furthermore, since the preferred phenotype descriptor may change over time, it is also critical for disease entities to have a stable unique identifier (e.g., Mondo Disease Ontology [Mondo] term) and a list of synonyms and their provenance for that disease entity. We also recognize that in the absence of a genetic etiology, or when referring to diseases with locus heterogeneity, that a more general disease term may be needed to collectively describe or define a patient's clinical diagnosis prior to obtaining a causal etiology.

In the process of defining the monogenic disease entity to be curated, it is frequent for a ClinGen expert panel or working group to conclude that existing assertions may need to be altered (lumped together, or split apart). Specific guidance has been developed by ClinGen to help expert panels determine the “curated disease entity” for a particular gene (https://clinicalgenome.org/working-groups/lumping-and-splitting/ and PMID:35754516). As a result of this process, ClinGen expert panels, working groups, and/or other disease experts may be faced with having to create and/or reorganize disease entities that subsequently require new identifiers and associated names, in order to proceed with curation. This document is intended to outline current recommendations and guidance for how to approach nomenclature updates to ensure decisions are made consistently across the clinical genomics field and should be harmonized with other existing nomenclature systems (e.g., OMIM, Mondo, etc).

To facilitate consistent disease naming, ClinGen formed the Disease Naming Advisory Committee (DNAC) which included representatives from other nomenclature systems and sought extensive feedback from many disease domains. **The guidance provided herein is required by ClinGen expert panels and working groups, however, these same recommendations may be useful to other groups that face the need to reconsider the name of an existing disease or new disease entities being described.** The major recommendation from the work is that when disease naming changes are necessary, ClinGen expert panels and/or working groups must specify a dyadic name.

2. Dyadic Naming Convention(s)

The dyadic naming convention means that the name for the curated monogenic disease entity should include labels representing both the gene and the disease phenotype. There are generally two approaches for delineating the gene that is involved in the monogenic disease entity. However, the phenotype label can be much more variable in terms of its organization and semantic content. Both concepts are described in further detail below.

2.1 Genetic label

Two common conventions for representing the gene involved in a monogenic disease are described here. The explicit dyadic relationship includes the current HGNC designated/approved gene symbol in the disease name, while an alphanumeric convention utilizes numbers and/or letters that correspond to the unique locus for a given monogenic disease. Expert panels and working groups are encouraged to consider designating both forms of the dyadic disease name (i.e., explicit and Alphanumeric) when updating names with Mondo.

2.1.1 - Explicit dyadic relationship

- Description: In this naming convention, the gene name and phenotypic descriptor are both specified in the disease name. Several variations may exist, such as using HGNC gene symbol in the name with a connecting adjective ("*GENE*-related *phenotype*") or phrase indicating causality ("*phenotype* due to *GENE*-deficiency" or "*phenotype* related to *GENE*"). **This is the preferred approach for ClinGen expert panels and/or working groups that need to create designations for disease entities due to application of the lumping and splitting guidelines.**
- Examples: *PTEN* hamartoma tumor syndrome; *DIAPH1*-related sensorineural deafness-thrombocytopenia syndrome; *SLC6A3*-related dopamine transporter deficiency syndrome
- Reference: Biesecker et al. (2019) PMID:31692258

2.1.2 - Alphanumeric dyadic relationship

- Description: In this convention, the disease entity is named using an alphabetical or numerical system to represent a specific gene or locus. For groups of diseases characterized by phenotypic similarity and locus heterogeneity (termed a "phenotypic series"), OMIM uses the naming convention of a phenotypic descriptor followed by an Arabic numeral that represents the gene or locus to which a specific entity has been associated. Often the Arabic numeral is set by the order of discovery of the genes/loci, which enables the naming of disease entities discovered through linkage analysis prior to the causal gene being definitively identified. Note that although many OMIM disease names therefore utilize an implicit dyadic form, certain monogenic entities in which only one locus has been implicated in a given phenotype do not contain an alphanumeric designation. This may lead to confusion regarding the above-noted suggestion on specifying phenotypes or affected individuals for which the gene is unknown. Additionally, numbers and letters have sometimes been used to designate certain disease characteristics (e.g., to encode inheritance pattern, and locus). Alphanumeric systems have also been used to differentiate among phenotypically similar conditions or between subgroups of patients with different degrees of severity. In some cases, these historical subtypes have been determined to arise

from a single genetic locus, e.g., both early and late onset disorders are due to variants in the same gene, and in other cases from distinct genetic loci. This can lead to confusion when numerical suffixes that designate phenotypic variation conflict with the locus designation. **Given these practical limitations, and in order to avoid confusion with existing nomenclature, we strongly encourage ClinGen expert panels not to develop their own set of alphanumeric descriptors. When using an existing nomenclature (e.g., OMIM), ClinGen expert panels should also specify an explicit dyadic name using the convention described above.**

- Examples: Autosomal dominant deafness 64; Spinocerebellar ataxia 15; Progressive myoclonic epilepsy 11; Retinitis pigmentosa 57; nephronophthisis 2; neuropathy, hereditary sensory, type 1D
    - For some of the phenotypes with alphanumeric descriptions, OMIM provides a disease symbol. For example Autosomal dominant deafness 64 has a disease symbol of DFNA64 (https://www.omim.org/entry/614152)
- Reference: Rasmussen et al. (2020) PMID:32555417

2.2 Phenotypic label

There have been a large number of historical approaches to describing disease manifestations, but the primary goal of the phenotypic descriptor should be to convey semantic information that helps the individual readily recognize and understand the observable features at some level. In some cases the phenotypic descriptor is specific and includes cardinal features or may represent biochemical and/or cellular abnormalities, while in other cases a more general clinical term or eponym is applied in order to use existing naming conventions for the disease entity, where appropriate (See Appendix 1). There are strengths and weaknesses of each descriptive approach, and certain conventions may be preferred in a given field. Therefore, ClinGen does not prescribe a particular approach to the descriptive/phenotypic label of the dyadic disease name, however, expert panels should be thoughtful, consistent, and collaborative with Mondo, OMIM and the clinical domain community when approaching the phenotypic label. **It is anticipated, and even expected, that disease names (genetic and phenotypic labels) will change over time, and that expert panels will need to revisit names as more information becomes available with future scientific publications and knowledge on any given monogenic disorder.** Use of the ClinGen GeneTracker and Mondo, which maintains persistent disease identifiers, will help to maintain provenance and clarity of the curated disease entities as names and definitions evolve.

2.2.1 - Nonspecific terms

ClinGen discourages the use of nonspecific terms such as “disorder(s)” or “dysfunction” alone as a phenotypic descriptor for dyadic naming of monogenic disease entities. We recognize that GeneReviews and other references may refer to, and use titles in the form of “Gene-related Disorders” (e.g., *NSDHL- related disorders*). In general, this terminology is to indicate that the review will focus on multiple phenotypes, or disorders, associated with the gene of interest rather than to suggest that the name represents a monogenic disease entity.

2.2.2 - What about inheritance pattern?

It is not necessary to include the inheritance pattern in the phenotypic label of the disease name. Instead, inheritance pattern can be indicated as part of the disease definition along with other relevant information about disease mechanism, natural history, presenting features, penetrance and expressivity. ClinGen discourages use of inheritance pattern in the disease name unless it is absolutely necessary for distinguishing a disease that otherwise cannot be differentiated at the level of the phenotypic label, or is consistently used within a disease area (e.g., autosomal dominant polycystic kidney disease/ADPKD).

2.2.3 - Use of terms like 'familial' or 'hereditary'

As with inheritance patterns, ClinGen generally discourages using terms like "familial" or "hereditary" in a phenotypic label, since by definition monogenic diseases have implied hereditary implications. "Familial" is a particularly challenging term given the increasing recognition of rare dominant disorders that almost always result from *de novo* variants in absence of family history. An exception would be if the well-recognized name already includes this descriptor and is accepted by the community (e.g., Familial Mediterranean Fever) or if this term is important to distinguish the monogenic disease from non-genetic form of the disease (e.g., Familial Hypercholesterolemia).

Table 1: Examples of acceptable approaches to dyadic naming conventions for monogenic disorders across disease entities.

Explicit Dyadic Gene-Phenotype Naming Approach		**Implicit Dyadic AlphaNumeric Naming Approach (e.g., OMIM)**	**MIM ID**	**Mondo ID**
Format 1	**Format 2**			
RYR1-related malignant hyperthermia	Malignant hyperthermia related to *RYR1*	Susceptibility to malignant hyperthermia-1	145600	MONDO:0007783
MYBPC3- related hypertrophic cardiomyopathy	Hypertrophic cardiomyopathy related to *MYBPC3*	Hypertrophic cardiomyopathy 4 (CMH4)	115197	MONDO:0007268
ACTG1-related Baraitser-Winter syndrome	Baraitser-Winter syndrome related to *ACTG1*	Baraitser-Winter syndrome 2 (BRWS2)	614583	MONDO:0013812
ACTG1-related non-syndromic hearing loss	Non-syndromic hearing loss related to *ACTG1*	Autosomal dominant deafness 20/26 (DFNA 20/26)	604717	MONDO:0011480
SLC19A3-related biotin-responsive basal ganglia disease	Biotin-responsive basal ganglia disease related to *SLC19A3*	Thiamine metabolism dysfunction syndrome 2 (THMD2)	607483	MONDO:0011841
PDHA1-related pyruvate dehydrogenase deficiency	Pyruvate dehydrogenase deficiency related to *PDHA1*	Pyruvate dehydrogenase E1-alpha deficiency (PDHAD)	312170	MONDO:0010717

3. Process for Updating Nomenclature (internal and external collaborations)

When does a ClinGen expert panel need to undergo a disease naming process? In general, necessary updates to disease names arise during the gene curation (GCEP) process of lumping and splitting, especially for genes that have been associated with multiple phenotypes (guidelines linked here). Initial goals should be to ensure that curated disease entities are well-defined, have a stable identifier, and that the selected descriptor is clear, consistent with the above guidance, and reflects the data that is included in the final gene-disease validity curation. **Most importantly, these entities should be correctly defined at the gene level before specification and curation of variant pathogenicity, and would be most beneficial for harmonization between these curation activities (gene and variant) and others.**

There is no requirement for groups to change disease names when an acceptable dyadic descriptor already exists. ClinGen expert panels are also not expected to develop new nomenclature systems or harmonize disease names across their entire scope of work. That being said, we also recognize that as leaders in their fields, expert panels and/or working groups will sometimes determine that there is a need to create a more coherent set of disease names for conditions within their purview. Therefore, expert panels must be aware of the potential ramifications of changing disease names for the curated disease entity and obtain support from their community (e.g., clinical, patient advocacy, foundations, etc.) in doing so. **Expert panels need to contact and collaborate closely with both OMIM and Mondo at the start of developing new nomenclature to ensure propagation across information systems, and to ensure the development of identical and/or equivalent accepted forms of the dyadic naming conventions for the disease.** It is also important for expert panels and/or working groups to consult among internal ClinGen groups when a shared gene of interest requires updating, and arrangements should be made to discuss any planned changes with affected curation groups (e.g., variant, actionability, dosage, etc.). Below is an outline of the steps to take to communicate changes to disease nomenclature.

Internal ClinGen collaborations:
Gene-disease validity curations set the stage for the disease nomenclature(s) and identifier(s) to be used in other downstream curation activities and reports, including variant classification and actionability curations. Harmonization across all ClinGen curation activities is important, however, it is acknowledged that in some curation activities (e.g., chromosomal dosage, polygenic risk) a dyadic approach is not feasible. ClinGen's informatics infrastructure is being updated to ensure appropriate designation of disease entities and tracking across different curation activities.

Any group(s) deciding to update nomenclature should assess the current curation landscape for the gene(s) of interest by reviewing the ClinGen website, accessing the ClinGen GeneTracker, or other ClinGen systems (VCI, CSpec, ACI, etc.). Communication among expert panels and/or working groups is essential to come to a common understanding and approval of the resulting name change.

If groups find they need assistance with nomenclature updates and/or discussions, the ClinGen Disease Naming Advisory Committee can be contacted by emailing diseasenaming@clinicalgenome.org.

External OMIM and Mondo Collaborations:

For groups that will undertake any nomenclature change, whether a single disease entity, multiple disease entities, or an entire family of disease names, collaboration and communication with OMIM and Mondo is required.

OMIM (Online Mendelian Inheritance in Man): To contact OMIM, groups should send a request via the OMIM "Contact Us" tab on OMIM's website (www.omim.org). Indicate in the comment section that you are requesting a nomenclature update, and indicate any other relevant information (e.g., gene and phenotype to update, ClinGen expert panel name). Representatives from OMIM are happy to participate in ClinGen discussions and play an active role in developing nomenclature.

Mondo (Monarch Disease Ontology): To contact Mondo for nomenclature updates, groups can use either of the following methods, below. Representatives from Mondo are happy to participate in ClinGen discussions and play an active role in developing this guidance.

1. Create a GitHub issue in their Mondo repository: https://github.com/monarch-initiative/mondo/issues
    a. This requires a GitHub login, registration is free: https://github.com/join
    b. Click the green "New Issue" button after accessing the direct link above (mondo/issues)
    c. Choose the appropriate sub ticket (e.g., "Add term-gene related syndrome" or "Add synonym")
    d. Fill out the required information on the ticket.
        i. **Include that this request is from a ClinGen expert panel or working group.**
    e. Detailed instructions for Mondo term requests can be found here: https://mondo.readthedocs.io/en/latest/editors-guide/c-make-good-term-request/
        i. For Bulk requests see: https://mondo.readthedocs.io/en/latest/editors-guide/bulk-request/
2. Send an email request to info@monarchinitative.org and indicate the need for a nomenclature update. Representatives from the Mondo curation team are happy to participate in ClinGen discussions. Alternatively, Mondo holds weekly meetings where ClinGen representatives are welcome to join.

4. Considerations for Nomenclature Changes and Example Scenarios

As mentioned above, ClinGen expert panel(s) and/or working group(s) may need to navigate among several existing naming conventions and the Mondo terminologies, due to “lumping” and “splitting” decisions that come with defining the curated disease entity. Outlined below are some examples and scenarios intended to help ClinGen expert panels and/or working groups think through some of the complex issues related to disease naming with a dyadic approach.

4.1 General recommendations

4.1.1 - Well-established disease names

ClinGen expert panels should minimize changing well-known and widely accepted phenotypic labels and/or names (e.g., Marfan syndrome, cystic fibrosis, sickle cell disease), regardless of the nomenclature conventions used. Only consider renaming if there is a very strong rationale such as reducing confusion or correcting major errors. For example, the initial paper (PMID:10617473) describing germline variants in *CHEK2* asserted that there were features similar to Li Fraumeni Syndrome (LFS, MIM:151623), which is caused by variants in *TP53*. This resulted in its designation as Li Fraumeni Syndrome 2 (LFS2, MIM:609265) although the cancer risks are now known to be much lower and the tumor spectrum distinctly different from *TP53*-related LFS, which may result in inappropriate patient care (PMID:18178638, PMID:12442270). This has led to consideration of renaming the condition by the relevant ClinGen expert panel(s) as “*CHEK2*-related tumor predisposition syndrome.”

4.1.2 - Clinical validity of the gene-phenotype relationship

ClinGen expert panels should avoid creating new phenotypic labels or nomenclature for gene-phenotype relationships that do not reach a Moderate, Strong, or Definitive classification. For disease entities that have already been cataloged in a resource such as OMIM, the exact phenotypic descriptor can be utilized as the disease entity. For newly reported disease entities, the expert panel can use a broad descriptor that best reflects the phenotype(s) observed (e.g., complex neurodevelopmental disorder) or the label that is proposed in the original manuscript. Given that the process of lumping and splitting and precuration may occur before a gene-disease validity classification is known, the process of assigning a name and ID for a given curated disease entity may be an iterative process. For gene-phenotype relationships that have a limited classification, the use of the explicit gene name in the curated disease entity is discouraged and prohibited, as the presence of the gene name in the curated disease entity may imply causation that is not supported by the clinical validity classification.

4.1.3 - A gene with a single phenotype assertion and well-established uniform naming

When a gene is only associated with a single phenotype assertion with no need to “lump” or “split,” ClinGen recommends using the current well-established phenotype nomenclature whenever possible. These labels can be readily incorporated into an explicit dyadic name to provide a synonym. Exceptions to this general principle include: (1) changes initiated in order to correct or clarify a disease name that has inconsistencies or has

caused clinical confusion, or (2) to update nomenclature based on current efforts within a clinical domain are being undertaken by professional societies.

4.1.4 - A single disease entity for which different names have been used

It is not uncommon for the same disease entity to be referred to by more than one name in the literature or by different authorities or medical specialties. In this case, the expert panel would need to determine which phenotypic label they prefer, and construct an explicit dyadic name accordingly. The ClinGen gene curation for this entity would then utilize the "preferred" name and communicate with other authorities (e.g., OMIM and Mondo) to represent the exact match between the ClinGen curated disease entity and its representation in other sources.

Example - Three ways that the same entity appeared when considered by expert panel**:**

- A Mondo identifier representing the terminal "leaf" of the ontology, that describes the distinct and unitary disease entity
    - *ADNP*-related multiple congenital anomalies - intellectual disability - autism spectrum disorder
    - Mondo – MONDO:0014379
- A name used in an OMIM phenotype entry with a unique Phenotype MIM number
    - Helsmoortel-van der Aa syndrome
    - OMIM – MIM:615873
- A GeneReviews disease name for a distinct and unitary disease entity
    - *ADNP*-related disorder

In this instance a ClinGen expert panel should establish their preferred dyadic name for this disease entity (e.g., "*ADNP*-related multiple congenital anomalies - intellectual disability - autism spectrum disorder") and communicate with OMIM and Mondo.

Example - A disease gene with a known biochemical function:

- When the disease entity involves deficiency of an enzyme or metabolic pathway, the disease name can incorporate that information. This may be especially important when curating a newly associated gene-phenotype relationship where there are too few cases to fully evaluate the phenotypic spectrum to create a meaningful label.
- For example "FAR1: Fatty acyl-coA reductase 1 deficiency" (curated by the Peroxisomal GCEP)
    - ClinGen *FAR1* curation link: https://search.clinicalgenome.org/kb/genes/HGNC:26222

4.1.5 - A gene with multiple disease or phenotype assertions, in which a "lumped" entity is created

When an expert panel is reorganizing a set of disease assertions for a given gene, it may be necessary to create a "lumped" disease entity that encompasses two or more of the existing assertions, thus requiring the development of a new dyadic name for the curated disease entity. The preferred phenotypic label could combine previously used phenotypic descriptors (e.g., eponyms, phenotypic features, or pathologic characteristics) or could be a newly created term that the expert panel agrees best represents the disease entity.

Examples From ClinGen expert panels -

- *MED12*
    - This gene is associated with four eponymously named phenotypes (https://www.omim.org/entry/300188), only 3 of the phenotypes were known at the time of evaluation.
    - An explicit dyadic name of "*MED12*-related intellectual disability syndrome" was developed and used for the curated disease entity that combines two or more of the associated phenotypes.
    - ClinGen *MED12* curation link: https://search.clinicalgenome.org/kb/genes/HGNC:11957
- *OPA1*
    - This gene is associated with multiple phenotypes (https://www.omim.org/entry/605290) with overlapping phenotypic spectra with differing severity.
    - An explicit dyadic name of "OPA1-related optic atrophy with or without extraocular features" was developed and used for a curated disease entity that combines two or more of the associated phenotypes
    - ClinGen *OPA1* curation link: https://search.clinicalgenome.org/kb/genes/HGNC:8140

Example from an external group -

- *NF2*
    - The *NF2* gene has been associated with Neurofibromatosis Type 2 and Schwannomatosis.
    - Expert consensus from an international consortia (outside of ClinGen) re-evaluated the diagnostic criteria for Neurofibromatosis Type 2 and schwannomatosis and found that they were part of a spectrum of disease, thus they combined the phenotype into a new disease entity and updated the nomenclature to *NF2*-related schwannomatosis (Plotkin et al. 2022 PMID:35674741).

4.1.6 - Phenotypically related conditions for which locus heterogeneity exists

In some cases it may be desirable to have consistent naming of phenotypically related conditions with heterogeneous, non-specific, and/or indistinguishable phenotypic features, for which a number of different genes have been implicated in different individuals or families. This locus heterogeneity can be represented implicitly as a "phenotypic series" (in OMIM alphanumeric disease names) or via the explicit "gene-related phenotype" naming convention. In some cases it may be the preferred approach of the expert panel to apply the same set of phenotypic labels for all disease entities within a given spectrum.

- If a phenotypic series with consistent nomenclature exists (e.g., OMIM alphanumeric names), an expert panel may choose to adopt that naming system or convert it to an explicit dyadic naming system using the same phenotypic label.

- The expert panel may prefer to utilize a broader term (e.g., Developmental and Epileptic Encephalopathy) as the phenotypic label for a large group of diseases using an explicit dyadic naming system.

4.1.7 - Disease entities in which inter-individual phenotypic variation has been asserted as "subtypes"

Many monogenic diseases have been described to have varying phenotypic expression between different affected individuals or families. These phenotypic differences (e.g. severity of disease, age of onset, presence or absence of specific features) have often resulted in delineation of "subtypes" of disease, possibly correlating with specific types of variation within a given gene. ClinGen expert panels must therefore grapple with "lumping versus splitting" to define the curated disease entity. When an expert panel decides to "lump" subtypes with overlapping phenotypic features that vary by degree of severity into one broader disease entity, it may be advisable to retain part of the existing disease name in order to maintain consistency with previous naming systems

- *IDUA*
  - This gene is associated with a form of mucopolysaccharidosis historically divided into three subtypes based on clinical severity: Hurler (early onset, most severe), Scheie (later onset, least severe), and Hurler-Scheie (intermediate).
  - The ClinGen expert panel determined that this phenotypic spectrum is continuous and depends on the specific combination of variants present in the affected individual, suggesting that it should be deemed a single disease entity.
  - Note that while this decision to "lump" may facilitate individual variant classification and clinical diagnostic testing, it may still be reasonable for clinicians to utilize the phenotypic subtypes for individual patients when they feel it is appropriate for prognosis or management.
  - ClinGen *IDUA* curation link: https://search.clinicalgenome.org/kb/genes/HGNC:5391

4.1.8 - Curating a disputed or refuted gene-disease relationship

Expert panels should avoid creating new disease names for entities created by a "split" for the purpose of disputing or refuting a particular gene-disease relationship. For instances in which the entity is being split to dispute, use the closest entity name to the one asserted by the authors of the manuscript(s). Use a higher level term (e.g., a "branch" of the Mondo ontology, examples below) rather than creating an explicit dyadic name.

Examples -

- *BMPR1A*: pulmonary arterial hypertension
  - https://search.clinicalgenome.org/kb/genes/HGNC:1076
- *AGTR2*: X-linked complex neurodevelopmental disorder
  - https://search.clinicalgenome.org/kb/genes/HGNC:338
- *TRIM63*: hypertrophic cardiomyopathy
  - https://search.clinicalgenome.org/kb/genes/HGNC:16007

APPENDIX: Review of Current and Historical Nomenclature Conventions

This section provides an overview of several historical and current practices that do not necessarily reflect the dyadic nature of the disease entity. These descriptions also provide key examples and a brief outline of strengths/weaknesses but do not critique each convention in detail. Note that these 'categories' represent descriptors at different levels of specificity - some are broad categories of groups of phenotypes, others are quite specific and there is overlap among them. It is the heterogeneity of these various approaches that the dyadic naming convention is intended to address.

Categories

1. Eponyms

A. Description: The entity is named after a person, typically the first person who reported it in Western medical literature, or one or more individuals that have contributed substantially to the understanding of the disease etiology. Sometimes the name or initials of a patient are used, or a combination of the clinician and patient. Alternatively, some names describe a particular cultural group in which the disease was first observed.
B. Examples: Marfan syndrome; Angelman syndrome; Noonan syndrome; Lou Gehrig's disease; Parkinson's Disease; Opitz GBBB syndrome; Lynch syndrome; Gaucher disease, Native American Myopathy (aka Bailey-Bloch congenital myopathy).
C. Strengths/Weaknesses: Common, well-known eponymous names can have wide recognition and be a stable nomenclature even as the understanding of the phenotypic features evolves over time based on new data. However, eponyms lack clues as to the nature of the disease phenotype or the gene involved and often the name is based on Western publications and may not encompass other nomenclatures used in different cultures or historic precedents. Further, eponyms do not always reflect a distinct and unitary monogenic disease but rather are examples with locus heterogeneity (e.g., Bardet-Biedl syndrome). In cases of individuals who were prolific in the characterization of syndromes (e.g., Opitz), the same eponym may be included in the names for multiple distinct diseases that are completely unrelated and could thereby cause confusion among experts and non-experts alike. Many indigenous groups did not participate in the decision to include the name of their group in the disease name and discourage this practice.

2. Physical or pathologic manifestations

A. Description: The entity is named based on a clearly defined and observable or detectable (grossly or microscopically) clinical phenotype. The name can include an overarching syndromic phenotypic spectrum or a specific, organ system-limited phenotype. Often the clinical descriptor applies to a well-known and specific gene-disease relationship. Of note, sometimes a clinical phenotype is combined with an eponymous name or biochemical phenotype.
B. Examples: Cystic Fibrosis; Sickle Cell Disease; Neurofibromatosis type I; Hereditary Breast and Ovarian Cancer Predisposition; Maple Syrup Urine Disease; Oral-facial-digital syndrome

C. Strengths/Weaknesses: May provide clinically relevant information about one or more manifestations. However, it often lacks information about the gene involved, may not properly describe a unitary and distinct monogenic disease entity, and often only includes the earliest recognized and most prominent feature, e.g., Hereditary Non-Polyposis Colon Cancer (does not reflect risk of other cancer types). Further, as the understanding of the phenotype evolves over time, the original phenotypic or symptom label may become obsolete.

3. Biochemical features

A. Description: The entity is named after an abnormal biochemical finding or the biochemical function to which it is attributed. In some ways, this convention is a subset of "phenotypic" naming given that an abnormally high or low metabolite is a phenotypic feature that is often measured clinically as part of the diagnostic evaluation. Disease names that use a biochemical feature could indicate the absence of an enzyme or protein, excess or absence of an intermediate product of a biochemical pathway measurable in bodily fluids or tissues, or disruption of a particular metabolic pathway. This could include being part of a family of genes and/or proteins that all have similar or identical biochemical functions, or part of a cellular signaling cascade. In some cases the biochemical naming system has emerged alongside and sometimes supplanted a previous clinical or eponymous nomenclature as the biochemical abnormality was characterized.

B. Examples: Phenylketonuria; Homocystinuria due to MTHFR deficiency; Mucopolysaccharidosis type II; Hypoparathyroidism; Intrinsic factor deficiency (pernicious anemia); Glycogen storage disease type V (McArdle disease)

C. Strengths/Weaknesses: May provide clinically relevant information about underlying pathophysiology; often lacks clues to the clinical features of disease or gene involved, even though that may be implicit for some entities that have locus homogeneity.

4. Acronym

A. Description: The entity is named using words or letters that represent components of the observed phenotypic spectrum. The name can serve as a mnemonic to help clinicians recall specific features, and is sometimes associated with variations that reflect when specific features are present. An acronym can include biochemical, anatomic, or clinical phenotypic features. In some cases, the name used for the disorder is interchangeable between a phenotypic descriptor and its associated acronym.

B. Examples: VACTERL/VATER; MELAS; CHARGE

C. Strengths/Weaknesses: Provides some information about key phenotypic features of the disease. Lacks information about the gene involved and may appear contrived or focus on only a subset of the overall symptoms. They are often difficult to computationally distinguish from non-genetic acronyms in the literature (e.g., CHARGE), and may become obsolete if the spectrum of the disorder evolves.

5. Molecular pathway and subcellular compartment

A. Description: The entity is named based on the molecular pathway that it affects. In these cases, it has been noted that genetic variation in one or more of the pathway members leads to similar, related or sometimes opposite constellations of phenotypic features. Clinicians may suspect a specific pathway

based upon these features and prioritize examining the genes in those pathways and regulatory factors affecting those genes.

B. Examples: RASopathy; Telomeropathy, Ciliopathy; WNT signaling; SHH pathway; Double stranded break repair (DSBR); Mismatch repair (MMR)

C. Strengths/Weaknesses: Familiarity and grouping of similar diseases and/or genes implicates a class or family of proteins. This approach may or may not include clues as to the symptoms or system affected depending on the pathway name. It may invoke a group of phenotypically similar diseases (e.g., "RASopathy" with Noonan-like phenotypes), or erroneously imply phenotypic similarity when none exists (e.g., "Caveolinopathy" where *CAV1* is associated with lipodystrophy and hypertension and *CAV3* is implicated in four phenotypes that have overlapping genetic variation and features of a muscle disease with heart involvement). Furthermore, the application of pathway-based nomenclature could be misleading with respect to disease associations for related genes that have not yet been linked to disease (e.g., "Caveolinopathy" and *CAV2*, which is not associated with monogenic disease as of 2023).

6. Geographic names

A. Description: The entity is named after the geographic location of its discovery or recognition, sometimes alone, or sometimes as a modifier to another descriptor.

B. Examples: Floating Harbor syndrome, Boston-type craniosynostosis, Naxos disease, Nijmegen Breakage Syndrome, Familial Mediterranean Fever

C. Strengths/Weaknesses: Historic value, mnemonic aid for experts in the field, and can represent a stable descriptor even if phenotypes evolve. Names may be poorly understood or convey little information to outsiders; often lack details about the gene involved or phenotypic features. May cause erroneous assumptions about risk for the disease (or lack thereof) based on an individual's place of origin or residence. May not accurately represent the history of the disorder, nor represent a unitary and distinct monogenic entity.