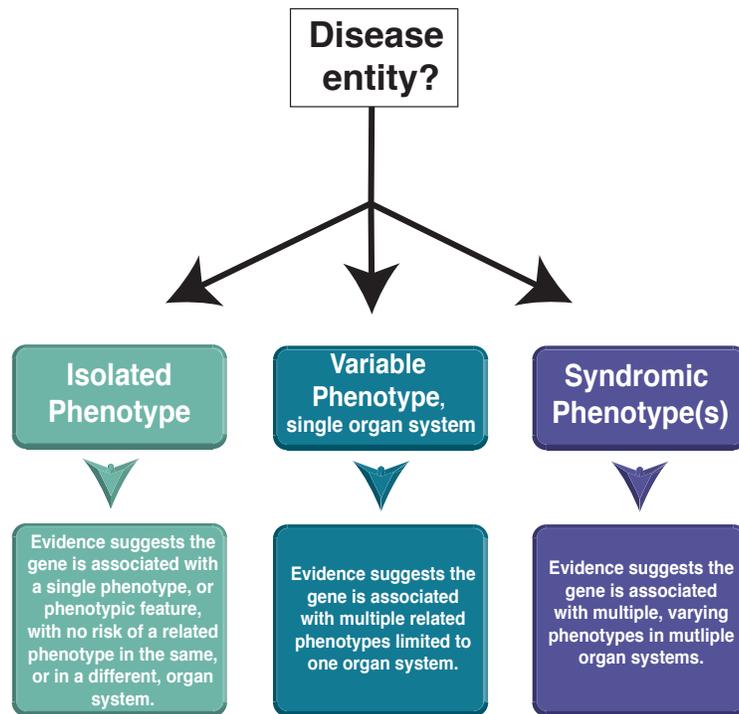


# Guidance for lumping and splitting in ClinGen gene:disease clinical validity curations.

Guidelines created by the ClinGen Lumping and Splitting Working Group<sup>1</sup>

This document is intended to assist Gene Curation Expert Panels (GCEP) and curators to decide upon the disease entity (Figure 1) used for a ClinGen gene:disease clinical validity classification when the gene is associated with one or more conditions, a variable single organ phenotype, or a syndrome.



**Figure 1. 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: (1) an isolated phenotype, where one phenotype (or phenotypic feature) arises in a single organ system with no risk of other phenotypes arising in that organ system or elsewhere; (2) a variable single organ phenotype, where multiple related phenotypes (or phenotypic features) arise in a single organ system; or (3) a syndromic phenotype, where multiple, varying phenotypes occur in multiple organs. Assessing the appropriate disease entity(ies) to curate can be challenging, thus requiring the use of defined criteria.

**In general:** Genes associated with a single published disease entity should only be curated for that condition (i.e. lumped) unless there are indications to split specific phenotypic features of a syndrome or variable phenotype into separate curation(s) based on the guidance provided in this document. If a gene is associated with multiple published disease entities this may or may not require multiple curations depending on the criteria below.

## **Key criteria to consider before lumping or splitting a gene:disease clinical validity curation**

### ❖ **Assertion/ Defining the Disease Entity:**

- Assess if one or more disease entities have been reported as being associated with a gene by nosological authorities or in the literature.
  - Check OMIM, MonDO (Monarch Initiative), Orphanet, and GeneReviews, supplemented by the primary literature.

### ❖ **Molecular Mechanism:**

- Assess whether differences in molecular mechanism(s) underlie each asserted disease entity.
  - Molecular mechanism(s) includes loss of function (LOF), gain of function (GOF), domain specific effects, isoform specificity, etc.

### ❖ **Phenotypic Variability:**

- Assess how the gene, the variant(s), and the phenotypic feature(s) present throughout a single pedigree (intrafamilial expressivity), or between two or more unrelated probands (interfamilial expressivity).

❖ **Inheritance Pattern:**

- Assess if the disease entities asserted for the gene follow different inheritance patterns.
  - It is important to determine if the disease entities are distinct (multiple varying phenotypes between them), or are part of a continuum of disease (same phenotypes, differing severity).

**Reasons to lump:**

- 1) An assertion for only one disease entity has been made in the literature.
- 2) No difference in molecular mechanism is observed among the disease entities.
- 3) Intrafamilial phenotypic variability is as, or more, pronounced than interfamilial variability.
- 4) The difference in the inheritance pattern for the disease entities is representative of a continuum of disease, i.e. mild carrier phenotypic features are observed in recessive disease(s) or dosage impacts are observed for dominant disease(s) (more severe phenotype in homozygotes).
  - a. **Note:** curate for the well-established inheritance pattern and note the additional manifestations in carrier state or homozygous state in the Gene Curation Interface (GCI).
- 5) 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.

**Reasons to split:**

- 1) An assertion for more than one distinct disease entity has been made in the literature.
- 2) A well-established difference in molecular mechanism(s) between two or more disease entities is observed.
- 3) The representative disease entities between differing inheritance patterns are distinguishable, with notable varying phenotypes and/or clinical management distinctions.
- 4) To dispute a disease entity asserted for the gene in question.
  - a. If the curator finds convincing evidence to dispute or refute the role of a gene in one of the asserted disease entities, then it may be useful to split out the additional disease, curate it separately, and record the dispute.
  - b. 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.

**Implementation of Lumping and Splitting Criteria**

**Pre-curation:** Curators should engage in a pre-curation process before beginning to curate a gene or set of genes. For GCEPs addressing a group of genes, ideally, this should occur after the gene list is composed and approved. During pre-curation, curators should collect data on the following: (1) the disease entities and/or phenotypic presentations associated with the gene, (2) the molecular mechanism(s) of the gene and variants asserted to cause each condition(s) (if available), (3) the variability of phenotypic presentations associated with each condition, and (4) the inheritance pattern for each disease entity asserted. This can be done by reviewing information on the gene and/or disease in OMIM, MonDO (Monarch Initiative), Orphanet, PubMed, and GeneReviews, and supplemented by the primary literature. ***This process is designed to give a quick overview, and is not meant to be an exhaustive search,*** but curators may find that questions arise as they progress through the curation

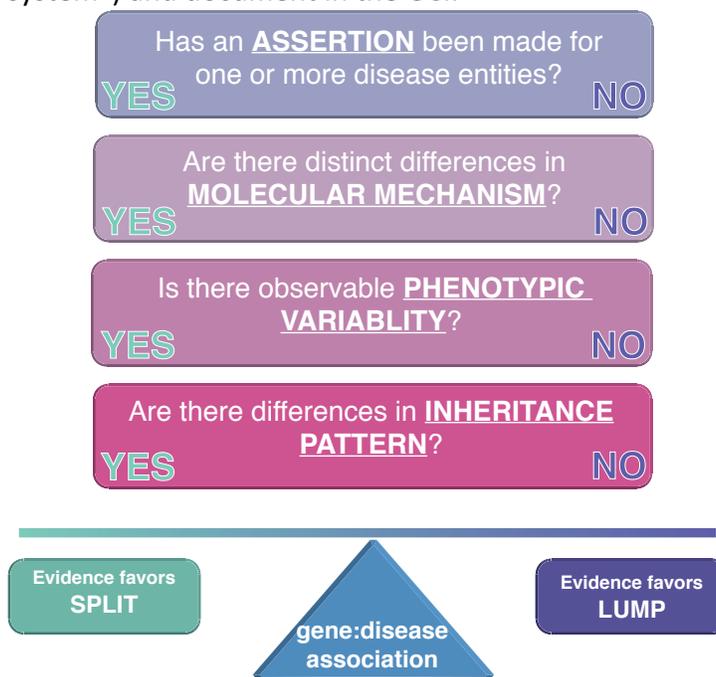
process and delve into the literature in more detail. If this occurs, then it may be necessary to revisit the criteria above to refine the disease entity(ies) appropriate for the curation.

**Binning:** Once pre-curation of genes is performed, curators will have a preliminary idea of any lumping and splitting issues and the probable disease entity(ies) to curate. Use of one or more bins, outlined below, may be useful to provide further clarification and to assist in defining the disease entity(ies) most applicable for the gene:disease curation(s). For more complex disease etiologies, the use of multiple bins may be appropriate. Furthermore, a gene may have more than one disease entity that could be curated, and thus fall into more than one of these bins. The use of binning is not to restrict gene:disease curations, but rather to provide guidance on the appropriate disease entities for which the gene in question could be curated based on the criteria listed above.

The binning strategy in its simplest form is binary, as any given gene may be involved in only an isolated phenotype (restricted to one phenotype in one organ), or in a syndrome (multiple, varying phenotypes manifesting in multiple organs). However, some genes may be involved a more complex phenotype, in which multiple related phenotypes emerge in a single organ system. For these “variable phenotypes” use of an additional bin(s), apart from the isolated and syndromic phenotype bins, may be helpful and represents a way to lump varying, but inter-related phenotypes occurring in a single organ system (see figure 1 above).

### **Final assessment for lumping or splitting:**

Deciding whether to lump or split is a balance of the criteria, i.e. a weighted scale. If the majority of criteria weigh towards splitting, then split into the relevant disease entity(ies)/conditions and record the criteria met for future review within your working group. If the criteria are weighted towards lumping, then define a single lumped condition (syndrome or “variable phenotype, single organ system”) and document in the GCI.



**Figure 3. A balance of criteria.** The four criteria for lumping and splitting should be assessed and weighed as a balance. If the evidence is equally balanced between lumping or splitting, experts should be consulted to compare the relevant weight of each piece of evidence.

### **Other considerations for gene-disease curation**

Sometimes a GCEP may wish to assess genes for their potential to be associated with a phenotypic feature (or phenotype) that has special testing, treatment or management distinctions, but which may not represent a truly distinct condition; i.e. the phenotypic feature is part of a known syndrome. For example, GCEPs may wish to identify which

syndromic genes have the potential to present with an apparent isolated phenotype or phenotypic feature (e.g. cardiomyopathy, hearing loss, aortic dissection) to ensure that the appropriate genes are

tested in patients presenting with that isolated feature. ***In these cases, the gene should be curated for the syndrome and not for the isolated phenotype (or phenotypic features) UNLESS the criteria above are met and suggest an appropriate split curation.*** In order to display the significance of a subset of isolated features of a syndrome, GCEPs may find it useful to generate a table to depict the possibility of presenting as an isolated phenotype as well as the presence, absence, or likelihood of individual features of interest for publication and testing purposes. This data can be displayed simply as an annotated table, without requiring a formal splitting for gene curation or use of the gene-disease Clinical Validity Classifications.

Finally, it may be useful to curate (or at least consider) the strongest gene-disease association first before deciding whether to lump or split out any additional, more limited disease association(s). By applying the evidence to the stronger association first, one can more effectively determine if there is sufficient independent evidence to split out, or define, additional disease associations for the gene. In some cases, one may wish to split out an additional disease entity to more clearly dispute a limited evidence claim that has been made.

**Please direct any questions or comments pertaining to this document to Courtney Thaxton:  
courtney\_thaxton@med.unc.edu**

**<sup>1</sup> *Members of the ClinGen Lumping and Splitting Working group are as follows:***

*Chair:* Courtney Thaxton, Ph.D., ClinGen Biocurator

*Co-chair:* Jennifer Goldstein, Ph.D., CGC, ClinGen Biocuration Supervisor

*Clinicians, Clinical Molecular Geneticist, Nosologists, Ontologists:*

Ada Hamosh, M.D., M.P.H, Director of OMIM

Melissa Haendel, Ph.D., Monarch Initiative PI

Jonathan S. Berg, M.D., Ph.D., ClinGen PI

Heidi Rehm, Ph.D., ClinGen PI

*Biocurators:*

Marina DiStefano, Ph.D., ClinGen Biocurator

Kathleen Wallace, B.A. (Hons), ClinGen Biocurator

P. Dane Witmer, Ph.D., Biocurator