

**Update Notes for version 2.1:**

An update to the segregations (meioses) counted in Question 9 was updated from the stated 13, to 14.

**Frequently Asked Questions for Gene Curation**

*The FAQ was updated to version 2 on October 1, 2024 by the ClinGen Gene Curation Working Group. Minor updates to wording were included.*

**Note: This is a view-only document, but please use the comment feature to add ideas for new questions if possible.**

**Table of Contents:**

**Getting Started**

- A. [Curation Shortcuts and General Advice](#)
- B. [Lumping & Splitting](#)

**Genetic Evidence**

- C. [Scoring Genetic Evidence from a Large Family](#)
- D. [Scoring Segregation Data](#)
- E. [Registering a Variant](#)
- F. [Scoring a Variant](#)
- G. [Up-scoring a Missense Variant](#)
- H. [Scoring a Nonsense or Frameshift Variant](#)
- I. [Scoring an Incompletely Penetrant Variant](#)

J. [Scoring a Case-Control Study](#)

Experimental Evidence

K. [Scoring Expression Data](#)

L. [Scoring Protein Interaction Data](#)

M. [Scoring Functional Alteration Data](#)

N. [Scoring a Model Organism](#)

O. [Scoring a Rescue Experiment](#)

**Getting Started:**

Curation Shortcuts and General Advice

**Question #1:** [Are there any websites used by curators as a starting point to quickly find many of the best articles for a gene curation?](#)

**Question #2:** [Are there any websites used by curators to quickly locate the best mouse model for a gene curation?](#)

Lumping & Splitting

**Question #3:** Under what circumstances would it be appropriate to lump an autosomal recessive disease entity and an autosomal dominant disease entity into a single gene curation with semi-dominant mode of inheritance?

**Genetic Evidence:**

**Scoring Genetic Evidence from a Large Family**

**Question #4:** When scoring genetic evidence from a large family (for example 14 individuals and 13 segregations over 4 generations), does each affected family member with the genotype count as a proband and receive a score for the variant(s) they harbor?

**Scoring Segregation Data**

**Question #5:** If a gene-disease curation has many large pedigrees with segregations but minimal functional evidence about their gene impact, can a proband be up-scored based on surplus segregation evidence (that is not necessary for scoring)?

**Question #6:** Do I need to input each individual family member separately? Is there a way to count all the variant points in a family (is there “family” scoring)? I can only see how to create the family, then to add the score I need to make an individual. If there is no family scoring available, could I create 1 individual, but increase points and describe that there are 13 others with this similar phenotype in the pedigree?

**Question #7:** Can I count individual variant points (the 0.5pt for missense for the proband and other affected members) as well as segregation data for the same family? I know we want to

avoid double counting individual cases and case-control (if the same case is individual and also in the case control), but it is double counting to score a proband variant + members of the family and then segregation too?

**Question #8:** Many papers do not identify a proband in a family (some will at least use an arrow if they don't mention the proband in the paper, but others just present the pedigree and don't mark a proband). What is the best practice for choosing a proband if it's not marked on the pedigree or mentioned in the paper?

**Question #9:** What is the procedure for counting the segregations within a family pedigree?

## Registering a Variant

**Question #10:** Is there another way to register a complex allele (harboring multiple variants) with the ClinGen Allele Registry, other than using the "Get Identifier" button (which generates an error message for this variant type)?

## Scoring a Variant

**Question #11:** When scoring a variant from a proband during a gene curation, is it necessary to perform a variant curation for each variant as a prerequisite for scoring it?

**Question #12:** When a paper describes a proband exhibiting phenotypes consistent with an autosomal recessive condition but identifies only one likely pathogenic variant (apparently heterozygous), is it okay to score the variant in a gene curation?

**Question #13:** [When performing a gene curation with an autosomal recessive mode of inheritance, is it acceptable to score variants from a proband who lacks phase data to establish that the two variants are in trans?](#)

**Question #14:** [In recessive conditions, is it possible to score homozygous occurrence of a variant in a proband with uniparental isodisomy of the chromosome that harbors the variant? \(Uniparental isodisomy is an unusual genotype in which both copies of a particular chromosome are duplicates of one of the chromosomes from a single parent. As a result, the proband will be homozygous at all loci along that chromosome.\)](#)

**Question #15:** [Is it possible to score variants from a proband who exhibits phenotypes not only of the disease entity of interest but also a second disease entity?](#)

**Question #16:** [Should missing phenotypic details from a proband be a reason to down-score their variant\(s\)?](#)

**Question #17:** [What is the role of genotyping method in determining the appropriate score of a proband's variant?](#)

**Question #18:** [What is the role of allele frequencies in gnomAD in determining the appropriate score of a proband's variant?](#)

**Question #19:** [If the majority \(or all\) of the genetic evidence for a gene-disease relationship comes from a single variant that has been identified in a large number of apparently unrelated probands, should all of the probands with the same genotype be scored, or is there a maximum number after which additional probands should not be scored?](#)

## Up-scoring a Missense Variant

**Question #20:** If an autosomal dominant gene-disease curation has many probands with missense variants but minimal experimental evidence of gene impact, are there any kinds of evidence that can be used to up-score a particular variant?

**Question #21:** If an autosomal recessive gene-disease curation has many probands with missense variants but minimal experimental/functional evidence of gene impact, are there any kinds of evidence that can be used to up-score a particular variant?

**Question #22:** If a missense variant is present in two (or more) different families (from different geographic areas of origin, with no haplotyping evidence to address the possibility of a co-inherited genomic region harboring a founder mutation), how much should the variant be up-scored in each proband for the recurrence?

### Scoring a Nonsense or Frameshift Variant

**Question #23:** For probands with nonsense or frameshift variants that are predicted to truncate the gene product but not to trigger nonsense-mediated decay, is it possible to score them as "null" variants?

**Question #24:** Regarding the 3 point limit for 1 variant introduced in SOP v8, does this limit mean that variant A in families a, b, and c can receive a maximum of 3 points collectively across families a, b and c? Or does it mean that variant A can receive a maximum of 3 points in family a, plus additional points in families b and c?

**Question #25:** When performing a gene curation with an autosomal recessive mode of inheritance, under what circumstances should a proband with homozygous null variants be scored as 3 points as opposed to 2 points?

**Question #26:** What is the appropriate score for a nonsense or frameshift variant predicted to undergo nonsense-mediated decay when performing a gene curation with an autosomal

[dominant mode of inheritance predominantly associated with gain-of-function rather than loss-of-function variants? Is 1.5 points no longer the default score for this variant type?](#)

### Scoring an Incompletely Penetrant Variant

[Question #27: For disease states known to have incomplete penetrance, are there adjustments made to the typical procedure for scoring genetic evidence?](#)

[Question #28: For X-linked dominant conditions in which heterozygous females exhibit milder disease and/or incomplete penetrance, can female probands be curated for genetic evidence?](#)

### Scoring a Case-Control Study

[Question #29: When is it appropriate to score a case-control study, and how many points should it be worth?](#)

### Experimental Evidence:

### Scoring Expression Data

[Question #30: Does the GCI permit multiple pieces of expression data to be scored for a single gene-disease curation?](#)

**Question #31:** [Is it possible to score multiple pieces of functional evidence of the same type \(for example, Expression Evidence Type A\) despite their similarity to one another?](#)

**Question #32:** [In what scenarios does expression evidence merit up-scoring past default levels?](#)

**Question #33:** [Are there scenarios in which Expression Evidence Type B \(expression altered in patients\) should not be scored but should instead be treated as variant-level evidence \(used for up-scoring variants\)?](#)

**Question #34:** [Does experimental evidence that the protein product localizes to a certain place belong in Expression A or Biochemical Function B category?](#)

**Question #35:** [Should experimental evidence that loss-of-function in the gene of interest causes a disease-related phenotype be used in the Experimental Evidence section or as variant-level functional evidence to support the scoring of a proband's variant\(s\)?](#)

## Scoring Protein Interaction Data

**Question #36:** [When the protein product from the gene-of-interest physically interacts with another protein, can this be scored as Experimental Evidence under the Protein Interaction category?](#)

## Scoring Functional Alteration Data

**Question #37:** [When a variant is exogenously expressed in a cell line and the protein product shows a functional difference from the wild-type control \(i.e. change in molecular weight, subcellular localization, or enzymatic activity\), should this be counted as evidence of Functional Alteration in non-patient cells?](#)

## Scoring a Model Organism

**Question #38:** [How many shared features with the human disease state are necessary to receive maximum scoring for a mouse model \(4 points\)?](#)

**Question #39:** [How many shared features with the human disease state are necessary for a mouse model to receive up-scoring \(~3 points\)?](#)

**Question #40:** [How many shared features with the human disease state are necessary to receive default scoring for a mouse model \(2 points\)](#)

**Question #41:** [How many shared features with the human disease state are necessary to receive reduced scoring for a mouse model \(1 point\)?](#)

**Question #42:** [What are some other factors or scenarios that can trigger down-scoring of a mouse model?](#)

**Question #43:** [How many points are appropriate for a unicellular model organism that is capable of recapitulating only the cellular-level features but not the organismal-level features of the human disorder?](#)

## Scoring a Rescue Experiment

**Question #44:** [When scoring a rescue experiment, what are some common criteria for up-scoring from the default score?](#)

## **Getting Started:**

### **Curation Shortcuts and General Advice**

**Question #1:** Are there any websites used by curators as a starting point to quickly find the best articles for a gene curation?

**Answer:** Yes, Online Mendelian Inheritance in Man (OMIM) has a website (<https://www.omim.org/>) that allows the curator to search for a gene of interest and access a specific page summarizing its function, molecular genetics, animal models, and examples of disease-associated alleles. References are provided, often including the critical publication that first established the gene-disease relationship. At the top of the OMIM page is also a table of “Genotype-Phenotype Relationships” that lists disease entities asserted by the literature to be associated with the gene of interest. This table can be a good starting point from which to evaluate multiple disease entities for lumping and splitting. ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/>) can also be a place to search for the gene of interest and inspect the variants listed as “Pathogenic”. The pages associated with these variants may list publications of cases harboring the variant.

The screenshot shows the OMIM (Online Mendelian Inheritance in Man) website. At the top, there is a navigation bar with links for About, Statistics, Downloads, Contact Us, MIMmatch, Donate, Help, and a search icon. Below the navigation bar is a search bar with the placeholder "Search OMIM..." and a "Display: Highlights" dropdown set to "Highlights". The main content area has a green header "180069" followed by "Table of Contents". To the right, the title "180069" is shown in bold. Below the title, the gene name "RETINOID ISOMEROHYDROLASE RPE65; RPE65" is displayed in bold. Underneath the gene name, the HGNC Approved Gene Symbol is listed as "RPE65". The cytogenetic location is given as "1p31.3" and the genomic coordinates as "GRCh38: 168,428,822-168,449,954 (from NCBI)". A section titled "Gene-Phenotype Relationships" contains a table with three rows of data. The columns are Location, Phenotype, View Clinical Synopses, Phenotype MIM number, Inheritance, and Phenotype mapping key. The data rows are:

Location	Phenotype	View Clinical Synopses	Phenotype MIM number	Inheritance	Phenotype mapping key
1p31.3	Leber congenital amaurosis 2		204100	AR	3
	Retinitis pigmentosa 20		613794	AR	3
	Retinitis pigmentosa 87 with choroidal involvement		618697	AD	3

**Question #2:** Are there any websites used by curators to quickly locate the best mouse model for a gene curation?

**Answer:** Yes, Mouse Genome Informatics (<http://www.informatics.jax.org/>) is a site that allows the curator to search for a gene of interest and pull up all associated mouse models. The curator could go to the “Alleles” tab and look through the list. By clicking on some of the candidate mouse models one by one, the curator can find out if any of them has an associated reference (PMID) that provides the genotype and phenotype details needed to document this evidence type. If the search lists any hits on the “Vocabulary Terms” tab, the curator can alternatively look there for any mouse models associated with the disease of interest.

**Example:** A PubMed search for a mouse model corresponding to the human [MOGS gene curation for MOGS-CDG \(autosomal recessive\)](#) yielded no results. However, a search through the MGI website revealed a *Mogs* knockout model that was generated and published as part of a high-throughput study ([PMID: 27626380](#)). The details on this

mouse model are limited to embryonic lethality; however, this matches the human MOGS-CDG patient phenotype of death in infancy and was recommended to receive a score of 0.5 points. Please note that Supplementary Table 1 from the above publication provides a list of 410 other genes associated with a lethal homozygous knockout, 198 genes associated with a subviable homozygous knockout, and 1143 genes associated with a viable homozygous knockout.

[informatics.jax.org/quicksearch/summary?queryType=exactPhrase&query=RPE65&submit=Search+Again](http://informatics.jax.org/quicksearch/summary?queryType=exactPhrase&query=RPE65&submit=Search+Again)

Score	Type	Symbol	Name	Chr
★★★☆	QTL allele	Rpe65 <sup>450L</sup>	retinal pigment epithelium 65; 450 Leucine	3
★★★☆	QTL allele	Rpe65 <sup>450M</sup>	retinal pigment epithelium 65; 450 Methionine	3
★★★☆	Endonuclease-mediated allele	Rpe65 <sup>em1Gpt</sup>	retinal pigment epithelium 65; endonuclease-mediated mutation 1, GemPharmatech Co., Ltd	3
★★★☆	Endonuclease-mediated allele	Rpe65 <sup>em2Gpt</sup>	retinal pigment epithelium 65; endonuclease-mediated mutation 2, GemPharmatech Co., Ltd	3
★★★☆	Targeted allele	Rpe65 <sup>tm1Lrcb</sup>	retinal pigment epithelium 65; targeted mutation 1, Christian Grimm	3
★★★☆	Targeted allele	Rpe65 <sup>tm2.1Tmr</sup>	retinal pigment epithelium 65; targeted mutation 2.1, T Michael Redmond	3
★★★☆	Targeted allele	Rpe65 <sup>tm1Tmr</sup>	retinal pigment epithelium 65; targeted mutation 1, T Michael Redmond	3
★★★☆	Endonuclease-mediated allele	Rpe65 <sup>em1Smoc</sup>	retinal pigment epithelium 65; endonuclease-mediated mutation 1, Shanghai Model Organisms Center	3
★★★☆	Endonuclease-mediated allele	Rpe65 <sup>em1Tfur</sup>	retinal pigment epithelium 65; endonuclease-mediated mutation 1, Takahisa Furukawa	3
★★★☆	Spontaneous allele	Rpe65 <sup>rd12</sup>	retinal pigment epithelium 65; retinal degeneration 12	3
★★★☆	Targeted allele	Rpe65 <sup>tm1.1(cre/ERT2)Kser</sup>	retinal pigment epithelium 65; targeted mutation 1.1, Philip Kiser	3

## Lumping & Splitting

**Question #3:** Under what circumstances would it be appropriate to lump an autosomal recessive disease entity and an autosomal dominant disease entity into a single gene curation with semi-dominant mode of inheritance?

**Answer:** Lumping would be appropriate when phenotype evidence and mechanistic evidence outweigh the difference in mode of inheritance. On the other hand, splitting is appropriate when key phenotypes are not shared between autosomal recessive and autosomal dominant disease entities, so that clinical management is quite different. The [webpage of ClinGen's Lumping & Splitting Working Group](#) provides additional details in the form of guidelines and training materials.

**Example:** In the [LDLR gene curation for hypercholesterolemia, familial, 1 \(semidominant mode of inheritance\)](#), the affected individuals harboring heterozygous variants have a later onset (adult) than the affected individuals harboring homozygous variants (infantile), with similarities in phenotypes and clinical management. In addition, the heterozygous individuals and homozygous individuals are present within the same families, so that the variants (and their mechanisms of pathogenicity) are shared between the autosomal recessive and autosomal dominant disease entities. On the other hand, the [ATM gene curation for hereditary breast carcinoma \(autosomal dominant\) and ataxia telangiectasia \(autosomal recessive\)](#) is an example in which the two disease entities were split in spite of the shared variants. The reason was that key phenotypes are not shared, so that clinical management was quite different.

### **Genetic Evidence:**

#### **Scoring Genetic Evidence from a Large Family**

**Question #4:** When scoring genetic evidence from a large family (for example 14 individuals and 13 segregations over 4 generations), does each affected family member with the genotype count as a proband and receive a score for the variant(s) they harbor?

**Answer:** No, in this situation, only the family member whose genotyping was the starting point for subsequent studies of the family is defined as a proband. Only that proband receives a score for their variant(s), while the family as a whole will be scored for segregation points instead. Scoring each family member individually would be considered double-counting, while scoring a single proband plus segregation evidence is considered acceptable. If the paper does not designate a proband in the pedigree, the curator should designate one family member who has been genotyped and phenotyped as the proband and count the rest as segregations.

## Scoring Segregation Data

**Question #5:** If a gene-disease curation has many large pedigrees with segregations but minimal functional evidence about their gene impact, can a proband be up-scored based on surplus segregation evidence (that is not necessary for scoring)?

**Answer:** No, since there is a danger of allowing too much weight to segregation when variants/genes may simply be in cis with the true cause of disease. Generally segregation data is not used to up-score variants. In an exceptional case where segregation data were used this way, it would be important to ensure that all other causes within the linkage region have been fully ruled out.

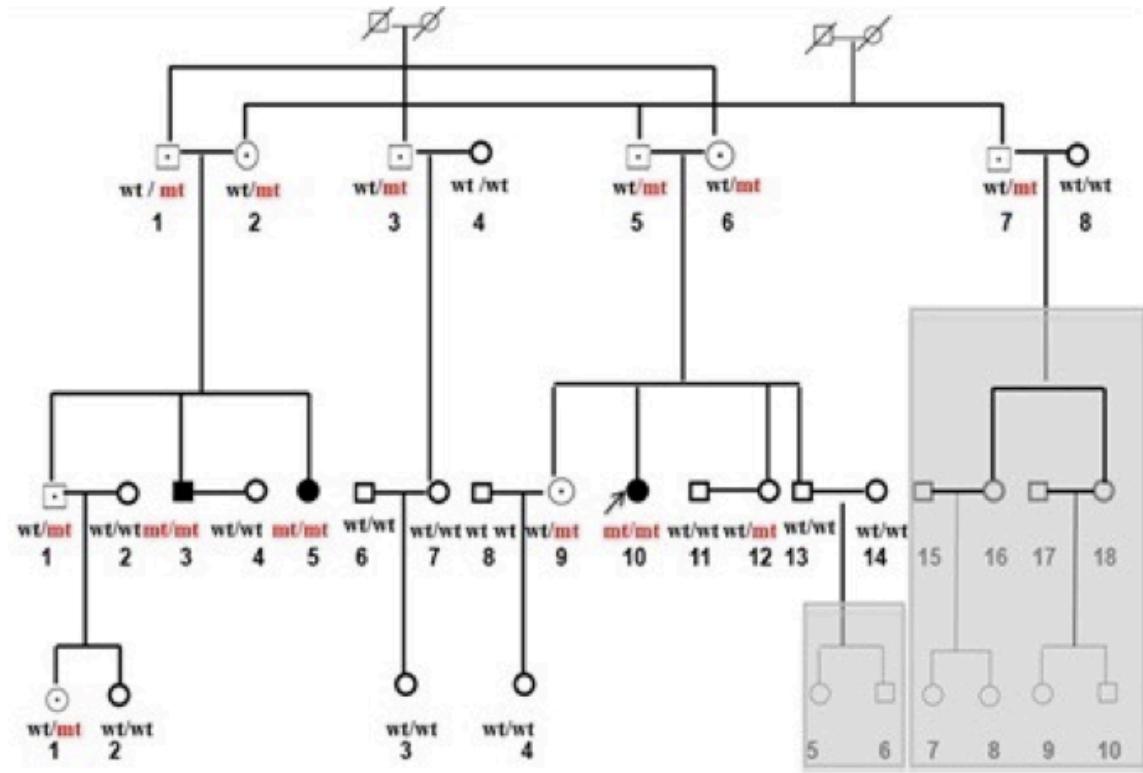
**Example:** In the [curation of TIMP3 for Sorsby fundus dystrophy \(autosomal dominant\)](#), the proband from [PMID:16989765](#) was scored at 0.1 points, despite the presence of segregation evidence from the pedigree in Figure 1. Instead, the eLOD score was

estimated and incorporated into the final calculation of segregation evidence (which had already been maximized by the combined effects of other large pedigrees).

**Question #6:** Within a family with multiple affected family members, is it necessary to input each individual family member into the gene curation separately?

**Answer:** No, for a large family, it can be a lot of work to enter the details for each family member and this is not necessary in most cases. Instead, when the curator creates an entry for the family, the GCI will ask for symptoms that the affected members have in common. This is the required level of detail in order to calculate segregation points from the family, so it is not necessary to enter details on every family member in order to get those points. Near the bottom of the family page, the GCI will ask for the proband label. The curator can then enter additional details about the proband and score the variant(s) they harbor.

**Example:** The family pedigree below (from PMID: 30839500) shows three affected individuals who have been well-characterized clinically. However, only the individual indicated by the arrow as the family proband should be documented in the GCI and scored in the *BBS7* gene curation. Her similarly affected cousins have not been documented in the curation in detail, but have contributed to the genetic evidence of the curation in the form of family segregation.



**Question #7:** Is it double-counting to score the variant in the proband and also score segregation data for the proband's family?

**Answer:** It is acceptable to count both case points for the proband and segregation data for the family. This is not considered to be double-counting because these are different types of evidence. For a large family that meets requirements for scoring segregation, the curator can count points for the variants from the proband and points for segregation (as calculated by the GCI).

**Question #8:** Many papers do not identify a proband in a family (some will at least use an arrow if they don't mention the proband in the paper, but others just present the pedigree and

don't mark a proband). What is the best practice for choosing a proband if it's not marked on the pedigree or mentioned in the paper?

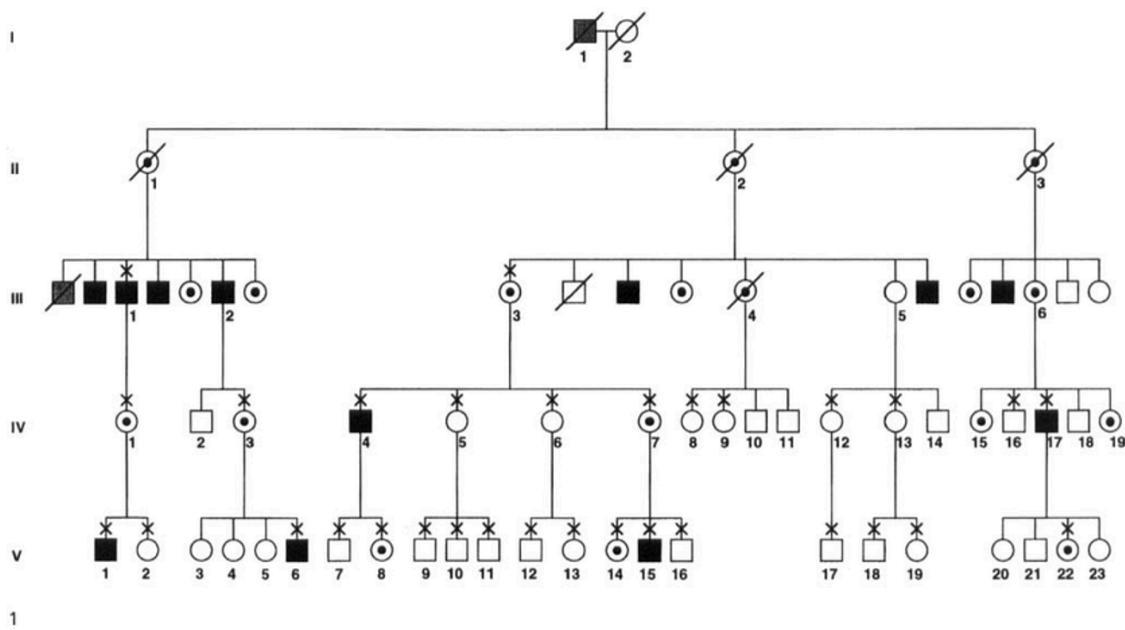
**Answer:** There is no strict rule for this. If a proband is not designated in the paper, you may choose an individual for whom there is the most clinical/laboratory information, and enter those details into the GCI. The curator can make a note to say "no proband designated in the pedigree, therefore the curator designated the proband" or something to that effect. If the pedigree is labeled with generation and individual numbers, it is possible to use that for the proband label e.g. II-V (designated by curator). If not, it will be necessary to use another way to designate who the proband is.

**Example:** In the partial table below, which has been copied from PMID: 25329329, clinical details have been provided from three separate family members, but the authors have not indicated one of them as the proband of the family. In this scenario, a curator might choose patient A.II.9 as the proband, based on the larger number of features present, which may help to establish the overall phenotype as a specific match to the disorder expected in association with the gene of interest.

<b>Patients</b>	<b>A.II.5</b>	<b>A.II.8<sup>†</sup></b>	<b>A.II.9<sup>†</sup></b>
<b>Year of birth</b>	<b>1959</b>	<b>1963</b>	<b>1965</b>
Diarrhea/Enteropathy	yes	yes	yes
GLILD	-	yes	yes
Respiratory infections <sup>#</sup>	yes	yes	yes
Splenomegaly	yes	-	yes
AI thrombocytopenia	-	-	-
AI hemolytic anemia	-	-	-
Lymphadenopathy	-	-	-
Bone marrow infiltration	nd	nd	yes
Growth retardation (<3 <sup>rd</sup> percentile)	-	-	-
Kidney disease*	-	-	-
Liver disease*	-	-	-
Psoriasis/other skin disease*	-	-	-

**Question #9:** What is the procedure for counting the segregations within a family pedigree?

**Answer:** Each segregation (also known as a meiosis) is the distance in the family pedigree between two individuals who are immediate family members. For example, there is one segregation between an individual and his parent, but two segregations separating the individual from his grandparent. When counting segregations within an affected family, the strategy is to count only the number of segregations / connections between individuals who are positive for both the genotype and the affected phenotype.



**Example:** In the example figure above, which comes from PMID: 9457748, there are 6 affected individuals (black squares / circles) who have also been genotyped (small x above the squares / circles); III:1, IV:4, IV:17, V:1, V:6, and V:15. For example, there are two segregations separating individual V:1 from his grandfather III:1, three more segregations separating individual III:1 from his brother's grandson V:6, etc. The total number of segregations separating the 6 affected individuals is 14. Other affected family members are not considered in this total number of segregations because they either have not been genotyped or are unaffected by the disease.

## Registering a complex variant allele with the ClinGen Allele Registry

**Question #10:** Is there another way to register a complex allele (harboring multiple variants) with the ClinGen Allele Registry, other than using the "Get Identifier" button (which generates an error message for this variant type)?

**Answer:** The semicolon used to separate the two variants in HGVS nomenclature appears to trigger a problem for the allele registration system. This is a scenario that requires contacting the ClinGen Allele Registry (at brl-allele-reg@bcm.edu). It may be helpful to send screenshots of the problem, along with the question.

**Example:** An example of a compound allele (harboring two different variants in cis) successfully registered on the ClinGen Allele Registry is [CA658653750](#) (also known as [ClinVar Variation ID: 375688](#), described in [PMID:24716661](#)).

## Scoring a Variant

**Question #11:** When scoring a variant from a proband during a gene curation, is it necessary to perform a variant curation for each variant as a prerequisite for scoring it?

**Answer:** No, it is not necessary to do a full variant curation for variants included in gene curations. While curators consider many of the same factors in deciding on scores for gene curation as they do for applying criteria in variant curation, gene curation does not require the rigor of selecting which variant curation criteria are met or not met for every single variant. On the other hand, gene curators can feel free to use the tools in the variant curation interface to examine the evidence for pathogenic or benign impact by a particular variant, including allele frequencies in control populations from gnomAD. This

info can be accessed after the variant is added to the gene curation record for the proband, which triggers the appearance of a link in the GCI entitled “view variant evidence in Variant Curation Interface”. For more information about the impact of allele frequencies on genetic evidence scoring, please see [this question](#).

Individual 🚑 – Variant(s) and Score(s) segregating with Proband

Check here if hemizygous:

[ClinVar](#) Preferred Title: NM\_012186.3(FOXE3):c.720C>A (p.Cys240Ter) [View variant evidence in Variant Curation Interface](#) [Clear variant selection](#)

[ClinVar](#) Variation ID:8448 [ClinGen Allele Registry](#) ID:CA119630

Variant Type: *	Predicted or proven null
Is this variant de novo ⓘ ?: *	No
If yes, is the variant maternity and paternity confirmed?:	Yes
Is there functional data to support this variant?: *	Yes
If yes, please describe functional data: * <i>(Required if selected yes above)</i>	Transfection of the variant into HEK-293 cells harboring the DNAJB1-GLuc-SEAP reporter vector was associated with decreased expression of luciferase, indicating loss of ability to transactivate
Select Status:	Supports
Explanation:	

**Question #12:** When a paper describes a proband exhibiting phenotypes consistent with an autosomal recessive condition but identifies only one likely pathogenic variant (apparently heterozygous), is it okay to score the variant in a gene curation?

**Answer:** Generally a single variant from an apparently heterozygous proband would not be scored for an autosomal recessive curation. This is particularly true if the phenotype is not highly specific for dysfunction in this exact gene. While there is a possibility that the proband harbors a second, unidentified variant in the other allele (such as a large deletion or a difficult-to-identify deep intronic variant with gene impact), it is also possible that there is a different mechanism causing disease (such as digenic inheritance, with the second pathogenic variant present within a completely different gene). The

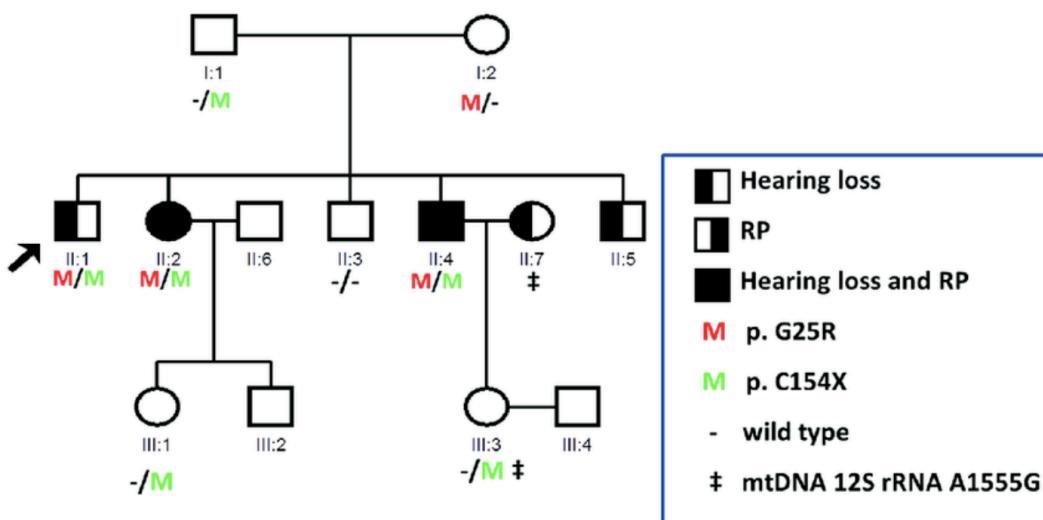
gene curation protocol has been developed for monogenic rather than digenic / polygenic disease. As a result, the norm is to avoid scoring this proband.

**Example:** A case that indicates a phenotype consistent with biallelic (i.e., autosomal recessive) inheritance, but in which only one variant within the asserted gene is indicated, and there is no parental testing to confirm homozygous *in trans* inheritance, nor additional testing to confirm a variation on the opposite allele. In this situation, the evidence is not supporting the mode of inheritance, and thus should be scored 0 or not even included in the curation. A group may choose to indicate this evidence in the final Evidence summary in order to note the presence of the data, but indicate that due to the inability to confirm the mode of inheritance and/or alleles that it is not included in the scoring at this time.

**Question #13:** When performing a gene curation with an autosomal recessive mode of inheritance, is it acceptable to score variants from a proband who lacks phase data to establish that the two variants are in trans?

**Answer:** Most GCEPs do not score variants with untested phasing due to the possibility that they may be harbored within the same allele. There are some exceptions where GCEPs choose to down-score the variants instead. It may be a good idea to consult the expert panel to weigh in on this question with regard to the specific variant and the evidence available. Successful scoring has sometimes been achieved by reaching out to the authors by email to ask for unpublished details about phasing. In curations with abundant probands / variants, it is best to omit those without information about the phase of the two variants. In curations where it would be very helpful to include these variants for scoring, it can be an option to write to the authors to request those details. “Per author communication” is an acceptable rationale to provide to justify scoring these variants. The GCI also now requires that the phasing status of autosomal recessive variants be specified. Please answer the GCI question accordingly based on the paper and/or the personal communication with the authors.

**Example:** The figure below from PMID: 26425852 has multiple affected individuals in the second generation, each of whom harbors two variants identified as originating from the paternal side (green) or maternal side (red) based on genotyping of the father and mother in the first generation. This genotyping of heterozygous family members makes it clear that the p.G25R (red) and p.C154X (green) variants are on two different alleles from two separate sides of the family, allowing the curator to confidently score both variants in the affected proband. In the hypothetical scenario where only the affected family members had been genotyped, and no unaffected individuals had been identified as heterozygous carriers, then it would be unclear whether the two variants were present in two different alleles (scorable) or the same allele (unscorable), so scoring of the proband would be avoided.



**Question #14:** In recessive conditions, is it possible to score homozygous occurrence of a variant in a proband with uniparental isodisomy of the chromosome that harbors the variant? (Uniparental isodisomy is an unusual genotype in which both copies of a particular chromosome are duplicates of one of the chromosomes from a single parent. As a result, the proband will be homozygous at all loci along that chromosome.)

**Answer:** Score these probands with caution as you would probands with variants in genes with similar phenotypes. Probands with uniparental isodisomy will be homozygous at all loci along that chromosome. As such, they may have multiple genetic findings and present with a complex phenotype. Consultation with your GCEP is recommended.

**Example:** The figure below, adapted from PMID: 12142464, depicts a proband of interest considered for the [DNAH11 gene curation for primary ciliary dyskinesia 7](#). The proband is difficult to evaluate because they have inherited two identical copies of chromosome 7 from the father and none from the mother, through an unusual mechanism called uniparental isodisomy. As a result, the proband is homozygous for all loci on chromosome 7, including the *DNAH11* locus (which harbors a nonsense variant that may cause disease) and the *CFTR* locus (which harbors a well-known disease causing variant, p.Phe508del). The proband exhibits a combination of clinical features, some associated with the disease of interest (ciliary dyskinesia, situs inversus totalis, and neonatal respiratory distress) and some associated with cystic fibrosis (elevated sweat chloride, end-stage lung disease at age 7). Since the proband likely has two different diseases that are both respiratory, it was difficult to evaluate which clinical features, if any, came from *DNAH11* vs. from *CFTR*. The curator was encouraged to find other *DNAH11* probands with less complex / confounding genetics.

**Bartoloni\_2001\_Proband:**

DNAH11 status:  
homozygous for p.Arg2845Ter

Genetic confounder:  
Uniparental isodisomy of chromosome 7,  
Homozygous *CFTR* p.Phe508del

DNAH11-related phenotypes present:

- Ciliary dyskinesia
- Situs inversus totalis
- Neonatal respiratory distress

CFTR-related phenotypes present:

- Elevated sweat chloride
- End-stage lung disease at age 7

**D.**

**Question #15:** Is it possible to score variants from a proband who exhibits phenotypes not only of the disease entity of interest but also a second disease entity?

**Answer:** In general, it is possible to score a proband when the phenotypes from the two disease entities are sufficiently distinguishable. However, it may not be possible to score the proband if the two inherited conditions affect the same organ system. It may be a good idea to consult the expert panel to weigh in on this question with regard to the specific proband and the evidence available.

**Example:** In the Motile Ciliopathy curation of *DNAH11* for Primary Ciliary Dyskinesia 7, the proband who was the founder case for the gene-disease association ([PMID: 12142464](#)) could not be scored due to co-inheritance of both primary ciliary dyskinesia and cystic fibrosis (both of which affect respiratory function).

**Question #16:** Should missing phenotypic details from a proband be a reason to down-score their variant(s)?

**Answer:** One of the scenarios where down-scoring of a variant may be recommended is when most of the phenotypes are consistent with the suspected disease, but one critical feature expected to always be present has been tested for and found to be absent. On the other hand, if a proband is found to lack a common disease-associated phenotype and the phenotype is known to demonstrate variable expressivity, down-scoring is NOT recommended.

If a patient is missing too many expected phenotypes, or a single necessary phenotype, the proband could receive a score of 0. Specifically which phenotypes are necessary in order for a proband to receive a score above 0 should be decided by each expert panel. However, any evidence showing that the proband does not have a different disease (such as exome sequencing-based genotyping) can strengthen the argument that the proband should receive a score greater than 0.

**Example:** In the General Inborn Errors of Metabolism group [SLC6A3 gene curation for SLC6A3-related dopamine transporter deficiency syndrome](#), elevated homovanillic acid (HVA) is expected in the urine or cerebrospinal fluid of patients with infantile-onset parkinsonism-dystonia 1. Only 1 of ~25 patients diagnosed with this disease has ever been found to have normal HVA concentration (or HVA:HIAA ratio) prior to treatment. In this curation, the proband was down-scored 0.5 pts for lacking an important/expected phenotype.

**Question #17:** What is the role of genotyping method in determining the appropriate score of a proband's variant?

**Answer:** Limitations of the method used to genotype a proband can be a rationale for down-scoring a variant from that proband when other genes that might explain the phenotype have not been examined. Sanger sequencing of a single candidate gene or a small number of candidate genes is a common rationale for down-scoring a variant. One exception might be when Sanger sequencing has been combined with a genome-wide screening method such as linkage mapping or homozygosity mapping that has already narrowed down the causal variant in the family to a small fraction of the genome. On the other hand, more comprehensive genotyping methods such as exome sequencing are generally not used as a rationale to up-score a variant.

**Example:** In the curation of [GPR143 for GPR143-related foveal hypoplasia with or without albinism](#), the heterozygous frameshift variant c.933\_934dup (p.Gly312AlafsTer22) was predicted to trigger nonsense-mediated decay, but was down-scored from 1.5 points to 1 point since the proband was genotyped by Sanger sequencing of the *GPR143* locus alone (PMID: 7647783). The 1995 publication did not genotype other loci such as *CACNA1F* that can harbor variants causing a similar phenotype, and *CACNA1F* in particular was not identified as a disease-associated gene until a later (1998) publication. This practice of down-scoring variants due to a limited genotyping method is common when scoring genetic evidence from older papers. The full 1.5 points would be considered more appropriate if all relevant genes had been included in genotyping (for example by exome sequencing), and only 1 gene was found to harbor a suspected disease-causing variant.

On the other hand, some GCEPs that down-score variants when consanguinity is known to be present might do so to a lesser degree when exome or genome sequencing has comprehensively ruled out alternative causes of disease in other genes of interest.

**Question #18:** What is the role of allele frequencies in gnomAD in determining the appropriate score of a proband's variant?

**Answer:** Allele frequency in gnomAD can be a rationale for down-scoring a proband's variant if it is too high to be consistent with the variant causing a rare disease, while low allele frequency in gnomAD is generally not used as a rationale to up-score. According to gnomAD's frequently asked questions (FAQ/Help) section, the resource has "made every effort to exclude individuals with severe pediatric diseases from the gnomAD data set, and do not expect the data set to be enriched for such individuals, but cannot rule out the possibility that a given participant has a disease of interest." Caution should be applied "before excluding variants as disease candidates when seen in one or a few individuals" (PMID: 34859531). Genotypes present in gnomAD may still cause later-onset or mild conditions, or may cause disease with incomplete penetrance. When using gnomAD, it is important to consider the inheritance pattern of the disease. Recessive conditions may have higher occurrences of disease causing variants in the heterozygous state given individuals are often unaffected. Many GCEPs will set allele frequency cutoffs; please consult with your GCEP.

**Example:** In the [curation of CFAP298 for primary ciliary dyskinesia 26 \(autosomal recessive\)](#), a nonsense variant predicted to cause C-terminal truncation of the protein product (p.Tyr245Ter) was down-scored from 0.5 points to 0.1 point for being [found in gnomAD v2.1.1 at a frequency that was nearly too high to believe as a cause of this rare disease](#). Some scoring was still considered appropriate because the highest allele frequency of the variant was associated with one of the bottlenecked populations within gnomAD (Ashkenazi Jewish) and because the variant only occurred in gnomAD in the heterozygous state (below).

SNV: 21-33974609-G-C(GRCh37)				<a href="#">Copy variant ID</a>	Dataset <a href="#">gnomAD v2.1.1</a>
Filters	Exomes	Genomes	Total	External Resources	
	Pass	Pass		• <a href="#">dbSNP (rs202094637)</a>	• <a href="#">UCSC</a>
<b>Allele Count</b>	20	1	21	• <a href="#">ClinVar (209002)</a>	• <a href="#">ClinGen Allele Registry (CA346951)</a>
<b>Allele Number</b>	251370	31402	282772		
<b>Allele Frequency</b>	0.00007956	0.00003185	0.00007426		
<b>Popmax Filtering AF ⓘ (95% confidence)</b>	—	—			
<b>Number of homozygotes</b>	0	0	0	<b>Feedback</b>	
<b>Mean depth of coverage</b>	67.2	32.7		<a href="#">Report an issue with this variant</a>	
<b>Population Frequencies ⓘ</b>					
Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency	
↳ Ashkenazi Jewish	18	10370	0	0.001736	
↳ Other	2	7224	0	0.0002769	
↳ European (non-Finnish)	1	129184	0	0.000007741	
↳ African/African American	0	24966	0	0.000	
↳ Latino/Admixed American	0	35440	0	0.000	
↳ East Asian	0	19954	0	0.000	
↳ European (Finnish)	0	25018	0	0.000	
↳ South Asian	0	30616	0	0.000	
XX	8	129430	0	0.00006181	
XY	13	153342	0	0.00008478	
<b>Total</b>	<b>21</b>	<b>282772</b>	<b>0</b>	<b>0.00007426</b>	

On the other hand, in the curation of the *CTLA4* gene, which causes severe early-onset disease through an autosomal dominant mechanism with approximately 50% penetrance, the frequency of 0.0004365 in gnomAD (below) was considered inconsistent with a disease-causing effect, so any proband harboring the variant should not be scored.

**SNV: 2-204735525-G-A(GRCh37)** [Copy variant ID](#) [Gene page](#) Dataset gnomAD v2.1.1

Filters	Exomes	Genomes	Total	External Resources
<b>Allele Count</b>	57	14	71	<ul style="list-style-type: none"> <li>dbSNP (rs144988077)</li> <li>UCSC</li> <li>ClinVar (542071)</li> <li>ClinGen Allele Registry (CA2067088)</li> </ul>
<b>Allele Number</b>	250600	31382	281982	
<b>Allele Frequency</b>	0.0002275	0.0004461	0.0002518	
<b>Popmax Filtering AF ⓘ (95% confidence)</b>	0.0002903	0.0004987		
<b>Number of homozygotes</b>	0	0	0	
<b>Mean depth of coverage</b>	94.6	31.3		<a href="#">Report an issue with this variant</a>

**Population Frequencies ⓘ**

Population	Allele Count	Allele Number	Number of Homozygotes	Allele Frequency
European (non-Finnish)	56	128304	0	0.0004365
Latino/Admixed American	8	35438	0	0.0002257
Other	1	7226	0	0.0001384
South Asian	4	30616	0	0.0001307
Ashkenazi Jewish	1	10362	0	0.00009651
African/African American	1	24964	0	0.00004006
East Asian	0	19950	0	0.000
European (Finnish)	0	25122	0	0.000
XX	31	129024	0	0.0002403
XY	40	152958	0	0.0002615
<b>Total</b>	<b>71</b>	<b>281982</b>	<b>0</b>	<b>0.0002518</b>

**Question #19:** If the majority (or all) of the genetic evidence for a gene-disease relationship comes from a single variant that has been identified in a large number of apparently unrelated probands, should all of the probands with the same genotype be scored, or is there a maximum number after which additional probands should receive 0 points?

**Answer:** If variant-level functional evidence is available indicating that the variant has a deleterious effect on the gene product that may be consistent with the disease state, it is acceptable to score as many apparently unrelated probands as needed to reach a definitive classification. While it is highly encouraged to prioritize diversity of variants and probands across multiple ancestral populations, it is still possible to reach a definitive classification scoring only a single variant. This could be relevant for gene-disease relationships in which a recurrent variant is the primary cause of the disease (e.g., sickle cell anemia, achondroplasia, etc.). In these cases, curators should evaluate whether consanguinity is present and determine whether the variant should be down-scored or not scored). Curators should also consider the possibility that the

variant may be in linkage with a second nearby variant that is also present in all of the affected probands.

**Examples:** Multiple occurrences of a single variant are responsible for all of the genetic evidence for the curation of [RPE65 for RPE65-related dominant retinopathy](#), and all 5 probands harboring the c.1430A>G variant have been scored despite their shared geographic origin in the U.K. The initial proband was scored at 1 point due to the presence of functional evidence demonstrating a deleterious effect of the variant, while the other four probands each received 0.5 points (half-scoring) in order to account for the potential origin of the variant as a founder mutation. Genetic evidence for the [VPS33A gene curation](#) is similarly heavily based on probands harboring the p.Arg498Trp variant in the homozygous state, with evidence of a destabilizing impact on the protein product.

### Up-scoring a Missense Variant

**Question #20:** If an [autosomal dominant](#) gene-disease curation has many probands with missense variants but minimal experimental evidence of gene impact, are there any kinds of evidence that can be used to up-score a particular variant?

**Answer:** Yes, possibly. Observation of the same variant in multiple apparently unrelated affected individuals can be a rationale for bumping up proband score. Another rationale for up-scoring a variant can be its compatibility with a pattern (such as occurrence within a well-known hot spot) for which you can cite a reference describing the pattern. It may be a good idea to share the specific details with the GCEP and request their guidance in defining a hot spot or recurring pattern.

**Examples:** In the [curation of TIMP3 for Sorsby fundus dystrophy \(autosomal dominant\)](#), the same missense variant (ClinVar:12677) was reported in [PMID:7894485](#) (Family M in

Figure 1A) and [PMID:27601084](#) (Family 1 in Figure 1). The two author lists and the two families were confirmed to be apparently distinct. Scoring was increased to 0.5 points for the first proband and to 0.25 points for the second. (Due to the absence of haplotyping data to evaluate whether the two probands were inheriting the same founder mutation, the second observation of the variant was only up-scored by half.) An example of a hot spot for disease-causing missense variants was identified in the [GUCY2D gene curation for GUCY2D-related dominant retinopathy](#), in which all 8 variants scored were missense variants occurring within a small region between codons 838 and 849. These variants were up-scored from 0.1 points to 0.25 points in the absence of any variant-level evidence of gene impact, or up-scored from 0.1 points to 1 point if experimental evidence of defective dimerization was also found.

**Question #21:** If an autosomal recessive gene-disease curation has many probands with missense variants but minimal experimental/functional evidence of gene impact, are there any kinds of evidence that can be used to up-score a particular variant?

**Answer:** Yes, possibly. Observation of the same variant in multiple apparently unrelated affected patients in the compound heterozygous state with multiple different variants in trans can be a rationale for bumping up variant scores. Another rationale for up-scoring a variant can be its compatibility with a pattern (such as occurrence within a well-known hot spot) for which you can cite a reference describing the pattern. It may be a good idea to share the specific details with the GCEP and request their guidance in defining a hot spot or recurring pattern.

**Example:** In the curation of [MOGS for CDG2B \(autosomal recessive\)](#), a single codon deletion variant ([p.Gln399del](#)) was reported in [PMID:33261925](#) in two separate probands, each with a different variant in trans ([p.Arg389Ter](#) in one patient and [p.Arg429Ter](#) in the other). Scoring was increased by 0.4 points for the first proband and

by 0.2 points for the second. (Due to the absence of haplotyping data to prove the probands were not inheriting the same founder mutation, the second observation of the variant was only up-scored by half.) As a second example, [PMID: 31227806](#) describes a pattern of variants in the *FBN1* gene that are associated with [Marfan syndrome \(autosomal dominant\)](#) and disrupt cysteine or glycine residues within critical domains of the protein product and have a consistently pathogenic effect.

**Question #22:** If a missense variant is present in two (or more) different families (from different geographic areas of origin, with no haplotyping evidence to address the possibility of a co-inherited genomic region harboring a founder mutation), how much should the variant be up-scored in each proband for the recurrence?

**Answer:** Some GCEPs recommend up-scoring one proband by the full amount (i.e. up-score from 0.1 to 0.5) and up-scoring the second proband by half (i.e. from 0.1 to 0.25). Also, try as much as possible to confirm that the two pedigrees do not represent different branches of the same family, or the same family in two different publications. For example, if the two different populations in which the variant recurs are historically related, be more conservative with up-scoring to acknowledge the possibility of a founder mutation. Please note that down-scoring is not necessary if the variant has been confirmed to occur on multiple different haplotypes, or if one of the cases has been identified as *de novo*. It may be a good idea to share the specific details with the GCEP and request their guidance on questions like this. One of the advantages of consulting the GCEP is to maintain consistency with how the group has handled this scenario in other gene curations.

**Example:** In the [curation of \*TIMP3\* for Sorsby fundus dystrophy \(autosomal dominant\)](#), the same missense variant (ClinVar:12677) was reported in [PMID:7894485](#) (Family M in Figure 1A) and [PMID:27601084](#) (Family 1 in Figure 1). The two author lists and the two

families were confirmed to be apparently distinct. Scoring was increased to 0.5 points for the first proband and to 0.25 points for the second.

### Scoring a Nonsense or Frameshift Variant

**Question #23:** For probands with nonsense or frameshift variants that are predicted to truncate the gene product but not to trigger nonsense-mediated decay, is it possible to score them as "null" variants?

**Answer:** The Gene-Disease Validity SOP v11 now indicates that the variant should be categorized based on the type of variant, and points be adjusted once evidence indicates the true mechanism. Therefore, missense variants should be categorized as "other type" of variant. If functional evidence supports the missense exhibits a loss of function mechanism, then the total points for the variant can be adjusted up to 1.5, which is the default for a "predicted or proven null" variant. Similarly, if you have a variant that is a frameshift or nonsense variant, it should be categorized as "predicted or proven null". If functional evidence indicates the mechanism is NOT loss of function, or if the variation occurs within 50bp of the penultimate exon and therefore not predicted to go through non-sense mediated decay, then we would suggest the score be decreased from the 1.5 default into the 0.1-0.5 range similar to the "other" variant default.

**Example:** In the curation of [MOGS for CDG2B \(autosomal recessive\)](#), the [p.Arg535Ter](#) (aka., ARG429Ter, non MANE transcript) nonsense variant occurs in the final exon of *MOGS* and is predicted to truncate the C-terminal ~35% of the enzyme, but lacks functional evidence to confirm loss-of-function ([PMID:29235540](#)). Thus, instead of being scored at 1.5 points as a "null", the variant was scored at 0.5 ("some evidence of gene impact", in the form of the predicted truncation).

**Question #25:** Regarding the 3 point limit for a single piece of genetic evidence, does this limit mean that variant A found in probands a, b, and c cannot receive more than the maximum of 3 points in any single proband? Or does it mean that variant A can only receive a maximum of 3 points collectively across families a, b and c?

**Answer:** The 3-point limit refers to the maximum number of total points for a single case. For a case with an autosomal recessive disorder who is compound heterozygous, and each variant scores the maximum number of points (3), the maximum for the case is still 3 points total. The GCI will automatically calculate this. On the other hand, because curators are permitted to score the same variant in multiple individuals (as long as they are not related), it is possible for a variant to score more than 3 points total. However, the recommendation is to try to score as many different variants as possible and also to consider the characteristics of the specific variant if you do end up scoring it more than once; the concern being that you could over-score if the variant is simply a "common" (and possibly benign) variant in a particular population.

**Example:** The curation of the [FGFR3 gene for achondroplasia \(autosomal dominant\)](#) is a relevant example in which the p.Gly380Arg variant is found in a large proportion of affected individuals and on multiple haplotypes, with some of the occurrences *de novo*. This reflects the need to up-score a variant that is a recurrent *de novo* variant, or is seen in individuals from different populations (and so may have arisen more than once). These situations support the pathogenicity of the variant and make it permissible to score more than one patient with the same variant. It may be a good idea to consult the expert panel to weigh in on this question with regard to the specific variant and the evidence available.

**Question #26:** When performing a gene curation with an autosomal recessive mode of inheritance, under what circumstances should the two null variants from a homozygous proband be scored as 3 points as opposed to 2 points?

**Answer:** Default scoring for this scenario is 2 points, in the absence of evidence that the two alleles arose independently. If haplotype analysis establishes that the maternal and paternal variants were not each inherited as part of a larger region from a shared common ancestor, or if parents are of different ethnic origins from one another, scoring for the proband may be increased to 3 points. The underlying logic is that it would be double-counting for the same variant to receive the maximum score (1.5 points) twice when the proband's parents each inherited it from a common ancestor.

**Example:** In the curation of [RLBP1 for inherited retinal dystrophy \(autosomal recessive\)](#), the null (frameshift) variants from a homozygous proband described in the supplementary materials of [PMID: 25356976](#) were scored at 1 point each (2 points total for the proband). This was motivated by the lack of detail about the haplotype, parental origins, or other evidence that the two alleles might have arisen independently. In a more clear-cut case where both parents of a homozygous proband originate from the same very small village, that would be a clear rationale for scoring 2 points total instead of 3 points total.

**Question #27:** What is the appropriate score for a nonsense or frameshift variant predicted to undergo nonsense-mediated decay when performing a gene curation with an autosomal dominant mode of inheritance predominantly associated with gain-of-function rather than loss-of-function variants? Is 1.5 points no longer the default score for this variant type?

**Answer:** Correct, 1.5 points is only the default score for a predicted null variant when the mechanism of disease is known to be caused by loss-of-function variants. In the context of a disease known to be associated with gain-of-function, discuss with the GCEP

whether reported putative loss of function variants have enough evidence to warrant scoring or should not be scored. If variant-level evidence is available showing that the functional impact of the variant is consistent with the known disease mechanism (i.e. gain-of-function), up-scoring can then be performed.

**Example:** In the curation of the [\*BEST1\* gene for \*BEST1\*-related dominant retinopathy](#), several affected cases were found harboring a heterozygous nonsense or frameshift variant predicted to undergo nonsense-mediated decay. These were not scored as part of the curation, as the disease mechanism is known to be mediated by missense variant proteins that exert dominant negative effects on wild-type *BEST1* proteins. The cases harboring an apparent heterozygous null variant were thought to be cases of recessive disease (a separate curation) in which a second *BEST1* loss-of-function variant had been missed, or cases unrelated to *BEST1* in which an alternative cause in another gene had been missed.

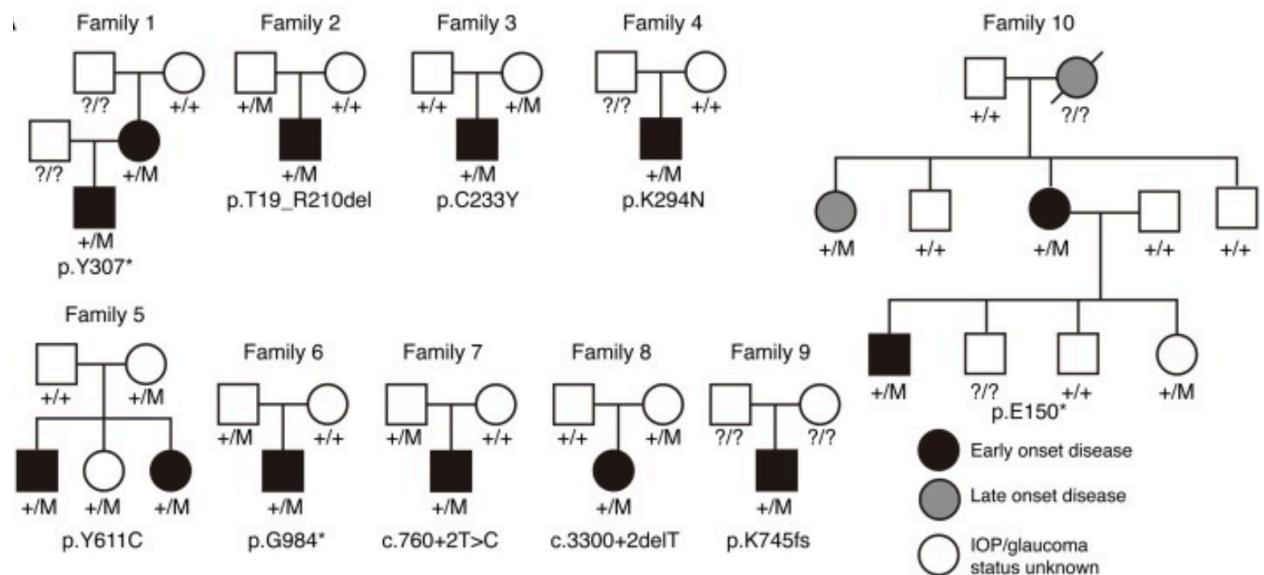
### Scoring an Incompletely Penetrant Variant

**Question #28:** For disease states known to have incomplete penetrance, are there adjustments made to the typical procedure for scoring genetic evidence?

**Answer:** Yes. When a disease is associated with incompletely penetrant variants, one of the adjustments is generally to avoid probands known to harbor a modifier variant in a different gene that is known to increase the penetrance of the disease. These probands can be avoided by searching the literature for modifier genes or cases of digenic inheritance, then prioritizing the scoring of probands who have been confirmed by genotyping to have no known modifier variants in these additional loci. A second adjustment for incomplete penetrance is to set lower expectations for disease-causing variants to be rare in gnomAD. Incompletely penetrant variants may occur in gnomAD at higher frequencies despite being disease-causing, or may have some occurrences in

gnomAD in the homozygous state despite being associated with autosomal recessive disease. While these gnomAD findings should trigger down-scoring for a variant expected to have complete penetrance, down-scoring may not be recommended for an incompletely penetrant variant. Finally, family segregation evidence for an incompletely penetrant condition can be counted even when genotype-positive family members lack the phenotype. These are not considered inconsistent segregations, but rather are excluded from the calculation of the estimated LOD score.

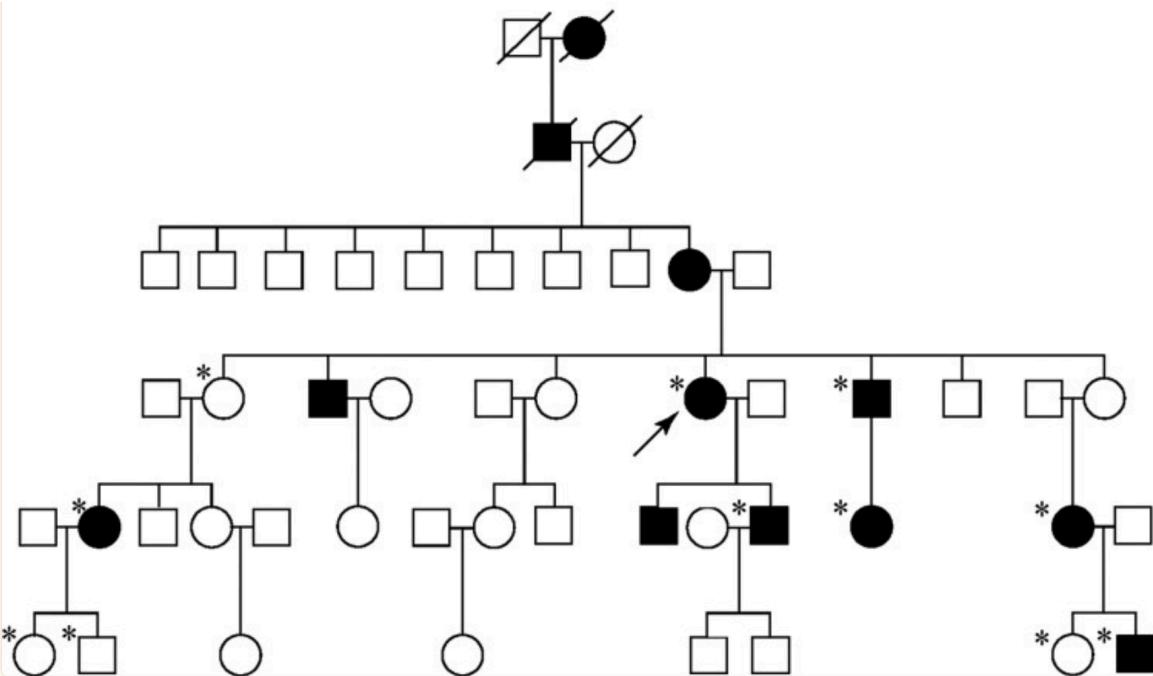
**Example:** In the curation of the [TEK gene for primary congenital glaucoma-3E](#), the disease is associated with early onset disease in some family members and late onset in others, indicating variable expressivity and incomplete penetrance (please see figure below, adapted from PMID: 27270174). A literature search revealed that penetrance of *TEK* variants is influenced in some cases by the presence of modifier alleles in the related genes such as *CYP1B1* (PMID: 28620713), *SVEP1* (PMID: 33027505), and *LTBP2* (PMID: 34956319), consistent with the possibility of digenic inheritance. As a result, the curation avoided such digenic cases and prioritized genetic evidence from recently published cases in which exome sequencing-based genotyping had ruled out modifier variants in these loci.



**Question #29:** For X-linked dominant conditions in which heterozygous females exhibit milder disease and/or incomplete penetrance, can female probands be curated for genetic evidence?

**Answer:** Yes, females can be curated for X-linked gene-disease relationships. Curators should keep in mind that the process of X-chromosome inactivation in the disease-relevant tissue(s) may vary widely from proband to proband, often resulting in a broad spectrum of presentation and severity of features in females. Although the pattern of X-chromosome inactivation doesn't necessarily need to be evaluated in order for a female proband to be counted for genetic evidence, the GCEP may be more interested or less interested to score the proband based on the presentation of phenotypes. It may be helpful for the curator to consult the GCEP members for scoring guidance if the female proband's presentation is less severe, less detailed, or less disease-specific.

**Example:** For the family pedigree below, which comes from PMID: 23372056, four females have been shaded black to indicate affected status and have been marked with asterisks to indicate a positive genotype for a variant in *RPGR*. These females can be included in the counting of segregation evidence and the calculation of the estimated LOD score. The three females who are genotype positive but unaffected (marked with white circles) should be omitted from the segregation evidence calculations. The female proband marked with the arrow may be suitable for inclusion in the genetic evidence as a proband as well, depending on the amount of phenotypic data available and the degree to which the presentation matches the phenotypes expected for the disorder.



### Scoring a Case-Control Study

**Question #30:** When is it appropriate to score a case-control study, and how many points should it be worth?

**Answer:** A case-control study should be scored if the phenotype is a good match for the disease of interest, and if there is a statistical comparison presented in cases with variants in the gene of interest vs controls (non-diseased individuals).

**Examples:** During the curation of [HTT for Huntington disease](#), the GCEP considered a case-control study linking *HTT* variants to higher incidence of major depressive disorder in cases vs. controls (PMID: 20360314). However, since major depressive disorder and psychological disorders in general are only a minor feature of Huntington disease, the GCEP decided to include this study under the Genetic Evidence category but to score it 0 points. During the [DHDSS gene curation for DHDSS-CDG](#), the GCEP considered a case-control study linking a particular *DHDSS* variant to autosomal-recessive retinitis

pigmentosa, a central feature of the disease entity (PMID: 21295282). The 132 affected retinitis pigmentosa cases included 15 individuals homozygous for the variant, while the 322 unaffected controls included 0 homozygous individuals. This disparity was considered highly statistically significant, and was scored at 3 points after a discussion with the GCEP. Please note that any patients included in the case-control study were not separately scored for genetic evidence as probands.

### **Experimental Evidence:**

#### **Scoring Expression Data**

**Question #31:** Does the GCI permit multiple pieces of expression data to be scored for a single gene-disease curation?

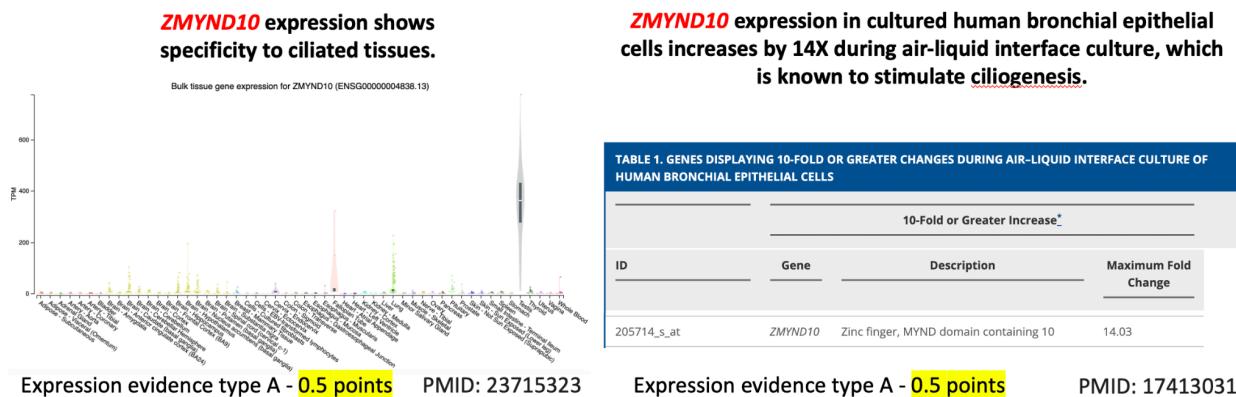
**Answer:** Yes, for example, expression evidence from Type A (Gene normally expressed in tissue relevant to the disease) and Type B (Altered expression in patients) could be scored for the same curation. Type A and Type B expression evidence offer different information about the gene-disease relationship, so scoring them both is not double-counting.

**Example:** The [curation of TIMP3 for Sorsby fundus dystrophy \(autosomal dominant\)](#) had scoring for both expression evidence Type A ([PMID:9006347](#), Figure 2B) and Type B ([PMID:9924344](#), Figure 2).

**Question #32:** Is it possible to score multiple pieces of functional evidence of the same type (for example, Expression Evidence Type A) despite their similarity to one another?

**Answer:** On one hand, similarities between two pieces of evidence indicate reproducibility by multiple sources. On the other hand, it is important to avoid over-scoring a single evidence type. It is common to curate two similar pieces of evidence together, as one strong piece of evidence, and to up-score accordingly.

**Example:** Multiple pieces of expression Type A evidence have been scored in the [ZMYND10 gene curation for primary ciliary dyskinesia 22](#) (PMID: 23715323, PMID: 17413031). The rationale was that one piece of evidence established a pattern in adult tissues that was consistent with a role in multi-ciliated cells (right, from PMID: 23715323), while the other established an onset of expression at a time point in differentiating cells that was consistent with the formation of cilia (left, from PMID: 17413031).



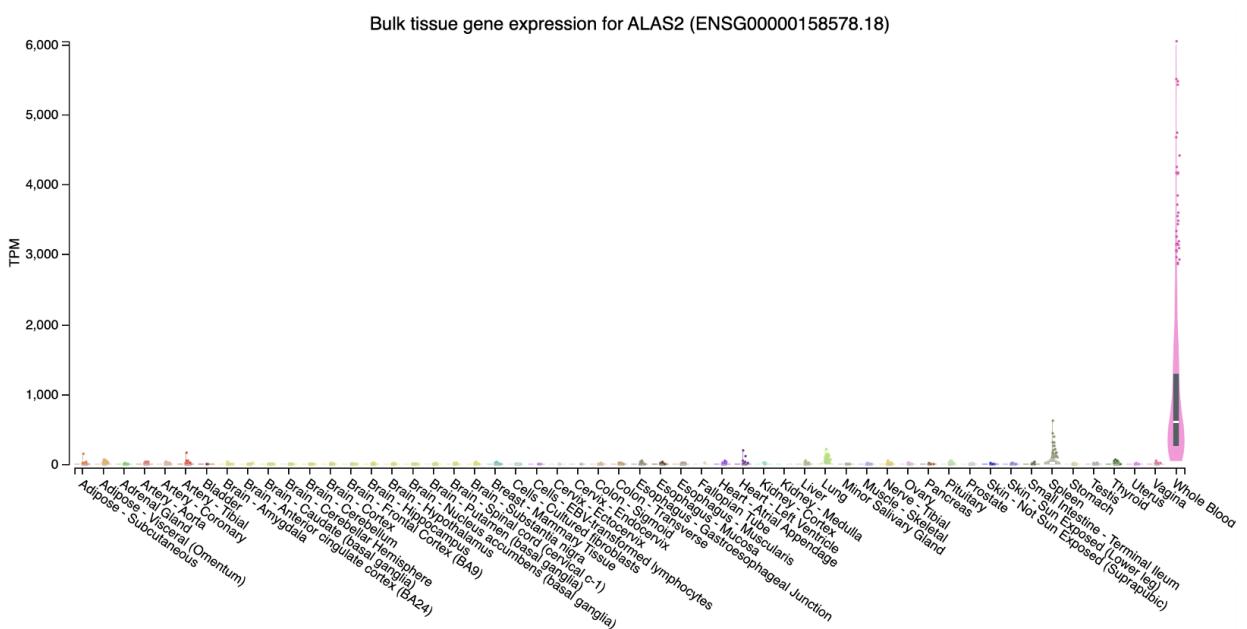
**Question #33:** In what scenarios does expression evidence merit up-scoring past default levels?

**Answer:** This is generally a decision made by the GCEP experts as a group. However, one rationale that has often motivated up-scoring by the GCEP is restriction of expression to only the tissue affected by the disease entity. Alternatively, evidence of relatively high

expression within the affected tissue (for example from the GTEx database) is usually scored at the default level.

**Example:** The curation of [FBN1 for Marfan syndrome \(autosomal dominant\)](#) assigned 0.5 points to expression evidence type A confirming the presence of FBN1 protein by immunohistochemistry in tissues relevant to this connective tissue disorder. The curation of [ALAS2 for X-linked erythropoietic protoporphyria](#) assigned 1 point to expression evidence type A showing the restricted expression of ALAS2 RNA almost exclusively in blood cells, consistent with this disorder of heme biosynthesis associated with erythroid cells. This example is summarized in the figure below (from PMID: 23715323).

**ALAS2 expression is highly specific to erythroid cells.**  
Expression Data – Default Score 0.5 – Tentative Score 1



**Question #34:** Are there scenarios in which Expression Evidence Type B (expression altered in patients) should not be scored but should instead be treated as variant-level evidence (used for up-scoring variants)?

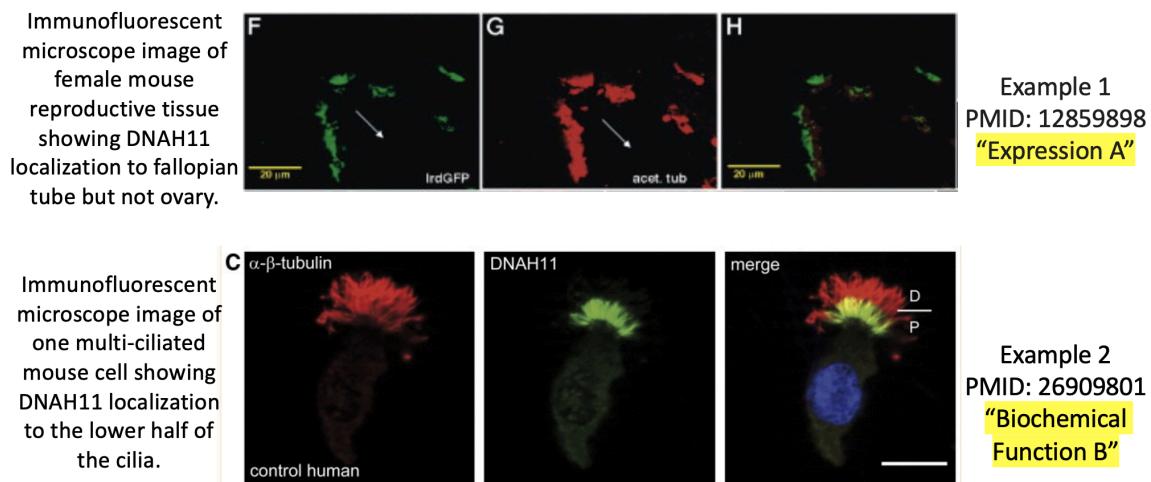
**Answer:** Yes, this is particularly true when the evidence is limited to a change in expression level alone, and is not clearly linked to the pathogenic mechanism. Type B evidence is scored only in particular scenarios, for example when there's a cohort study of individuals with a shared phenotype but unknown molecular background. In that situation, it's informative if all individuals with a variant in that locus show an expression effect in their patient tissue samples, regardless of the particular variant. On the other hand, many other scenarios that sound compatible with Expression Type B should instead be considered information about the variant only / primarily, without much insight into the disease itself.

**Example:** The curation of *SMAD3* for aneurysm-osteoarthritis syndrome (autosomal dominant) scored Expression Evidence Type B for the finding of increased *SMAD3* expression in aortic tissue from multiple unrelated patients harboring *SMAD3* variants (PMID: 21217753).

**Question #35:** Does experimental evidence that the protein product localizes to a certain place belong in Expression A or Biochemical Function B category?

**Answer:** If the localization is at the tissue level and establishes the presence of the protein product within a specific tissue or cell type relevant to the disease, Expression A might be the best fit. On the other hand, if the localization is at the subcellular level and establishes the presence of the protein at a specific part of the cell known to have a function relevant to the disease (such as within the cilia or mitochondria), Biochemical Function B might be the better fit.

**Example:** In the three top panels below (from PMID: 12859898), the localization of the DNAH11 protein (green) to tissues known to contain multiciliated cells (red) supports its association to primary ciliary dyskinesia as a form of Expression A evidence. In the three bottom panels below (from PMID: 26909801), the subcellular localization of the DNAH11 protein (green) to the cell's cilia (red) supports its association to primary ciliary dyskinesia as a form of Biochemical Function B evidence.



**Question #36:** Should experimental evidence linking a variant in the gene of interest to a disease-related phenotype be used in the Experimental Evidence section or as variant-level functional evidence to support the scoring of a proband's variant(s)?

**Answer:** If the experiment shows a defect of the variant at the RNA level or protein level in cells from the proband, it should be used as variant-level functional evidence to justify up-scoring of the variant. This is because it is part of the overall picture of that particular proband, and cannot be separated from the rest of that individual's genetic background. On the other hand, there are a few situations where studies of a variant may be used for the Experimental Evidence category instead, such as when a large number of different disease-associated variants are expressed in a non-patient cell type, triggering a defect

relevant to the disease. This evidence could either be used to justify up-scoring of the variant in a proband, or could be scored as Experimental Evidence under the category of Functional Alteration in Non-Patient Cells.

**Example:** If a missense variant shows instability / degradation in the proband's cells, or mislocalizes to an incorrect part of the cell, this can be listed as variant-level functional evidence and can justify up-scoring of the missense variant from 0.1 points (default) to 0.5 points (or higher, depending on the GCEP and the disease mechanism). Evidence in the proband's cells showing the variant to be absent from the protein level or to undergo nonsense-mediated decay at the RNA level can similarly justify up-scoring of the variant in the proband. On the other hand, if a variant-level experiment is performed in non-patient cells and has not been used to up-score genetic evidence from a proband, it may be used for Experimental Evidence instead. For example, in the curation of [SLC6A6 for hypotaurinemic retinal degeneration and cardiomyopathy](#), several missense variants expressed in oocytes showed significant taurine uptake impairment, consistent with the mechanism of disease, and was scored as evidence of Functional Alteration in Non-Patient Cells (PMID:24790004).

## Scoring Protein Interaction Data

**Question #37:** When the protein product from the gene-of-interest physically interacts with another protein, can this be scored as Experimental Evidence under the Protein Interaction category?

**Answer:** Physical interaction with another protein is generally scored under the Protein Interaction category only if the second protein is encoded by a gene involved in the same disease or a very similar disease to the gene of interest.

**Example:** In the curation of the *BBS7* gene for Bardet-Biedl syndrome 7, an immunoprecipitation experiment showing the physical interaction of the *BBS7* protein with the *BBS4* protein, which is similarly encoded by a gene associated with the very similar Bardet-Biedl syndrome 4. The evidence of their interaction was counted for 0.5 points under the Protein Interaction category. Ideally, the interacting protein should be encoded by a gene that has been previously curated by the GCEP. If not, the GCEP may be consulted to approve scoring of this evidence. On the other hand, when the interacting protein has not been associated with any disease, or has only been associated with a condition unrelated to the disease of interest, it may be appropriate not to score this protein interaction evidence.

### Scoring Functional Alteration Data

**Question #38:** When a variant is exogenously expressed in a cell line and the protein product shows a functional difference from the wild-type control (i.e. change in molecular weight, subcellular localization, or enzymatic activity), should this be counted as evidence of Functional Alteration in non-patient cells?

**Answer:** If the functional alteration recapitulates a feature of the disease, this experiment would be counted as Functional Alteration in non-patient cells. The [Gene Curation Standard Operating Procedure](#) says that this category of evidence would only be the appropriate choice if the cultured cells “have a phenotype that is consistent with the human disease process”. On the other hand, if there is no direct link to the disease, this experiment should be used instead as a rationale to up-score the variant itself. (Note: The experiment should not be double-counted as both Functional Alteration evidence and as a rationale for up-scoring the variant.)

**Example:** An experiment that shows mislocalization of a missense variant protein should only be counted as Functional Alteration if protein mislocalization is a known

mechanism underlying the human disease state. On the other hand, if protein mislocalization is not a known underlying cause of disease, then the experiment should be used instead as a rationale to up-score the missense variant from 0.1 point to 0.5 points. Ideally, the non-patient cell type itself should also match the disease context, as illustrated by an example from the evaluation of an [\*FHL1\* variant for Emery-Dreifuss Muscular Dystrophy \(X-linked recessive\)](#). Mouse cardiomyocytes were induced to express wild-type and variant FHL1, which revealed mislocalization of the variant and perturbed localization of other proteins known to be aberrantly localized in hypertrophic cardiomyopathy (HCM). Because HCM is a key phenotype of the patients, the experiment was scored as Functional Alteration in non-patient cells.

### Scoring a Model Organism

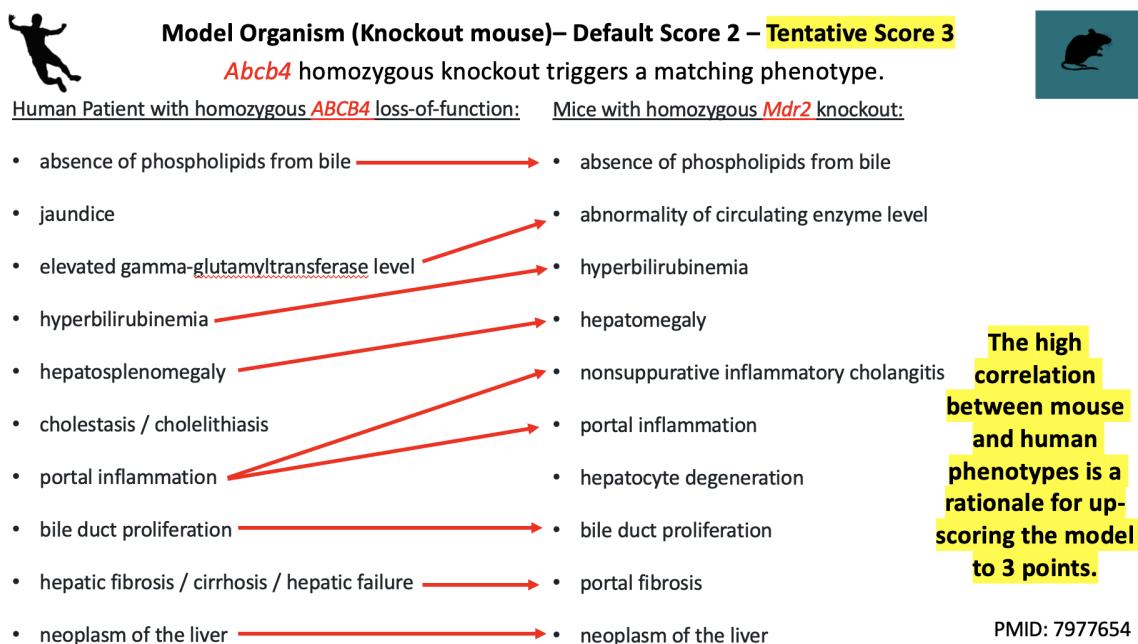
**Question #39:** How many shared features with the human disease state are necessary to receive maximum scoring for a mouse model (4 points)?

**Answer:** Typically this requires a knock-in of a human variant into the endogenous mouse locus and near-total recapitulation of the various human phenotypes (at both the organ / organismal level and the molecular / cellular level). One caveat is that the mode of inheritance must match, so that a dominant disorder must have a monoallelic mouse model.

**Question #40:** How many shared features with the human disease state are necessary for a mouse model to receive up-scoring (~3 points)?

**Answer:** Recapitulation of most human phenotypes both at the organ / organismal level and at the molecular / cellular level, with consistent mode of inheritance.

**Example:** A mouse model that recapitulates most phenotypes of a monoallelic / dominant human condition can up-scored to 3 points, provided that the phenotypes occur only in the homozygous animals and not in the heterozygotes. Please see the summary figure below from the [ABCB4 gene curation for progressive familial intrahepatic cholestasis type 3 \(PMID:7977654\)](#).

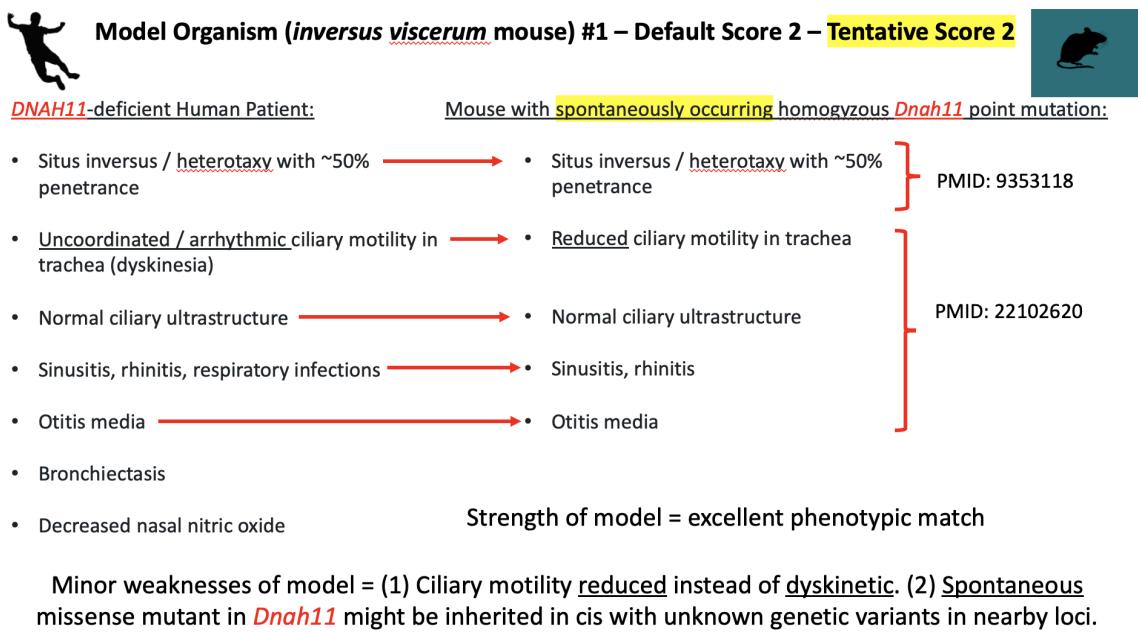


**Question #41:** How many shared features with the human disease state are necessary to receive default scoring for a mouse model (2 points)?

**Answer:** Typically it is sufficient for the mouse model to recapitulate a subset of the major phenotypes from the human patients (some at the organ / organismal level and some at the molecular / cellular level), with consistent mode of inheritance (i.e. autosomal recessive) and mechanism (i.e. loss of function). There is a caveat that some human phenotypes can't be recapitulated in mice, and vice versa. One example is that

motile ciliopathies in mice can recapitulate human upper but not lower respiratory infections. Each model should be evaluated only for the phenotypes that can reasonably be recapitulated across the two species.

**Example:** Please see the summary figure below from the [DNAH11 gene curation for primary ciliary dyskinesia 7](#) (PMID: 9353118, PMID: 22102620).



**Question #42:** How many shared features with the human disease state are necessary to receive reduced scoring for a mouse model (1 point)?

**Answer:** Typically it is sufficient for the mouse model to match the human patients only at the molecular or cellular level but not in phenotypes at the organ / organismal level, with consistent mode of inheritance (i.e. autosomal recessive) and mechanism (i.e. loss of function). There is a caveat that some human phenotypes can't be recapitulated in mice, and vice versa. One example is that motile ciliopathies in mice can recapitulate

human upper but not lower respiratory infections. Each model should be evaluated only for the phenotypes that can reasonably be recapitulated across the two species.

**Example:** In the curation of [CCDC40 for primary ciliary dyskinesia 15 \(autosomal recessive\)](#), the mouse model ([PMID: 21131974](#)) matched the human patients in ciliary ultrastructure (the cellular phenotype), loss-of-function mechanism, and autosomal recessive mode of inheritance, but only weakly matched a single organ-level phenotype. The model received 1 point of scoring (please see summary figure below). A second example from the [RDH5 curation for inherited retinal dystrophy \(autosomal recessive\)](#) is a knockout mouse that was down-scored due to the milder phenotype in the mice than in the human patients. Yet this model was kept at a score of at least 1 point due to the explanation that other *RDH* paralogs such as *RDH11* may compensate for the loss of *RDH5* due to an overlapping role in the mouse tissue of interest ([PMID: 10825191](#)).



#### Model Organism (mouse) #1 – Default Score 2 – Tentative Score 1



##### *CCDC40*-deficient Human Patient:

- Situs inversus / heterotaxy with ~50% penetrance → • Situs inversus / heterotaxy with ~39% penetrance
- Abnormal axonemal structure (affecting inner dynein arms, central pair, peripheral MT doublet, radial spokes, nexins) → • Shorter cilia in nodal pit of embryo
- Rigid, fast-flicking, low-amplitude ciliary motility
- Sinusitis / Rhinorrhea
- Chronic cough
- Pneumonia / Bronchitis
- Bronchiectasis
- Otitis media
- Immotile sperm / female infertility

##### *Inks* mouse harboring p.Ser792Ter nonsense mutation in *Ccdc40*:

Strength = recapitulates some phenotypes of human patients, as well as mode of inheritance (autosomal recessive) and variant type (loss-of-function).

Weakness = generated by mutagenesis, not targeted disruption of *Ccdc40*. Upper respiratory phenotypes were not characterized.

PMID: 21131974

**Question #43:** What are some other factors or scenarios that can trigger down-scoring of a mouse model?

**Answer:** One rationale for down-scoring an animal model is a design that makes the animal less relevant to the disease state of interest, (but more relevant to some other disease instead). A second is tissue-specific disruption or over-expression of the gene of interest in a cell type that is less relevant to the disease of interest can be a scenario that fits this pattern. A third trigger for down-scoring is mismatch in mode of inheritance.

**Example:** Examples can include incomplete phenotypic characterization of a mouse model in a way that does not comment on the presence or absence of key phenotypes present in the human patients of interest, but rather focuses on other areas of the body. In addition, down-scoring can occur when patient phenotypes from an autosomal dominant disorder are only recapitulated by a biallelic mouse model. The [NF2 gene curation for neurofibromatosis type 2 \(autosomal dominant\)](#) includes a mouse model with tissue-specific disruption of the gene of interest in a cell type that was designed to circumvent embryonic lethality but does not permit full recapitulation of the disease (PMID: 9553042), resulting in only 1.5 points of scoring.

**Question #44:** How many points are appropriate for a unicellular model organism that is capable of recapitulating only the cellular-level features but not the organismal-level features of the human disorder?

**Answer:** Some GCEPs working on diseases with a combination of cellular-level and organ-level features have decided that a unicellular model organism deserves a lower default score for modeling only the cellular-level features of the human patients. Successful recapitulation of the cellular-level defects is often rewarded with 1 point in these GCEPs. When the curator is unsure of the precedent / policy for their specific GCEP, it may be valuable to ask the GCEP members for scoring advice in this scenario.

**Examples:** The Motile Ciliopathy GCEP has frequently scored *Chlamydomonas* models (unicellular algae) at 1 point, as they generally recapitulate cellular defects of human

motile ciliopathy patients, but cannot model the respiratory issues or other organ-level phenotypes. Please note that scoring may be adjusted up or down depending on how well or poorly the model matches the human disease at the cellular level. Other model organisms may be inappropriate for a particular disease entity, for example, zebrafish are considered poor models for dilated cardiomyopathy due to the tendency for any morpholino-treated zebrafish to exhibit this phenotype, regardless of the gene targeted. It may be helpful to have a discussion with the GCEP members about whether or not the model organism is an appropriate model for the disease.



### ***Chlamydomonas* – Default Score 2 – Tentative Score 0.5**



Human patients with biallelic *IFT140* loss-of-function:      Mutant *Chlamydomonas* with disruption of *IFT140* ortholog:

- Retinal dystrophy / Reduced visual acuity
- Short stature / bell-shaped thorax / short ribs
- Respiratory insufficiency
- Skeletal abnormalities
- Kidney disease
- Absence of cilia from fibroblasts

- Short flagella
- Slower flagellar regeneration / assembly
- Partner proteins mislocalized

PMID: 28207750

**Strength = recapitulates human patients' cellular level defect**

**Weakness = unable to model organ-level phenotypes**

## Scoring a Rescue Experiment

**Question #45:** When scoring a rescue experiment, what are some common criteria for up-scoring from the default score?

**Answer:** Use of CRISPR technology to repair the variant can be a design feature that leads to up-scoring of the rescue experiment. However, a known concern is that CRISPR has off-target effects. Thus, up-scoring can be reasonably performed if the authors of the paper have checked for / ruled out off-target CRISPR effects. Knock-in to the mouse locus is another great rescue design that should be up-scored since it reduces the risk of

over-expression (off-target) effects. Human enzyme replacement (with the wild-type gene product) is another design that has historically received scores as high as 4. In addition to the design, the other important caveat is the degree to which rescue is successful. It may be a good idea to consult the expert panel to weigh in on this question with regard to the specific evidence available.

**Example:** In the curation of [GAA for Pompe disease \(autosomal recessive\)](#), 3 points were assigned to the clinical trial on enzyme replacement therapy ([PMID: 16860134](#)), reporting improvement of clinical symptoms in humans infused regularly with the gene product.