Genetic Evidence Summary Matrix

A matrix used to categorize and quantify the genetic evidence curated for a gene-disease relationship is provided below.

	Evidence Type	Case Information Type (Suggested Starting Score)		Suggested Upgrades		Scoring Range	Points Given	Max Score
				Functional Data	*De Novo*			
Case-Level Data [1]	Variant Evidence*[2]	Predicted or proven null variant (1.5 points)[3]		+0.5 points	+0.5 points	0-3 points (per variant)		12 points
		Other variant type (0.1 points)**[4]		+0.4 points	+0.4 points	0-1.5 points (per variant)		
	Segregation Evidence	Evidence of segregation in one or more families		Sequencing Method		0-3 points		3 points
			Total LOD Score[5]	Candidate Gene Sequencing	Exome/Genome or all genes sequenced in linkage region			
			2-2.99	0.5 points	1 point			
			3-4.99	1 point	2 points			
			≥5	1.5 points	3 points			
Case-Control Data [6]	**Case-Control Study Type [7]**	**Case-Control Quality Criteria [8]**		**Suggested Points/Study**			**Points Given**	**Max Score**
	Single Variant Analysis [7a]	• Variant Detection Methodology [8a] • Power [8b] • Bias and Confounding Factors [8c] • Statistical Significance [8d]		0-6 points				12 points
	Aggregate Variant Analysis [7b]			0-6 points				
Total Allowable Points for Genetic Evidence								**12 points**

*In the case of AR conditions, evaluate each variant (in trans) independently, then combine for the final score.
**For 2 missense variants without functional data, score at 0.25. For all other scenarios, round to the nearest 0.5 point.

Case-Level Data

1. With the updates to the SOP, some important changes have been made to the genetic evidence scoring system. Baseline points have been assigned per variant type, guidance is available on when to add points to the baseline for common upgrades, and lastly, there are no category maximums. Each case may be given points for A) the variant observed and B) segregation evidence, if applicable.

2. Variant evidence is considered based on properties of the variant in the context of the mode of inheritance for the disorder being evaluated. For a variant to be considered potentially disease-causing, its frequency in the general population should be consistent with phenotype frequency, inheritance pattern, disease penetrance, and disease mechanism, if it is known. It is ideal that variants have some evidence suggesting disease causality.
3. Variants can be scored at the following increments: 0, 0.1, 0.25, 0.5, and 1. Predicted/proven null variants in general should start at 1.5 points. Variants may be upgraded based on the quality of evidence, consistency and/or specificity of the phenotype, and/or the likelihood of a putative null variant actually leads to loss of function. An upgrade may also be considered if a variant is de novo and has supportive functional information. A variant may be downgraded based on the lack in quality of evidence, a nonspecific and/or genetically heterogeneous phenotype, insufficient prior testing to rule out other potential causes of disease, etc.
4. "Other" variant types, missense for example, should start at 0.1 points. Points may be upgraded by the default point amounts listed in the table if A) functional data is available to support the variant's effect on the underlying protein and/or B) the variant is de novo. Evaluate each observed variant in trans independently, then sum for the final score when evaluating biallelic variants in the context of autosomal recessive disease. Always round to the nearest 0.5 points, except when there are biallelic missense variants without supportive functional data.
5. Segregation evidence is categorized by LOD scores. LOD scores provided by authors can be used however, if a LOD score is not provided, one can be estimated across multiple families by counting the number of times a variant segregates with affected individuals in a family, excluding the proband.
    - For dominant or X-linked disorders, the estimated LOD score should be calculated using only families with 4 or more segregations present.
    - For recessive disorders, the estimated LOD score should be calculated using only families with at least 3 affected individuals.
    - More points should be awarded if the variants were identified using genome-wide approaches, such as whole exome or whole genome sequencing.
    - Fewer points should be awarded if variants were identified using candidate-gene sequencing.
    - Expert reviewers may choose to specify the most appropriate way to approach segregation scoring within their disease domain, including enacting more formal, rigorous LOD score calculations.

Case-Control Data

6. Each case-control study should be independently assessed to evaluate the quality of the study design. Consensus with a clinical domain expert group is highly recommended.

7. Case-control studies are classified based on how variation in cases and controls is evaluated: single variant analysis or aggregate variant analysis. Studies presenting both types of analyses may be counted in either category at the discretion of the curator/expert, but the same variants should not be counted in both categories.
    a. Single variant analysis studies are those in which individual variants are evaluated for statistical enrichment in cases compared to controls. More than one variant may be analyzed, but the variants should be independently assessed with appropriate statistical correction for multiple testing.
    b. Aggregate variant analysis studies are those in which the statistical enrichment of two or more variants as an aggregate is assessed in cases compared to controls. This comparison could be accomplished by genotyping specific variants or by sequencing the entire gene.
8. Points for case-control studies may be assigned at the discretion of expert opinion based on the overall quality of each study. The following should be considered when evaluating case-control study quality:
    a. *Variant Detection Methodology*: Cases and controls should ideally be analyzed using methods with equivalent analytical performance (e.g. equivalent genotype methods, sufficient and equivalent depth and quality of sequencing coverage, correction for batch effects).
    b. *Power*: The study should analyze a sufficient number of cases and controls given the prevalence of the disease, the allele frequency, and the expected effect size in question to provide appropriate statistical power to detect an association.
    c. *Bias and Confounding Factors*: The manner in which cases and controls were selected for participation and the degree of case-control matching may impact the outcome of the study. The following are some factors that should be considered:
        - Are there systematic differences between individuals selected for study and individuals not selected for study?
        - Are the cases and controls matched by demographic information (e.g., age, ethnicity, location of recruitment, etc.)?
        - Are the cases and controls matched for genetic ancestry, if not did investigators account for genetic ancestry in the analysis?
        - Have the cases and controls been equivocally evaluated for presence or absence of a phenotype, and/or family history of disease?
    d. *Statistical Significance*: The level of statistical significance should be weighed carefully. When an odds ratio is presented, its magnitude should be consistent with a monogenic disease etiology. When p-values or 95% confidence intervals (CI) are presented, the strength of the statistical association can be weighed in the final points assigned. Factors, such as multiple testing, that might impact that interpretation of uncorrected p-values and Cis should be considered when assigning points.