

# The ClinGen framework for defining the clinical validity of monogenic disease associations for use in research and clinical analyses

J.S. Berg, O. Birsoy, A.J. Butte, M. Giovanni, K. Goddard, A. Hamosh, L. Milko, M.F. Murray, E. Riggs, P.N. Robinson, W.S. Rubinstein, A. Santani, A. Scott, T. Sneddon, N. Strande, H.L. Rehm, S.E. Plon, C.L. Martin; ClinGen Consortium

## Abstract

The rapid pace at which new gene-disease associations are reported for monogenic disorders poses a tremendous challenge to the clinical interpretation of genome-scale sequencing data in patients and research participants. The NIH-funded ClinGen consortium is working with the NCBI to generate a publicly available resource for delineation of clinically relevant genes and variants. As part of this effort, a workgroup was assembled to develop standardized procedures for curating genes and their relationship to monogenic disorders, in order to provide well curated, up-to-date information for use in research and clinical analyses. The workgroup first addressed the problem of categorizing the clinical validity of gene-disease associations according to strength of evidence.

The workgroup established seven categorical designations for strength of evidence for a causal role in disease: Definitive, Strong, Moderate, Limited, No Reported Evidence, Disputed, and Evidence Against. These categories are defined according to the type and strength of evidence provided in the published literature or other public sources, the presence or absence of contradictory evidence, and whether the initial report of disease-gene causality has been replicated.

This framework is initially being applied to genes relevant to three clinical domains: hereditary cancer, cardiology, and inborn errors of metabolism. We are actively curating target genes within these areas and evaluating results of the categorization, consistency between reviewers and frequency of category change. Our overall goal is to provide the categorization and supporting evidence for every gene implicated in monogenic disorders, and to develop a standardized system for community curation in order to support updated gene-disease relationships and associated evidence as new information is published. This framework will be useful to researchers investigating human genetic disorders, so that they can apply consistent criteria when claiming evidence of disease association, and to highlight those claims of disease association that require more data. Categorization of the clinical validity of gene-disease associations will also be of substantial clinical benefit given the increasing use of large multi-gene panels and genome-scale analysis in diagnostic genetic testing.

## Categories used to define clinical validity evidence level

<b>DEFINITIVE</b>	<p>The role of this gene in this particular disease has been repeatedly demonstrated in both the research and clinical diagnostic settings, and has been upheld over time (in general, at least 3 years). No valid evidence has emerged that contradicts the role of the gene in the specified disease.</p> <p>Selected examples: CFTR (Cystic fibrosis); NF1 (Neurofibromatosis, type I); HBB (Sickle cell anemia); FGFR3 (Achondroplasia); PKD1 (Polycystic kidney disease, adult type); MSH2 (Colon cancer and Lynch syndrome); MECP2 (Rett syndrome)</p>
<b>STRONG</b>	<p>There is strong evidence, typically by at least two independent studies, to support a causal role for this gene in this disease, such as:</p> <ul style="list-style-type: none"> <li>Strong statistical evidence demonstrating an excess of pathogenic variants in affected individuals as compared to appropriately matched controls</li> <li>Multiple pathogenic variants within the gene in unrelated probands with several different types of supporting experimental data. The number and type of evidence might vary (eg. fewer variants with stronger supporting data, or more variants with less supporting data)</li> </ul> <p>In addition, no valid evidence has emerged that contradicts the role of the gene in the noted disease.</p> <p>Selected examples: NGLY1 (Congenital disorder of glycosylation); PALB2 (breast cancer and pancreatic cancer), SDHC (Hereditary Pheochromocytoma and Paraganglioma syndrome); SMAD3 (Loeys-Dietz syndrome, type 3)</p>
<b>MODERATE</b>	<p>There is moderate evidence to support a causal role for this gene in this disease, such as:</p> <ul style="list-style-type: none"> <li>At least 3 unrelated probands with pathogenic variants within the gene with some supporting experimental data.</li> </ul> <p>The role of this gene in this particular disease may not have been independently reported, but no valid evidence has emerged that contradicts the role of the gene in the noted disease.</p> <p>Selected examples: RAD51D (Ovarian cancer); RAD51C (Breast and Ovarian cancer); SLX4 (Fanconi anemia); WRAP53 (Dyskeratosis congenita); DYRK1B – Abdominal Obesity-Metabolic Syndrome 3; ASXL3 - Bohring–Opitz-like disease</p>
<b>LIMITED</b>	<p>There is limited evidence to support a causal role for this gene in this disease, such as:</p> <ul style="list-style-type: none"> <li>Fewer than three observations of a pathogenic variant within the gene</li> <li>Multiple variants reported in unrelated probands but without sufficient evidence for pathogenicity</li> </ul> <p>Selected examples: RAD51C (Fanconi anemia); CEACAM16 (Deafness, AD); AGRN (Congenital myasthenic syndrome); GPHN (Hyperekplexia); APBB1 (Alzheimer disease ); ERBB3 (Schizophrenia); CHD11 (Congenital anomalies of the kidney)</p>
<b>NO EVIDENCE</b>	No evidence reported for a causal role in disease.
<b>DISPUTED</b>	<p>Valid evidence refuting a role for this gene in this disease is equivalent to or stronger than existing evidence supporting this role.</p> <p>Selected examples: RAD51D (Breast cancer); TMPO (Dilated cardiomyopathy); CTFI (Dilated cardiomyopathy)</p>
<b>EVIDENCE AGAINST</b>	<p>Evidence refuting the role of the gene in the specified disease has been reported and significantly outweighs any evidence supporting the role.</p> <p>Selected examples: AGTR2 (X-linked MR 88); UPK3A (Renal adysplasia); CDKAL1 (Noninsulin-dependent diabetes)</p>

## Notes regarding evidence types and examples

In this framework, the term “pathogenic variants” refers to variants that meet the ACMG 2014 revised criteria for pathogenicity in the context of the disorder in question.

Examples of types of supporting experimental data (modified from MacArthur et al. 2014):

- Gene burden:** the affected gene shows statistical excess of rare (or de novo) probably damaging variants segregating in cases compared to control cohorts or null models.
- Experimental Protein interactions:** the gene product interacts with proteins previously implicated (genetically or biochemically) in the disease of interest.
- Biochemical function:** the gene product performs a biochemical function shared with other known genes in the disease of interest, or consistent with the phenotype.
- Expression:** the gene is expressed in tissues relevant to the disease of interest and/or is altered in expression in patients who have the disease.
- Gene disruption:** the gene and/or gene product function is demonstrably altered in patients carrying candidate mutations.
- Model systems:** non-human animal or cell-culture models with a similarly disrupted copy of the affected gene show a phenotype consistent with human disease state.
- Rescue:** the cellular phenotype in patient-derived cells or engineered equivalents can be rescued by addition of the wild-type gene product

The selected examples given here represent conclusions based on available evidence at a particular point in time and should be considered provisional. Since the clinical validity classification for a gene-disease pair may change over time as new evidence becomes available, all categorizations made by this group will be dated/versioned.

## Utilization for different clinical applications

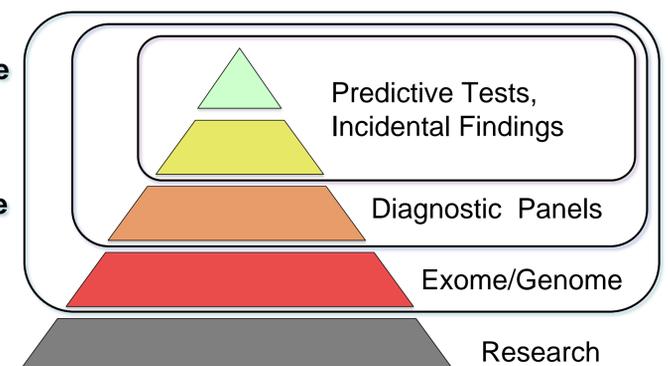
**Definitive evidence**

**Strong evidence**

**Moderate evidence**

**Limited evidence**

**Disputed / no evidence**



This framework could be used to choose which genes to analyze based on the clinical validity threshold appropriate for a given application.

## Funding Sources & Disclosures



National Human Genome Research Institute

U41 HG006834, U01 HG007437, U01 HG007436

The following relationship(s) exist related to this presentation:  
No relationships to disclose.