ClinGen
Clinical Genome Resource

Assessing Gene-Disease Validity

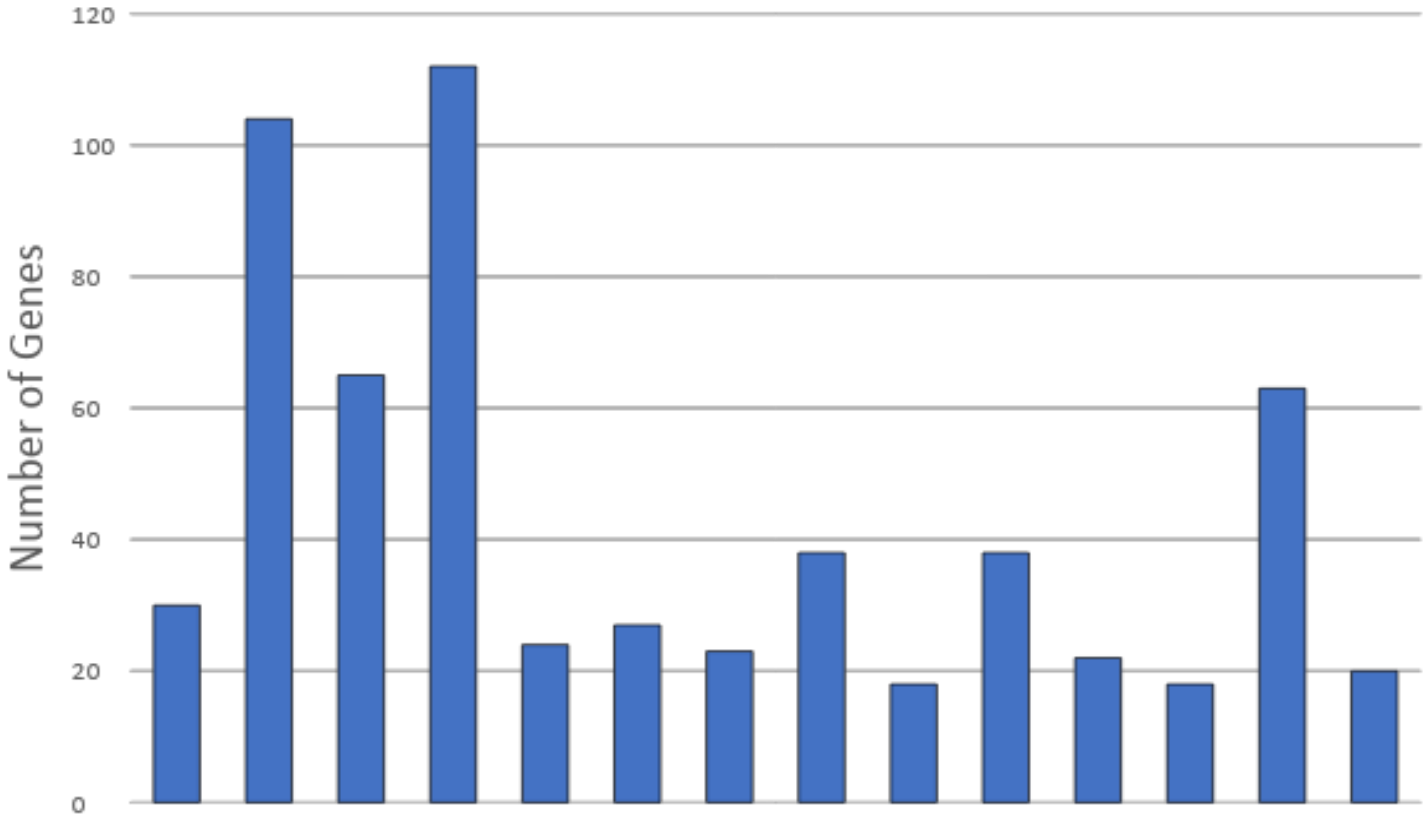
Erin Rooney Riggs, MS, CGC
Geisinger
eriggs@geisinger.edu

Clinical validity…

- *…of a laboratory test:*
  - Involves a test's ability to "consistently and accurately detect or predict the outcome of interest"*
  - In terms of genetic tests, requires correctly identifying the causative variant within the *appropriate gene*
- *…of a gene-disease pair:*
  - How strong is the evidence that variation in that gene causes the disease in question?

*Haddow, J., Palomacki, G. ACCE: A Model Process for Evaluating Data on Emerging Genetic Tests. in *Human Genome Epidemiology: A Scientific Foundation for Using Genetic Information to Improve Health and Prevent Disease* (ed. Khoury, M., Little, J., Burke, W.) 217-233 (Oxford University Press, 2003).

Why is this important?



Hypertrophic cardiomyopathy tests from 14 different US clinical labs

Standards and consensus are needed to determine which genes are valid disease genes ready for clinical testing.

Slide courtesy of Marina DiStefano, PhD

Circulation

Reappraisal of Reported Genes for Sudden Arrhythmic Death

Evidence-Based Evaluation of Gene Validity for Brugada Syndrome

S. Mohsen Hosseini, Raymond Kim, Sharmila Udupa, Gregory Costain, Rebekah Jobling, Eriskay Liston, Seema M. Jamal
Marta Szybowska, Chantal F. Morel, Sarah Bowdin, John Garcia, Melanie Care, Amy C. Sturm, Valeria Novelli, Michael J. Ackerman,
James S. Ware, Ray E. Hershberger, Arthur A.M. Wilde, Michael H. Gollob
and On behalf of the National Institutes of Health Clinical Genome Resource Consortium

Originally published 29 Jun 2018 | Circulation. 2018;138:1195–1205

Results:

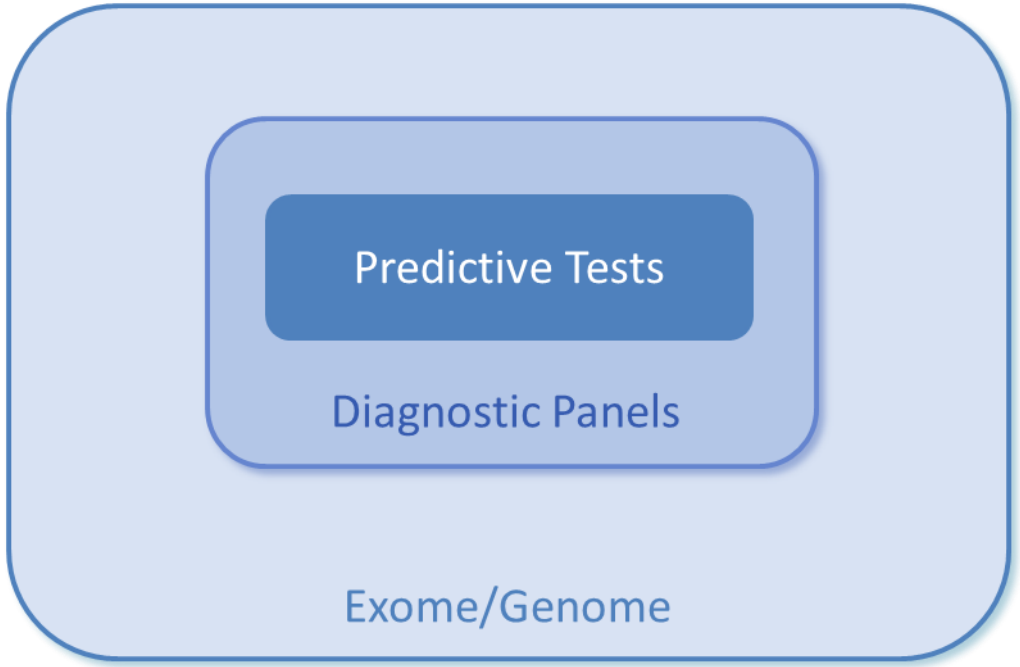
Of 21 genes curated for clinical validity, biocurators classified only 1 gene (*SCN5A*) as definitive evidence, whereas all other genes were classified as limited evidence. After comprehensive review by the clinical domain Expert panel, all 20 genes classified as limited evidence were reclassified as disputed with regard to any assertions of disease causality for BrS.

Conclusions:

Our results contest the clinical validity of all but 1 gene clinically tested and reported to be associated with BrS. These findings warrant a systematic, evidence-based evaluation for reported gene-disease associations before use in patient care.

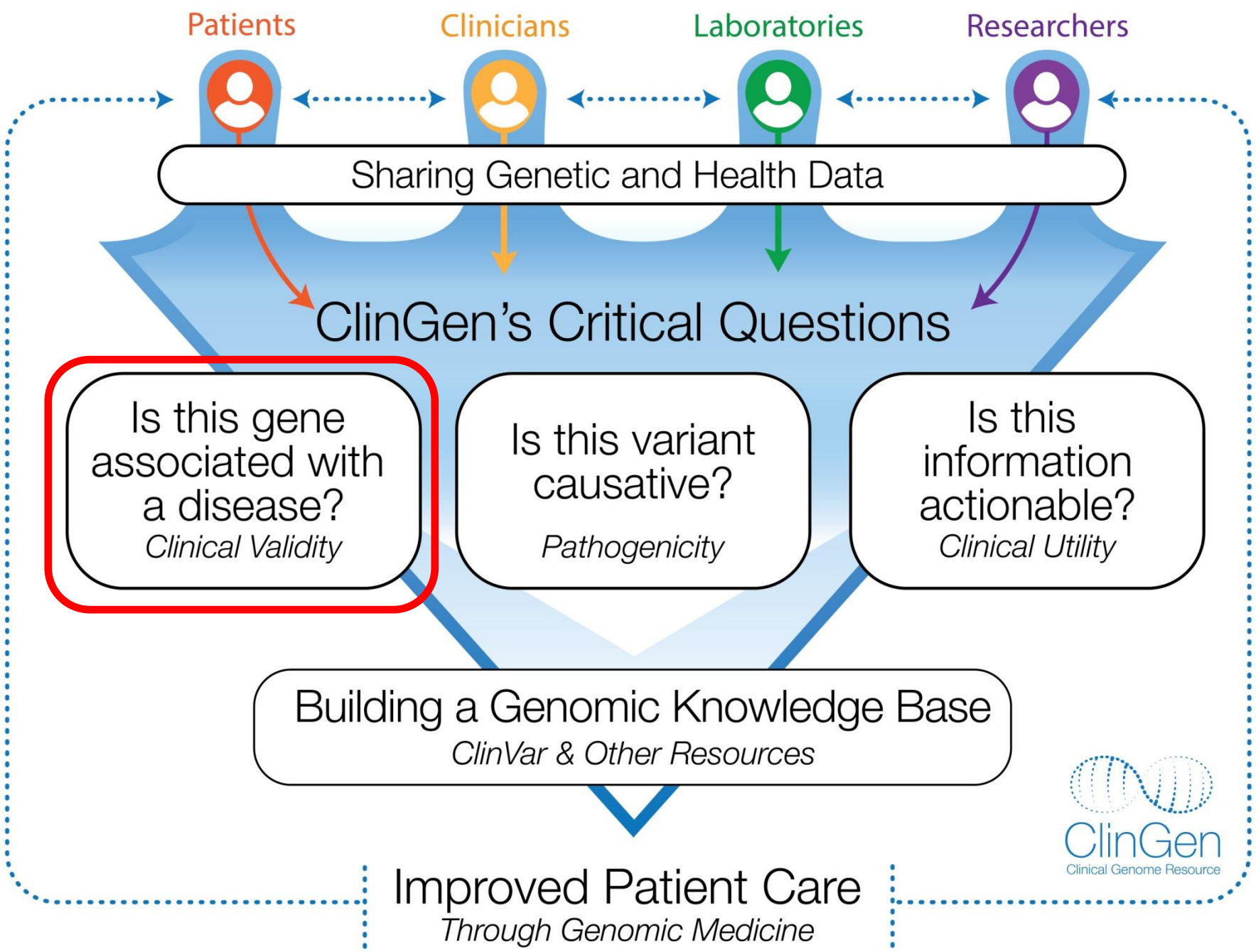
Why is this important?

- Clinical testing issues:
  - ACMG/AMP sequence variant interpretation guidelines were developed for use in the context of "recognized genotype-phenotype associations."
    - Using these guidelines for genes of uncertain significance is not appropriate
  - Given this, should genes with little evidence supporting their role in disease be included in multi-gene testing panels? Should variants in these genes be returned on clinical ES/GS?
- Research issues:
  - What evidence is necessary to solidify the role of a "candidate gene" in disease?



Different levels of evidence are needed for different clinical uses.

Graphic courtesy of Marina DiStefano, PhD


Patients
Clinicians
Laboratories
Researchers
Sharing Genetic and Health Data
ClinGen's Critical Questions
Is this gene associated with a disease?
Clinical Validity
Is this variant causative?
Pathogenicity
Is this information actionable?
Clinical Utility
Building a Genomic Knowledge Base
ClinVar & Other Resources
ClinGen
Clinical Genome Resource
Improved Patient Care
Through Genomic Medicine

Systematic Method to Evaluate Clinical Validity of Gene-Disease Pairs

- Defines the criteria needed to assess gene-disease validity
- Describes the strength evidence supporting a gene-disease association in a semi-quantitative manner
- Allows users to methodically classify the validity of a given gene-disease pair
- NOT intended to replace well-established statistical thresholds used for GWAS, or validated methods to define multifactorial disease risk

Definitive	Role has been repeatedly demonstrated in research & clinical diagnostic settings • Upheld over time (in general, at least 3 years) • No convincing contradictory evidence
Strong	≥2 independent studies with: • Multiple pathogenic variants in unrelated probands • AND • Several different types of supporting experimental data • OR • Excess of pathogenic variants in cases vs. controls • No convincing contradictory evidence
Moderate	≥1 independent study with: • Several unrelated probands with pathogenic variants • Some supporting experimental data • No convincing contradictory evidence
Limited	≥1 independent study with: • <3 unrelated probands with pathogenic variants • OR • Multiple variants reported in unrelated probands but *without* sufficient evidence for pathogenicity • No convincing contradictory evidence
No Known Disease Relationship	No evidence reported for a causal role in disease (candidate genes, etc.), therefore no pathogenic variants have been identified in humans to date.
Conflicting Evidence Reported	
Disputed	Convincing evidence disputing a role for this gene in this disease has arisen • Disputing evidence need not outweigh existing evidence supporting the gene/disease relationship
Refuted	Evidence refuting the role of the gene in the specified disease has been reported and significantly outweighs any evidence supporting the role • Applied at the discretion of clinical domain experts after thorough review of available evidence

Assertion criteria	Genetic Evidence (0-12 points)	Experimental Evidence (0-6 points)	Total Points (0-18)	Replication Over Time (Y/N)
Description	Case-level, family segregation, or case-control data that support the gene-disease association	Gene-level experimental evidence that support the gene-disease association	Sum of Genetic & Experimental Evidence	> 2 pubs w/ convincing evidence over time (>3 yrs)
Assigned Points				

CALCULATED CLASSIFICATION	**LIMITED**	**1-6**
	MODERATE	**7-11**
	STRONG	**12-18**
	DEFINITIVE	**12-18 AND replication over time**

Valid contradictory evidence? (Y/N)	List PMIDs and describe evidence:
CURATOR CLASSIFICATION	
FINAL CLASSIFICATION	

Step 1: Selecting the Gene

- ClinGen's gene curation efforts are focused around several different clinical domains
- Each domain has one or more gene curation expert panels (GCEPs)
  - Currently ~30 ClinGen GCEPs
- Each GCEP is comprised of:
  - Clinicians
  - Clinical laboratory staff
  - Researchers

  Expertise in the disease area of interest
  - Biocurators → expertise in the ClinGen curation process
- Typically focus on genes included on clinical testing panels
  - Scope ranges from 10s to 100s of genes!

Gene Curation Expert Panels

Aminoacidopathy
Arrhythmogenic Right Ventricular Cardiomyopathy
Brain Malformations
Breast/Ovarian Cancer
Brugada Syndrome
Charcot-Marie-Tooth Disease
Colon Cancer
Congenital Myopathies
Craniofacial Malformations
Dilated Cardiomyopathy
Epilepsy
Familial Thoracic Aortic Aneurysm and Dissection
Fatty Acid Oxidation Disorders
Hearing Loss
Hemostasis/Thrombosis
Hereditary Cancer
Humoral Defects
Hypertrophic Cardiomyopathy
Intellectual Disability and Autism
Kidney Cystic and Ciliopathy Disorders
Limb-Girdle Muscular Dystrophy
Long QT Syndrome
Mitochondrial Diseases
Monogenic Diabetes
Parkinson's Disease
Peroxisomal Disorders
Pulmonary Hypertension
RASopathy
Retina
SCID-CID
Skeletal Disorders
Syndromic Disorders

Step 2: Selecting the Disease

* 300188

MEDIATOR COMPLEX SUBUNIT 12; MED12

Alternative titles; symbols

MEDIATOR OF RNA POLYMERASE II TRANSCRIPTION, SUBUNIT 12, S. CEREVISIAE, HOMOLOG OF
TRINUCLEOTIDE REPEAT-CONTAINING GENE 11; TNRC11
THYROID HORMONE RECEPTOR-ASSOCIATED PROTEIN, 230-KD SUBUNIT; TRAP230
HUMAN OPPOSITE PAIRED GENE; HOPA
KIAA0192

HGNC Approved Gene Symbol: MED12

Cytogenetic location: Xq13.1 *Genomic coordinates (GRCh38): X:71,118,595-71,142,449* (from NCBI)

Gene-Phenotype Relationships View clinical synopses as a table

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key
Xq13.1	Lujan-Fryns syndrome	309520	XLR	3
	Ohdo syndrome, X-linked	300895	XLR	3
	Opitz-Kaveggia syndrome	305450	XLR	3

- Not always straightforward
- Are these 3 different diseases that require 3 different curations?
- ClinGen has a detailed precuration process to guide these decisions focusing on:
  - Molecular mechanism
  - Inheritance Pattern
  - Phenotypic Variability
- In general, if multiple disease entities can't be distinguished on a molecular level, we "lump" into a single disease entity

To learn more, please attend the "Lumping and Splitting" workshop hosted by Courtney Thaxton on Friday!

Gene-Disease Clinical Validity Training Materials

Gene-Disease Validity | Training Materials | Documents | Interface | Browse Curations

Interested in Gene Curation? The following documents and presentations are available to help people learn and understand the Gene-Disease Validity curation process. In order to get involved with our Gene Curation activities, please fill out our volunteer survey. For questions about existing materials or requests for new materials, contact us at clingen@clinicalgenome.org.

Training Modules | Additional Supporting Materials

Training Modules

https://clinicalgenome.org/curation-activities/gene-disease-validity/training-materials/

1

Required

The Publication

Read the publication describing the ClinGen gene-disease validity evaluation process.

→ Evaluating the Clinical Validity of Gene-Disease Associations: An Evidence-Based Framework Developed by the Clinical Genome Resource.

Start →

2

Required

Standard Operating Procedures

Detailed documentation outlining the gene-disease validity process.

→ Gene-Disease Validity Standard Operating Procedures, Version 8

Start →

Step 3: Assessing the Evidence (Genetic)

	Evidence Type	Case Information Type (Suggested Starting Score)		Suggested Upgrades		Scoring Range	Points Given	Max Score
				Functional Data	*De Novo*			
Case-Level Data	Variant Evidence*	Predicted or proven null variant (1.5 points)		+0.5 points	+0.5 points	0-3 points (per variant)		12 points
		Other variant type (0.1 points)**		+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	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	Case-Control Study Type	Case-Control Quality Criteria		Suggested Points/Study			Points Given	Max Score
	Single Variant Analysis	• Variant Detection Methodology • Power • Bias and Confounding Factors • Statistical Significance		0-6 points				12 points
	Aggregate Variant Analysis			0-6 points				
						Total Allowable Points for Genetic Evidence		12 points

Modified from Strande et al. *Am J Hum Genet* 2017; Gene Curation SOP

Scoring Case-Level Data

- Each case may be given points for:
  - Variant evidence
  - Segregation evidence (if applicable)

Case-Level Variant Evidence

Evidence Type	Case Information Type (Suggested Starting Score)	Suggested Upgrades		Scoring Range	Points Given	Max Score
		Functional Data	*De Novo*			
Variant Evidence*	Predicted or proven null variant (1.5 points)	+0.5 points	+0.5 points	0-3 points (per variant)		12 points
	Other variant type (0.1 points)**	+0.4 points	+0.4 points	0-1.5 points (per variant)		

- Baseline number of points assigned per variant type
  - Predicted/proven null = 1.5 points
  - Other variant type (e.g., missense) = 0.1 points
- Add points to the baseline score for common upgrades as described in the table
  - Functional data
  - *De novo* status
  - Other upgrades/downgrades may be appropriate at the discretion of your GCEP

Case-Level Variant Evidence

Evidence Type	Case Information Type (Suggested Starting Score)	Suggested Upgrades		Scoring Range	Points Given	Max Score
		Functional Data	*De Novo*			
Variant Evidence*	Predicted or proven null variant (1.5 points)	+0.5 points	+0.5 points	0-3 points (per variant)		12 points
	Other variant type (0.1 points)**	+0.4 points	+0.4 points	0-1.5 points (per variant)		

- For AR diseases: assess each variant independently, then sum for final score
- Round up to the nearest 0.5
  - EXCEPTION: For 2 missense variants w/o functional data, round to 0.25
- No single proband may score more than 3 points

Caveats

- Default scores assume the variant type is consistent with the expected disease mechanism. If this is not the case, downgrade or do not score unless there is compelling rationale for partial scoring. Document this rationale in the GCI.
  - For example, if the disease mechanism is known to be gain of function, consider not scoring null variants and upgrading gain of function missense variants.
- Variants may be up- or downgraded beyond the values suggested here (but within the scoring range) based on the strength of evidence.
  - For example, a missense variant may score at the top of its range if robust functional assays demonstrate that the missense is acting in a manner consistent with the expected disease mechanism.

Caveats

- Variants may be up- or downgraded for other reasons beyond those listed in this chart at the discretion of the GCEP. Always document the rationale for up- or downgrading variants in the GCI.
  - For example, one may opt to upgrade missense variants if they are within a known functional domain, if they appear to be clustering in the same area of the gene, etc.
  - Consider upgrading based on consistency and/or specificity of the phenotype, the likelihood that a putative null variant actually leads to loss of function, etc.
- When assigning points for *de novo* status, consider further upgrades if statistical evidence shows that *de novo* variation in a particular gene is rare. Use caution (and consider not upgrading or not scoring) if a gene is known to have a high rate of *de novo* variation (e.g. *TTN*).

Case-Level Segregation Evidence

Evidence Type		Case Information Type				Suggested points/case		Points Given	Max Score
						Default	Range		
	Segregation Evidence	Evidence of segregation in one or more families[6]		Sequencing Method			0-3		3
			Total LOD Score	Candidate Gene Sequencing	Exome/ Genome or all genes sequenced in linkage region				
			2-2.99	0.5	1				
			3-4.99	1	2				
			≥5	1.5	3				

- Segregation evidence can be used to support a gene-disease relationship, though it must be considered carefully, as segregation implicates a particular locus, not necessarily a particular gene.
- Consider the methods used to identify candidate variants when assigning points
- Guidance for scoring (based on LOD scores) for reasonably penetrant Mendelian disorders is outlined in our SOP

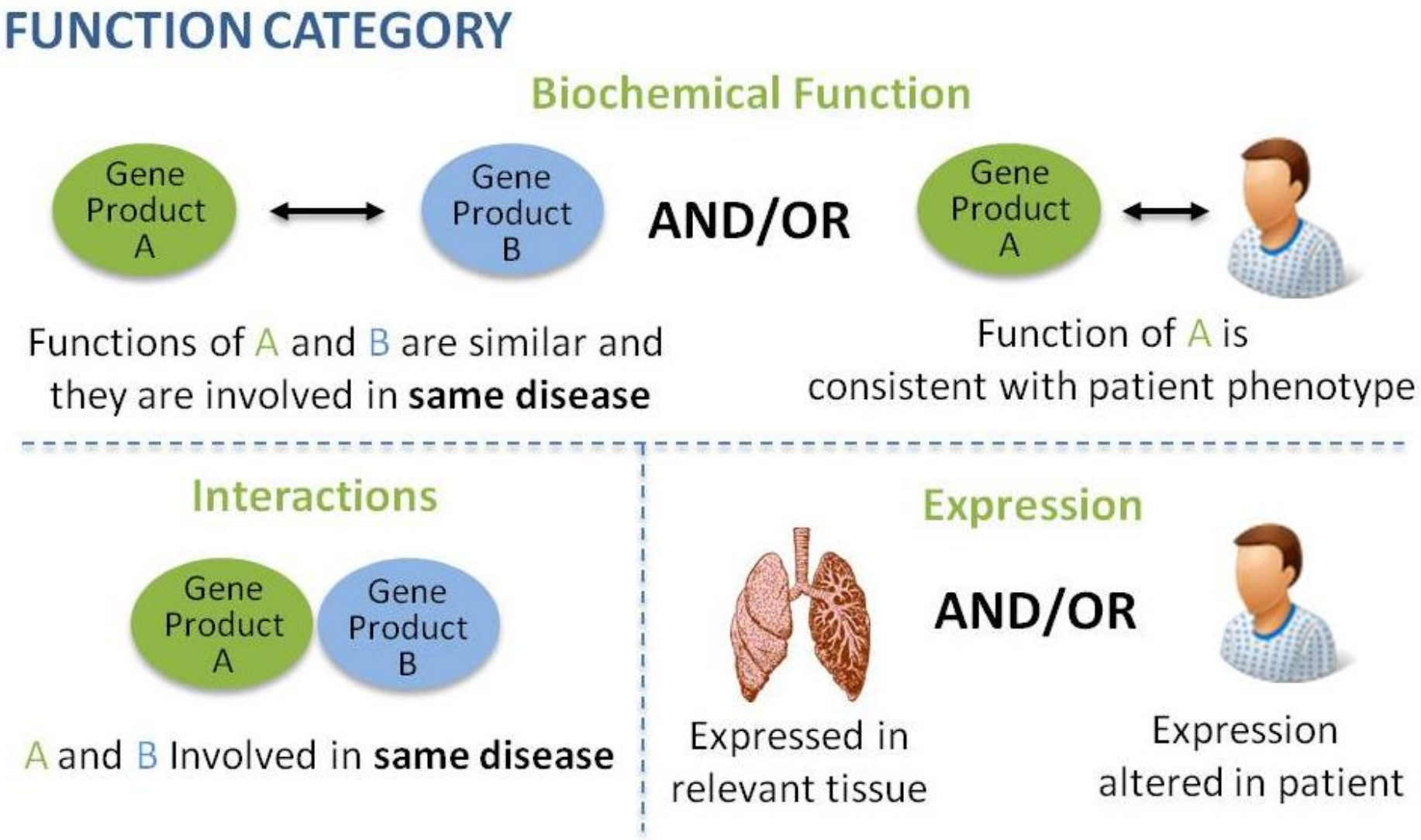
Strande et al. *Am J Hum Genet* 2017

Scoring Case-Control Data

Case-Control Study Type[7]	Case-Control Quality Criteria[8]	Suggested points/study	Max Score
Single Variant Analysis[7a]	• Variant Detection Methodology[8a] • Power[8b] • Bias and Confounding[8c] • Statistical Significance[8d]	0-6	12
Aggregate Variant Analysis[7b]		0-6	12

- Case-control studies are classified into two types:
  - Single variant analysis: studies 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.
  - Aggregate Variant Analysis: studies in which the statistical enrichment of two or more variants as an aggregate is assessed in cases compared to controls.
- 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.

Step 3: Assessing the Evidence (Experimental)



Consistent with MacArthur et al. Nature. 2014 Apr 24;508(7497):469-76

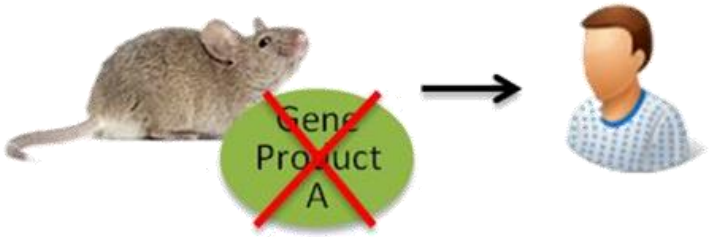
Experimental Evidence (cont'd)

FUNCTIONAL ALTERATION CATEGORY

Gene or gene product function is demonstrably altered in patients carrying candidate mutations

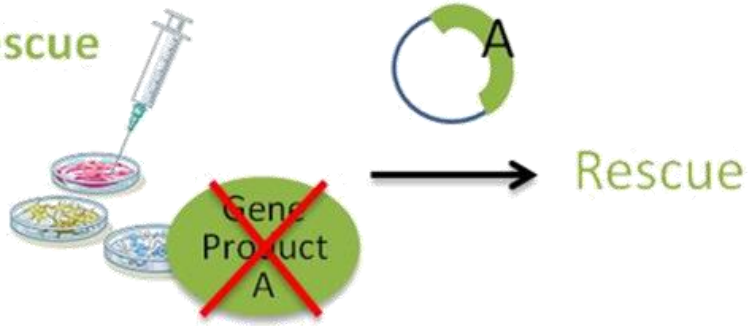
MODEL SYSTEMS & RESCUE CATEGORY

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

Consistent with MacArthur et al. Nature. 2014 Apr 24;508(7497):469-76

Experimental Evidence Scoring

Evidence Category	Evidence Type	Suggested Points/		Points Given	Max Score
		Default	Range		
Function	Biochemical Function	0.5	0-2		2
	Protein Interaction	0.5	0-2		
	Expression	0.5	0-2		
Functional Alteration	Patient cells	1	0-2		2
	Non-patient cells	0.5	0-1		
Models	Non-human model organism	2	0-4		4
	Cell culture model	1	0-2		
Rescue	Rescue in human	2	0-4		
	Rescue in non-human model organism	2	0-4		
	Rescue in cell culture model	1	0-2		
	Rescue in patient cells	1	0-2		
Total Allowable Points for Experimental Evidence					6

Strande et al. *Am J Hum Genet* 2017

Step 3: Assessing the Evidence (Contradictory)

- Not quantified in the summary matrix
- Manual review and expert input is needed to accurately assess this type of information in the context of the available supporting evidence
- No “score” will be generated in these situations
- Summary matrix can still be used to organize and display both types of evidence for further review.

Step 4: Assigning a Classification

Assertion criteria	Genetic Evidence (0-12 points)	Experimental Evidence (0-6 points)	Total Points (0-18)	Replication Over Time (Y/N)
Description	Case-level, family segregation, or case-control data that support the gene-disease association	Gene-level experimental evidence that support the gene-disease association	Sum of Genetic & Experimental Evidence	> 2 pubs w/ convincing evidence over time (>3 yrs)
Assigned Points				
CALCULATED CLASSIFICATION		LIMITED	1-6	
		MODERATE	7-11	
		STRONG	12-18	
		DEFINITIVE	12-18 & Replicated Over Time	
Valid contradictory evidence (Y/N)*	List PMIDs and describe evidence:			
CURATOR CLASSIFICATION				
FINAL CLASSIFICATION				

All curations publicly available at *www.clinicalgenome.org*

ClinGen
Clinical Genome Resource

Get Started | About Us | Curation Activities | Working Groups | Expert Panels | Documents & Annoucements | Tools

Gene | Start typing a gene symbol... | Search

All Curated Genes | Gene-Disease Validity | Dosage Sensitivity | Clinical Actionability | Curated Variants | More

Gene-Disease Validity

1219	968	35
Total Curations	Unique Genes	Expert Panels

Search in table

Showing 1 to 25 of 1219 rows 25 rows per page

‹ 1 2 3 4 5 ... 49 ›

Gene	Disease	MOI	Expert Panel	SOP	Classification	Last Eval.
A2ML1	Noonan syndrome with multiple lentigines	AD	RASopathy	SOP5	No Known Disease Relationship	06/07/2018
A2ML1	Noonan syndrome-like disorder with loose anagen hair	AD	RASopathy	SOP5	No Known Disease Relationship	06/07/2018
A2ML1	Noonan syndrome	AD	RASopathy	SOP5	Disputed	06/07/2018
A2ML1	Costello syndrome	AD	RASopathy	SOP5	No Known Disease Relationship	06/07/2018
A2ML1	cardiofaciocutaneous syndrome	AD	RASopathy	SOP5	No Known Disease Relationship	06/07/2018
AARS1	undetermined early-onset epileptic encephalopathy	AR	Epilepsy	SOP6	Limited	11/20/2018

Gene-Disease Validity Classification Summary

Return to Listing

CD3E - immunodeficiency 18

Gene:	CD3E (HGNC:1674)
Disease:	immunodeficiency 18 (MONDO:0014278)
Mode of Inheritance:	Autosomal recessive inheritance (HP:0000007)

Definitive

Classification - 01/26/2021

Replication over time:	YES	Contradictory Evidence:	NO
Expert Panel:	SCID-CID		

Evidence Summary: CD3E was first reported in relation to autosomal recessive immunodeficiency 18 in 1993 (Soudais et al., PMID: 8490660). CD3ε deficiency results in an absence of mature TCR αβ and γδ T cells in the periphery whereas natural killer and B lymphocytes were normally detected, causing T−, B+, NK+ SCID. At least 8 variants predicted to cause a loss of function of the protein have been reported, suggesting biallelic loss of function is the molecular mechanism. Evidence supporting the gene-disease relationship includes the identification of at least 8 probands with CD3E variants who have been reported in at least 8 publications (PMID: 8490660, 28916186, 15546002, 27807805, 28597365, 24515816, 31589898, 32016651). This gene-disease relationship is also supported by animal models (PMIDs: 9843989, 9885898, 7588594), rescue experiments (PMID: 9843989), and expression studies (PMIDs: 3012357, 2410254). More evidence is available in the literature, but the maximum score for genetic evidence has been reached. In summary, CD3E is definitively associated with autosomal recessive immunodeficiency 18. This has been repeatedly demonstrated in both the research and clinical diagnostic settings and has been upheld over time. This classification was approved by the ClinGen SCID/CID Working Group on 01/21/2021.
Gene Clinical Validity Standard Operating Procedures (SOP) - SOP8

Genetic Evidence			Evidence Type	Case Information Type	Guidelines Default	Guidelines Range	Guidelines Max	Guidelines Count	Points Total	Points Counted	PMIDs/Notes
	Case-Level Data	Variant Evidence	Autosomal Dominant or X-linked Disorder	Variant is de novo	2	0-3	12				
				Proband with predicted or proven null variant	1.5	0-2	10				
				Proband with other variant type with some evidence of gene impact	0.5	0-1.5	7				
			Autosomal Recessive Disease	Two variants in trans and at least one de novo or a predicted/proven null variant	2	0-3	12	8	16	12	Soudais C et al. 1993 Jan (PMID:8490660); de Saint Basile G et al. 2004 Nov (PMID:15546002); Fuehrer M et al. 2014 May (PMID:24515816); Engelhardt KR et al. 2017 Jan (PMID:27807805); Firtina S et al. 2017 Oct (PMID:28597365); Abolhassani H et al. 2018 Apr (PMID:28916186); Shahbazi Z et al. 2019 Dec (PMID:31589898); Erman B et al. 2020 Apr (PMID:32016651);
				Two variants (not predicted/proven null) with some evidence of gene impact in trans	1	0-1.5					
		Segregation Evidence				Summed LOD		Family Count			
				Candidate gene sequencing							
				Exome/genome or all genes sequenced in linkage region							
				Total Summed LOD Score							

Case-Control Data	Case-Control Study Type	Case-Control Quality Criteria	Guidelines Points/Study	Guidelines Max	Guidelines Count	Points Points	Points Counted	PMIDs/Notes
	Single Variant Analysis	1. Variant Detection Methodology 2. Power 3. Bias and confounding 4. Statistical Significance	0-6	12				
	Aggregate Variant Analysis		0-6					
	Total Genetic Evidence Points (Maximum 12)						12	

	Evidence Category	Evidence Type	Guidelines Default	Guidelines Range	Guidelines Max	Count	Points Total	Points Counted	PMIDs/Notes
Experimental Evidence	Function	Biochemical Function	0.5	0 - 2	2			1	
		Protein Interaction	0.5	0 - 2					
		Expression	0.5	0 - 2		2	1		Gold DP et al. 1986 May 22-28 (PMID:3012357); Pessano S et al. 1985 Feb (PMID:2410254);
	Functional Alteration	Patient cells	1	0 - 2	2				
		Non-patient cells	0.5	0 - 1					
	Models	Non-human model organism	2	0 - 4	4	3	2	3	DeJarnette JB et al. 1998 Dec 8 (PMID:9843989); Wang N et al. 1998 Dec (PMID:9885898); Malissen M et al. 1995 Oct 2 (PMID:7588594);
		Cell culture model	1	0 - 2					
	Rescue	Rescue in human	2	0 - 4					
		Rescue in non-human model organism	2	0 - 4		1	1		DeJarnette JB et al. 1998 Dec 8 (PMID:9843989);
		Rescue in cell culture model	1	0 - 2					
		Rescue in patient cells	1	0 - 2					
	Total Experimental Evidence Points (Maximum 6)							4	

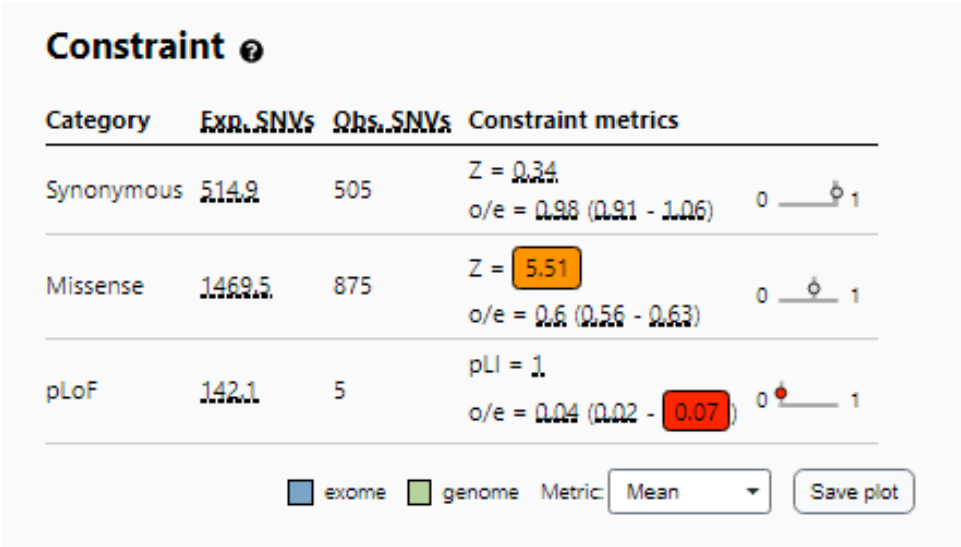
Assertion criteria	Genetic Evidence (0-12 points)	Experimental Evidence (0-6 points)	Total Points (0-18)	Replication Over Time (Y/N)
Description	Case-level, family segregation, or case-control data that support the gene-disease association	Gene-level experimental evidence that support the gene-disease association	Sum of Genetic & Experimental Evidence	> 2 pubs w/ convincing evidence over time (>3 yrs)
Assigned Points	12	4	16	YES
CALCULATED CLASSIFICATION		LIMITED	1-6	
		MODERATE	7-11	
		STRONG	12-18	
		DEFINITIVE	12-18 AND replication over time	
Valid contradictory evidence (Y/N)*	NO			
	CALCULATED CLASSIFICATION	Definitive		
	EXPERT CURATION (DATE)	**Definitive**	01/26/2021	

A Tale of 2 Genes: Examples from the ClinGen ID/Autism GCEP

- ClinGen ID/Autism GCEP began evaluating genes in 2017
- Initial focus: genes included on multi-gene testing panels marketed specifically for intellectual disability (ID) and/or autism
- Queried the Genetic Testing Registry (GTR) in 2017 and 2019 for the genes included on these panels
  - Excluded panels that were too broad (e.g. “Neurology Panel”)
  - Final list: 30 panels and 973 unique genes
- Continue to work through this list, but also evaluating “new” ID/Autism genes

Example 1: NBEA and Complex Neurodevelopmental Disorder

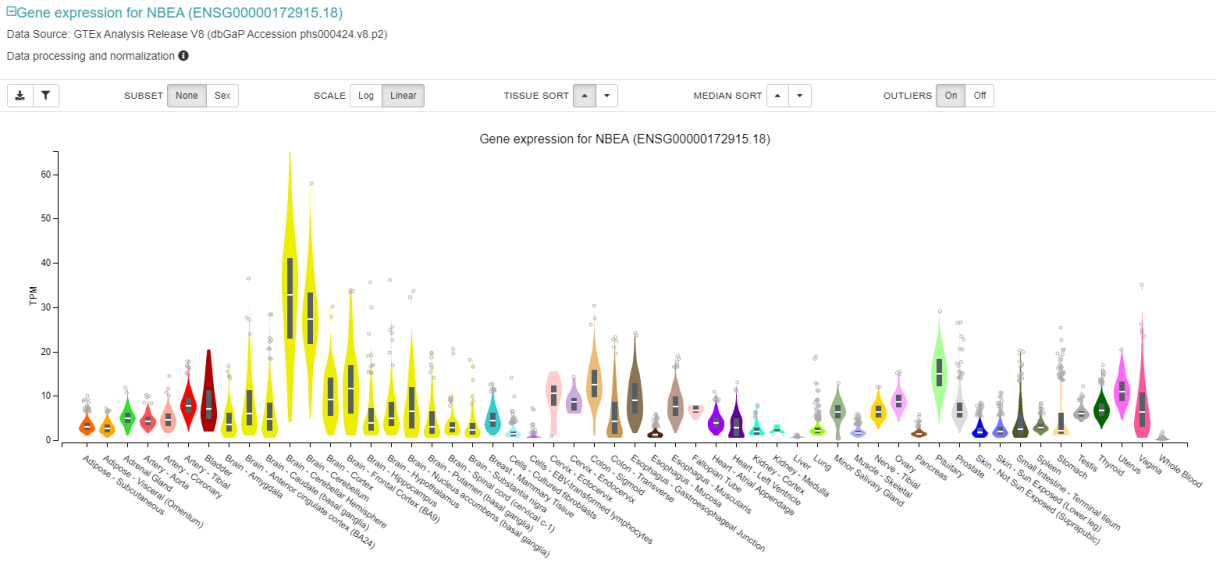
- Neurobeachin
- 13q13.3
- Brain-specific multi-domain scaffolding protein that plays a role in vesicle trafficking and dynamics
- Thought to regulate synaptic structure and function
- Initial evaluation by ClinGen: December 2020
  - At the time, no disease entries in OMIM, Orphanet, or MONDO
  - Not on any ID/Autism testing panels per 2019 pull of GTR data



Constraint

Category	Exp. SNVs	Obs. SNVs	Constraint metrics
Synonymous	514.9	505	Z = 0.34 o/e = 0.98 (0.91 - 1.06)
Missense	1469.5	875	Z = 5.51 o/e = 0.6 (0.56 - 0.63)
pLoF	142.1	5	pLI = 1 o/e = 0.04 (0.02 - 0.07)

exome genome Metric: Mean Save plot



Initial Suggestion of Disease Relationship: Castermans *et al.* 2003

- Describe a *de novo* balanced translocation t(5;13)(q12.1;q13.2) in a male with a diagnosis of autism and negative family history
  - Breakpoint at Chr13 located within a 2.8kb area in the intron between exons 2-3 in *NBEA*
  - Breakpoint at Chr5 within a 125 kb area of the centromere "devoid of known or predicted genes"
  - No functional data provided
- Other previous reports of microdeletions in 13q13.3 involving *NBEA* all included other genes; this was one of the first reports where *NBEA* alone was implicated
- Case ultimately not scored for curation purposes, but this report spurred further investigation of this gene as a potential cause for neurodevelopmental disorders (NDDs).

Total: 0 points

In the interim…

- Reports of variation in this gene in larger cohorts of NDDs
  - Iossifov *et al.* 2014 (Simons Simplex Collection - autism):
    - 1 *de novo* putative loss of function (LOF) variant (c.6829C>T, p.Arg2277Ter) (2 points)
  - De Rubeis *et al.* 2014 (Autism Sequencing Consortium):
    - 1 *de novo* missense variant (c.7381G>A, p.Val2461Met), no functional data provided (0.5 points)
  - Bowling *et al.* 2017 (ID/DD cohort):
    - 1 *de novo* putative LOF variant (c.6637C>T, p.Arg2213Ter) (2 points)
    - 1 *de novo* missense variant (c.2836C>T, p.His946Tyr), no functional data provided (0.5 points)
  - All variants absent from gnomAD
  - No detailed phenotypic information available

Total: 5 points

Seminal Paper: Mulhern *et al.* 2018

- Collection of 24 *de novo* NBEA variants in patients with NDDs identified in multiple laboratories (clinical and research) around the world, facilitated by GeneMatcher
- Phenotypic spectrum includes:
  - Developmental disability (100%)
  - Autistic features/autism (50%)
  - Epilepsy (62.5%) (not previously described!)
- Variant types included:
  - Nonsense (33%)
  - Frameshift (21%)
  - Missense (17%)
  - Intragenic deletion (21%)
  - Splice site (4%)
  - Large, multi-gene deletion (4%)

Genotype and phenotype details for individuals with *NBEA* variants

	Variant (NM_015678.4); variant type	Sex, Age (years)	Development	Autism	Seizure Types	Age at 1st seizure
1	c.1006C>T; p.Arg336*; NS	M, 6	No motor delay; mild speech delay	–	Myoclonic, Astatic, Absence, Myoclonic-Clonic-Tonic, GTC, Tonic, Focal unaware, SE	3 yr
2	c.6829C>T; p.Arg2277*; NS	F, 21	W=18 m.; NV	++	Tonic, GTC	3 yr
3	c.3994C>T; p.Pro1332Ser; MS	M, 4	W=30 m.; speaks some two word phrases	+	None	N/A
4	c.4484del; p.Asn1495Ilefs*17; FS	M, 19	W=2.5 yr; speaks words at 16 yr	++	unknown	3.5-4 yr
5	c.6313G>T; p.Glu2105*; NS	M, 18	W=15-17 m; slightly delayed speech; regression at 2 yr.	–	Myoclonic, Atonic, Atypical Absence, Clonic, GTC	2 yr
6	c.7294_7295dup; p.Glu2433Argfs*3; FS	F, 13	W=prior to 15 mo.; first word=prior to 2 yr.	–	none	N/A
7	c.7707+2T>C; SS	F, 18	W-unavailable, speech delay	–	Febrile, Absence	<1 yr
8	c.6868C>T; p.Gln2290*; NS	M, 3	W=19 m; speaks <10 words at 3 yr	+	none	N/A
9	c.7462G>T; p.Glu2488*; NS	F, 19	W=12 m; speech delay	–	none	N/A
10	c.7230del; p.Asp2411Ilefs*21; FS	M, 9	W=11 m; first word=16 m.	++	Myoclonic, Myoclonic-atonic, Atypical Absence, GTC	19 m
11	c.3183delA; p.Glu1062Argfs*8; FS	M, 19	Non-ambulatory; NV	+	Myoclonic, Myoclonic-atonic, GTC, Focal unaware, SE	19 m
12	c.1448C>T; p.Ala483Val; MS	F, 11	W=15 m; first word=30 m	–	Febrile, hemi-convulsive, GTC	1 yr
13	c.6637C>T; p.Arg2213*; NS	F, 23	normal until regression at 26 m.	–	Atonic, Tonic, GTC	26 m
14	c.2836C>T; p.His946Tyr; MS	F, 20	W=18 m; words only (no phrases) at 2 yr	–	Nocturnal frontal lobe seizures	2 yr
15	c.3832C>T; p.Arg1278*; NS	F, 5	W=1 yr.; severe speech delay	–	None	N/A
16	c.3362del; p.Asn1121Metfs*9; FS	F, 24	W=2 yr; first word=2 yr	++	Generalized	19 yr
17	c.8401G>A; p.Glu2801Lys; MS	M, 11	W=15 m; 10 words at 24 m	+	Myoclonic, Tonic, Focal, Focal to GTC, Febrile	14 m
18	c.4715C>A; p.Ser1572*; NS	F, 3	W=14 m; 10-15 words at 2 yr	+	None (paroxysmal spells not confirmed to be seizures)	N/A
19	chr13:33957317-36828237 ×1; MGD	M, 15	W=3.5 yr; first word=14 m, stagnation until 4 yr	–	Myoclonic, GTC, epileptic spasms/tonic	18 m
20	chr13:35590335-35940429 ×1; IGD	M, 16	W=15 m; first word=3 yr	–	None	N/A
21	chr13:35574513-36163037 ×1; IGD	M, 16	W=2 yr; NV at 11 yr	+	None	N/A
22	chr13:36038249-36141224 ×1; IGD	M, 11	W-unavailable, IQ=~60	+	Myoclonic, GTC, Tonic	8 m
23	chr13:35963197-36125577 ×1; IGD	M, 5	W=18 m; delayed speech	+	Possible Absences	4 yr
24	chr13:35700830-35887000 ×1; IGD	F, 9	W=17 m; first word=18–20 m	–	Non-convulsive SE, generalized	2 yr

From: Mulhern *et al.* 2018, PMID: 30269351

All absent from gnomAD

Total: >>12 points (max)

Supportive gene-level experimental data

- Nbea (+/-) mice exhibit ASD-like features, including changes in self-grooming behavior, social behaviors, conditioned fear responses, and spatial learning and memory (Nuytens *et al.* 2013).
- Nbea null mice demonstrated a complete block of evoked synaptic transmission at neuromuscular junctions (Su *et al.* 2004).
- Nbea is required for electrical synapse formation and function, and is required to maintain dendritic complexity in zebrafish. Null zebrafish have broad behavioral deficits (Miller *et al.* 2016).
- In Drosophila, adult LOF mutants exhibit defective social interactions, impaired habituation, aberrant locomotion, and hyperactivity (Wise *et al.* 2015).

4 points in Model Systems/Rescue (max)
Total: >>16 points

NBEA/Complex Neurodevelopmental Disorder

- ClinGen classification: Definitive
- Recently recognized as being associated with disease in OMIM
- As of January 2021, now being included on ID and Epilepsy testing panels

* 604889

NEUROBEACHIN; NBEA

Other entities represented in this entry:

FRAGILE SITE FRA13A, INCLUDED

HGNC Approved Gene Symbol: NBEA

Cytogenetic location: 13q13.3 *Genomic coordinates (GRCh38):* 13:34,942,269-35,672,735 (from NCBI)

Gene-Phenotype Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key
13q13.3	Neurodevelopmental disorder with or without early-onset generalized epilepsy	619157		3

Contributors: Cassandra L. Kniffin - updated : 01/19/2021
Creation Date: Patti M. Sherman : 4/28/2000
Edit History: alopez : 01/26/2021
ckniffin : 01/19/2021
alopez : 03/13/2006
terry : 3/7/2006
alopez : 8/5/2005
tkritzer : 11/10/2003
terry : 11/6/2003
ckniffin : 5/8/2002
mgross : 2/21/2001
mcapotos : 2/20/2001
mcapotos : 2/16/2001
terry : 2/16/2001
mcapotos : 5/12/2000
mcapotos : 5/8/2000
psherman : 5/1/2000

Example 2: *SHROOM4* and X-linked Complex Neurodevelopmental Disorder

- Shroom family member 4
- Xp11.22
- Localized to a wide range of cell types during development, including vascular endothelium and the polarized endothelium of neural tube and kidney
- May regulate cytoskeletal structure by modulating the spatial distribution of myosin II
- Associated with “Stocco dos Santos X-linked mental retardation syndrome” per OMIM
- Included on at least 9 ID/Autism testing panels per 2019 pull of GTR data
- Last evaluated by ClinGen: January 2021

Constraint

Category	Exp. SNVs	Obs. SNVs	Constraint metrics
Synonymous	218.2	212	Z = 0.33 o/e = 0.97 (0.87 - 1.09)
Missense	565.7	502	Z = 0.95 o/e = 0.89 (0.82 - 0.95)
pLoF	40.5	4	pLI = 1 o/e = 0.1 (0.05 - 0.23)

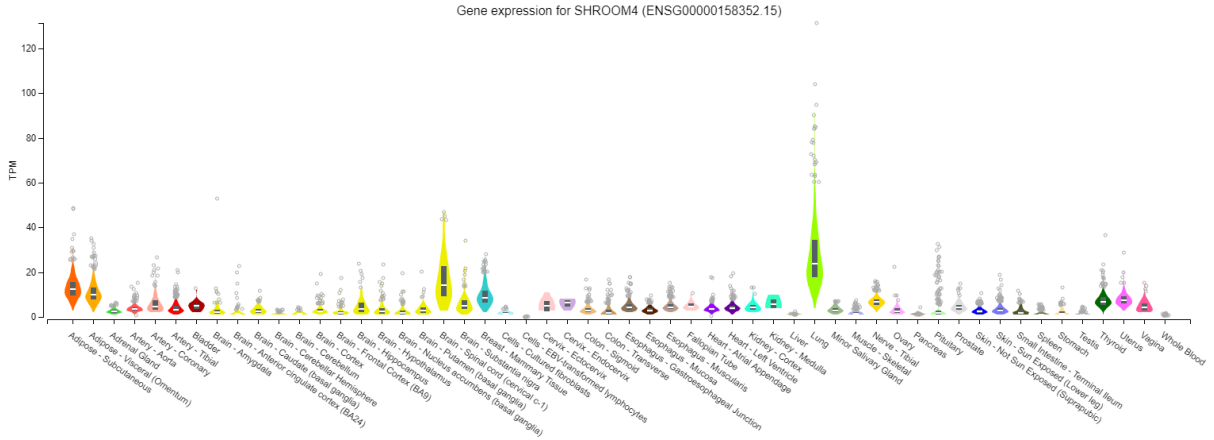
exome genome Metric: Mean Save plot

Gene expression for SHROOM4 (ENSG00000158352.15)

Data Source: GTEx Analysis Release V8 (dbGaP Accession phs000424.v8.p2)

Data processing and normalization

SUBSET None Sex SCALE Log Linear TISSUE SORT MEDIAN SORT OUTLIERS On Off

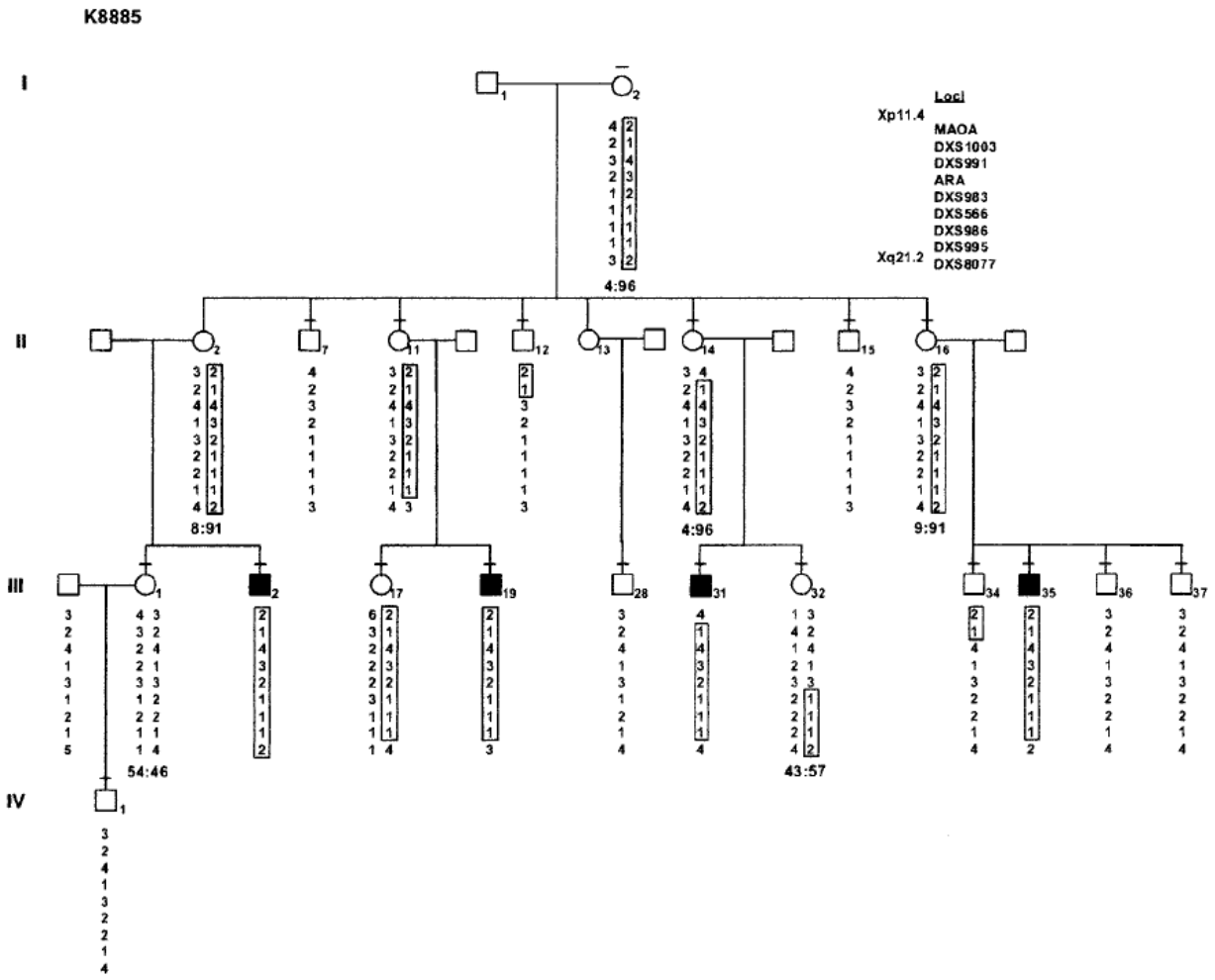


Initial Report: Hagens *et al.* 2006

- Reported 3 cases with ID and variants involving *SHROOM4*
- Two unrelated females had balanced translocations disrupting *SHROOM4* (as well as other genes) → not scored
- This finding prompted the authors to sequence *SHROOM4* in a cohort of 220 X-linked ID families
- Identified 1 missense variant in a Brazilian family with 4 affected males (c.3266C>T, p.Ser1089Leu)
  - Family initially described by Stocco dos Santos *et al.* in 1991, with linkage analysis conducted in 2003
  - No functional data provided
  - Is observed in 1/179989 alleles in gnomAD (heterozygote; no hemizygote observations)

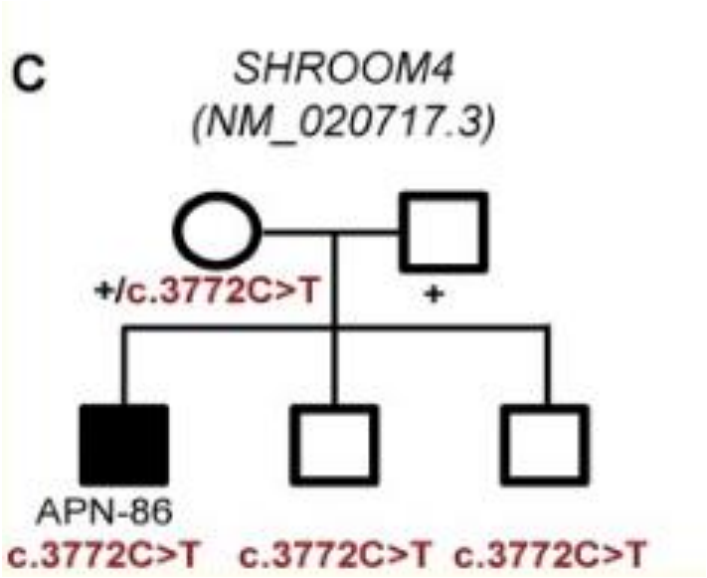
Stocco dos Santos Family

- Affected males had:
  - Severe ID
  - Delayed or absent speech
  - Seizures
  - Hyperactivity
  - Mothers reported to have seizures and periods of depression.
- Linkage analysis narrowed the region to Xp11.3-Xq21.3
  - Variant identified in Hagens found by candidate gene sequencing only – did not sequence all genes in region
- LOD score reported by the authors: 3.02



Second Report: Redin *et al.* 2014

- Analyzed a cohort of 106 patients with intellectual disability using a panel of 217 proposed ID/Autism genes
- Identified a maternally inherited putative LOF variant (c.3772C>T, p.Gln1258Ter) in an affected male *as well as his 2 unaffected brothers*
- Not scored



Third report: Lopes *et al.* 2016

- Report a 14-year-old male with dyspraxic gait; “no language;” kyphosis; and peripheral vasomotor disturbances
  - No reported family history
- Exome sequencing revealed 2 maternally inherited X-linked variants:
  - c.409G>A (p.Asp137Asn) in *ZFX* (no proposed gene relationship in OMIM as of Jan 2021)
  - c.436C>T (p.Arg146Trp) in *SHROOM4*
    - Observed in 135/203469 alleles in gnomAD, including 1 homozygote and 53 hemizygotes)
  - Not Scored

SHROOM4/ X-Linked Complex Neurodevelopmental Disorder

- ClinGen Classification: Disputed
  - Only 3 total variants have been reported.
  - 2 have been ruled out due to presence in the general population and/or non-segregation in a family.
  - Evidence supporting the only remaining variant (Stocco dos Santos family) is weak:
    - Missense with no supportive functional data; gene is not constrained for missense variation
    - Other possible causes of disease not effectively ruled out
    - No other plausible variants reported in this gene in ~15 years
    - For these reasons, the ID/Autism GCEP opted not to count this variant or its segregation and DISPUTE this gene-disease relationship
    - Intentionally did not use the eponymous name used in OMIM in the curation due to lack of evidence

Global Collaboration: The Gene Curation Coalition (GenCC)

- Understand the approaches and classification systems of the different curation efforts
- Develop consistent terminology for validity assessment as well as inheritance, allelic requirement, mechanism of disease
- Post gene curations from any group willing to share with the public
    - The GenCC DB is like a ClinVar for genes!
- Resolve differences and collaborate on gene curation projects

Slide courtesy of Marina DiStefano, PhD

GenCC PUBLIC BETA
The Gene Curation Coalition

Genes | Submitters | Statistics | Download | FAQ | Contact | About Us | Stay Informed

Genes

https://TheGenCC.org

Help

2281 Genes with classifications based on your filters

Filter by gene symbol... | Filter by submitted disease... | Filter by Submitter

About GenCC Classifications

			Definitive	Strong	Moderate	Supportive	Limited	Disputed	Refuted	Animal	No Known	
A2M HGNC:7	1 Disease Equivalents	1 Submitters	0	0	0	0	0	0	0	0	1	Details
A2ML1 HGNC:23336	5 Disease Equivalents	2 Submitters	0	0	0	0	1	1	0	0	4	Details
AAAS HGNC:13666	1 Disease Equivalents	1 Submitters	1	0	0	0	0	0	0	0	0	Details
AADAC HGNC:17	1 Disease Equivalents	1 Submitters	0	0	0	0	0	0	0	0	1	Details
AAGAB HGNC:25662	1 Disease Equivalents	1 Submitters	0	1	0	0	0	0	0	0	0	Details
AASS HGNC:17366	1 Disease Equivalents	3 Submitters	0	2	1	0	0	0	0	0	0	Details

AASS

Return to list

Help

Gene Symbol: **AASS**
HGNC:17366

Locus Group: Protein-Coding Gene

Locus Type: Gene With Protein Product

By Classification | By Disease | By Submitter

Filters: Classifications 2 | Diseases 1 | MOI 1 | Submitters 3

Strong classifications

Strong	AASS HGNC:17366	hyperlysinemia (disease) MONDO:0009388 Submitted as: OMIM:238700	AR	07/22/2015 Evaluated 11/25/2020 Submitted	TGMI|G2P Assertion Criteria | More Details
Strong	AASS HGNC:17366	hyperlysinemia (disease) MONDO:0009388 Submitted as: OMIM:238700	AR	03/27/2017 Evaluated 12/04/2020 Submitted	Invitae Assertion Criteria | More Details

Moderate classifications

Moderate	AASS HGNC:17366	hyperlysinemia (disease) MONDO:0009388	AR	06/29/2020 Evaluated 10/27/2020 Submitted	ClinGen Public Report | Assertion Criteria | More Details

How can you get involved?

- Incorporate a gene-disease validity process into your workflow to augment information available from public resources
  - Utilize this information when designing tests/filtration pipelines and when reporting results
- Contribute to efforts designed to aggregate evidence
  - Matchmaker Exchange (matchmakerexchange.org)
  - GenomeConnect (genomeconnect.org) → Learn more about this Friday (Juliann Savatt)
- Volunteer with ClinGen
  - Consider forming your own expert panel → Learn more about this Friday (Laura Milko)
  - Curate for an existing gene curation expert panel (ccdb.clinicalgenome.org)

Acknowledgements

ClinGen Gene Curation Working Group

Chairs
Erin Rooney Riggs
Courtney Thaxton

Past Chairs
Jonathan Berg
Christa Lese Martin

Danielle Azzariti
Taylor Bingaman
Marina DiStefano
May Flowers
Ada Hamosh
Steven Harrison
Madeline Hughes
Meera Jairath
Liz Kearns
Xi Luo
Megan Mayers
Laura Milko

Shruthi Mohan
Mayher Patel
Natalie Pino
Sharon Plon
Christine Preston
Erin Ramos
Heidi Rehm
Justyne Ross
Tasha Strande
Meredith Weaver
Emma Wilcox
Matthew Wright

Additional questions?
eriggs@geisinger.edu

ClinGen is primarily funded by the National Human Genome Research Institute (NHGRI) through the following three grants: U41HG006834, U41HG009649, and U41HG009650