

# The Gene Curation Coalition: A global effort to harmonize gene-level resources

Curating the Clinical Genome 2019

Marina DiStefano PhD

Clinical Molecular Fellow, HMSGTP

Partners HealthCare Laboratory for Molecular Medicine

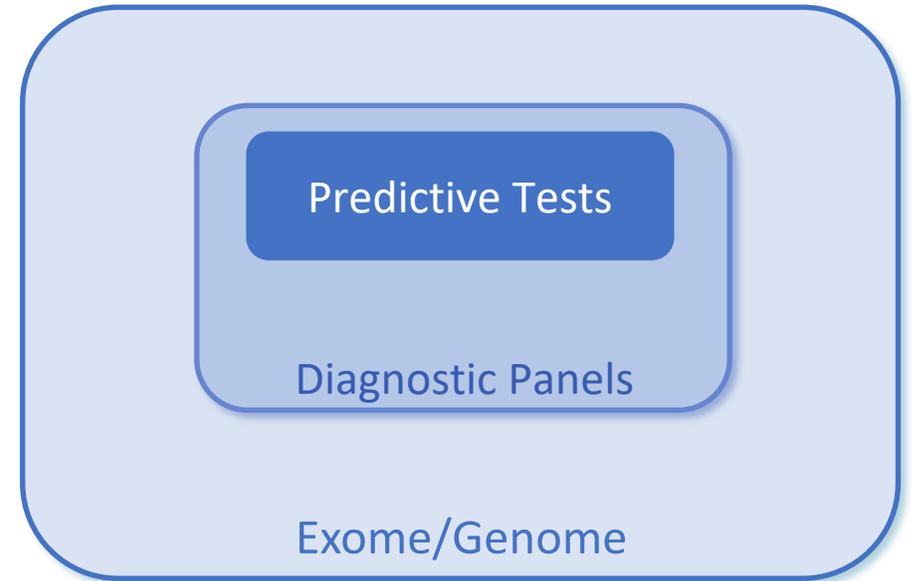
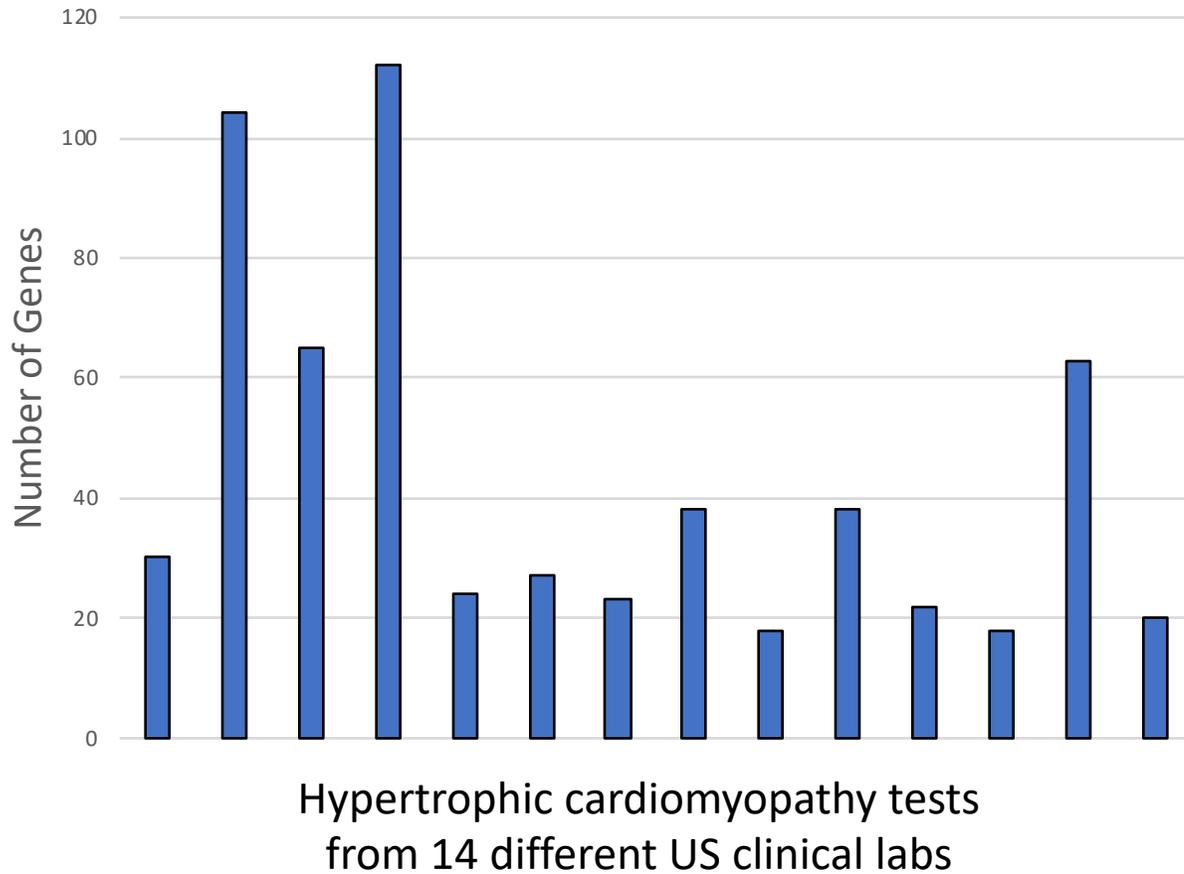
The Broad Institute

ClinGen Curator

**Gene Curation Coalition**



# The Need for Curated Gene Databases



Different levels of evidence are needed for different clinical uses

Standards and consensus are needed for which genes are valid disease genes that are ready for clinical testing

# Some reasons for gene curation discordance

- Definition of disease (high penetrance versus inclusion of lower penetrance phenotypes)
- Purpose of curation (validity versus panel inclusion based on phenotype match)
- Differences in disease/phenotype assignment (genes often have claims for multiple diseases)
- Date of evidence evaluation
- Understanding the differences and focusing on resolving them requires harmonization of terms and definitions

# Gene Curation Coalition (GenCC)

Orphanet/ORDO

Catalog and ontology – presence in a publication is sufficient for entry

OMIM

Descriptive entries for reported gene-disease implications (requires a minimum level of evidence to enter database, includes “?” entries)

G2P/DECIPHER/TGMI

Database of cases with curation of the evidence for implicated genes

Rahman Gene-Disease Map

Rapid review of gene evidence for highest disease implication  
Completed for all genes July 2018, will not be updated

ClinGen Gene-Disease Validity

Expert consensus review of evidence for gene-disease implications

Genomics England PanelApp

Crowdsourcing review tool and curated evidence resource for documenting which genes are valid for use in diagnostic panels

Genetics Home Reference

Accessible educational resource for genes and disease

# A Few Goals of GenCC

- Clarify the overlap between gene curation efforts
- Understand the aims, processes, information used, classification systems, and users of the different curation efforts
- **Develop consistent terminology for validity assessment** as well as inheritance, allelic requirement, mechanism of disease
- Collaborate on gene curation projects

# GenCC Term Delphi Survey Process

Draft harmonized definitions for gene curation categories



Solicit term suggestions from the GenCC members and draft Delphi Survey



Round 1: Survey GenCC members (N=33)

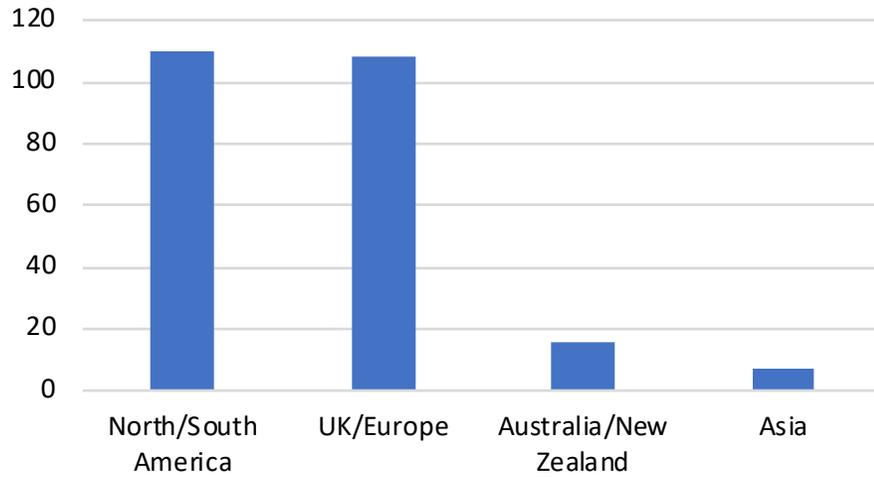
Round 2: Survey extended membership of GenCC groups (N=38)

Round 3: Survey the international genetics community (N=241)

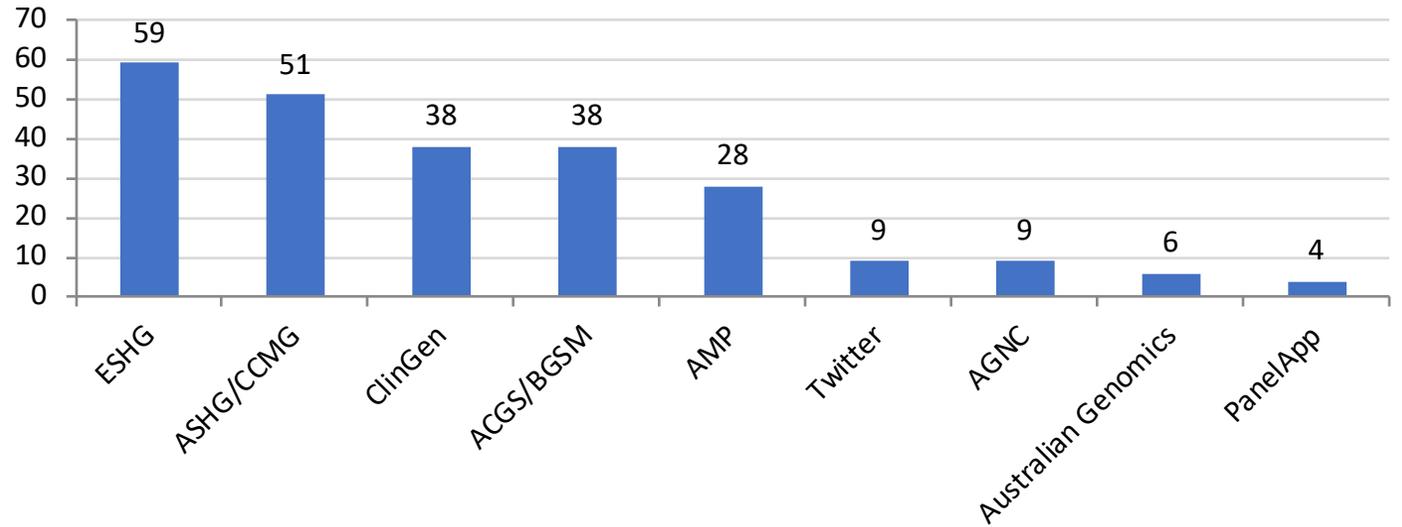
Optional Explanatory  
Video

# Survey Round 3 Demographics

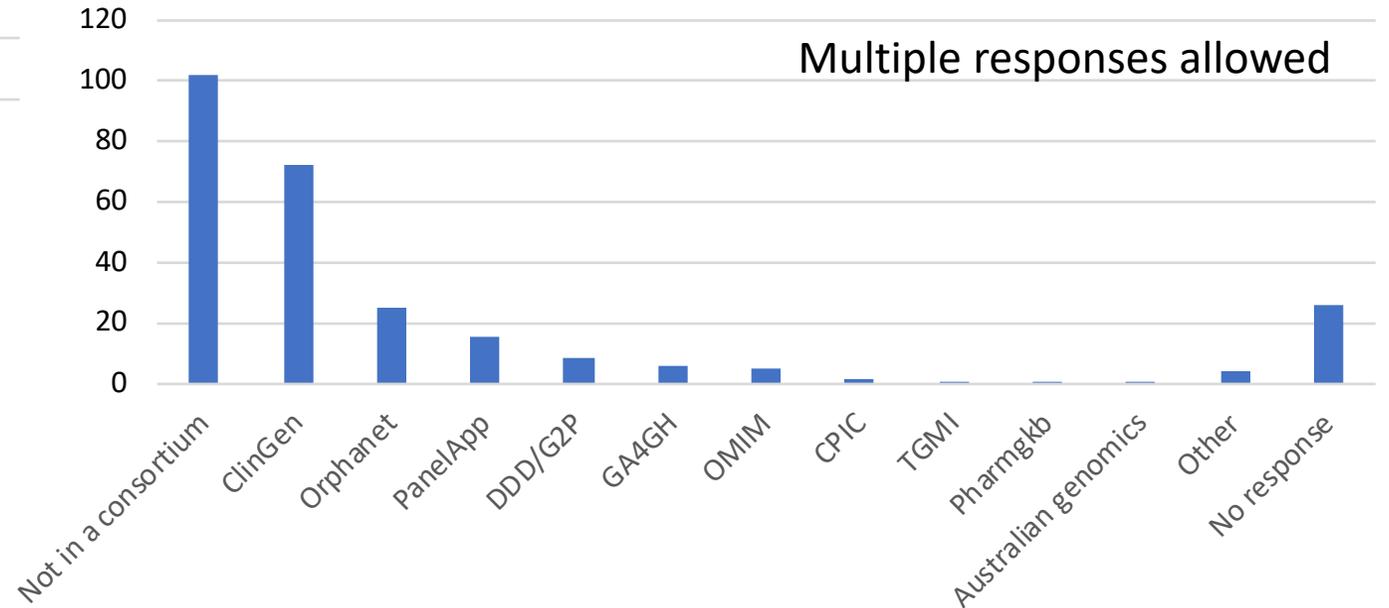
Respondant Location (N=241)



Surveyed Populations (n=241)

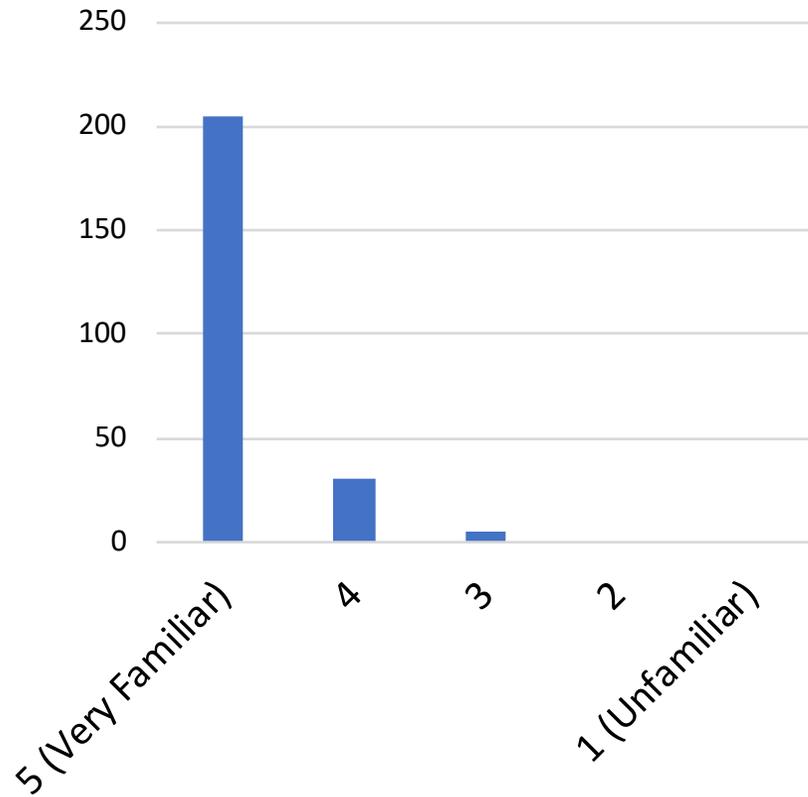


Consortium Participation (n=270)

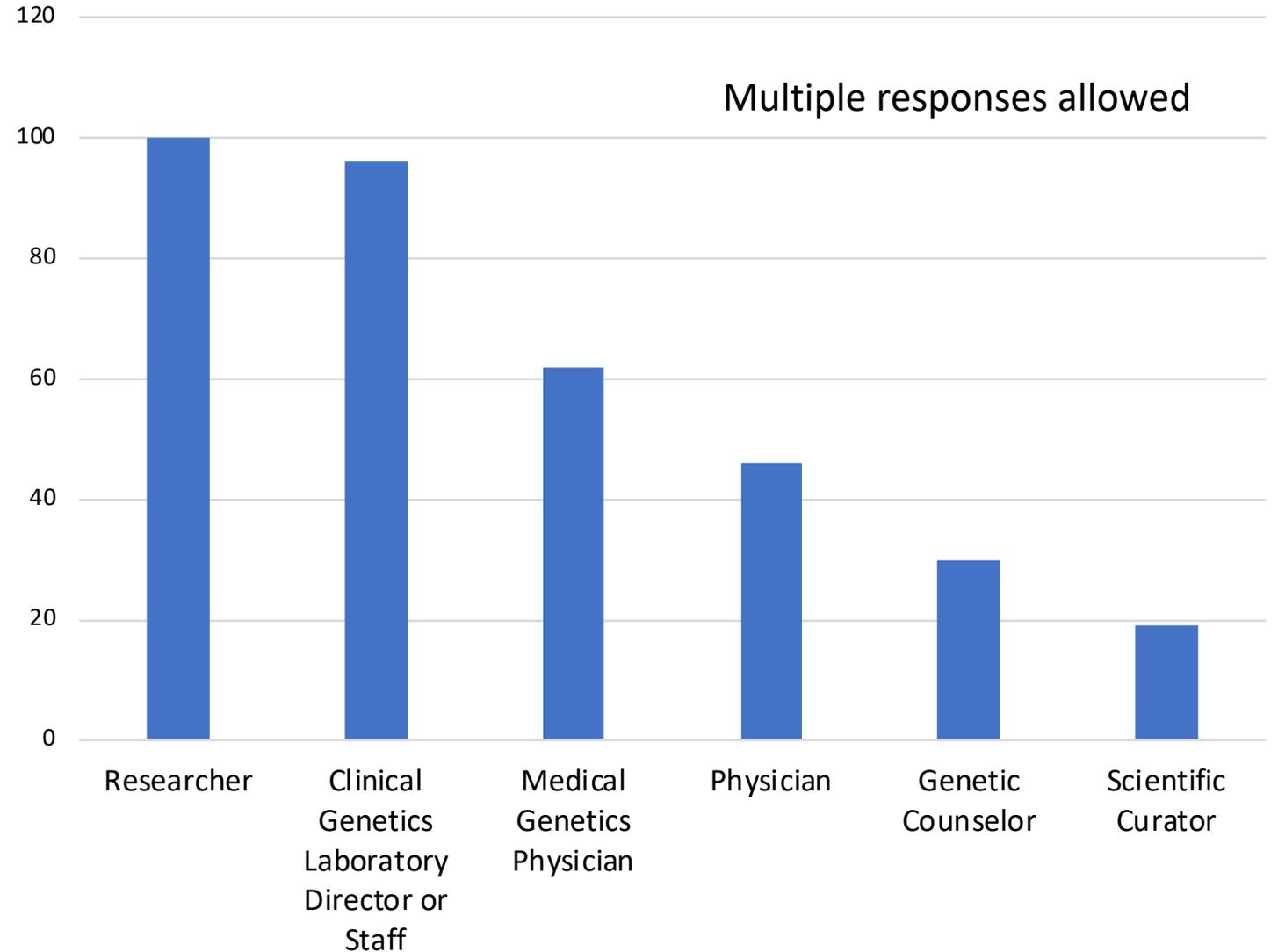


# Demographics

## Genetics Familiarity (n=241)



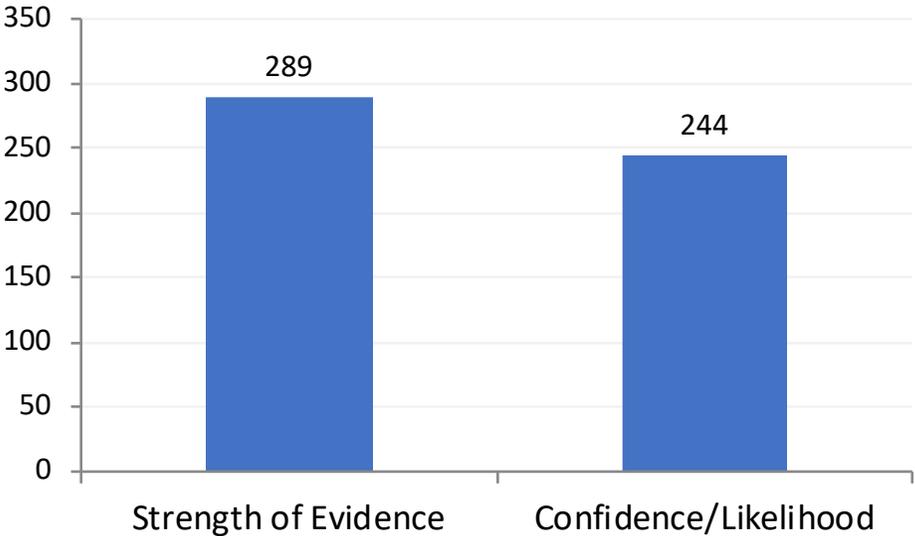
## How would you describe yourself? (N= 353)



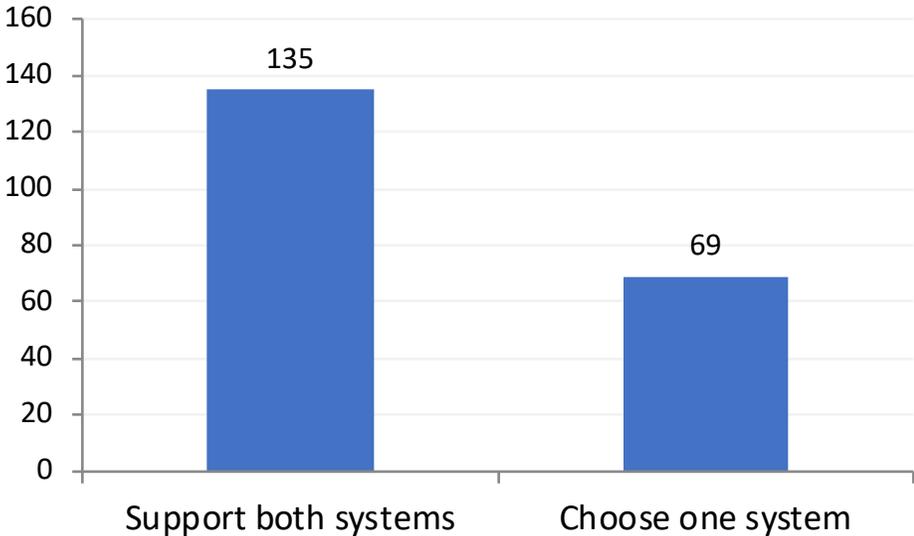
# Strength of Evidence vs. Confidence/Likelihood

Example: Strong/Weak vs. Likely/Possible

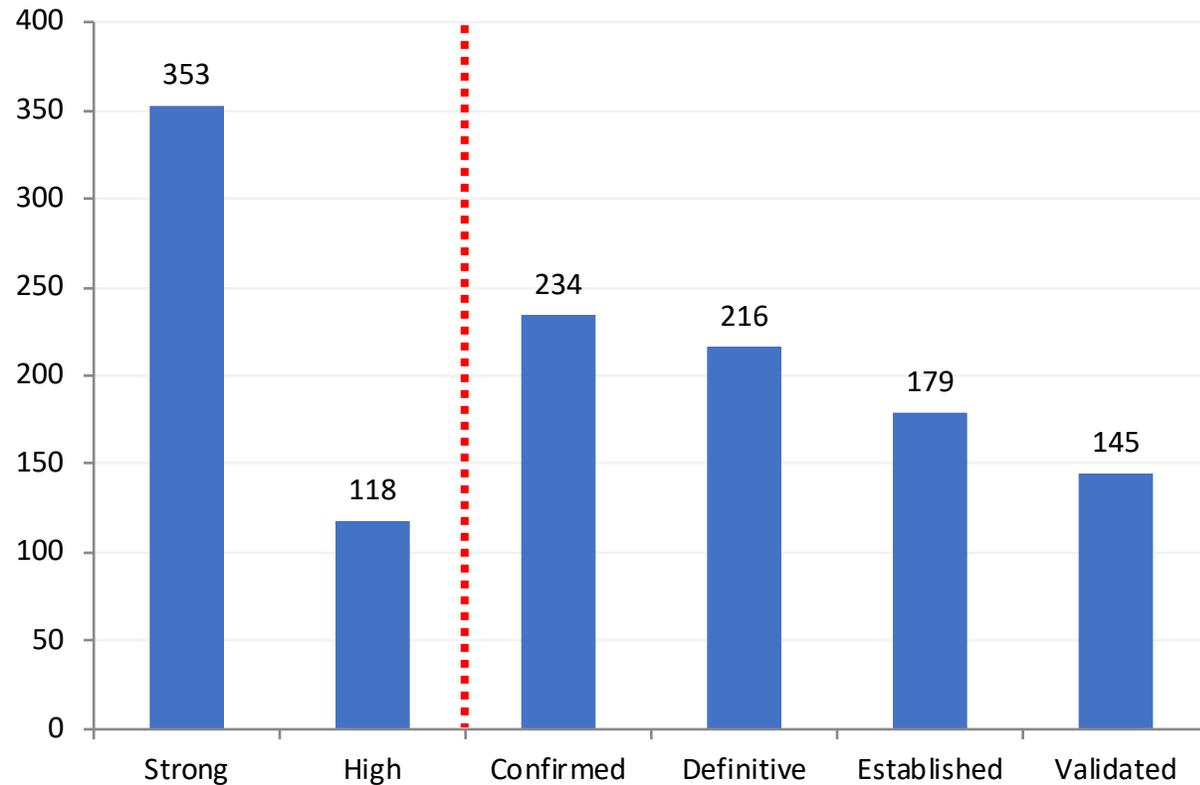
Which system is more agreeable?



Choose one or support both systems?



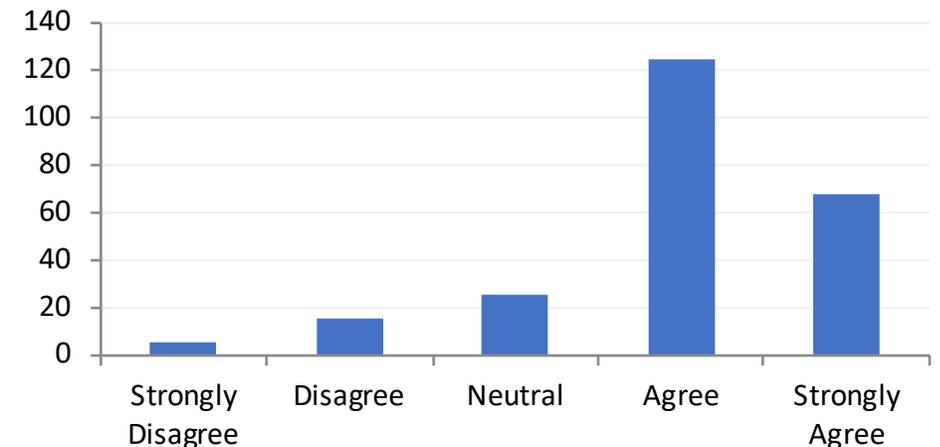
...a gene that has unequivocally been implicated in disease



Evidence Terms

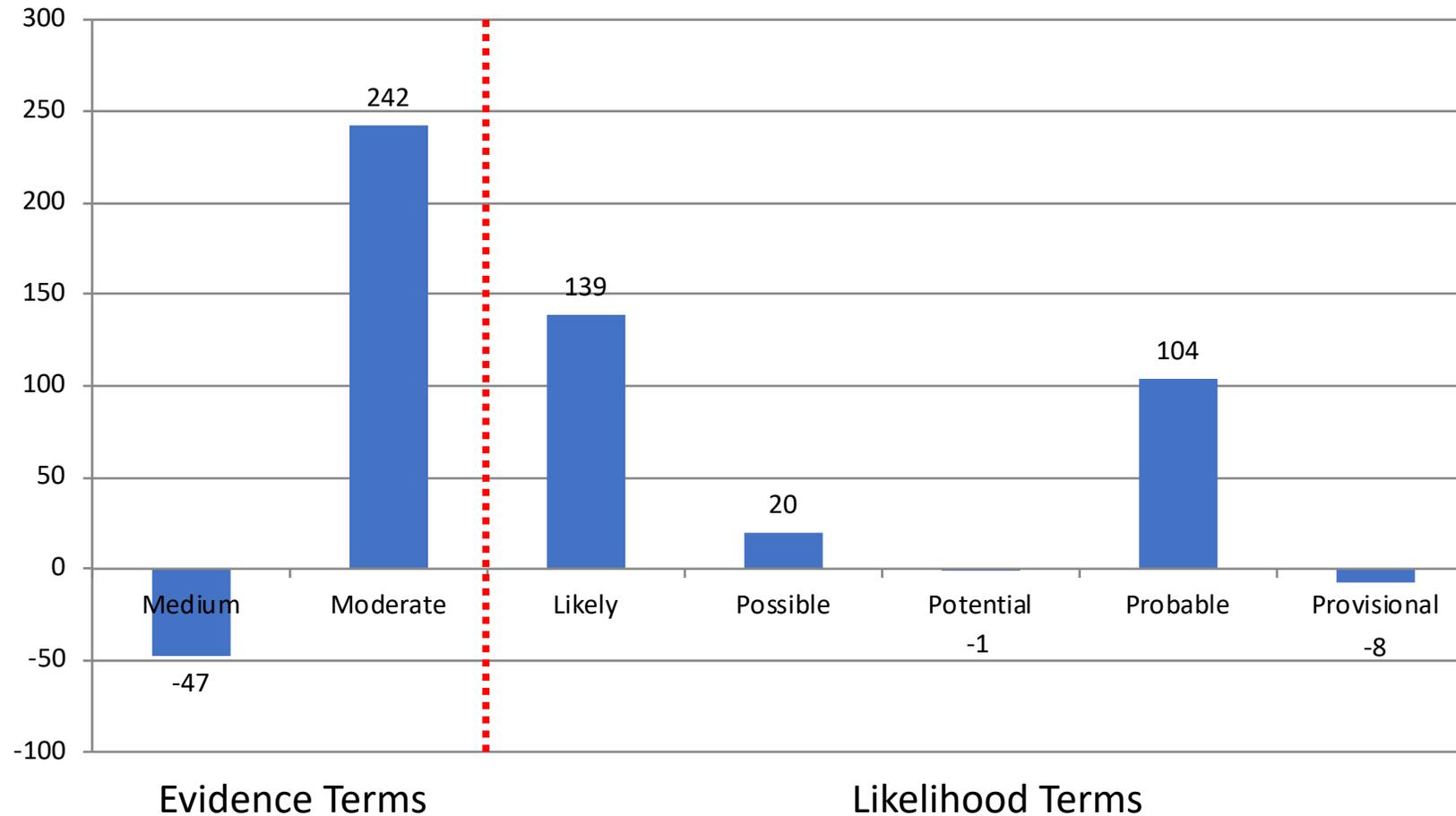
Likelihood Terms

**Distuingish between "Definitive" vs "Strong" genes? (n=241)**

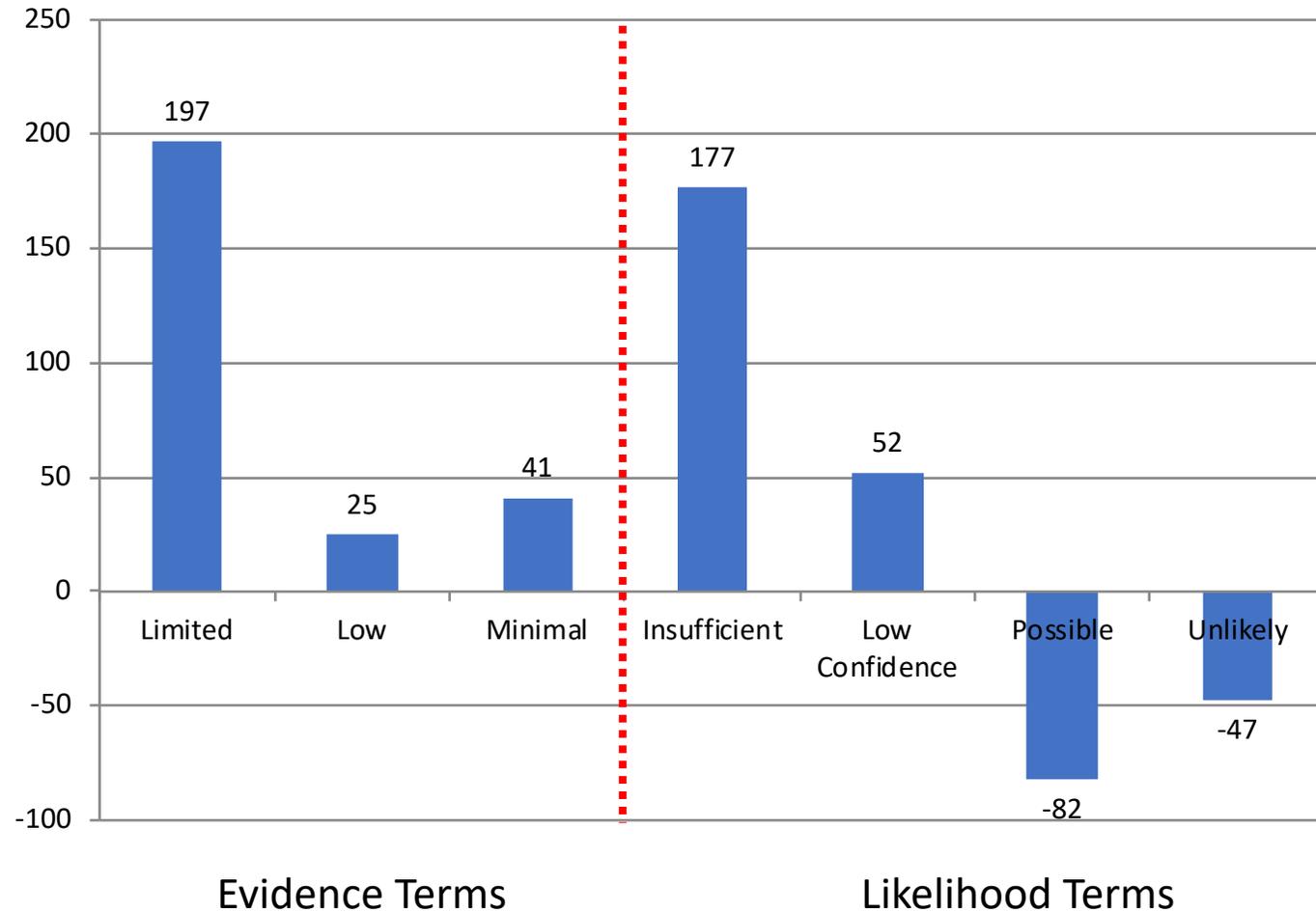


Strongly Agree=2pts, Agree=1pt, Neutral=0pts, Disagree=-1pts, Strongly Disagree=-2pts

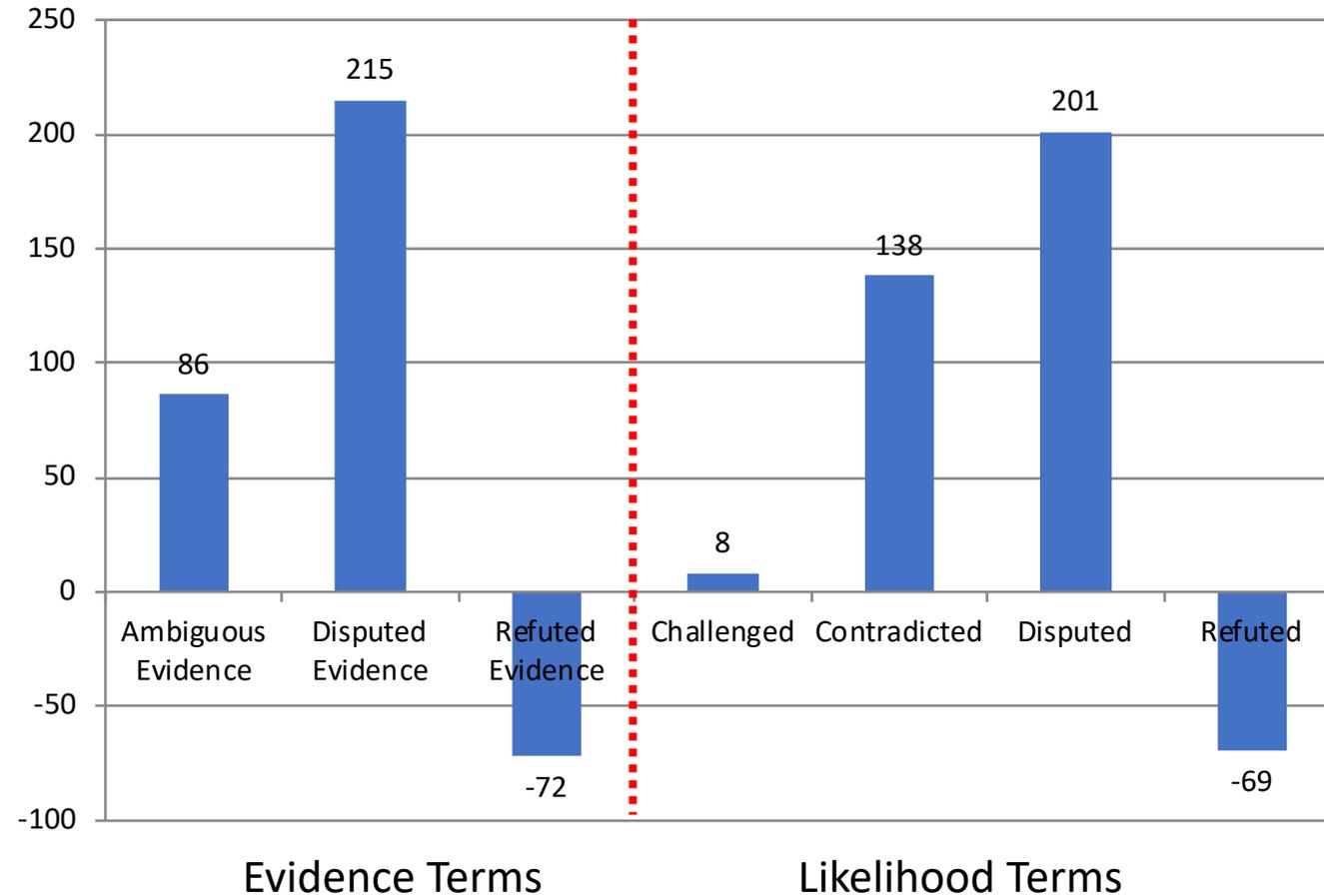
...a gene where there is an intermediate amount of evidence in humans to support a causal role



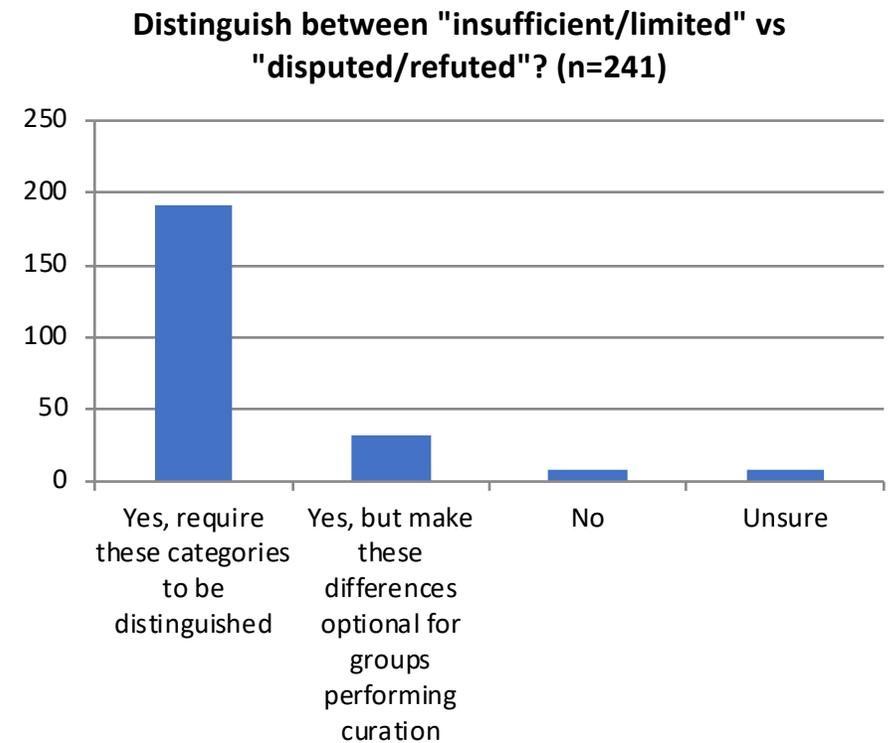
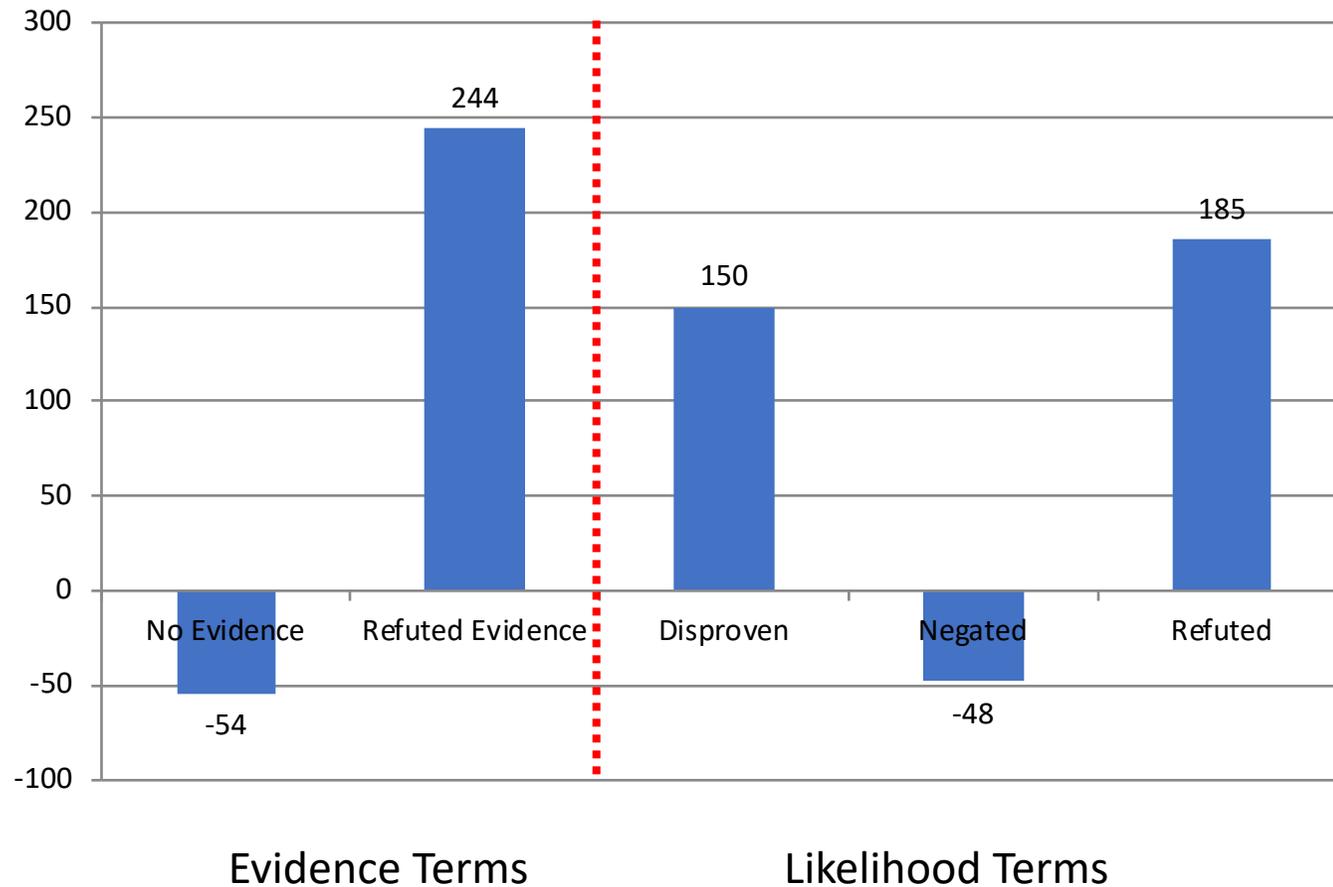
...a gene where little human evidence exists to support a causal role



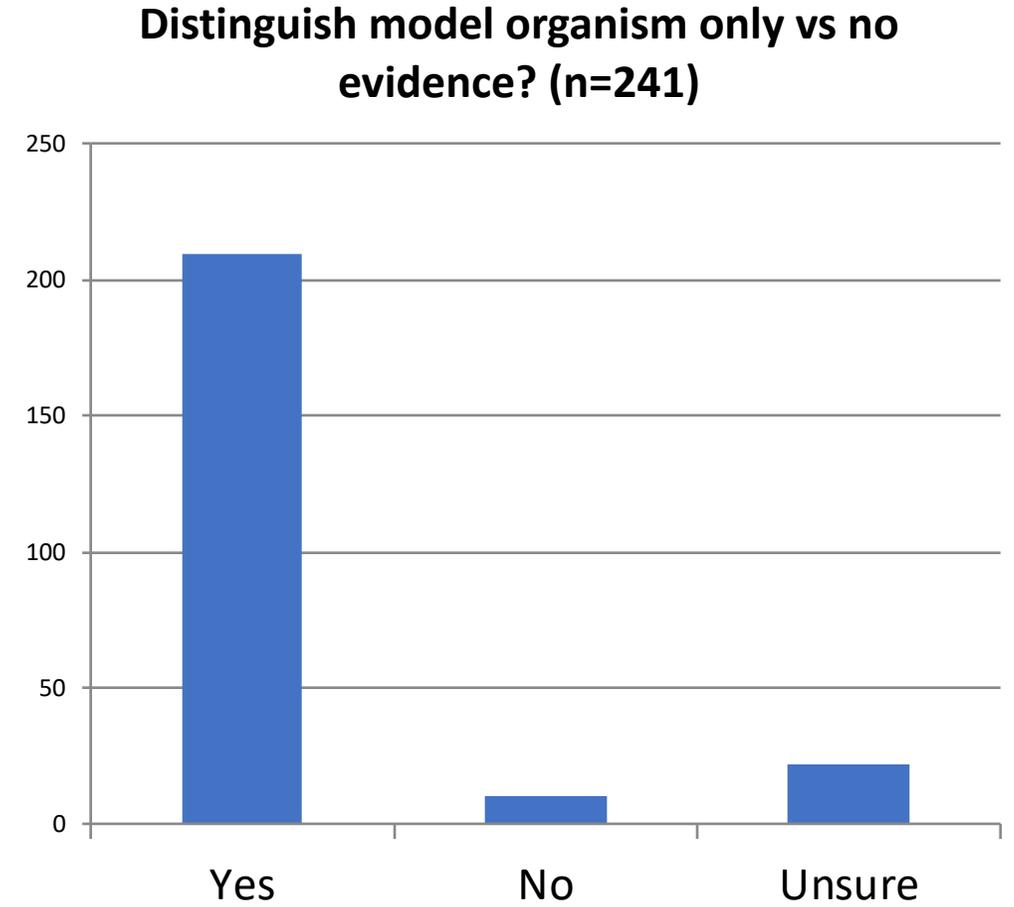
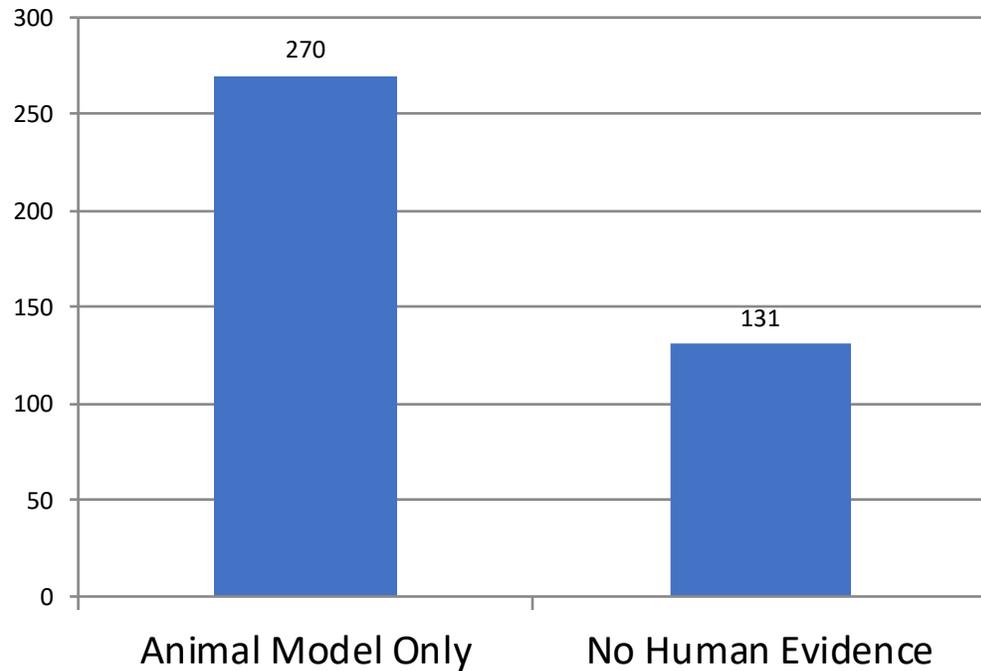
...a gene that, although evidence has been reported, other evidence of equal weight challenges the claim



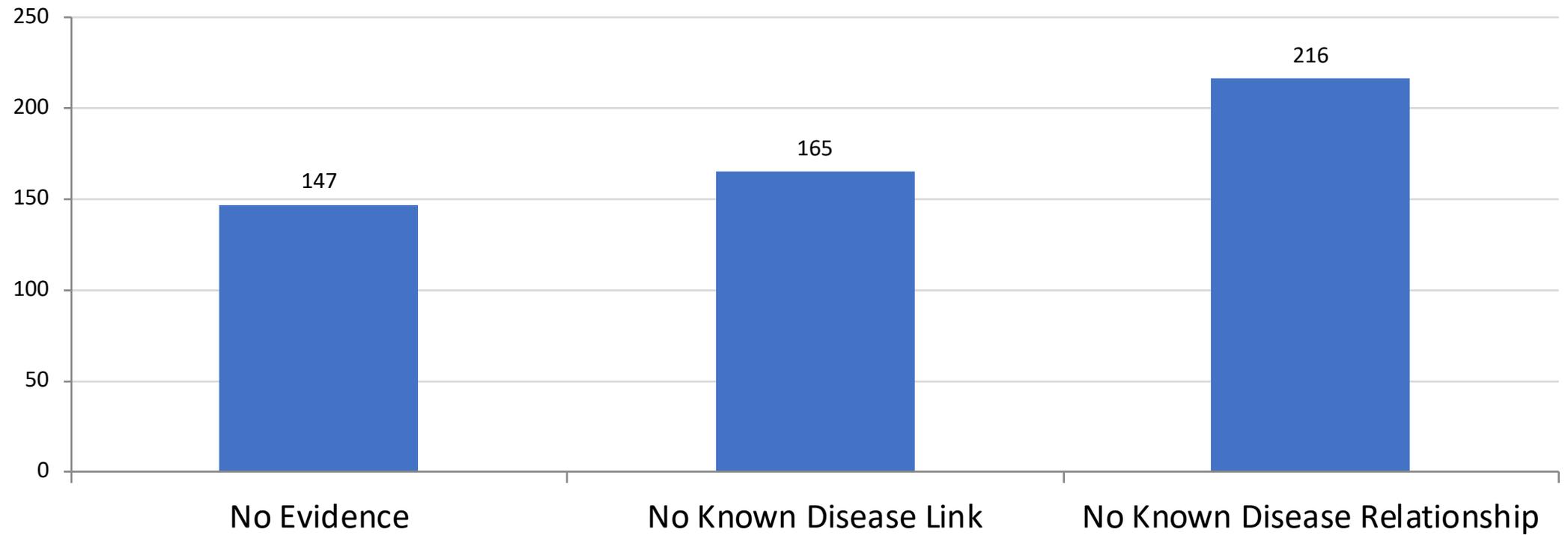
...a gene with a reported gene-disease relationship, but new valid evidence has arisen that overturns the original body of evidence



...a gene where no human disease evidence exists, but a convincing animal model of the disease exists



...a gene where no disease claim in any organism has ever been made



# Term List Ranking

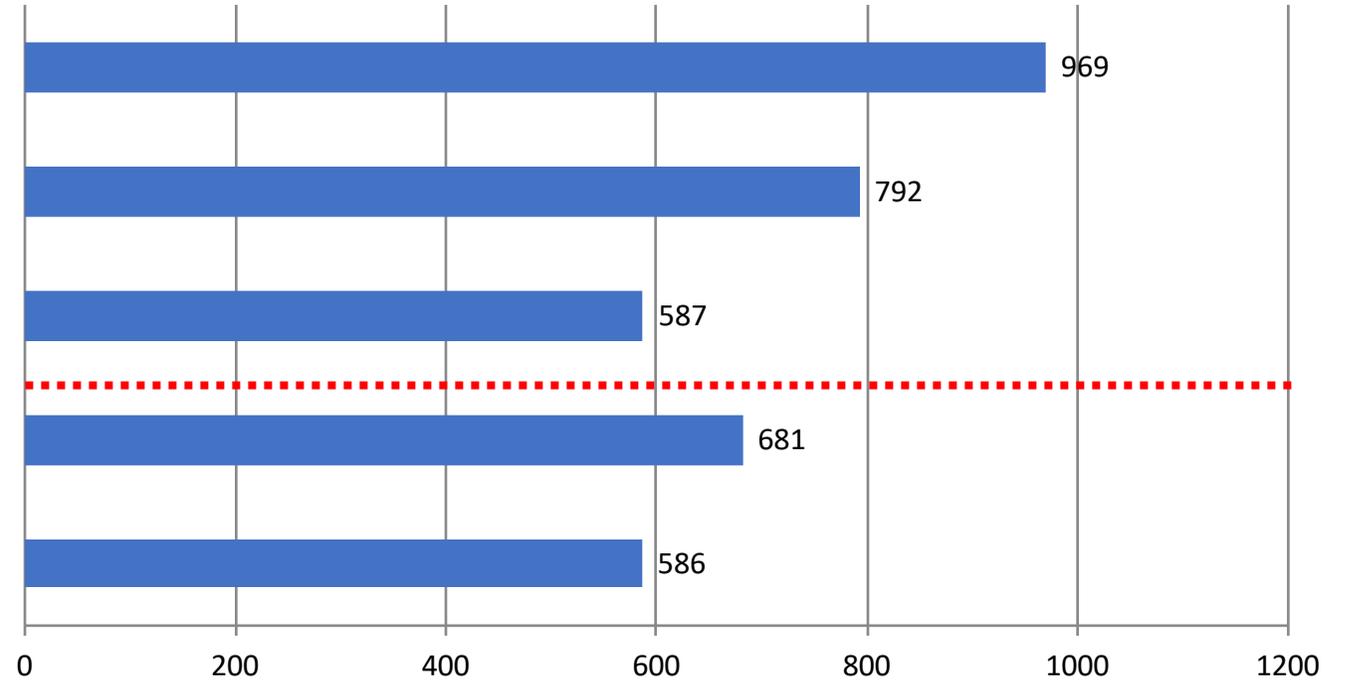
**Definitive**/Strong/Moderate/Limited/Disputed  
evidence/Refuted evidence/Animal model only/No known  
disease relationship

Strong/Moderate/Limited/Disputed evidence/Refuted  
evidence/Animal model only/No known disease relationship

Strong/Moderate/**Low**/Disputed evidence/Refuted  
evidence/Animal model only/No known disease relationship

Confirmed/Likely/Insufficient/Disputed/Refuted/Animal model  
only/No known disease relationship

Confirmed/**Probable**/Insufficient/Disputed/Refuted/Animal  
model only/No known disease relationship



1<sup>st</sup>=5pts, 2<sup>nd</sup>=4pts, 3<sup>rd</sup>=3pts, 4<sup>th</sup>=2pts, 5<sup>th</sup>=1pt

# Conclusions and continuing steps toward harmonization

- Delphi survey round 3 had **93% term agreement** (13/14 terms) with round 2, suggesting the genetics community is harmonious with GenCC
- One final survey was conducted to assess the practical considerations for adopting the consensus term set among the GenCC member groups

# Term Adoption Responses

	ClinGen	G2P/TGMI	PanelApp	GHR	OMIM	Orphanet
Will you use terms in your own efforts?	YES	YES	YES	NO, but may use terms in text write-ups	NO, but will display ClinGen classifications	YES
Will you adopt (evidence set) or map?	ADOPT	ADOPT	ADOPT	Not applicable	Not applicable	MAP
Do you intend to use the secondary consensus (likelihood) set?	NO	MAYBE	MAYBE	MAYBE (if useful in text summaries)	Not applicable	MAYBE
Retroactive implementation ?	YES, likely	To be discussed	NO	NO	YES	N/A

Overall preferred Terms/Set	Top Confidence Terms (lower score overall)
Definitive*	Confirmed
Strong	
Moderate	Likely
Limited	Insufficient
Disputed Evidence	Disputed
Refuted Evidence	Refuted
Animal Model Only	Animal Model Only
No Known Disease Relationship	No Known Disease Relationship

ClinGen	G2P
Definitive	Confirmed
Strong	
Moderate	Probable
Limited	Possible
Disputed	
Refuted	
No Evidence	

Note: Other GenCC groups mostly used colors or no formal terms

\*The highest scoring term set was the evidence set with “Definitive” added and most respondents felt it was important to distinguish Strong versus Definitive states. However, we did not have a separate question to ask about what term serves the “definitive” category best.

# Acknowledgements

Member groups of the GenCC:

- ClinGen
- TGMI/G2P
- Genetics Home Reference
- OMIM
- Orphanet
- PanelApp
- Gene-Disease Map

## Gene Curation Coalition

