criTRia Standard Operating Procedure

Macayla Ann Weiner[1], Laurel Hiatt[3], Pamela Ajuyah[2], and Harriet Dashnow[1]

1. Department of Biomedical Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO, USA
2. Broad Institute of MIT and Harvard, Cambridge, MA, USA
3. Department of Human Genetics, University of Utah, Salt Lake City, UT, USA

Table of Contents

Background

The criTRia tandem repeat locus curation process is a framework to evaluate the strength of a monogenic locus-disease relationship based on publicly available evidence. While established protocols exist for evaluating monogenic gene-disease relationships and variant interpretation[1,2,3,9,10,11,12,13], these do not account for the specific considerations of tandem repeat disease loci. In particular, here we consider the evidence that a specific tandem repeat locus is associated with a disease rather than considering the gene as a whole. criTRia builds on ClinGen's established gene-disease relationship standard operating procedure[4], and adapts and extends it to specifically address tandem repeat diseases through additions of tailored criteria and adjusted score weighting. Much of the procedure remains unchanged, ensuring consistency and only modifying what was necessary. It is important to note that there are emerging loci of the same gene, some very close to one-another, that require special care.

Overview

The criTRia scoring process uses the following criteria to assign a score and classification for a locus-disease relationship:

- Establishing the locus-disease mode of inheritance
- Evidence collection
    - Genetic Evidence
    - Experimental Evidence
- Evaluation and scoring of evidence
- Review, classification, and approval of a locus-disease relationship
- Publication of the final classification and supporting evidence
- Re-evaluation as indicated

In the subsequent sections of this document, each step will be outlined in detail and general recommendations provided.

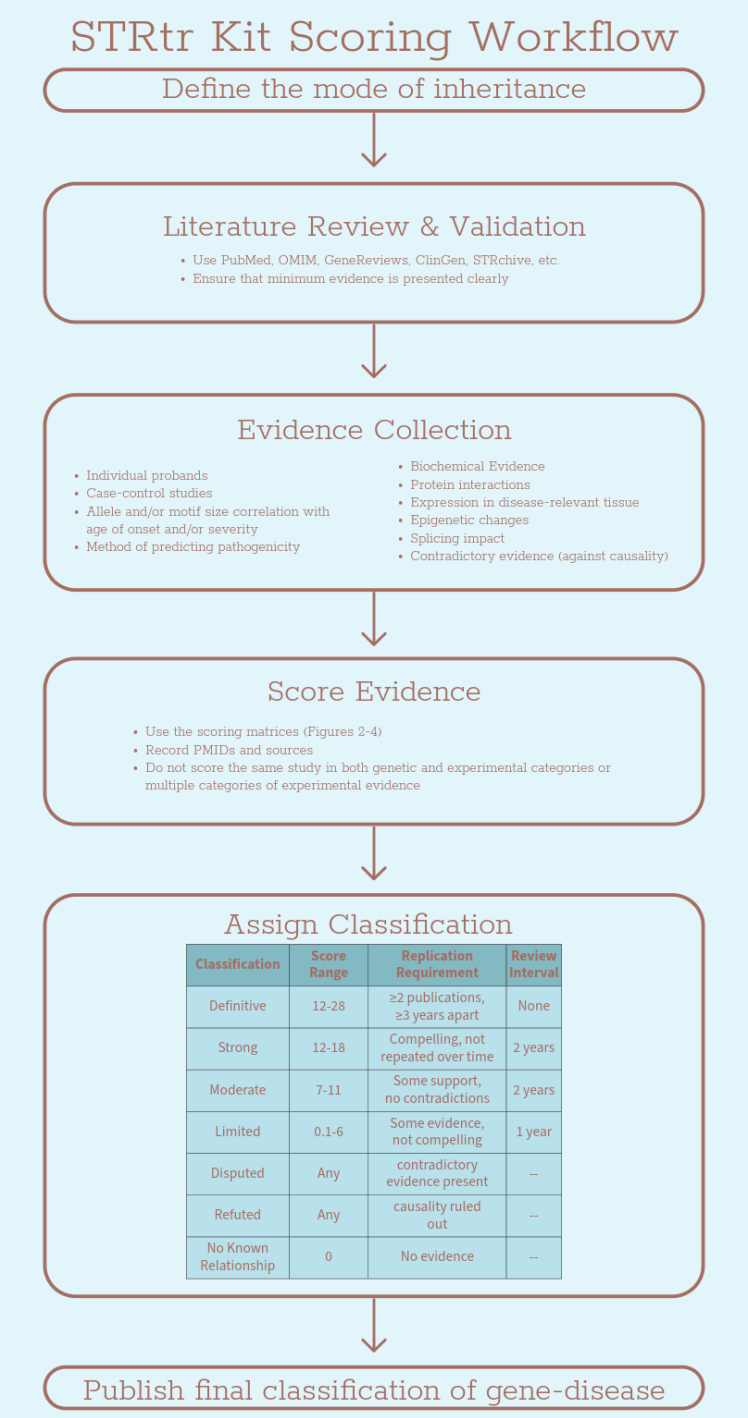

Figure 1. *Flow chart detailing the steps to discovering and publishing new loci.*

Categorical Classifications

Definitive

- Score: 12-18
- Replicated over time

The role of the evaluated locus and gene in the evaluated disease has been repeatedly demonstrated in both research and clinical settings, and *has been replicated over time* (at least 2 publications with convincing evidence at least 3 years apart). Repeat expansions that disrupt gene function and are supported by strong genetic and population data (e.g, absence of the repeat expansion in study controls/unaffected individuals, *de novo* expansion, segregation with disease in families, or consistent linkage to a genomic region) are considered convincing of disease causality in this process.

- Interval for re-evaluation: no set time point
- Re-evaluation may be considered if contradictory information has arisen or a need has been expressed for another reason.

Strong

- Score: 12-18
- Not replicated over time

The role of the evaluated locus in the evaluated disease has been independently demonstrated in at least two separate studies providing strong supporting evidence for this gene's role in the disease. Locus-disease pairs with strong evidence demonstrate considerable genetic evidence (numerous unrelated probands harboring variants with sufficient supporting evidence for disease causality). Compelling evidence from different types of supporting experimental data is typically also present, but is not necessarily required to reach this designation if substantial convincing genetic evidence is present. In addition, no convincing evidence has emerged that contradicts the role of the locus in the noted disease.

- Interval for re-evaluation: two years after last scoring

Moderate

- Score 7-11

There is moderate evidence to support a causal role for the evaluated locus in the evaluated disease. Locus-disease pairs with moderate evidence typically demonstrate some convincing genetic evidence (unrelated probands harboring variants with sufficient supporting evidence for disease causality with or without experimental data supporting the locus-disease relationship). The role of this locus in the disease may not have been directly reported, but no convincing evidence has emerged that contradicts the role of the locus in the noted disease.

- Interval for re-evaluation: two years after last scoring

Limited

- Score: less than 7

In this category there is a small amount of evidence supporting the locus-disease relationship. The Limited category should not be applied in circumstances where none of the presented evidence is compelling; in these circumstances, the Disputed category should be considered.

- Interval for re-evaluation: one year after last scoring

No Known Disease Relationship

Evidence for a causal role in the monogenic disease of interest has not been reported within the literature (published, pre-published and/or present in public databases [e.g. ClinVar, etc.]).

These genes might be “candidate” genes based on linkage intervals, animal models, implication in pathways known to be involved in human disease, etc., but no reports have directly implicated the gene or locus in the specified disease. If a claim of a relationship with the specified disease has been reported, but the evidence is minimal or not compelling, consider Limited, Disputed, or Refuted.

A notable difference in this procedure from ClinGen’s: A tag designating “animal model only” is applied on clinicalgenome.org for those gene-disease pairs in which no human genetic evidence has been asserted, but an animal model exists.

Contradictory
A locus-disease pair may be classified as Disputed or Refuted when significant contradictory evidence emerges or the strength of previously reported supporting evidence diminishes. While genetic and experimental data may initially support the relationship, conflicting evidence calls the association into question. The choice between these classifications is based on the strength of the data. It is best to use Disputed when supporting and contradictory evidence are both convincing. Whereas, Refuted should be used when the contradictory evidence substantially outweighs any supporting evidence. Input from experts is essential when dealing with contradictory evidence and using one of these classifications. More details about these categories and example scenarios are described below.
Disputed:

- Presented cases have disparate phenotypes.
    - Notably distinct phenotypes between cases, such as phenotypes in completely different patient systems (neuromuscular versus respiratory symptoms)
    - Negligible overlap in symptoms between cases, such that cases are not convincingly connected
- Proposed pathogenic variants are observed in unaffected individuals or control populations at a rate higher than expected based on the disease prevalence and proposed penetrance.
- There are cases present in the research with differing or unclear modes of inheritance
- *Alternative causes have NOT been identified in the original proband and there is contradictory evidence.*

Refuted:

- The evidence against the locus-disease relationship outweighs the evidence for, or the locus-disease relationship and evidence that it has been ruled out.
- *Alternative causes for the disease are identified in the original proband.*
- The proband is later determined not to have the disease in question.
- Statistically rigorous case-control shows no enrichment.

No TR-Disease Relationship
This distinction should only be used when the relationship appears polygenic rather than Mendelian.

Evidence Collection

Evidence is collected primarily from published peer-reviewed literature, but can also be present in publicly accessible resources, such as variant databases, and pre-published literature (preprints) with compelling evidence used with discretion in scoring. When determining whether a case is appropriate for use, consider the following:

- Sufficient evidence is provided.
- Case must be publicly accessible. For example, do not include cases that are only available to authorized users.
- Case is well-described with appropriate phenotype, testing, statistical analysis when appropriate, and other variant information.
- Case is not otherwise believed to be described in already scored literature – for example, if a proband is included across multiple studies, only score them once.
- If a locus is being proposed pathogenic for the first time, the case must have supporting evidence that the variant is causative beyond its presence in an affected individual.
- Gene curation is not a prerequisite. As we have seen, some TRs are clearly pathogenic but there are no other variants in the gene associated with the disease
    - I.e. Huntington's Disease (*HTT*) or Myotonic Dystrophy 1 (*DMPK*)

Evidence that has an associated persistent identifier is preferred, typically a PMID. Check whether the database includes citations with PMIDs, as this will allow you to submit evidence more easily, and submit the DOI of anything without.

Lumping and Splitting

Give special care to cases where multiple tandem repeat loci are reported in the same gene. Many of the loci like this have only recently been defined and, in turn, much of the publication on these loci may not be as accurate as they appear. Unfortunately, these issues vary greatly so there is no set direction for going about this, instead, we urge scorers to make note of phenotypic and bioinformatic evidence and to use their best professional judgment. Isolating evidence from a phenotypic spectrum may be determined through symptoms or age of onset. Evidence from repeats on the same gene should be defined by coordinate; however, when that information is not available or viable, they may be defined by the motif, repeat range, or other positional clues. A common practice is scoring any unclear publications only under experimental evidence and not genetic. The following examples have been included to aid in deciding how to approach such diseases.

Hand-foot-genital (HFG) syndrome of *HOXA13*: This phenotype has been linked to tandem repeat expansions in three distinct polyalanine tracks, each separated by ~60 base pairs. This has led to the proposed three conditions: HFG-I, HFG-II, and HFG-III associated with the loci *HOXA13-I, HOXA13-II* and *HOXA13-III*, respectively. For most papers, the pathogenic range of the repeat gave us a clear sign of which type of HFG the publications were studying. Any of the papers that were still unclear were scored across all three types and only scored in the experimental evidence category. It should be noted that the pathogenic repeat range is usually not a viable option and should be used as a last attempt.

Diseases of *FMR1* (gene): Fragile X syndrome (FXS), fragile X-associated tremor/ataxia syndrome (FXTAS), and fragile X-associated primary ovarian insufficiency (FXPOI/POF1). Identifying FXS-focused papers was fairly straightforward, as most defined the repeat range, which is specific to the one disease. It became increasingly clear that FXTAS and FXPOI were quite entangled, as there were overlapping symptoms in XX individuals with both diseases, and the bioinformatic evidence was similar (if not the same). For this reason, we decided to lump FXTAS and FXPOI together and score FXS separately.

The *ARX* phenotypic spectrum: Expansions in *ARX* vary in length and location on the gene and cause a range of diseases. Partington syndrome (PRTS) and early infantile epileptic encephalopathy (EIEE1) were focused on specifically. Most studies were easily identified with one disease, but for those that were not, it was decided that they would only be scored in experimental evidence, but given to both diseases if the findings were relevant to both phenotypes. It is worth noting that there was an attempt to define the disease based on the coordinates, but that was not particularly fruitful.

Minimum Evidence

We strongly urge scorers and researchers to ensure that sufficient evidence is provided for a newly presented locus-disease relationship. Having this evidence will allow for future replication and clarity. Following is the minimum genetic and bioinformatic evidence:

- Proposed inheritance pattern
- Genomic coordinates
    - Chromosome, start, end, and reference genome build
    - For both controls and cases, coordinates should be either explicit or easy to derive
- Size in repeats of all alleles analyzed (control and case)
- Motif of repeat, including alternate motifs/interruptions and flanking repeats (if any)
    - Are there distinct motifs present in data?
    - What is the role of interruptions?
- Methodology of discovery/validation
    - Sequencing technology
    - Genotyping tool

Our minimum clinical evidence includes:

- Proband/patient medical history
    - Age of onset for first symptoms
    - Familial history of disease (positive or negative)
    - Frequency of symptoms if the studied population is in cohort
- Proposed allelic thresholds (benign and pathogenic ranges)
    - Generally based on population data and/or segregation analysis to define what is ‘not normal’, then molecular testing is used to see when something functionally changes

We also suggest the following evidence (although it is not required):

- Genomic context (coding, 5'UTR, etc)
- Gene strand
- Proband/patient ancestry
- HPO terms

Literature Search

The "Literature Search" section in ClinGen's most recent SOP[4] provides more guidance on how to do a specific and effective search for literature for gene-disease relationships, which will be very useful as a broad search. But, we recommend the scorer uses their best professional judgement to ensure that they are also finding research on locus-phenotype relationships. After a broad and inclusive search, it may be useful to limit the search with additional terms such as "tandem repeat" or "repeat expansion". For a list of useful websites, see Appendix A in ClinGen's most recent SOP[4].

Broad considerations for scoring

The default scoring of a piece of evidence can be increased by the strength of the evidence, as well as the addition of multiple pieces of evidence in the same publication. For instance, the default scoring of 0.5 for a proven biochemical alteration may be increased to 1 if there is additional protein characterization in the study. Scorers should try to score a range of different and independent publications where possible. Ensure that scoring follows the category's maximum score as well as the maximum overall score for experimental evidence. If the abnormal function of a variant in a non-human model was used to upgrade/augment the variant under genetic evidence, experimental data from the same study should not be used under an animal model in the experimental evidence section. Genetic evidence should receive priority over experimental evidence. If a study includes both a patient cell line and a non-human model, it may contribute to proband evidence under genetic evidence and to the non-human model organism category under experimental evidence; It should not receive a score under the patient cells category of experimental evidence unless another study has already filled this category. However, allele size/anticipation data, case-control information, and segregation evidence can be taken from the same study where probands are evaluated if there is proband-specific information available.

Additionally, evidence that can be attributed to a gene but not to a tandem repeat-specific locus can be given half the scoring value in experimental data and should not be used for genetic evidence. The highest priority of cataloging evidence should be given to locus-specific studies where there is an association made between the TR and phenotype, but gene-level findings can be informative in proving experimentally that there are phenotype consequences of alterations in the gene/region.

Evidence that the reviewer finds to be specious or questionable should be scored as 0, regardless of evidence type. Such evidence can lead to a limited or disputed classification.

Scoring Evidence

Scores of genetic and experimental evidence categories are added together. Total evidence may sum to a maximum of 18 (a maximum of 12 to genetic evidence and 6 to experimental), with maximum scores within categories as well. We suggest genetic evidence be prioritized over experimental. It is important to document where the evidence came from (PMIDs, ClinVarIDs, etc.); If evidence came from a review article or site like OMIM or GeneReviews, cite both the review and the original research. While we do offer default scores, we suggest an increase if there is multiple forms of evidence for the same category. Examples include having multiple animal cell models or evidence of association between allele length/motif and age of onset, penetrance, and/or severity. A piece of evidence being especially convincing is not grounds to increase the score, as there is no way to standardize this. It is also important to make note of why a certain score was given (i.e. “Study details two animal models, so a score of 4 was granted”).

Scoring Matrix for all Evidence

	Evidence Category	Evidence Type	Description	Default Score	Suggested Upgrades	Scoring Range	Maximum Score	Maximum Category Score
Genetic Evidence	Singular Evidence	Probands	Unrelated Probands	0.5	+0.5 for inheritance/de novo evidence, +0.5 for functional evidence	0-1.5 points per proband	6.0	6.0
	Collective Evidence	Allele	Relationship between allele and/or motif vs. age of onset and/or severity	1.0	Multiple forms of evidence: I.e Motif length affects both age of onset and severity	0-2	2.0	3.0
		Computational	Computational Pathogenicity Prediction	0.5	Multiple methods of prediction	0-1.5	3.0	
		Segregation	Linkage Region for disease	1.5	--	--	1.5	
	Statistics?	Case-Control Data	1. Variant detection methadology 2. Power 3. Bias and Confounding 4. Statistical significance	--	6: appropriately matched case and controls, no biases/confounding factors, highly statistically significant 4: no appropriately matched, p-value mdoerately statistically significant 2: aggregate/population data, p-value not very significant 0: detection method differs (may be case-level)	0-6	12.0	12.0
						Maximum Overall Score for all Genetic Evidence		12.0
	Evidence Category	Evidence Type	Description	Default Score	Suggested Upgrades	Scoring Range	Maximum Score	Maximum Category Score
Experimental Evidence	Function	Biochemical function	Enzymatic activity, DNA/RNA binding, structural roles, etc.	0.5	Multiple foms of evidence	0-2	2.0	2.0
		Protein Interaction	Protein product interacts with other disease-relevant proteins or complexes.	0.5	Multiple foms of evidence	0-2	2.0	
		Regulatory Impact	Including gene expression, epigenetic changes, or impacts on splicing.	0.5	Multiple foms of evidence	0-2	2.0	
	Functional Alteration	Patient Cells	Patient cells display disease-relevant phenotypes	1.0	Multiple patient cells studied in different experiments	0-2	2.0	2.0
		Non-patient Cells	Test the impact of candidate repeat variants in controlled conditions.	0.5	Multiple control sample cells studied in different experiments	0-1	1.0	
	Models	Non-Human Model Organism	Used to support genetic and phenotypic findings and should mimic the environment of the diseased human tissue.	2.0	Multiple models studied: I.e. Mouse and zebrafish models in the same experiment	0-4	4.0	4.0
		Cell Culture		1.0	Multiple different types of cells cultured	0-2	2.0	
	Rescue	Human treatment	Restoring gene function reverses disease phenotypes in humans	2.0	Rescues in multiple different healthy human cells	0-4	4.0	
		Non-Human Model Organism	Reversal of disease-like phenotypes by removing the repeat, restoring gene expression, expressing a normal allele or silencing the mutant allele.	2.0	Rescues in multiple moodels: I.e. Mouse and zebrafish	0-4	4.0	
		Cell Culture		1.0	Rescues in different types of cultured cells	0-2	2.0	
		Patient Cells		1.0	Rescues in multiple difernt patient cells or methods of rescue	0-2	2.0	
						Maximum Overall Score for All Experimental Evidence		6.0

Figure 2. *Scoring matrix for all evidence, genetic and experimental.*

Genetic Evidence

Genetic evidence may be derived from singular data (studies describing individuals or families with variants in the gene of interest), collective evidence, and/or statistical data (studies in which statistical analysis is used to evaluate enrichment of variants in cases compared to controls), but case-control data is preferred. While a single publication may include all three, individual

cases should not exceed the categories' maximum scores. For example, if a case from a case-control study were singled out for detailed discussion within the publication, and familial inheritance and pedigree information were provided, this case could be evaluated up to the maximum score and will not be used for any further evidence. The scorer should determine which is the stronger piece of evidence, and include those in the curation.

Singular Evidence

Proband evidence from case or family studies: Scoring is per unrelated proband, so a family with multiple affected individuals can only have one proband selected for scoring. The data may provide inheritance or functional evidence, but is not required to be awarded a score. This segment has a range of 0-6 total with 0.5 awarded for each proband and an additional 0.5 if there is functional data and 0.5 if there is data confirming that parent alleles were assessed and are consistent with the mode of inheritance—whether the variant is *de novo*, or other such familial information—for a total of 1.5 possible for each proband/family.

Collective Evidence

Repeat Length/Motif Consequences:

Data shows a relationship between allele size versus age of onset, anticipation, and/or between allele size versus disease severity. This segment has a range of 0-2 with a default of 1. A suggested increase in score may be evidence of correlations in age of onset and severity with allele size.

Computational Pathogenicity:

Predicted pathogenicity reflects evidence that leverages computational predictive strategies to infer the likelihood of pathogenicity for the locus. Such evidence may be assessed using appropriate constraint metrics or tandem-repeat specific tools such as RExPRT or SISTR[8]. Recent studies have proposed that the degree of repeat polymorphism in the general population is correlated with pathogenicity[5]. This segment has a range of 0-1.5 with a default score of 0.5. Suggested upgrades include evidence of multiple of the stated assessment measures[7].

Segregation

Evidence can be supported by the presence of a linkage region that overlaps with the candidate locus and tracks with disease status within families based on LOD scores. A score of 1.5 is granted to any study with a significant LOD score. For information about calculating an LOD score, please see the Segregation Analysis section of ClinGen's most recent SOP[4].

Statistical Data

Case-Control Data

This segment has a range of 0-6. Data that receives the maximum score should have appropriately matched cases and controls, with no biases or confounding factors, and should be highly statistically significant. Experiments where the samples are not appropriately matched and the p-value is moderately significant should receive a score of 4. Data using aggregate or population databases and with a hardly significant p-value, 2 should be awarded. In any

experiment in which the method of detection differs (such as case-level data), 0 should be awarded. Four types of evidence should be examined:

1. Variant Detection Methodology: Short-read or long-read sequencing, targeted PCR-based assays, or bioinformatic tools that infer repeat sizes from alignment patterns.
2. Power: The ability of a study to detect a true association between a genetic variant and a phenotype, depending on sample size, effect size, allele frequency, and phenotype specificity.
3. Bias and Confounding: Systematic errors that skew the results, or the presence of external variables that distort the true relationship between the variant and disease.
4. Statistical Significance: A significant p-value is reported in the literature.

Genetic Evidence Scoring Matrix

	Evidence Category	Evidence Type	Description	Default Score	Suggested Upgrades	Scoring Range	Maximum Score	Maximum Category Score
Genetic Evidence	Singular Evidence	Probands	Unrelated Probands	0.5	+0.5 for inheritance/de novo evidence, +0.5 for functional evidence	0-1.5 points per proband	6.0	6.0
	Collective Evidence	Allele	Relationship between allele and/or motif vs. age of onset and/or severity	1.0	Multiple forms of evidence: I.e Motif length affects both age of onset and severity	0-2	2.0	3.0
		Computational	Computational Pathogenicity Prediction	0.5	Multiple methods of prediction	0-1.5	3.0	
		Segregation	Linkage Region for disease	1.5	--	--	1.5	
	Statistics?	Case-Control Data	1. Variant detection methadology 2. Power 3. Bias and Confounding 4. Statistical significance	--	6: appropriately matched case and controls, no biases/confounding factors, highly statistically significant 4: no appropriately matched, p-value mdoerately statistically significant 2: aggregate/population data, p-value not very significant 0: detection method differs (may be case-level)	0-6	12.0	12.0
						Maximum Overall Score for all Genetic Evidence		12.0

Figure 3. *Scoring matrix for Genetic/Bioinformatic Evidence* (link to full-size figure)

Experimental Evidence

There are several forms of experimental and functional assays to elucidate gene function. For clinical validity classifications, only evidence that supports the role of a locus in a disease, or phenotypic features related to the disease entity of interest should be scored. Validated functional assays should be identified by expert panels or, if they are curator-identified, confirmed by expert review.

Functional Evidence

Biochemical Function

Evidence that the gene or repeat locus has a known molecular or cellular function relevant to the disease process. This may include enzymatic activity, DNA/RNA binding, transcriptional regulation, or structural roles. Disruption of this function by repeat expansion or mutation likely supports pathogenicity. This segment has a range of 0-2 with a default score of 0.5. Suggested upgrades may be multiple forms of evidence (i.e. both RNA binding and transcriptional regulation).

Protein Interaction

Data showing that the protein product of the gene interacts with other disease-relevant proteins or complexes. Repeat expansions that affect these interactions – either by disrupting binding or creating toxic interactions – support a functional role in disease. This segment has a range of 0-2 with a default score of 0.5, which can be increased based on having multiple experimental

methods of assessing protein interaction or by compelling simulations leveraging computational strategies.

Regulatory Impact

Evidence that the repeat expansion causes some regulatory impact, including gene expression, epigenetic changes, or impacts of splicing. Expression evidence may also show that the gene is expressed in disease-relevant tissues, or that expression levels change with the size of the repeat expansion. Evidence of epigenetic changes may prove that the repeat expansion leads to epigenetic modifications, such as DNA methylation, histone changes or chromatic remodeling. These may result in transcriptional silencing or altered expression patterns. Splicing impact may show evidence that a repeat expansion or nearby variant alters pre-mRNA splicing, leading to exon skipping, intron retention, or cryptic splice site use. This may be shown via RT-PCR, RNA seq, or splice reporter assays. This category has a range of 0-2, with a default score of 0.5, with a suggested upgrade for multiple forms of evidence supporting this category.

Evidence of Functional Alteration

Patient Cells

Functional studies using cells derived from affected individuals. These reflect disease-relevant genotypes and can show expression changes, aggregation, or other molecular phenotypes. This segment has a range of 0-2 with a default score of 1 and this may be upgraded if there are multiple patient cell types (i.e. hiPSCs and living patient cells).

Non-Patient Cells

Studies in generic or engineered cell lines that do not originate from patients, but are used to test the impact of repeat expansions or candidate variants in controlled conditions. This segment has a range of 0-2 with a default score of 0.5. This may be upgraded when there are multiple forms of these cells, that are NOT cultures. A prominent example would be when there are multiple control groups, such as looking at unaffected family members and unaffected subject data from a database.

Model-based Evidence

Evidence under this category should be used to support genetic and phenotypic findings and should mimic the environment of the diseased human tissue[12].

Non-human Model Organism

In vivo models used to test the effects of repeat expansions or gene loss/gain. These models may show behavioral, anatomical, or molecular phenotypes that mimic human disease. This segment has a range of 0-4 with a default score of 2 per model. This is the most commonly upgraded form of experimental evidence, upgrading for each model with a maximum of 4.

Cell Culture Model

Any *in vitro* cell-based assay, including patient-derived, immortalized, or stem-cell-derived models. Cell Culture Models differ from the Cell categories above in that they recapitulate features of the disease tissue environment. Generally used to study molecular mechanisms,

expression, and toxicity in a controlled setting. This segment has a range of 0-2 with a default score of 1.

Rescue Evidence
Suggested upgrades for the following evidence categories will be rescues in multiple types of cell categories or multiple rescues for the same cell type.

Human Treatment
Evidence that restoring gene function reverses disease phenotypes in human controls. This is strong evidence of causality but rarely available; therefore this segment has a range of 0-4 with a default score of 2.

Non-human model organism
Reversal of disease-like phenotypes in a non-human model by removing the repeat, restoring gene expression, or expressing a normal allele. Supports functional relevance of the locus and repeat. This segment has a range of 0-4 with a default score of 2.

Cell Culture Model
Evidence that introducing a normal gene or silencing the mutant allele in a cell line can reverse pathological features like aggregation, toxicity, or mis-splicing. This segment has a range of 0-2 with a default score of 1.

Patient Cells
Similar to above, but specifically in cells derived from affected individuals. Rescue may involve gene editing, knockdown of repeat-containing transcripts, or antisense approaches to restore cellular function. This segment has a range of 0-2 with a default score of 1.

	Evidence Category	Evidence Type	Description	Default Score	Suggested Upgrades	Scoring Range	Maximum Score	Maximum Category Score
Experimental Evidence	Function	Biochemical function	Enzymatic activity, DNA/RNA binding, structural roles, etc.	0.5	Multiple foms of evidence	0-2	2.0	2.0
		Protein Interaction	Protein product interacts with other disease-relevant proteins or complexes.	0.5	Multiple foms of evidence	0-2	2.0	
		Regulatory Impact	Including gene expression, epigenetic changes, or impacts on splicing.	0.5	Multiple foms of evidence	0-2	2.0	
	Functional Alteration	Patient Cells	Patient cells display disease-relevant phenotypes	1.0	Multiple patient cells studied in different experiments	0-2	2.0	2.0
		Non-patient Cells	Test the impact of candidate repeat variants in controlled conditions.	0.5	Multiple control sample cells studied in different experiments	0-1	1.0	
	Models	Non-Human Model Organism	Used to support genetic and phenotypic findings and should mimic the environment of the diseased human tissue.	2.0	Multiple models studied: I.e. Mouse and zebrafish models in the same experiment	0-4	4.0	4.0
		Cell Culture		1.0	Multiple different types of cells cultured	0-2	2.0	
	Rescue	Human treatment	Restoring gene function reverses disease phenotypes in humans	2.0	Rescues in multiple different healthy human cells	0-4	4.0	
		Non-Human Model Organism	Reversal of disease-like phenotypes by removing the repeat, restoring gene expression, expressing a normal allele or silencing the mutant allele.	2.0	Rescues in multiple moodels: I.e. Mouse and zebrafish	0-4	4.0	
		Cell Culture		1.0	Rescues in different types of cultured cells	0-2	2.0	
		Patient Cells		1.0	Rescues in multiple diferrt patient cells or methods of rescue	0-2	2.0	
						Maximum Overall Score for All Experimental Evidence		6.0

Experimental Evidence Scoring Matrix

Figure 4. *Scoring matrix for Experimental Evidence*

References

1. Alhathal, Naif, Sateesh Maddirevula, Serdar Coskun, Hamed Alali, Mirna Assoum, Thomas Morris, Hesham A. Deek, et al. 2020. “A Genomics Approach to Male Infertility.” *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 22 (12): 1967–75.
2. Balzotti, Marie, Linyan Meng, Dale Muzzey, Katherine Johansen Taber, Kyle Beauchamp, Myriad Genetics Curation Team, Baylor Genetics Curation Team, Rebecca Mar-Heyming, Bethany Buckley, and Krista Moyer. 2020. “Clinical Validity of Expanded Carrier Screening: Evaluating the Gene-Disease Relationship in More than 200 Conditions.” *Human Mutation* 41 (8): 1365–71.
3. Ceyhan-Birsoy, Ozge, Jaclyn B. Murry, Kalotina Machini, Matthew S. Lebo, Timothy W. Yu, Shawn Fayer, Casie A. Genetti, et al. 2019. “Interpretation of Genomic Sequencing Results in Healthy and Ill Newborns: Results from the BabySeq Project.” *The American Journal of Human Genetics* 104 (1): 76–93.
4. Clinical Genome Resource. (n.d.). Gene-disease validity standard operating procedure - ClinGen. Retrieved October 3, 2025, from https://clinicalgenome.org/docs/gene-disease-validity-standard-operating-procedure/
5. Danzi, Matt C., Isaac R. L. Xu, Sarah Fazal, Egor Dolzhenko, David Pellerin, Ben Weisburd, Chloe Reuter, et al. 2025. “Detailed Tandem Repeat Allele Profiling in 1,027 Long-Read Genomes Reveals Genome-Wide Patterns of Pathogenicity.” *bioRxivorg*. https://doi.org/10.1101/2025.01.06.631535.
6. Ellingford, Jamie M., Joo Wook Ahn, Richard D. Bagnall, Diana Baralle, Stephanie Barton, Chris Campbell, Kate Downes, et al. 2022. “Recommendations for Clinical Interpretation of Variants Found in Non-Coding Regions of the Genome.” *Genome Medicine* 14 (1): 73.
7. Fazak, S., et al., *RExPRT: a machine learning tool to predict pathogenicity of tandem repeat loci.* Genome Biology. 25:39.
8. Huang, Bonnie, Arun Durvasula, Nima Mousavi, Helyaneh Ziaei-Jam, Mikhail Maksimov, Kirk E. Lohmueller, and Melissa Gymrek. 2022. “Genome-Wide Selection Inference at Short Tandem Repeats.” *bioRxiv*. https://doi.org/10.1101/2022.05.12.491726.
9. “Illumina Clinical Services Laboratory Assertion Criteria for Gene Curation.” n.d. GeneCC. Accessed July 17, 2025. https://thegencc.org/docs/ICSL-Assertion-Criteria-for-Gene-Curation.pdf.
10. Machini, Kalotina, Ozge Ceyhan-Birsoy, Danielle R. Azzariti, Himanshu Sharma, Peter Rossetti, Lisa Mahanta, Laura Hutchinson, et al. 2019. “Analyzing and Reanalyzing the Genome: Findings from the MedSeq Project.” *The American Journal of Human Genetics* 105 (1): 177–88.
11. “Procedural Document Inventory of Genes Related to Rare Diseases (Version 3).” 2023. Orpha.net. October 2023. https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/Orphanet_Genes_inventory_R1_Ann_gen_EP_02.pdf.
12. Rehm, Heidi L., Jonathan S. Berg, Lisa D. Brooks, Carlos D. Bustamante, James P. Evans, Melissa J. Landrum, David H. Ledbetter, et al. 2015. “ClinGen--the Clinical Genome Resource.” *The New England Journal of Medicine* 372 (23): 2235–42.

13. Smith, Erica D., Kelly Radtke, Mari Rossi, Deepali N. Shinde, Sourat Darabi, Dima El-Khechen, Zöe Powis, et al. 2017. “Classification of Genes: Standardized Clinical Validity Assessment of Gene-Disease Associations Aids Diagnostic Exome Analysis and Reclassifications.” *Human Mutation* 38 (5): 600–608.