Supplemental Information

Supplemental Methods

Data collection methods: The HGMD (https://portal.biobase-international.com/hgmd/pro/start.php), ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/), GoogleScholar (https://scholar.google.com) and MasterMind (https://mastermind.genomenon.com) were utilized to identify relevant articles.

Case Information: Results from either the Caffeine Halothane Contracture Test (CHCT) or the *In Vitro* Contracture Test (IVCT) were considered valid diagnostic tests for determination of malignant hyperthermia (MH) status. Both tests rely on isolating muscle fibers from affected individuals and determining the strength of contraction when the fiber is exposed to caffeine and halothane. Individuals are determined to be MH-susceptible (MHS), or MH-negative (MHN). Only individuals who were diagnosed as MHS were considered to have a positive test for the purpose of variant assessment.

Explanation for criteria dropped from *RYR1* specific guidelines: Because the *RYR1/*MH pathogenic variant spectrum consists almost entirely of missense variants, with just a handful of small, in-frame deletions and no loss of function variants the following criteria were dropped; PVS1 (putative loss of function variants), PM4 (protein length change), and BP1 (missense variant in gene where predominantly loss of function variants cause disease). Because the majority of possible *RYR1* missense variants (~30,000 variants) are not represented in gnomAD v2.1.1 (2,800 *RYR1* missense variants) and many known pathogenic variants (assessed without the use of PM2) are present in gnomAD we dropped PM2. Because MH typically follows an autosomal dominant inheritance pattern, we dropped PM3 (in *trans* to pathogenic variant for recessive disorder). Because *RYR1* missense variants are not uncommon in gnomAD, we dropped PP2 (low rate of benign missense variants in gene). Because multiple loci contribute to MH, we dropped PP4 (patient's phenotype specific for disease with single genetic etiology). It has been reported that up to 12% of family members that test positive by IVCT/CHCT are negative for the familial *RYR1*/MH variant. Therefore, we dropped BS4 (lack of segregation). While it has been suggested that the CHCT/IVCT diagnostic test has a high false positive rate, it is possible that other MH pathogenic variants exist in these families that account for this finding (Miller et al. 2018) and it was not felt to be a reliable criterion for this disorder. Because *RYR1* does not have regions of sequence repeats, we dropped BP3 (variant in repetitive region without known function). There have been reports of individuals with MHS who have been found to have two pathogenic *RYR1* variants (Kraeva et al. 2011) and at least one family has been identified with pathogenic variants in both *RYR1* and *CACNA1S*.(Monnier et al. 2002) Based on these reports, we dropped the BP5 criterion (alternate molecular basis for disease).

Rules For Combining Pathogenic Criteria
The Bayesian Classification Framework developed by Tavtigian et al. 2018 is used.

Sum all criteria that are applicable to the variant. Calculate Odds of Pathogenicity using formula below, calculate posterior probability, use posterior probability to determine pathogenicity.

Odds of Pathogenicity = $2.1^{\#\text{Total Supporting}} * 4.3^{\#\text{Total Moderate}} * 18.7^{\#\text{Total Strong}} * 350^{\#\text{Total V Strong}} * 0.4808^{\#\text{Benign Supporting}} * 0.2326^{\#\text{Benign Moderate}} * 0.0535^{\#\text{Benign Strong}}$

Posterior Probability = (Odds Path * 0.1) / (Odds Path - 1) * 0.1+1)

Assignment of Pathogenicity based on Posterior Probability:

Posterior Probability < 0.001	Benign
Posterior Probability ≥ 0.001 < 0.1	Likely Benign
Posterior Probability ≥ 0.10 < 0.9	VUS
Posterior Probability ≥ 0.9 to < 0.99	Likely Pathogenic
Posterior Probability ≥ 0.99	Pathogenic

How to evaluate evidence strength for Bayesian pathogenicity criteria using likelihood ratios.

The general form of the Bayes equation is:

P(A|B) = P(B|A)*P(A)/P(B)

In pathogenicity assessments:

A is the pathogenicity
B is the evidence

So, P(Path|Evid) is “The probability of pathogenicity given the applied evidence”, which is what we are setting out to determine. The vertical bar (pipe) means ‘given’. This is also called the posterior probability, that is the pathogenicity of the variant after applying the evidence.

To do that we make a calculation, which is based on three factors:

P(Evid|Path) is “The probability of the evidence, given a pathogenic variant”
P(Path) is “The probability of a pathogenic variant”
P(Evid) is “The probability of observing the evidence”

To make things even easier, this equation can be rearranged and simplified using some unfamiliar, but very handy terms.

Prior Probability – the likelihood of observing something before a piece of evidence is observed.
Conditional Probability – this is P(Evid|Path) – the likelihood of observing the evidence if the variant is pathogenic. This can be expressed in the form of the Odds of Pathogenicity for a given piece of evidence, which we call OddsP

Posterior Probability – this is P(Path|Evid)

Prob(Path|Evid) = (OddsP*Prior)/((OddsP-1)*Prior+1)

From Tavtigian et al, the 'OddsP' is equal to the product of the Odds of pathogenicity of all of the criteria that one would use in the assessment.
In the current Bayesian formulation of the ACMG/AMP criteria, the prior probability is 0.1.

Richards et al implicitly set relative OddsP for each criterion, using the categories of 'Very Strong', 'Strong', 'Moderate', and 'Supporting'. Tavtigian et al transformed those into OddsP of 350:1, 18.7:1, 4.3:1, and 2.08:1, respectively

It is worth noting that these conditional OddsP can also be expressed as probabilities. The formula for converting odds to probabilities is:
Probability = Odds/((1 + Odds)
So, 'Very Strong' is 350:1, or 99.7% conditional probability, 'Strong' is 18.7:1, or 94.9% conditional probability, 'Moderate' is 81.1% conditional probability, and 'Supporting' is 67.5% conditional probability.

In this system, the OddsP of each criterion is multiplied together to determine the overall OddsP. Here are some examples of how to estimate the strength of these OddsP for each of several criteria.

***In silico* data analyses**

Let us say we are trying to evaluate the strength of evidence for an *in silico* predictor for variants in *RYR1* because we want to know how to use that evidence. For simplicity I will illustrate this using SIFT. The simplest way to do this is to just use the categorical output of SIFT as recommended by the programmers – either 'Damaging' or 'Tolerated'. To evaluate the SIFT predictor strength for PP3/BP4, we look at a set of variants we are confident are pathogenic and another set that we are confident are benign that were evaluated without using *in silico* predictors. For 20 known pathogenic *RYR1* variants, ~100% of the time (20 variants) the *in silico* predictor reads out damaging, and ~0% of the time (0 variants) it reads out 'Tolerated'. For known benign variants, 41% of the time it reads out damaging (11 variants), and 59% of the time (16 variants) it reads out tolerated.

SIFT for the *RYR1* sample set of 20 pathogenic variants and 27 benign variants.

	Path	Benign
Damaging	20	11
Tolerated	1*	16

The likelihood ratio for a Damaging readout (LR+) or a Tolerated readout (LR-) for a variant of unknown pathogenicity requires use of the Sensitivity and Specificity of the *in silico* predictor.

*We need a “1” for this position to allow for the calculations to work. Note that this heuristic leads to a slight underestimate of the utility of the predictor but is necessary, otherwise these data would suggest that a prediction of ‘Tolerated’ is a perfect predictor of pathogenicity, which we know it not to be.

Sensitivity is:
True Positives/(True Positives + False Negatives) = 20/(20+1) = 0.952

Specificity is:
True Negatives/(True Negatives + False Positives) = 16/(16+11) = 0.593

LR+ = Sensitivity/(1-Specificity) =
{(TP/(TP+FN)}/[{(1-{TN/(TN+FP)}] =
{(20/(20+1)}/[{(1-{16/(16+11)}] = 2.34, or 2.34:1

Note that the online LR calculators allow you to derive this directly from the 2 x 2 table.

So, the OddsP for a SIFT readout of ‘Damaging’ is between Supporting (2.08:1) and Moderate (4.3:1) and should be used as Supporting. Note that an LR+ of 1.0 would be no data for or against pathogenicity – the odds are 1:1. When you use the online LR calculators, they also give you the 95% confidence interval of the LR, which in this case is 1.47-3.72. It is a bit worrisome that the lower bound of the CI is below the specified LR for supporting evidence (2.08:1), suggesting that this limited dataset does not provide robust support for implementing this as supportive evidence. There is debate about this as to whether one should use point estimates for such data or the conservative bound of a 95% CI, but that is beyond the scope of this guide.

The likelihood ratio for a Tolerant readout (LR-) for a variant of unknown pathogenicity would be:

LR- = (1-Sensitivity)/Specificity = [1-{TP/(TP+FN)}]/{TN/(TN+FP)} = [1-{20/(20+1)}]/{16/(16+11)} = 0.08

So, the OddsP for a SIFT readout of ‘Tolerant’ is~ 0.08:1, or about 1:13 *for* pathogenicity, or 13:1 *against* pathogenicity, or 13:1 *for* benign. This would lie between Moderate (4.3:1) and Strong (18.7:1). Note that Richards et al did not have an evidence level of Moderate for Benign evidence. So, in that framework one would have to use this as benign supporting, or perhaps two benign supporting pieces of evidence. In Tavtigian et al, Formula 5 could be amended from the published form:

$$OP = O_{VSt}^{\left(\frac{N_{PSu}}{8}+\frac{N_{PM}}{4}+\frac{N_{PSt}}{2}+\frac{N_{PVSt}}{1}-\frac{N_{BSu}}{8}-\frac{N_{BSt}}{2}\right)}$$

To:

$$OP = O_{VSt}^{\left(\frac{N_{PSu}}{8}+\frac{N_{PM}}{4}+\frac{N_{PSt}}{2}+\frac{N_{PVSt}}{1}-\frac{N_{BSu}}{8}-\frac{N_{BMo}}{4}-\frac{N_{BSt}}{2}\right)}$$

By adding the term highlighted in yellow, which counts Moderate Benign evidence, this evidence strength can then be calculated into the overall pathogenicity (one could accomplish the same thing by simply coding one moderate benign evidence code as two supporting evidence codes.)

One can also consider a trichotomization* of *in silico* data, rather than dichotomization. In this approach one defines three categories of: 1) above a certain threshold (pathogenic evidence), 2) below a different threshold (benign evidence) and 3) in between the two thresholds (no evidence). Here we are using REVEL analysis of the same 20 pathogenic and 27 benign *RYR1* variants. The REVEL output is numerical, which facilitates trying different thresholds with a gene of interest and the known variants. After trying several different sets of thresholds we set an upper threshold of ≥0.85 and a lower threshold of ≤0.50. To derive Odds ratios for these data we use what is called a 2 x k odds ratio calculator (several are available online – I use https://www.scistat.com/statisticaltests/likelihoodratios.php) where we set k=3, since the data will be in the form of a 2 x 3 table. Note again that we have to insert a value of 1 in the cell for the 'Path;≤0.50' cell even though REVEL did not report any of the 22 known path variants with a score below the lower threshold. This is important because using zero in any of the cells makes the calculations problematic. As noted above, if there is a zero in that cell, the LR for the ≤0.50 category would come out as zero, which means infinite or perfect classification, which is, as we know, impossible. By putting a one in this cell of the LR table I am in effect making my data worse, and thus giving a more conservative, but more useful, estimate of the likelihoods.

	Path	Benign	LR+	95% CI
≥0.85	17	1*	22.68	3.27-157.08
>0.50-<0.85	3	8	0.50	0.15-1.66
≤0.50	1*	19	0.07	0.01-0.48

The LR+ for a readout of ≥0.85 would therefore be 22.68:1, above strong. Note that the 95% confidence interval calculated here ranges from above supporting to above strong. The LR for the ≤0.50 category is 0.07:1 in favor of pathogenicity, which is 14.29:1 in favor of benign. This is above moderate, although the 95% CI lies between 2.08:1 benign (between supporting and moderate) to 100:1 benign, which is between strong and very strong.

The intermediate range of REVEL output (0.5-0.85) gives an LR of 0.50:1 pathogenic, or 2.00:1 benign. Note here that the 95% CI ranges from 0.15:1 pathogenic (6.62:1 benign) to 1.66:1

* One could further subdivide this into more than three categories using a 2 x 4 or 2 x 5 table. The yet more sophisticated approach would be to perform linear regression on the data and generate LR+ outputs directly from the REVEL readout for each variant and then add that in to the Tavtigian et al equation 5 above, but that is beyond the scope of this analysis.

pathogenic. This should give one pause that it is not good evidence either way – if the 95% CI broadly flanks 1.0 (no evidence) you are in a middle gray zone and should not use the evidence at all. This makes intuitive sense – there are thresholds for evidence such that the extremes are used as evidence (one end for benign and the other for pathogenic) and the middle range is not.