REPORT

In silico versus functional characterization of genetic variants: lessons from muscle channelopathies

Vinojini Vivekanandam, Rebecca Ellmers, Dipa Jayaseelan, Henry Houlden, Roope Männikkö and Michael G. Hanna

Abstract

Accurate determination of the pathogenicity of missense genetic variants of uncertain significance is a huge challenge for implementing genetic data in clinical practice. In silico predictive tools are used to score variants' pathogenicity. However, their value in clinical settings is often unclear since they have usually not been validated against robust functional assays. We compare nine widely used in silico predictive tools including more recently developed tools (EVE and REVEL) with detailed cell-based electrophysiology for 126 *CLCN1* variants discovered in patients with the skeletal muscle channelopathy myotonia congenita. We found poor accuracy for most tools. The highest accuracy was with Mutation Taster (84.58%) and REVEL (82.54%). However, both scores have poor specificity. EVE has better specificity. Combined methods based on concordance, improved performance overall but still lacked specificity. Our calculated statistics for the predictive tools are different to reported values for other genes in the literature suggesting that utility of the tools varies between genes. Overall, current predictive tools for this chloride channel are not reliable for clinical use and tools with better specificity are urgently required. Improving the accuracy of predictive tools is a wider issue and a huge challenge for effective clinical implementation of genetic data.

Author Affiliations:

Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London UK

Correspondence to: Vinojini Vivekanandam
Box 102
The National Hospital for Neurology and Neurosurgery
Queen Square, London
WC1N 3BG
E-mail: v.vivekanandam@ucl.ac.uk

Running title: Genetic variant prediction in channelopathies

Keywords: muscle channelopathies; genetic variants; variants of uncertain significance; myotonia

Abbreviations: SNVs = single nucleotide variants; ACMG = American College of Medical Genetics; ACGS = Association for Clinical Genomic Science; MC = myotonia congenita

Introduction

The advent of next generation sequencing and whole genome sequencing is generating unprecedented volumes of genetic data. Accurate interpretation of novel variants of uncertain significance in the clinical context is arguably one of the biggest challenges in genomic medicine. Accurate classification is paramount. Falsely rejecting pathogenic variants leads to unnecessary ongoing search for the underlying genetic cause and a missed diagnosis. While attributing pathogenicity incorrectly has significant consequence for patients and their family. Several in silico predictive algorithms have been developed to assist in determining pathogenicity of missense single-nucleotide variants and are routinely used. However, their efficacy and reliability in specific genes requires assessment.

The prediction tools considered in the variant scoring framework from the American College of Medical Genetics (ACMG) include PolyPhen-2, SIFT, Align-GVGD and MutationTaster2[1]. These tools consider the nature of the substituting amino acid and the conservation of the substituted amino acid residue. More recently, metapredictors such as REVEL have been developed, that predict pathogenicity based on a combination of individual tools[2]. In 2021, EVE, a predictive model developed with deep generative models based on evolutionary data was released[3]. The Association for Clinical Genomic Science (ACGS) and diagnostic laboratory guidelines consider concordance of tools, in building support for pathogenicity of novel variants[4,5].

While studies comparing efficacy of these tool in specialities such cancer, audiology and cardiology have been performed, few studies have been conducted in neurology[1,6]. Moreover, several previously performed studies compare in silico predictive algorithms to databases such as ClinVar, which introduces concerns regarding circular comparisons - ClinVar variant characterisations already take in silico predictions into consideration[7,8]. To our knowledge, aside from validation performed by the authors of EVE, there have been no other comparison to EVE scores.

Ion channels provide an attractive model system for comparison of predicted and recorded measures of pathogenicity as electrophysiological data assessing function of channel variants is often readily available. In particular, as part of the diagnostic platform for myotonia we routinely characterize the function of *CLCN1* variants identified in patients with myotonia. *CLCN1*

encodes skeletal muscle chloride voltage gated channel 1 (CLC-1) that regulates electrical excitability of the muscle[9]. Variants that lead to reduction in chloride conductance increase muscle membrane excitability causing myotonia[10]. Myotonia can be caused by several conditions, of which myotonia congenita is the most common form of non-dystrophic myotonia. Myotonia congenita can be inherited in an autosomal dominant or autosomal recessive manner. We compare in silico predictive tools to the pathogenicity as determined by functional in vivo classification of variants in *CLCN1*.

Materials and methods

Our dataset includes 126 CLC-1 missense variants functionally characterised as a part of the diagnostic platform of skeletal muscle channelopathies. Assessment of pathogenicity for several of these variants was recently reported[11].

In silico prediction

Alamut Visual 2.15 -64bit (SOPHiA GENETICS, Lausanne, Switzerland) was utilised to determine pathogenicity scores and classifications with the tools PolyPhen-2, Align-GVGD (a-GVGD), SIFT and Mutation Taster[1,12–16]. GnomAD frequencies and Grantham distance were also extracted. Ensembl was utilised to determine pathogenicity scores and classifications for REVEL, MetaLR, CADD and Mutation Assessor[17–22]. EVE scores were taken from the EVE platform[3].

Functional determination

Methods for generation of channel variants, expression of channel variants in *Xenopus* oocytes, electrophysiological analysis using two-electrode voltage-clamp and criteria for determination of pathogenicity was recently described in Suetterlin et al[11]. Briefly, if the voltage of half-maximal activation was positive to cut-off value of -18.6 mV or if the channel variant did express no or only minimal ClC-1 currents the variant was considered pathogenic (Figure 1A). Variants with other loss-of-function features as reported in Suetterlin et al. were also considered pathogenic.

Statistical analyses were performed using Excel version 16.65 and IBM SPSS version 26 and expressed as specificity, sensitivity, positive predicative value, negative predictive value, accuracy and receiver operating characteristics (ROC) curve. The following equations were used:

Sensitivity = True positive (TP)/(TP + False negative (FN))

Specificity = True negative (TN)/(TN + False positive (FP))

Positive predictive value (PPV) = TP/(TP + FP)

Negative predictive value (NPV) = TN/ (TN + FN)

Accuracy = (TP + TN)/(TP + TN + FP + FN)

Data availability

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Results

Of a total of 126 *CLCN1* variants, based on in vivo functional characterisation, 91 were pathogenic and 35 were benign (Figure 1B). The variants were considered pathogenic if the half-maximal voltage dependence of activation was positive to -18.6 mV and the peak tail-current amplitude at -100 mV was smaller than -2.5 µA. For the variants with reduced current amplitude most of the cells did not show any currents[11]. In addition, variants with other loss-of-function features that could not be characterised in terms of voltage of half maximal activation or current amplitude were also considered pathogenic[11].

Comparing the prediction tools, Mutation Taster, REVEL, EVE and PolyPhen had above 80% accuracy. Sensitivity, specificity, positive and negative predictive values as well as accuracy for each tool is shown in Table 1.

Assessing ROC curves on sensitivity vs. specificity plots demonstrated that better predictive tools are EVE, Mutation Taster, MetaLR and REVEL for *CLCN1* (Figure 1B). The highest area under the curve (AUC) score was for REVEL (Table 2).

When considering concordance of different tools as is done using ACMG criteria, ACGS recommendations as well as diagnostic lab consensus, with three of four tools requiring concordance to be accepted, we found that 79 variants were classified correctly, 12 were incorrectly classified and 34 were unable to be classified due to a lack of concordance, Table 3. The tools commonly used when applying the ACMG criteria are PolyPhen, SIFT, Mutation Taster and aGVGD.

We looked at concordance with REVEL, MetaLR, Mutation Taster and EVE as these four scores had good AUC and specificity based on our data. Three of four scores required concordant predictions for their predictions to be included. Using these scores, 100 variants were classified correctly, 14 incorrectly classified and 12 were unable to classified due to a lack of concordance. Although more variants were able to be classified using concordance of these 4 scores (REVEL, MetaLR, Mutation Taster and EVE) with good accuracy and sensitivity, the specificity was reduced to 0.48, Table 3. When MetaLR was no longer included, due to its poor individual specificity, the resultant concordant specificity for the three scores (REVEL, Mutation Taster and EVE) was improved to 0.65, as shown in Table 3.

In *CLCN1*, location of variants has been previously shown to be important[9,11]. Variants in the intracellular domain are more likely to be benign while those in the transmembrane domains are more likely to be pathogenic. In our data set, 27 variants were intracellular and 99 in the transmembrane domain. Looking at our concordance analysis, using REVEL + Mutation Taster + EVE, 12 of 27 (44.44%) variants in the intracellular domain were predicted correctly and 88 of 99 (88.89%) in the transmembrane domain were predicted correctly. When using the ACMG/ACGS guidelines based tools (Polyphen + SIFT + Mutation Taster + aGVGD) 17 of 27 (62.96%) variants in the intracellular domain were predicted correctly and 63 of 99 (63.63%) variants in transmembrane domain.

Discussion

In silico prediction tools are commonly used to score novel variants but their validity is often unclear. To assess this requires comparison against robust datasets assessing clinical and

functional features of the variants. We perform a comparison of functional features of ClC-1 variants against in silico tools. While Mutation Taster, REVEL, EVE and PolyPhen had above 80% accuracy and relatively good sensitivity over 0.8, the specificity for all four tools is poor. Of these four, EVE has the best specificity at 0.7. This specificity is far from ideal for clinical application but remains much better than the specificity of the other three tools, with good accuracy and sensitivity. EVE is trained only on evolutionary sequences which lends itself to having a higher degree of specificity[3].

The AUC of EVE is 0.8. While this is a good score, it is below REVEL, Mutation Taster, MetaLR and SIFT. The REVEL AUC score is high at 0.89. This is not surprising given the mechanism of REVEL which combines several individual tools as a meta-predictor. The best scores based on AUC for in silico prediction in *CLCN1* were REVEL, Mutation Taster and MetaLR.

The AUC values we report for *CLCN1* are lower than other reported AUC values in the literature. When EVE was compared to ClinVar datasets an AUC of 0.91 was reported[3]. Similarly REVEL was compared to SwissVar with an AUC of 0.908[18]. MetaLR has a reported AUC of 0.883. However, such comparison are inherently circular as such databases (ClinVar, SwissVar) incorporate in silico predictive algorithms in categorising variants as benign or pathogenic[8,23].

Few studies compare predictive algorithms to variants that have been functionally characterised in vitro. Similar characterisation can be performed with genes responsible for cardiac

channelopathies causing long QT syndromes[24]. When in silico prediction tools were compared to *KCNQ1*, *KCNH2* and *SCN5A* variants characterised in vivo or by co-segregation, AUC for PolyPhen was 0.77 for all genes combined and 0.715 for SIFT. When looking at individual genes the AUC varied from 0.63 to 0.94 using the same score (PolyPhen).

Comparing PolyPhen, Sift and Mutation Taster to functional characterisation of *RYR1* variants using in vitro contracture tests on muscle biopsies, demonstrated an AUC of 0.94 (PolyPhen), 0.98 (Sift) and 0.92 (Mutation Taster)[25]. These values are much higher than the AUC values we demonstrated in *CLCN1*.

These studies demonstrate clear differences in the AUC for in silico predictive tools for different genes. It is likely that this is due to variations in complex factors such as penetrance and pattern of inheritance. This is important to consider when interpreting a novel variant. Reported AUC, specificity and sensitivity for in silico predictive tools should not be applied generally to all genes.

Using concordance of several tools appears to improve performance. Concordance (all three in agreeance) between the REVEL, Mutation Taster and EVE improved accuracy, sensitivity, positive and negative predictive value compared to the ACMG or diagnostic lab based tools (3 of 4 concordant out of Polyphen, Sift, Mutation Taster and aGVGD). The specificity was slightly reduced which is a recurring issue across all in silico predictive tools. However, an accuracy of 90% makes a compelling case for considering the use of the newer predictive tools and concordance in the interim, while better tools are developed. Additionally, these tools appear to

be in line with differentiating variant pathogenicity based on variant location within the gene. Domains and loci of variants are likely to be useful aspects to include in the design of future predictive tools[26]. However, pathogenicity of variants in some functional domains may not be accessible with certain functional analyses, for example some CLC-1 intracellular variants that assert pathogenicity by disrupting muscle-specific protein interactions[27]. Practically, we suggest that variants in domains that are less well conserved are those that particularly require functional studies. In ClC-1 for example, variants outside the transmembrane domain are less well conserved[11].

The correlation of functional features with clinical characteristics such as inheritance patterns is not 100% and is expected for skeletal muscle channelopathies where variants show variable clinical features within and between pedigrees. Also, depending on the type of functional analysis only certain forms of pathogenicity can be detected – for example exonic variants affecting splicing or tissue specific interaction will not be picked using heterologous expression and electrophysiological analysis. However, functional expression is a strong indicator of pathogenicity, and is classified as such in the ACMG criteria. This creates a robust dataset, in particular compared to ClinVar based datasets, where some variants are reported without any indicators of pathogenicity.

Additional limitations to functional analysis include the time taken, the labour-intensive process and technical expertise needed. Functional expression of a new variant can take months depending on the assay. Not all genes, and indeed not all channel genes can be expressed. High throughput electrophysiology platforms which utilise automated multi-channel patch-clamping

may overcome some barriers, time in particular. However, the initial purchase cost and cost per data point of these platforms is significant. High throughput platforms may be an option in the future as access and costs are reduced. To perform traditional functional expression, significant equipment and technical experience is required. In contrast, more accurate in silico tools could be applied by clinicians, geneticists and bioinformaticians.

As per the ACGM guidelines multiple lines of in silico predictions provide supporting or moderate evidence of a variant being benign or pathogenic while functional analyses can provide strong indications. Thus, currently, in the case where a functional assay is available, it should be sought. In silico predictions can provide preliminary estimation that may precede functional analyses by months or years, and following the functional analysis provide supportive evidence for pathogenicity of the variant. In the absence of functional assays or other strong indicators the in silico predictive tools are part of the main pathway to assess the pathogenicity. Developing improved predictive tools that are more specific is a key area of need in genomics, particularly for genes without a method for functional assessment. At present, genes without robust expression systems are more limited to accuracy achieved with traditional parameters such as conservation, nature of mutation, mutation hotspots or clinical validation such as segregation testing which may not always be possible.

Ideally, improved algorithms would be developed that can be rapidly applied to new variants and newer machine learning techniques may see this happen. Machine learning techniques such as multi-task learning on channel data sets have been utilised to develop models to predict variant pathogenicity. In addition, the algorithms may incorporate homology modelling approaches[28].

However the key challenge remains in using a large enough data set to train an algorithm without compromising the validity of the data included[29,30]. Larger data sets with more inclusive data tend to incorporate unvalidated data points. For example, the multi-task learning support vector machine (MTL-SVM) model for potassium channels is trained on some data that is non-human and may not appear in a disease context[30].

At present, clinical assessment incorporating functional and in silico predictions is imperative. Other causes of myotonia need to be considered and excluded. In patients with other causes of myotonia, for example myotonic dystrophy, pathogenic *CLCN1* variants can alter the phenotype and must be considered in clinical assessment. For some variants, electrophysiological patterns may not be able to determine mode of inheritance and clinical assessment will be important for genetic counselling.

Our study in *CLCN1* using a robust data set and comparing to newer predictive models supports data in other fields of medicine illustrating the poor utility of current in silico predictive tools. Overall, tools with improved specificity while maintaining good sensitivity are urgently required with assessment in the future performed against robust data sets that have been functionally validated. Importantly, AUC, specificity and sensitivity of the predictive tools varies between genes and requires independent assessment for each gene. While the predictive tools may support in scoring a variant, functional assessment of the variant is warranted where possible.

Acknowledgements

MGH is supported by an MRC strategic award for International Centre for Genomic Medicine in Neuromuscular Diseases. VV is supported by the HSS for Channelopathies, Mission Possible (National Brain Appeal) and The Jon Moulton Charity. Our work is supported by the UCLH NIHR Biomedical Research Centre. We provide the only NHS England commissioned national diagnostic service and references laboratory for children and adults muscle channelopathies led from University College London NHS Foundation Trust – for further details contact m.hanna@ucl.ac.uk.

Competing interests

The authors report no competing interests.

References

1. Cubuk C, Garrett A, Choi S, et al. Clinical likelihood ratios and balanced accuracy for 44 in silico tools against multiple large-scale functional assays of cancer susceptibility genes. *Genet Med*. 2021;23(11):2096-2104. doi:10.1038/s41436-021-01265-z
2. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet*. 2016;99(4):877-885. doi:10.1016/j.ajhg.2016.08.016
3. Frazer J, Notin P, Dias M, et al. Disease variant prediction with deep generative models of evolutionary data. *Nature*. 2021;599(7883):91-95. doi:10.1038/s41586-021-04043-8

4. Karczewski K, Francioli L, Tiao G, et al. ACGS Best Practice Guidelines for Variant Classification in Rare Disease 2020. *bioRxiv*. 2019:531210.
5. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. doi:10.1038/gim.2015.30
6. Ernst C, Hahnen E, Engel C, et al. Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics. *BMC Med Genomics*. 2018;11(1):1-10. doi:10.1186/s12920-018-0353-y
7. Gunning AC, Fryer V, Fasham J, et al. Assessing performance of pathogenicity predictors using clinically relevant variant datasets. *J Med Genet*. 2021;58(8):547-555. doi:10.1136/jmedgenet-2020-107003
8. Landrum MJ, Lee JM, Benson M, et al. ClinVar: Public archive of interpretations of clinically relevant variants. *Nucleic Acids Res*. 2016;44(D1):D862-D868. doi:10.1093/nar/gkv1222
9. Vivekanandam V, Männikkö R, Matthews E, Hanna MG. Improving genetic diagnostics of skeletal muscle channelopathies. *Expert Rev Mol Diagn*. 2020;00(00):725-736. doi:10.1080/14737159.2020.1782195
10. Cannon SC. Pathomechanisms in Channelopathies of Skeletal Muscle and Brain. *Annu Rev Neurosci*. 2006;29(1):387-415. doi:10.1146/annurev.neuro.29.051605.112815
11. Suetterlin K, Matthews E, Sud R, et al. Translating genetic and functional data into clinical practice: a series of 223 families with myotonia. *Brain*. 2021;(2021). doi:10.1093/brain/awab344

12. Vaser R, Adusumalli S, Leng SN, Sikic M, Ng PC. SIFT missense predictions for genomes. *Nat Protoc*. 2016;11(1):1-9. doi:10.1038/nprot.2015.123
13. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010;7(4):248-249. doi:10.1038/nmeth0410-248
14. Mathe E, Olivier M, Kato S, Ishioka C, Hainaut P, Tavtigian S V. Computational approaches for predicting the biological effect of p53 missense mutations: A comparison of three sequence analysis based methods. *Nucleic Acids Res*. 2006;34(5):1317-1325. doi:10.1093/nar/gkj518
15. Tavtigian S V., Deffenbaugh AM, Yin L, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet*. 2006;43(4):295-305. doi:10.1136/jmg.2005.033878
16. Schwarz JM, Cooper DN, Schuelke M, Seelow D. Mutationtaster2: Mutation prediction for the deep-sequencing age. *Nat Methods*. 2014;11(4):361-362. doi:10.1038/nmeth.2890
17. Howe KL, Achuthan P, Allen J, et al. Ensembl 2021. *Nucleic Acids Res*. 2021;49(D1):D884-D891. doi:10.1093/nar/gkaa942
18. Ioannidis NM, Rothstein JH, Pejaver V, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet*. 2016;99(4):877-885. doi:10.1016/j.ajhg.2016.08.016
19. Liu X, Jian X, Boerwinkle E. dbNSFP: A lightweight database of human nonsynonymous SNPs and their functional predictions. *Hum Mutat*. 2011;32(8):894-899. doi:10.1002/humu.21517
20. Liu X, Li C, Mou C, Dong Y, Tu Y. dbNSFP v4: a comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site

SNVs. *Genome Med*. 2020;12(1):1-8. doi:10.1186/s13073-020-00803-9

21. Dong C, Wei P, Jian X, et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet*. 2015;24(8):2125-2137. doi:10.1093/hmg/ddu733

22. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886-D894. doi:10.1093/nar/gky1016

23. Mottaz A, David FPA, Veuthey AL, Yip YL. Easy retrieval of single amino-acid polymorphisms and phenotype information using SwissVar. *Bioinformatics*. 2010;26(6):851-852. doi:10.1093/bioinformatics/btq028

24. Leong IUS, Stuckey A, Lai D, Skinner JR, Love DR. Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. *BMC Med Genet*. 2015;16(1):1-13. doi:10.1186/s12881-015-0176-z

25. Hoppe K, Jurkat-Rott K, Kranepuhl S, et al. Relevance of pathogenicity prediction tools in human RYR1 variants of unknown significance. *Sci Rep*. 2021;11(1):1-11. doi:10.1038/s41598-021-82024-7

26. Samocha KE, Kosmicki JA, Karczewski KJ, et al. Regional missense constraint improves variant deleteriousness prediction. *bioRxiv*. 2017:148353. https://www.biorxiv.org/content/10.1101/148353v1%0Ahttps://www.biorxiv.org/content/10.1101/148353v1.abstract.

27. Park E, Mackinnon R. Structure of the CLC-1 chloride channel from Homo sapiens. 2018. doi:10.7554/eLife.36629.001

28. Sallah SR, Sergouniotis PI, Barton S, et al. Using an integrative machine learning approach utilising homology modelling to clinically interpret genetic variants: CACNA1F as an exemplar. *Eur J Hum Genet*. 2020;28(9):1274-1282. doi:10.1038/s41431-020-0623-y

29. Heyne HO, Baez-Nieto D, Iqbal S, et al. Predicting functional effects of missense variants in voltage-gated sodium and calcium channels. *Sci Transl Med*. 2020;12(556):8-11. doi:10.1126/SCITRANSLMED.AAY6848

30. Boßelmann CM, Hedrich UBS, Müller P, et al. Predicting the functional effects of voltage-gated potassium channel missense variants with multi-task learning. *eBioMedicine*. 2022;81:104115. doi:10.1016/j.ebiom.2022.104115

Figure legends

Figure 1 Functional assessment of pathogenicity of ClC-1 variants and comparison to in silico predictive tools. A) Voltage of half-maximal activation ($V_{1/2}$) is plotted against current amplitude for 126 ClC-1 variants. Please note change of scale at 0 mV. The vertical red line and horizontal pink line represent the cut-off voltage (vertical (-18.6 mV)) and current amplitude (horizontal (-2.5 µA)). Data for wild-type channel is shown in red and all the variants in the wild-type channel quadrant defined by the cut-off lines were considered benign. Several variants showed no currents or showed currents that could not be characterised only in terms of $V_{1/2}$ and current amplitude[11]. The $V_{1/2}$ of these variants was not assessed but are plotted in the graph with 0 current amplitude in blue. Variants in orange show wild-type-like voltage dependence of activation and current amplitude, but the rate of activation differed from wild-type. Based on the

cut-off criteria these were not classified as pathogenic. **B)** ROC curves for in silico predictive tools + Grantham distance.

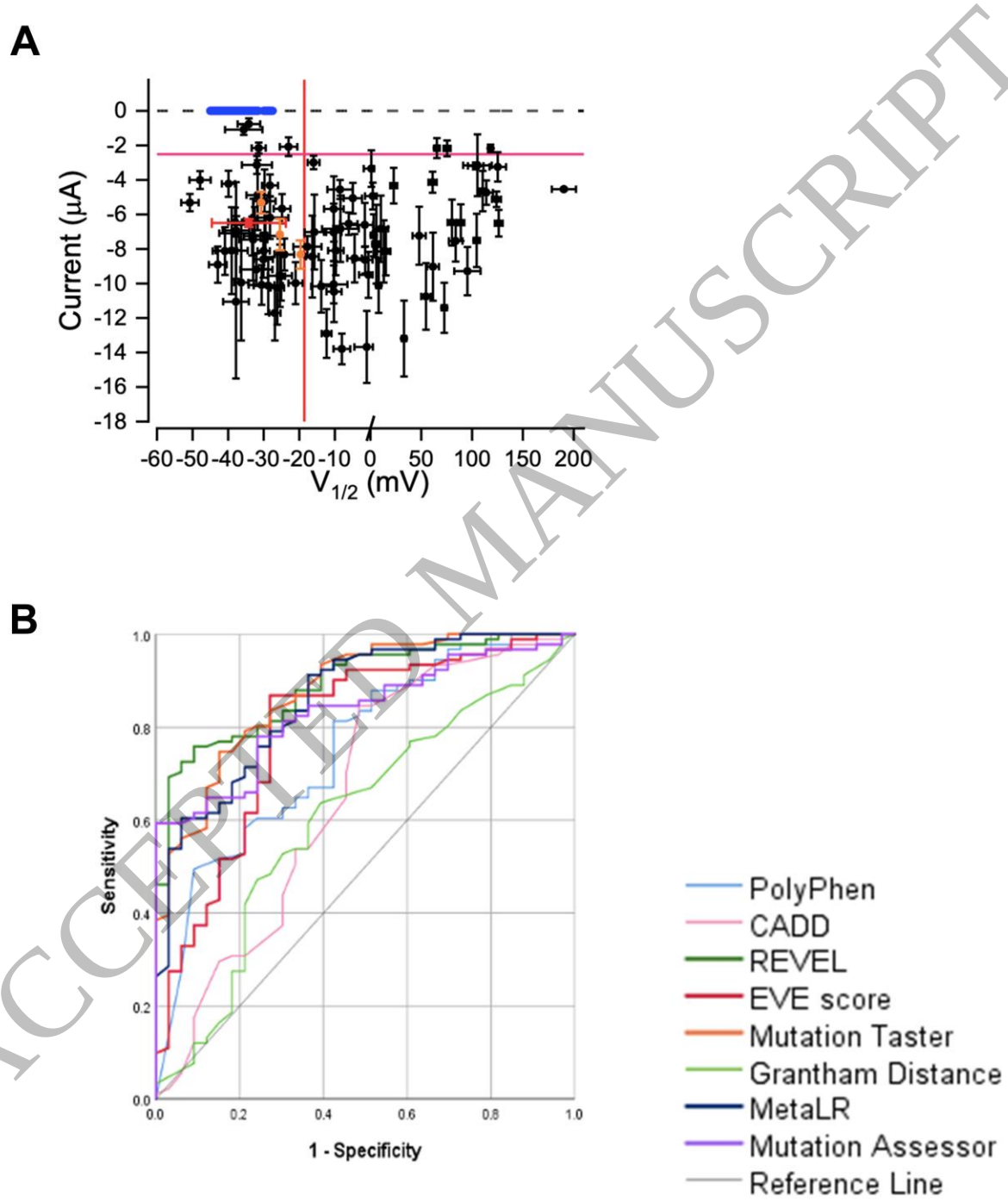

Figure 1
165x197 mm (.11 x DPI)

Table 1 Results for each *in silico* predictive tool

	Mutation Taster	REVEL	EVE	PolyPhen	MetaLR	SIFT	aGVGD	Mutation Assessor	CADD
True Positive (% of total pathogenic)	88 (97)	87 (96)	59 (65)	74 (81)	91 (100)	62 (68)	26 (29)	39 (43)	13 (14)
False Positive	15	18	8	15	28	9	0	18	3
Total Actual Pathogenic	91	91	91	91	91	91	91	91	91
True Negative (%)	18 (51)	17 (49)	19 (54)	10 (29)	7 (20)	26 (74)	31 (86)	17 (49)	32 (91)
False Negative	3	4	9	5	0	29	40	12	78
Total Actual Benign	35	35	35	35	35	35	35	35	35
Uncertain (n)	2	0	31	22	0	0	29	40	0
Accuracy	85.48%	82.54%	82.11 %	80.77%	77.78%	69.84 %	58.76%	65.12%	35.71%
Sensitivity	0.97	0.96	0.87	0.94	1.00	0.68	0.39	0.76	0.14
Specificity	0.55	0.49	0.70	0.40	0.20	0.74	1.00	0.49	0.91
Positive Predictive Value	0.85	0.83	0.88	0.83	0.76	0.87	1.00	0.68	0.81
Negative Predictive Value	0.86	0.81	0.68	0.67	1.00	0.47	0.44	0.59	0.29

Most accurate tool listed on the left and least accurate on the right.

Table 2 AUC for *in silico* prediction tools

	REVEL	Mutation Taster	MetaLR	Mutation Assessor	SIFT	EVE	PolyPhen	CADD
AUC (SE)	0.89 (0.3)	0.88 (0.03)	0.86 (0.04)	0.83 (0.04)	0.82 (0.04)	0.80 (0.05)	0.75 (0.05)	0.66 (0.06)
95% CI	0.83–0.95	0.81–0.94	0.79–0.93	0.76–0.9	0.74–0.89	0.7–0.89	0.66–0.85	0.54–0.77

Table 3 Data based on concordance of a combination of in silico predictive tools

	Tools used in the ACMG/ACGS guidelines (PolyPhen, SIFT, Mutation Taster, aGVGD)	Tools performing highly based on AUC and specificity (REVEL, Mutation Taster, MetaLR, EVE)	(REVEL, Mutation Taster, EVE)
Sensitivity	0.92	0.99	0.97
Specificity	0.74	0.48	0.65
Accuracy	86.95%	87.72%	90.00%
Positive Predictive Value	0.90	0.87	0.91
Negative Predictive Value	0.80	0.92	0.85
Number of concordant scores (% of all variants)	39 (30.95%) [4 of 4 scores concordant]	70 (55.56%) [4 of 4 scores concordant]	0 [3 of 3 scores concordant]
Number of concordant scores (% of all variants)	53 (42.06%) [3 of 4 scores concordant]	44 (34.92%) [3 of 4 scores concordant]	80 (63.49%) [2 of 3 scores concordant]
Number of variants unclassified	34	12	46