Author/s:
Pham, DH;Pitman, MR;Kumar, R;Jolly, LA;Schulz, R;Gardner, AE;de Nys, R;Heron, SE;Corbett, MA;Kothur, K;Gill, D;Rajagopalan, S;Kolc, KL;Halliday, BJ;Robertson, SP;Regan, BM;Kirsch, HE;Berkovic, SF;Scheffer, IE;Pitson, SM;Petrovski, S;Gecz, J

Title:
Integrated in silico and experimental assessment of disease relevance of PCDH19 missense variants

Date:
2021-08-01

Citation:
Pham, D. H., Pitman, M. R., Kumar, R., Jolly, L. A., Schulz, R., Gardner, A. E., de Nys, R., Heron, S. E., Corbett, M. A., Kothur, K., Gill, D., Rajagopalan, S., Kolc, K. L., Halliday, B. J., Robertson, S. P., Regan, B. M., Kirsch, H. E., Berkovic, S. F., Scheffer, I. E. ,... Gecz, J. (2021). Integrated in silico and experimental assessment of disease relevance of PCDH19 missense variants. Human Mutation, 42 (8), pp.1030-1041. https://doi.org/10.1002/humu.24237.

Duyen Pham ORCID iD: 0000-0003-0664-4133

Lachlan Jolly ORCID iD: 0000-0003-4538-2658

Mark Corbett ORCID iD: 0000-0001-9298-3072

Stephen Robertson ORCID iD: 0000-0002-5181-7809

Jozef Gecz ORCID iD: 0000-0002-7884-6861

INTEGRATED *IN SILICO* AND EXPERIMENTAL ASSESSMENT OF DISEASE-RELEVANCE OF PROTOCADHERIN 19 (*PCDH19*) MISSENSE VARIANTS

Duyen H. Pham[1,2], Melissa R. Pitman[3], Raman Kumar[1,2], Lachlan Jolly[1,2], Renee Schulz[1], Alison Gardner[1,2], Rebekah de Nys[1,2], Sarah E. Heron[1,2], Mark A. Corbett[1,2], Kavitha Kothur[4,5], Deepak Gill[4,5], Sulekha Rajagopalan[6], Kristy Kolc[1,2], Benjamin J. Halliday[7], Stephen P. Robertson[7], Brigid M. Regan[8], Heidi E. Kirsch[9], Samuel F. Berkovic[8], Ingrid E. Scheffer[8,10,11], Stuart M. Pitson[1,3,12], Slave Petrovski[13,14], Jozef Gecz[1,2,12,15]

Affiliations:

[1] Adelaide Medical School, The University of Adelaide, Adelaide, SA 5005, Australia.

[2] Robinson Research Institute, The University of Adelaide, Adelaide, SA 5005, Australia.

This is the author manuscript accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/humu.24237.

[3] Centre for Cancer Biology, University of South Australia and SA Pathology, Adelaide, SA 5000, Australia.

[4] Kids Neuroscience Centre, The University of Sydney, Sydney, NSW 2006, Australia.

[5] TY Nelson Department of Neurology and Neurosurgery, The Children's Hospital at Westmead, Sydney, NSW 2145, Australia.

[6] Department of Clinical Genetics, Liverpool Hospital, Liverpool, NSW 1871, Australia.

[7] Department of Women's and Children's Health, Dunedin School of Medicine, University of Otago, Dunedin 9016, New Zealand.

[8] Department of Medicine, Austin Health, The University of Melbourne, Heidelberg, VIC 3084, Australia.

[9] Department of Neurology, University of California, San Francisco, CA 94122, USA.

[10]Department of Paediatrics, Royal Children's Hospital, University of Melbourne, Flemington, VIC 3031, Australia

[11]Florey Institute of Neuroscience and Mental Health, Heidelberg, VIC 3084, Australia.

[12] School of Biological Sciences, The University of Adelaide, Adelaide, 5005, Australia.

[13] Centre for Genomics Research, Precision Medicine and Genomics, IMED Biotech Unit, AstraZeneca, Cambridge CB2 0RE, UK.

[14] Department of Medicine, University of Melbourne, Parkville, VIC 3052, Australia.

[15] South Australian Health and Medical Research Institute, Adelaide, 5000, Australia.

Correspondence to Jozef Gecz, Adelaide Medical School, The University of Adelaide, SOUTH AUSTRALIA, 5005; Ph: +618 8313 2453; E-mail: jozef.gecz@adelaide.edu.au

Running Title: Pathogenicity assessment of *PCDH19* missense variants

ABSTRACT:

PCDH19 is a non-clustered protocadherin molecule involved in axon bundling, synapse function and transcriptional co-regulation. Pathogenic variants in *PCDH19* cause an infantile onset epilepsy known as *PCDH19*-clustering epilepsy or *PCDH19*-CE. Recent advances in DNA sequencing technologies have led to a significant increase in the number of reported *PCDH19*-CE variants, many of uncertain significance, or VUS. We aimed to determine the best approaches for assessing the disease relevance of missense variants in *PCDH19*. Application of the American College of Medical Genetics and Association for Molecular Pathology (ACMG-AMP) guidelines was only 50% accurate. Using a training set of 322 known benign or pathogenic missense variants, we identified MutPred2, MutationAssessor and GPP as the best performing *in silico* tools. We generated a protein structural model of the extracellular domain and assessed 24 missense variants. We also assessed 24 variants using an *in vitro* reporter assay. A combination of these tools was 93% accurate in assessing known pathogenic and benign *PCDH19* variants. We increased the accuracy of the ACMG-AMP classification of 45 *PCDH19* variants from 50% to 94%, using these tools. In summary, we have developed a robust toolbox for the assessment of *PCDH19* variant pathogenicity to improve the accuracy of *PCDH19*-CE variant classification.

Key Words: *PCDH19*; epilepsy; VUS; variant assessment, functional test

MAIN TEXT

INTRODUCTION

Assessment of variants of uncertain significance (VUS) in genes known to be associated with disease is an increasingly important task, particularly where precision medicine is an option. However, determining the clinical relevance of these variants can be challenging, especially when functional assays for the affected protein are unavailable or scarce. While the clinical interpretation of DNA variation has improved significantly through the implementation of high international standards in the field (MacArthur et al., 2014; Richards et al., 2015), there is a mounting number of VUS that need ongoing and often very specialized assessment.

Here we aimed to systematically assess reported and novel *PCDH19* variants causing clustering epilepsy (*PCDH19*-CE; previously known as Epilepsy and Mental Retardation Limited to Females; EFMR, or Female-Limited Epilepsy; FLE; MIM# 300088) (Dibbens et al., 2008). *PCHD19*-CE is an early-onset neurodevelopmental disorder (NDD) characterised by drug-resistant epilepsy, intellectual disability (ID), autism spectrum disorder (ASD) and other neuropsychiatric disturbances (Depienne et al., 2009; Dibbens et al., 2008; Kolc et al., 2020; Scheffer et al., 2008). *PCDH19* pathogenic variants are among the most frequent causes of female developmental and epileptic encephalopathy (Kolc et al., 2019). *PCDH19*-CE is an X-chromosome disorder with unusual genetics, where girls are affected in heterozygous state, but hemizygous boys are spared (Gecz & Thomas, 2020). Cellular mosaicism resulting from either random X-chromosome inactivation in girls, or postzygotic *PCDH19* mutation in boys, is postulated to be the driver of the disorder, although the precise disease mechanism

remains largely unknown (Gecz & Thomas, 2020). Penetrance of *PCDH19*-CE is estimated to be 80%, with ~20% of girls with a pathogenic *PCDH19* variant are either not affected or mildly affected (Kolc et al., 2019).

Next generation sequencing has led to a significant increase in the number of reported *PCDH19* missense variants, both potentially pathogenic or VUS in patients with *PCDH19*-CE and NDDs (>150 unique published and unpublished variants) (Perucca et al., 2017). In this study, we used sets of benign and pathogenic *PCDH19* variants to develop an assessment toolbox combining *in silico* predictive and experimental approaches. We then applied our toolbox to a set of novel *PCDH19* variants. Overall, the combination of our tools achieved very high discriminatory power in *PCDH19* variant assessment.

MATERIALS AND METHODS

Editorial Policies and Ethical Considerations

The study was approved by the Human Research Ethics Committees of the Women's and Children's Health Network and the University of Adelaide, Adelaide. Patient clinical information was obtained from the Epi4K and the Epilepsy Phenome/Genome Project or personal communication in a manner conforming to the institutional review board (IRB) and/or granting agency ethical guidelines.

Pathogenicity assessment

We used 16 selected *in silico* pathogenicity prediction tools (Supp. Table S1) including MutPred2 (URL: http://mutpred.mutdb.org/) and ANNOVAR tool-package to annotate *PCDH19* variants (Yang & Wang, 2015) (URL: http://wannovar.wglab.org/). InterVar (based on 28 criteria from 2015 ACMG-AMP guidelines) was also used to intepret the clinical significance of these variants (URL: http://wintervar.wglab.org/).

PCDH19 variant selection

We used two different *PCDH19* missense variant datasets. The first dataset included 238 *PCDH19* missense variants selected from the general population (female only) gnomAD v2.1.1 database (URL:gnomad.broadinstitute.org/) (Lek et al., 2016). The second dataset included 84 reported *PCDH19*-CE disease-causing missense variants (Kolc et al., 2018; Smith et al., 2018).

A total of 45 *PCDH19* (43 missense) variants were selected for further integrated analysis: 1) *PCDH19*-CE disease-causing (*n*=9); 2) Novel variants (gnomAD allele frequency $\leq 1x10^{-4}$) identified in patients with *PCDH19*-CE and NDDs through the PCDH19 survey, genome sequence data or personal communication (*n*=27); and 3) More frequent population variants (ExAC/gnomAD allele frequency $>1x10^{-4}$) (*n*=9) (Suppl. Figure S1). Refer to Supporting Information for more information. For all cDNA and protein annotations we used the longest *PCDH19* reference transcript NM_001184880.2 and its corresponding protein NP_ 001171809.1.

In silico analysis of the effect of PCDH19 variants on protein structure

Three-dimensional protein structure of human PCDH19 EC1-5 prediction was performed by homology modelling using the zebrafish (Cooper, Jontes, & Sotomayor, 2016) and mouse (Goodman et al., 2016) PCDH19 crystal structures as a template on the ICM-Pro platform (version 3.8-6 Molsoft, San Diego, CA, USA) (Abagyan & Totrov, 1994). For more detail see Supporting Information.

Generation of expression constructs for in vitro functional assays

Myc-tagged human *PCDH19* (NM_001184880.2, which represents the longest isoform) variant expression constructs were generated by site-directed mutagensis

(GenScript) or overlap extension PCR of the full length wild-type (WT) Myc-*PCDH19* expression plasmid as reported previously (Pham et al., 2017). For more detail see Supporting Information.

Cell culture and transfection

Human breast adenocarcinoma MCF-7 cells (ATCC HTB-22) were cultured, transfected, treated, and harvested as described previously (Pham et al., 2017).

ERE-LUC reporter assay and western blotting

Dual-luciferase reporter assays (Promega) and western blotting was performed as previously described (Pham et al., 2017).

RESULTS AND DISCUSSION

Evaluation of 16 *in silico* tools for predicting pathogenicity of *PCDH19* missense variants

We compared the performance of publicly available *in silico* prediction tools for differentiating a set of likely benign from likely pathogenic *PCDH19* missense variants. A total of 84 unique and likely pathogenic missense variants (identified across 326 *PCDH19*-CE cases) were obtained from published reports. Incorrectly annotated variants were excluded (Supp. Table S2). The 84 selected variants were all located within the N-terminal region of the protein. A total of 238 unique and likely benign missense variants were obtained from gnomAD, which captures genetic variation in the general population. 121 of these were located within the N-terminal extracellular domain (Extracellular Domain; EXD) and 117 within the C-terminal region (Cytoplamic Region; CyR) of the PCDH19 protein (Supp. Figure S2). We used these two datasets (*PCDH19*-CE=84 and

gnomAD=238) to evaluate the performance of 16 different prediction tools including those using functional scores (MutPred2, SIFT, Polyphen2, MutationAssessor, PROVEAN, VEST3 and FATHMM), conservation scores (GERP++, phyloP_mammalian, phyloP_vertebrate, phastCons_mammalian and phastCons_vertebrate), human-lineage intolerance metrics (MTR), and 'three multidimensional prediction algorithms' which integrate several tools (CADD, DANN and GPP); (Traynelis et al., 2017; Yang & Wang, 2015). Conservatively, we ranked the tools based on their ability to statistically discriminate only PCDH19 N-terminal EXD variants as this is the only region where *PCDH19*-CE missense variants exist in our data set (*PCDH19*-CE=84 and gnomAD =121). We found that from among these tools, the functional and the multidimensional tools were superior. Using this training set, MutPred2 performed the best to discriminate pathogenic and benign variants (p=4.705E^{-25}), followed by MutationAssessor (p=9.763E^{-24}), GPP (p=1.491E^{-22}), SIFT (p=8.266E^{-21}), PROVEAN (p=9.863E^{-20}), VEST3 (p=1.672E^{-15}), CADD (p=3.465E^{-10}) and PolyPhen2 (p=8.703E^{-07}) (Figure 1; Supp. Figure S2 & Table S3). The MTR tool (p=0.146), and tools using conservation scores, did not perform as well (Supp. Figure S2 & Table S3).

We then evaluated pairwise combinations of the best performing tools to again assess variants specifically in the PCDH19 N-terminal EXD (n=84 likely pathogenic *PCDH19*-CE and n=121 likely benign gnomAD variants). We have done these comparisons with either including or excluding the ultra-rare gnomAD variants (allele frequency of <1x10^{-5}; Supp. Figure S2-S4). The combination of MutPred2 and GPP performed the best, with a prediction accuracy of 93% (190/205 correct predictions, Supp. Figure S3). Combination of MutationAssessor-GPP, and MutPred2-MutationAssessor gave similar results (92% and 89%, respectively; Supp. Figure S5).

The MutPred2-GPP combination achieved 93% accuracy in predicting pathogenicity (78/84 PCDH19-CE variants predicted correctly) and 93% accuracy in predicting non-pathogenicity (112/121 gnomAD EXD variants predicted benign) (Supp. Figure S5). We also found that nine *PCDH19* CyR variants out of the 121 (7.4%) obtained from gnomAD were predicted to be pathogenic by this combination.

Selection of the test set of *PCDH19* missense variants

We next applied multiple approaches to conduct in-depth assessment of 45 different *PCDH19* DNA variants which result in 43 unique amino acid changes. Two different DNA variants led to the same amino acid change i.e. c.1335C>G and c.1335C>A resulting in p.Asp445Glu; and c.1671C>A and c.1671C>G resulting in p.Asn557Lys (Supp. Figure S6; Supp Table S4). These 45 variants fell into three categories: pathogenic *PCDH19*-CE (*n*=9), benign population (gnomAD; *n*=9) and VUS (*n*=27) variants, respectively. Two *PCDH19*-CE variants were recurrent (c.1019A>G; p.Asn340Ser and c.1322T>A; p.Val441Glu) (Kolc et al., 2018). The *PCDH19*-CE variants were selected from individuals and families with genetic (i.e. segregation within families, or arising *de novo*) and clinical evidence supporting pathogenicity. These were all located within the conserved cadherin repeats (EC2-EC6) making up the EXD of PCDH19. The benign 'population' variants were selected from gnomAD exhibiting an allele frequency of $>1x10^{-4}$. The VUS were selected from the exome sequence data generated by the Epi4K consortium (Epi et al., 2013), the Epilepsy Phenome/Genome Project, patients captured in a recent survey of *PCDH19* families (Kolc et al., 2020) or gnomAD with allele frequencies of $\leq 1x10^{-4}$. One of these variants, c.1330A>G; p.Thr444Ala, a VUS, was identified in two individuals. As outlined below, we embarked on determining the robustness of different approaches to assess the pathogenicity of these

variants. This included the use of (1) clinical DNA diagnostic guidelines (2) *in silico* predictive tools (3) protein structural modelling and (4) *in vitro* luciferase reporter assays.

Assessment of *PCDH19* variants using ACMG-AMP guidelines

First, we used InterVar, a web-based tool that leverages criteria derived from the 2015 American College of Medical Genetics – Association of Medical Pathlogists (ACMG-AMP) clinical DNA variant interpretation guidelines (Q. Li & Wang, 2017), to assess the significance of these 45 missense variants. We adopted the default setting on this web-based platform, meaning 28 different ACMG-AMP criteria were used in scoring. At this point the scoring did not include evidence from our subsequent *in silico* and functional assays, as outlined below. Of the 45 variants, InterVar identified seven as likely pathogenic (including three *PCDH19*-CE and four VUS), another six were classified as likely benign (all of these were population variants); and the remaining 32 were classified as VUS (Supp. Table S4). Thus the accuracy of InterVar, in the absence of functional data, was 33% (3/9) for the likely pathogenic *PCDH19*-CE variants, and 66.6% (6/9) for the benign (population) variants. Given the relatively poor accuracy of InterVar (9/18 or 50% in total), we investigated additional approaches for variant assessment.

Assessment of *PCDH19* variants using *in silico* tools

Second, we assessed these 45 *PCDH19* DNA variants (43 missense variants) using the three-top performing *PCDH19 in silico* prediction tools as prioritised above (MutPred2, MutationAssessor, and GPP). Twenty nine of the 43 protein missense variants (67%, or 31 from all 45 variants, 69%) showed consistent predictions across the three tools, that is 1:B or 4:P (Supp. Table S4). Of the 14 variants with inconsistent results, three were *bona fide PCDH19*-CE variants (that is p.Val191Leu, p. Phe206Cys and p.Pro567Leu). To reconcile and integrate this *in silico* approach, we argued a variant to be likely pathogenic if it was predicted as such by at least two of the three *in silico* tools. We applied the same criteria for the likely benign variants and VUS. Using this

approach, 78% (7/9) *PCDH19*-CE variants were correctly predicted likely pathogenic, and 100% (9/9) population variants were correctly predicted to be likely benign, for a total accuracy of 88% (16/18; Supp. Table S4). Of the VUS, 63% (15/25) were predicted to be likely pathogenic (Supp. Table S4). To further support the integrated *in silico* approach, an additional set of 12 recently published novel *PCDH19* pathogenic missense variants were also assessed. All of these novel 12 PCDH19-CE variants (100%) were classified as likely pathogenic while InterVar assigned only 7 of these as likely pathogenic and five as VUS (Supp. Table S5) (Lindy et al., 2018).

Subsequently, we applied this *in silico* approach to test an additional 8 unpublished and novel *PCDH19* VUS found in nine individuals. Three individuals were referred to us through personal communication (Supporting Information) whilst the remaining six were recruited from the updates of the Epi4K and the Epilepsy Phenome/Genome Projects (Epi et al., 2013). One VUS (c.219C>G; p.Asp73Glu) was identified in a female with periventricular nodular heterotopia and epilepsy (Supp. Table S6). The *in silico* prediction showed it to be likely benign (Supp. Table S7) and was not present in gnomAD. This individual also carries a different variant, in another gene, possibly explaining her clinical presentations (data not shown). Overall, only two (c.824A>G; p.Tyr275Cys and c.1495G>T; p.Asp499Tyr) of the six VUS identified from Epi4K and the Epilepsy Phenome/Genome Project (Epi et al., 2013) were predicted to be likely pathogenic using our PCDH19 variant trained *in silico* set of tools. The affected female with the c.824A>G; p.Tyr275Cys variant had non-lesional parietal lobe epilepsy, mild developmental delay, and speech and language problems and the female with the c.1495G>T; p.Asp499Tyr variant had focal, tonic seizures and mild-borderline developmental delay, with onset at 8 years of age (Supp. Table S6-S7). The latter variant

was further assessed (see below) by protein structure modelling and *in vitro* reporter assays and suggested to be benign.

Assessment of *PCDH19* variants using protein structure modelling.

Third, we investigated the potential functional effect of the 45 *PCDH19* DNA variants (43 missense variants) using protein structure modelling. We generated a homology model of human PCDH19 EXD consisting of the conserved cadherin repeats (EC1-EC5) and modelled the effect of selected PCDH19 variants (Figure 2). Although the EC6 repeat was absent from the model, variants located within this repeat were assessed for their likely impact on local amino acid environment and secondary structures based on alignment with mouse PCDHγB4 (Supp. Figure S9) (S. C. Li, Goto, Williams, & Deber, 1996). In addition to visual inspection of the variant models for gross structural defects and hydrogen bonding interactions, we assessed calcium binding regions and subjected models to protein stability predictions (Supp. Table S8 & Figure S7-S8). Due to limitations of the protein structure modelling, we could only investigate the effect of 24/43 missense variants (*PCDH19*-CE=9, population=3, VUS=12; Supp. Figure S1). Of these 24 variants, 21 were located within cadherin repeats EC1-5 and 3 were located in the EC6 repeat.

Twelve missense variants (12/24) were predicted to be pathogenic due to major structural disruption. This included p.Arg198His, p.Ser276Pro, p.Asn340Ser, p.Pro343Gln, p.Asp375Asn, p.Val441Glu, p.Asn557Lys, p.Asp558Asn, p.Pro561Arg, p.Pro567Leu, p.Pro567Arg and p.Ala654Pro. These variants predominantly clustered in the EC3 and EC4 repeats (Figure 2b-c, Supp. Table S8). Although the EC5 variants p.Asn557Lys and p.Asp558Asn were located outside the regions that could be reliably modelled, they were predicted to cause structural disruption due to their location in the

conserved calcium-binding motifs ($_{464}$QENN $_{555}$DVNDN $_{595}$DYDE) (Nagar, Overduin, Ikura, & Rini, 1996) (Supp. Table S8, Supp. Figure S9). Of the 12 structurally disruptive variants (Supp. Table S8), calcium binding site variants (5/12, 42%) were prominent together with variants involving a proline residue (50%, 6/12).

Calcium ions bind at the junctions between EC domains and are coordinated by negatively charged residues (Cooper et al., 2016). The alterations in calcium binding sites were predicted for p.Arg198His (Supp. Figure S7b), p.Asn340Ser (Supp. Figure S8a), p.Pro343Gln (Supp. Figure S8b), and p.Asp375Asn variants (Supp. Figure S8c). Loss of negatively charged side-chains (p.Asp375Asn, Asp558Asn) or disruption of calcium-coordinating residues (p.Arg198His, p.Asn340Ser, pAsn557Lys) were predicted to disrupt calcium-binding, adversely affecting PCDH19 function. The disruptive nature of these variants is further supported by loss of homotypic interaction of PCDH19 with such variants (e.g. p.Asn340Ser) (Pederick et al., 2018).

Disruptions involving proline residues were relevant for p.Ser276Pro, p.Pro343Gln, p.Pro561Arg, p.Pro567Leu, p.Pro567Arg and p.Ala654Pro. Proline residues introduce kinks to the amino acid chain and can have either subtle or drastic effects on secondary structure depending on the surrounding residues and domains (Chou & Fasman, 1974). For example, the p.Ser276Pro variant was predicted to cause a drastic conformational shift (34.8Å) of the EC1-EC2 repeats due to the insertion of a proline kink at the junction of EC2:EC3 domains (Supp. Figure S7a). Thus, in line with previous studies, we propose that variants involving proline, are likely to produce defective mis-folded and/or unstable PCDH19 proteins.

On aggregate, protein structural modelling accurately predicted 78% (7/9) of the *PCDH19*-CE and 100% (3/3) of the population variants, for a total accuracy of 83% (10/12). Approximately 42% (5/12)

of the VUS were predicted to alter the PCDH19 protein structure. Structural information allowed us to interrogate the functional significance of PCDH19 variants, albeit with limitations. While we can accurately model PCDH19 EXD domains including EC1-5 using homologous templates from existing crystal structures (Cooper et al., 2016; Goodman et al., 2016), and to a degree, EC6 based on sequence homology to PCDHγB4, it is not possible to model intracellular domains (PCDH19 CyR) due to lack of template structures with sufficient resolution. In addition, very few protein-protein interaction motifs have been identified across the protein (Chen et al., 2014; Tai, Kubota, Shiono, Tokutsu, & Suzuki, 2010). Further knowledge of these PCDH19 domains and motifs would significantly strengthen this approach. Indeed, seven of the 43 missense variants were located in the C-terminal intracellular region of the protein, and as such not able to be assessed using our PCDH19 protein model. However, these variants were predominantly found with a high frequency in the population (gnomAD), suggesting they are unlikely to be pathogenic.

Assessment of *PCDH19* variants using *in vitro* reporter assays

We then turned to our recently developed PCDH19 functional co-transcriptional reporter assays to assess select variants. We previously showed that PCDH19 interacts with DNA and RNA binding protein NONO/P54NRB to stimulate estrogen receptor (ER) α-mediated gene expression (Pham et al., 2017). The WT-PCDH19 expression enhances transcription from an ER luciferase based reporter (ERE-LUC), whilst the *PCDH19*-CE variant expression does not, revealing loss of function. We thus used the *in vitro* ERE-LUC reporter assay to determine the functional effect of *PCDH19* variants as measured by comparison to the WT-PCDH19 in the presence or absence of estrogen ligand (E2).

We assessed 23 (of the 43) *PCDH19* missense variants (8/9 *PCDH19*-CE, 6/9 population, and 9/25 VUS) using the ERE-LUC reporter assay (Supp. Figure S1). Seven of the eight PCDH19-CE variants showed significantly ($p < 0.01$) reduced the ERE-LUC activity compared to WT-PCDH19 (Figure 3). The remaining variant (p.Phe206Cys)

showed a smaller reduction on the ERE-LUC activity; that was significant ($p<0.05$) only in the presence of E2 ligand. This data cast some doubt on the variants pathogenicity. Interestingly, it is this variant that has been found co-occurring with another PCDH19 variant, p.His203Pro, in two patients (Marini et al., 2012). One could speculate that these two variants may contribute additively to the function of the PCDH19 protein (Marini et al., 2012), however, we have not tested this possibility.

The *in vitro* functional analysis achieved 100% (8/8) accuracy for *PCDH19*-CE variant assessment, a superior accuracy over the other approaches tested. Population variants with allele frequencies of $>1x10^{-4}$ were predominantly shown not to have an effect (i.e. p.Met432Val, p.Leu540Phe, p.Arg958Gln and p.Asn1134His) or some effect on ERE-LUC reporter (i.e. p.Val257Leu and p.Arg1107Gly; Figure 3). Notably the four population variants without an effect were predicted by InterVar to be likely benign, while the two with some effect were predicted as VUS (Table 1). In total, the accuracy of the functional assay was 82% (14/17)

The results for nine VUS, including three unpublished variants (p.Asn75Ser, p.Asp499Tyr, p.Thr444Ala) identified in patients with NDDs and epilepsy, two from the Epi4K study and Epilepsy Phenome/Genome Project (p.Val441Leu, p.Thr444Ala), and five rare variants (p.Arg198His, p.Pro343Gln, p.Asp558Asn, p.Pro567Arg, p.Ser276Ala) with allele frequencies of $\leq 1x10^{-4}$ selected from gnomAD, were variable, with four of these nine (44%) being assessed as pathogenic. The p.Thr444Ala variant was identified in two patients; a female with complex NDD and epilepsy (unpublished data) and a male with sporadic non-acquired focal epilepsy (NAFE) from the Epi4K study. However, this variant is present in gnomAD at a higher allele frequency than would be expected for a pathogenic variant (Supp. Table S7). The *in vitro* ERE-LUC reporter assay indicates that the p.Thr444Ala variant is unlikely to be damaging, and this is also supported by the integrated *in silico* approach. The VUS p.Asp499Tyr, absent in gnomAD, was identified in a female patient. Results from *in vitro* ERE-LUC reporter assays indicates that the p.Asp499Tyr variant is also unlikely to be damaging. This was supported by protein structural modelling, but not by the *in silico* assessment. Segregation studies indicated that p.Asp499Tyr was inherited from the

father and his unaffected mother. The individual also had a 5.5 kb deletion in exon 4 of the methyl-CpG binding protein 2 (*MECP2*) gene, which was considered pathogenic and aligned with refractory epilepsy, ataxia, dyskinesia, microcephaly and progressive neurological deterioration resulting in demise at the age of 16 years. Unlike classical Rett syndrome, she lacked typical features of breathing irregularity, hypotonia, and hand stereotypies. Based on the above findings it is not possible to resolve the effect of this p.Asp499Tyr *PCDH19* VUS albeit such blended phenotypes due to multiple and independent DNA variants have been reported (Karaca et al., 2018).

Reconciliation of *PCDH19* variants effect using a combined approach.

In general, the integrated *in silico*, protein structural and *in vitro* transcriptional approaches assigned pathogenicity of known pathogenic variants, and the benign nature of known benign variants accurately (all >80%). However, no one approach achieved 100% accuracy. We thus investigated if a combination of these approaches might be beneficial. Of the 43 missense variants, 20 were assessed using all three approaches and the results were combined (Figure 4). Using this approach, 100% (8/8) of *PCDH19*-CE associated variants were correctly classified as likely pathogenic or pathogenic, and 100% (3/3) of population variants were classified as benign for a total of 100% accuracy (11/11).

It is not possible to assess all variants across all three approaches as shown above. There also exists a degree of circularity in the approach with respect to over-weighting of evidence obtained from structural disruptions (i.e. accumulating scores from structural models and the predictive tools which leverage structural impact). Thus we sought to illustrate the conservative and practical application of these tools by using them as lines of evidence in the framework of the ACMG-AMP classification system. We re-analysed the 45 variants we initially assessed using InterVar alone, by incorporating the outcomes of our tools within the ACMG-AMP guidelines (Table 1). Of the 8 *PCDH19*-CE variants

in this data set, the initial InterVar assignment suggested six were VUS and only two as likely pathogenic. Adding evidence provided by the approaches in our toolbox improved InterVar prediction for each one of these eight variants (Table 1). As for the nine population variants in this data set, six (66%) were initially classified as likely benign with the remainder as VUS. Adding evidence derived from our approaches resulted in eight of nine (89%) being classified as benign or likely benign. Intriguingly, the remaining population variant (p.Val257Leu) was classified as likely pathogenic, based primarily on its effect in the ERE-LUC reporter assay, but not using multiple *in silico* prediction tools. This variant is found in gnomAD with a relatively high frequency of 0.00067 (136 alleles; 91 females and 45 males). This finding raises an interesting possibility of a variant with mild protein function effect, which may not be *PCDH19*-CE disease relevant. In total, the application of our approaches within our toolbox applied within existing ACMG-AMP guidelines improved the total accuracy from 50% to 94% (16/17). Given this outcome, we turned to the remaining 27 variants in the data set of 45 which were obtained as VUS. Our initial InterVar analysis classified 5 of these as likely pathogenic, and 22 as VUS. Our reanalysis using our toolbox within ACMG-AMP guidelines resulted in only 14 assessed as of uncertain significance, with 5 classified as benign or likely benign, and 11 as either likely pathogenic or pathogenic. Application of our tool box led to the reclassification of several VUS: five variants (i.e. p.Asn75Ser, p.Val441Leu, p.Thr444Ala, p.Asp499Tyr and p.Thr626Ile) were re-classified from VUS to likely benign, whilst three variants (p.Asp445Glu, p.Asp558Asn and p.Leu640Arg) were assigned as likely pathogenic (Table 1). In total, of all 45 variants initially assed using ACMG-AMP guidelines, the additional evidence provided by our tool-box of approaches resulted in the reclassification of 51% of them (23/45; Table 1)

Conclusion

It is important to accurately distinguish pathogenic from benign DNA variation in the diagnosis *PCDH19*-CE, whether in females or in somatic mosaic males. *PCDH19*-CE is a drug resistant epilepsy (Lotte et al., 2016) where accurate diagnosis will facilitate precision medicine (Sadleir et al, 2020). While early and accurate diagnosis is the immediate need, better understanding of the underlying cellular pathways affected by *PCDH19* variants that accelerate design of targeted and timely interventions, remains the ultimate goal. We believe that our *PCDH19* variant assessment investigations and experience will facilitate this critical translational process. Our data suggest that combining *in silico* predictions, structual modeling and functional genomic *in vitro* approaches significantly improves classification of *PCDH19* variants, beyond the aplication of AMCG-AMP guidelines alone (Q. Li & Wang, 2017). In summary, by identifying and applying multiple *in silico* assessment tools and an *in vitro* functional assay for *PCDH19* variant assessment, we achieved improvement in the classification of 51% of missense variants. Our toolbox presents the most robust current approach for *PCDH19* variant assessment.

CONFLICT OF INTEREST

Ingrid Scheffer has served on scientific advisory boards for UCB, Eisai, GlaxoSmithKline, BioMarin, Nutricia and Xenon Pharmaceuticals; editorial boards of the *Annals of Neurology*, *Neurology* and *Epileptic Disorders*; may accrue future revenue on pending patent WO61/010176 (filed: 2008): Therapeutic Compound; has received speaker honoraria from GlaxoSmithKline, Athena Diagnostics, UCB, BioMarin, and Eisai; has received funding for travel from Athena Diagnostics, UCB, Biocodex, GlaxoSmithKline, Biomarin and Eisai; and receives/has received research support from

the National Health and Medical Research Council of Australia, National Institutes of Health, Australian Research Council, Health Research Council of New Zealand, CURE, and March of Dimes. The remaining authors declare no conflict of interest.

ACKNOWLEDEMENTS

We thank the families for their participation in this study and Epi4K Consortium and the Epilepsy Phenome/Genome Project for their collaboration. This work was supported by the WCH Foundation Research Project Grant to D.H.P., Australian NHMRC Grant 1091593 to S.F.B., I.E.S., and J.G. and NHMRC Fellowship 1155224 to J.G. and 1042589 to S.M.P., Practitioner Fellowship to I.E.S., and RAH Research Fund Florey Fellowship to M.R.P. We are also grateful for the ongoing support from the PCDH19 Alliance (USA) as well as the Onlus Insieme per la ricerca PCDH19 (Italy).

CONTRIBUTIONS

J.G. initiated and directed the project. D.H.P., A.G., M.A.C., and S.P. performed the *in silico* predictions. D.H.P. and R.Sc. performed molecular and functional experiments. M.R.P. and S.M.P. performed protein structural modelling. I.E.S., B.R., S.F.B., H.K., A.K., K.K., D.G., S.R., B.H., and S.R. provided clinical information for VUS. R.K., L.J. and R.D.N. critically revised and reassessed the variant data. K.K. provided VUS information from the *PCDH19* Survey. D.H.P., J.G., and M.R.P. wrote the paper. I.E.S, S.E.H., S.P., R.K. and L.J. provided critical editing and revisions of the manuscript. All authors discussed the results and revised the manuscript.

DATA AVAILABILITY STATEMENT

The majority of *PCDH19* missense variants tested in the study were found in the human population selected from the gnomAD database (URL:gnomad.broadinstitute.org/) (Lek

et al., 2016) or published literature as referenced in the paper. The remaining novel *PCDH19* missense variants were from the *PCDH19* Survey developed in our group or through genome sequence data generated by the Epi4K consortium (Epi et al., 2013) (URL: www.epi4k.org/) and the Epilepsy Phenome/Genome Project (URL: www.epgp.org/) or personal communication (refer to Supporting Information). All new variants have been submitted to ClinVar and their references numbers are given in the Supporting Information.

REFERENCES:

Abagyan, R., & Totrov, M. (1994). Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J Mol Biol, 235*(3), 983-1002. doi:10.1006/jmbi.1994.1052

Chen, B., Brinkmann, K., Chen, Z., Pak, C. W., Liao, Y., Shi, S.,... Rosen, M. K. (2014). The WAVE regulatory complex links diverse receptors to the actin cytoskeleton. *Cell, 156*(1-2), 195-207. doi:10.1016/j.cell.2013.11.048

Chou, P. Y., & Fasman, G. D. (1974). Prediction of protein conformation. *Biochemistry, 13*(2), 222-245.

Cooper, S. R., Jontes, J. D., & Sotomayor, M. (2016). Structural determinants of adhesion by Protocadherin-19 and implications for its role in epilepsy. *Elife, 5*. doi:10.7554/eLife.18529

Depienne, C., Bouteiller, D., Keren, B., Cheuret, E., Poirier, K., Trouillard, O.,... Leguern, E. (2009). Sporadic infantile epileptic encephalopathy caused by mutations in PCDH19 resembles Dravet syndrome but mainly affects females. *PLoS Genet, 5*(2), e1000381. doi:10.1371/journal.pgen.1000381

Dibbens, L. M., Tarpey, P. S., Hynes, K., Bayly, M. A., Scheffer, I. E., Smith, R.,... Gecz, J. (2008). X-linked protocadherin 19 mutations cause female-limited epilepsy and cognitive impairment. *Nat Genet, 40*(6), 776-781. doi:10.1038/ng.149

Epi, K. C., Epilepsy Phenome/Genome, P., Allen, A. S., Berkovic, S. F., Cossette, P., Delanty, N.,... Winawer, M. R. (2013). De novo mutations in epileptic encephalopathies. *Nature, 501*(7466), 217-221. doi:10.1038/nature12439

Gecz, J., & Thomas, P. Q. (2020). Disentangling the paradox of the PCDH19 clustering epilepsy, a disorder of cellular mosaics. *Curr Opin Genet Dev, 65*, 169-175. doi:10.1016/j.gde.2020.06.012

Goodman, K. M., Rubinstein, R., Thu, C. A., Bahna, F., Mannepalli, S., Ahlsen, G.,... Shapiro, L. (2016). Structural Basis of Diverse Homophilic Recognition by Clustered alpha- and beta-Protocadherins. *Neuron, 90*(4), 709-723. doi:10.1016/j.neuron.2016.04.004

Karaca, E., Posey, J. E., Coban Akdemir, Z., Pehlivan, D., Harel, T., Jhangiani, S. N.,... Lupski, J. R. (2018). Phenotypic expansion illuminates multilocus pathogenic variation. *Genet Med, 20*(12), 1528-1537. doi:10.1038/gim.2018.33

Kolc, K. L., Sadleir, L. G., Depienne, C., Marini, C., Scheffer, I. E., Møller, R. S.,... Gecz, J. (2020). A standardized patient-centered characterization of the phenotypic spectrum of PCDH19 girls clustering epilepsy. *Transl Psychiatry, 10*(1), 127. doi:10.1038/s41398-020-0803-0

Kolc, K. L., Sadleir, L. G., Scheffer, I. E., Ivancevic, A., Roberts, R., Pham, D. H., & Gecz, J. (2018). A systematic review and meta-analysis of 271 PCDH19-variant individuals identifies psychiatric comorbidities, and association of seizure onset and disease severity. *Mol Psychiatry*. doi:10.1038/s41380-018-0066-9

Kolc, K. L., Sadleir, L. G., Scheffer, I. E., Ivancevic, A., Roberts, R., Pham, D. H., & Gecz, J. (2019). A systematic review and meta-analysis of 271 PCDH19-variant individuals identifies psychiatric comorbidities, and association of seizure onset and disease severity. *Mol Psychiatry, 24*(2), 241-251. doi:10.1038/s41380-018-0066-9

Lek, M., Karczewski, K. J., Minikel, E. V., Samocha, K. E., Banks, E., Fennell, T.,... Exome Aggregation, C. (2016). Analysis of protein-coding genetic variation in 60,706 humans. *Nature, 536*(7616), 285-291. doi:10.1038/nature19057

Li, Q., & Wang, K. (2017). InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am J Hum Genet, 100*(2), 267-280. doi:10.1016/j.ajhg.2017.01.004

Li, S. C., Goto, N. K., Williams, K. A., & Deber, C. M. (1996). Alpha-helical, but not beta-sheet, propensity of proline is determined by peptide environment. *Proc Natl Acad Sci U S A, 93*(13), 6676-6681. doi:10.1073/pnas.93.13.6676

Lindy, A. S., Stosser, M. B., Butler, E., Downtain-Pickersgill, C., Shanmugham, A., Retterer, K.,... McKnight, D. A. (2018). Diagnostic outcomes for genetic testing of 70 genes in 8565 patients with epilepsy and neurodevelopmental disorders. *Epilepsia, 59*(5), 1062-1071. doi:10.1111/epi.14074

Lotte, J., Bast, T., Borusiak, P., Coppola, A., Cross, J. H., Dimova, P.,... Kluger, G. (2016). Effectiveness of antiepileptic therapy in patients with PCDH19 mutations. *Seizure, 35*, 106-110. doi:10.1016/j.seizure.2016.01.006

MacArthur, D. G., Manolio, T. A., Dimmock, D. P., Rehm, H. L., Shendure, J., Abecasis, G. R.,... Gunter, C. (2014). Guidelines for investigating causality of sequence variants in human disease. *Nature, 508*(7497), 469-476. doi:10.1038/nature13127

Marini, C., Darra, F., Specchio, N., Mei, D., Terracciano, A., Parmeggiani, L.,... Guerrini, R. (2012). Focal seizures with affective symptoms are a major feature of PCDH19 gene-related epilepsy. *Epilepsia, 53*(12), 2111-2119. doi:10.1111/j.1528-1167.2012.03649.x

Nagar, B., Overduin, M., Ikura, M., & Rini, J. M. (1996). Structural basis of calcium-induced E-cadherin rigidification and dimerization. *Nature, 380*(6572), 360-364. doi:10.1038/380360a0

Pederick, D. T., Richards, K. L., Piltz, S. G., Kumar, R., Mincheva-Tasheva, S., Mandelstam, S. A.,... Thomas, P. Q. (2018). Abnormal Cell Sorting Underlies the Unique X-Linked Inheritance of PCDH19 Epilepsy. *Neuron, 97*(1), 59-66 e55. doi:10.1016/j.neuron.2017.12.005

Perucca, P., Scheffer, I. E., Harvey, A. S., James, P. A., Lunke, S., Thorne, N.,... Kwan, P. (2017). Real-world utility of whole exome sequencing with targeted gene analysis for focal epilepsy. *Epilepsy Res, 131*, 1-8. doi:10.1016/j.eplepsyres.2017.02.001

Pham, D. H., Tan, C. C., Homan, C. C., Kolc, K. L., Corbett, M. A., McAninch, D.,... Gecz, J. (2017). Protocadherin 19 (PCDH19) interacts with paraspeckle protein NONO to co-regulate gene expression with estrogen receptor alpha (ERalpha). *Hum Mol Genet, 26*(11), 2042-2052. doi:10.1093/hmg/ddx094

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J.,... Committee, A. L. Q. A. (2015). 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, 17*(5), 405-424. doi:10.1038/gim.2015.30

Scheffer, I. E., Turner, S. J., Dibbens, L. M., Bayly, M. A., Friend, K., Hodgson, B.,... Berkovic, S. F. (2008). Epilepsy and mental retardation limited to females: an under-recognized disorder. *Brain, 131*(Pt 4), 918-927. doi:10.1093/brain/awm338

Smith, L., Singhal, N., El Achkar, C. M., Truglio, G., Rosen Sheidley, B., Sullivan, J., & Poduri, A. (2018). PCDH19-related epilepsy is associated with a broad neurodevelopmental spectrum. *Epilepsia, 59*(3), 679-689. doi:10.1111/epi.14003

Tai, K., Kubota, M., Shiono, K., Tokutsu, H., & Suzuki, S. T. (2010). Adhesion properties and retinofugal expression of chicken protocadherin-19. *Brain Res, 1344*, 13-24. doi:10.1016/j.brainres.2010.04.065

Traynelis, J., Silk, M., Wang, Q., Berkovic, S. F., Liu, L., Ascher, D. B.,... Petrovski, S. (2017). Optimizing genomic medicine in epilepsy through a gene-customized approach to missense variant interpretation. *Genome Res, 27*(10), 1715-1729. doi:10.1101/gr.226589.117

Yang, H., & Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc, 10*(10), 1556-1566. doi:10.1038/nprot.2015.105

FIGURES

Figure 1: Performance of the three top, stand alone *in silico* tools in the assessment of *PCDH19* missense variants. 322 *PCDH19* missense variants: 238 from gnomAD [121 located within N-terminal EXD and 117 within the C-terminal CyR] and 84 *PCDH19*-CE (CE) variants in the N-terminal EXD region. We have assessed our data and confirmed that the assumptions of normality have been met based on visual inspection of histograms and generation of skew and kurtosis statistics. Statistical analysis were conducted by Student's two-tailed, unpaired *t* test, assuming unequal variance, and significance set as $p < 0.05$. p=4.705E^{-25} comparing MutPred2 score for gnomAD EXD variants versus MutPred2 score for CE EXD variants, p=9.763E^{-24} comparing MutationAssessor score for gnomAD EXD variants versus MutationAssessor score for CE EXD variants, p=1.491E^{-22} comparing GPP score for gnomAD EXD variants versus GPP score for CE EXD variants using Bonferroni adjusted planned comparisons.

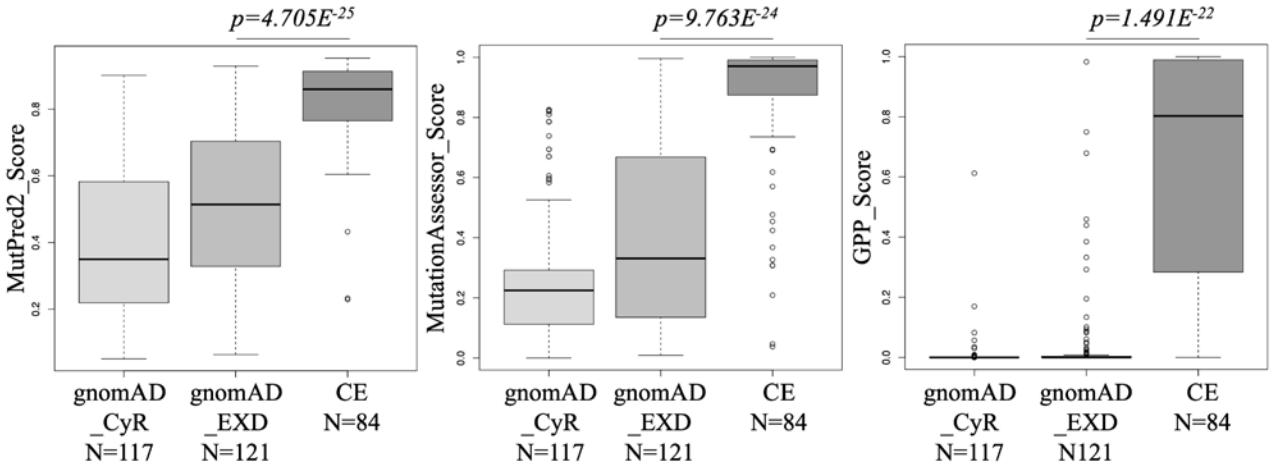

Figure 2: Functional assessment of the PCDH19 protein variants using protein structure analysis. (**a**) Overall structure of human PCDH19 homology model EC1-5 monomer shown in surface view (LHS) and ribbon view (RHS) in light blue. Structurally disruptive variants are highlighted in orange. Glycosylation modifications are shown in dark blue sticks and calcium ions as red spheres. A zoomed view of structurally disruptive variants are shown in PCDH19 ribbon view and highlighted in orange sticks. Red labels denote residues that cause defective calcium binding. Two variants, p.Asn557Lys and p.Asp558Asn were outside the EC1-5 region that could be accurately modelled, but are also predicted to disrupt calcium binding (data not shown). Purple labels denote structurally destabilizing mutations (**b**) EC3 domain mutants (**c**) EC4 domain mutants.

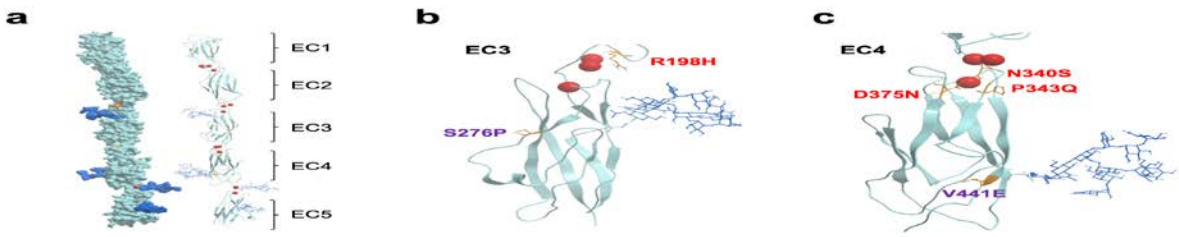

Figure 3: The effect of 23 *PCDH19* variants on ERE-LUC activity. (**a**) Dual luciferase reporter activity was assayed in MCF-7 cells transfected with a reporter

plasmid containing three copies of vitellogenin Estrogen Response Element (3× ERE TATA luc) and either control, wild-type (WT) or the variant Myc-PCDH19 expression vectors. Cells were initially cultured in charcoal-stripped FCS medium for 16 h and then for 6 h in the presence or absence of 10 nM estradiol (E2). The data is expressed as relative luciferase activity ± SD from 3 or more independent experiments for vector or WT and 6 or more independent experiments for the PCDH19 variants. *PCDH19*-CE variants (black circle), VUS (orange) and population variants (green). *** $p < 0.005$ or ** $p < 0.01$ comparing -E2 vs +E2 vector control; ***[1] $p < 0.005$ or **[1] $p < 0.01$ or *[1] $p < 0.05$ comparing –E2 vector control vs –E2 WT; ***[2] $p < 0.005$ or **[2] $p < 0.01$ or *[2] $p < 0.05$ comparing +E2 vector vs +E2 vector; ***[3] $p < 0.005$ or **[3] $p < 0.01$ or *[3] $p < 0.05$ comparing –E2 WT vs –E2 PCDH19 variant; **[4] $p < 0.01$ *[4] $p < 0.05$ comparing +E2 WT vs +E2 PCDH19 variant using Bonferroni adjusted planned comparisons. (**b**) Levels of PCDH19 protein were determined by western blotting with anti-Myc antibody. β-Actin was used as a loading control.

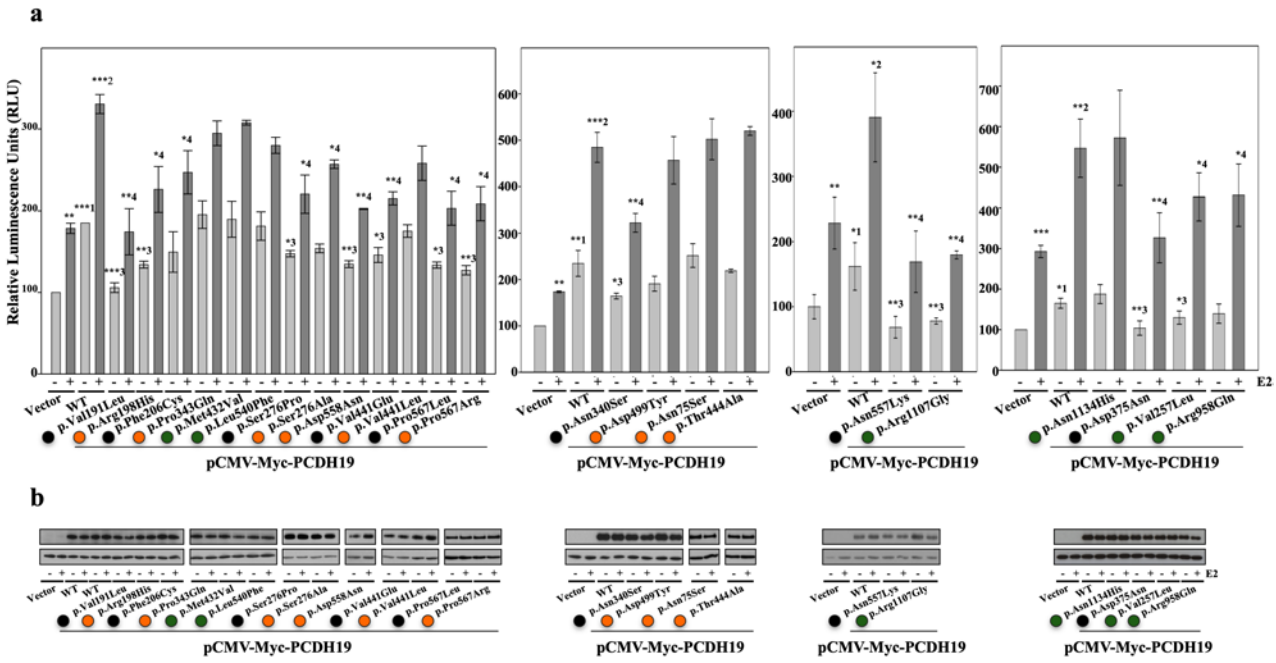

Figure 4: Stacked bar graph showing assessment of 20 *PCDH19* missense variants using three independent prediction tools. Each tool scored 1 to 4, that is Benign-1, Likely benign- 2, Likely Pathogenic-3 and Pathogenic-4. A score of ≥3 indicates that the variant is predicted to be Likely pathogenic from each tool: integrated *in silico* tool (black colour), protein structural prediction (dark grey) and *in vitro* functional analysis (light grey). Ultimately, the variants were classified as pathogenic, if they were predicted to be likely pathogenic by at least 2 of the 3 tools used and had a cumulative score of ≥7. Abbreviation: CE = *PCDH19*-CE, VUS = Variant of uncertain significance, Population = refers to three gnomAD variants with allele frequencies of 0.0015 for p.Val257Leu, 0.0012 for p.Met432Val and 0.0007 for p.Leu540Phe.

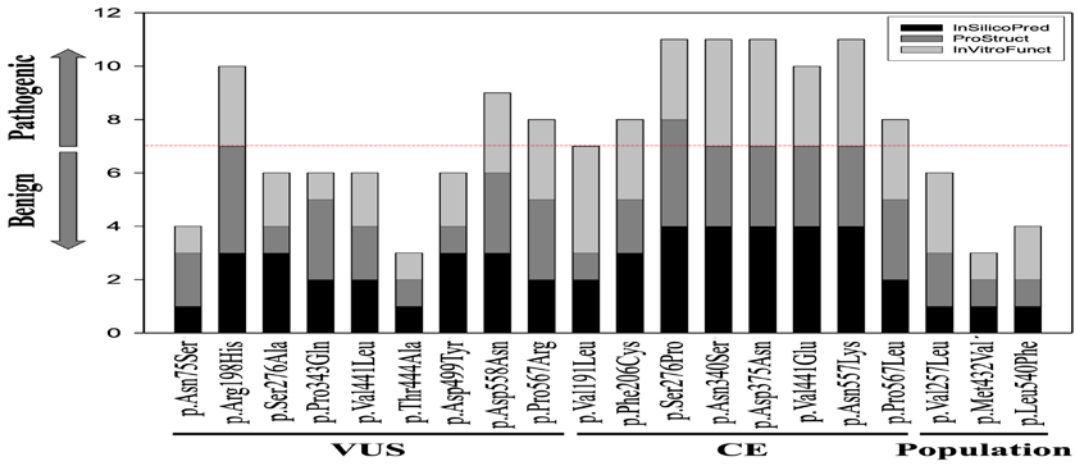

Table 1: Integrated *in silico* and experimental assessment combined with InterVar to classify 43 *PCDH19* missense (45 DNA) variants. Assessment of *PCDH19* variants using InterVar alone, integrated *in silico*, protein structural prediction, and functional analysis either alone and combined with InterVar. Abbreviations: CE = *PCDH19*-CE, P = pathogenic, B = benign, and N/A = not available.

cDNA	Protein	Location	Classes of Variants	Automated InterVar (ACMG/AMP guideline)	Integrated (Combined) *In Silico* Prediction Outcome P≥3	Protein structural Modelling P≥3	*In Vitro* Functional Assays P≥3	Final Prediction Outcome: Automated InterVar + Our study
219C>G	Asp73Glu	EC1 domain	VUS	Uncertain significance	B	N/A	N/A	Uncertain Significance
224A>G	Asn75Ser	EC1 domain	VUS	Uncertain Significance	B	B	B	Likely Benign
571G>C	Val191Leu	EC2 domain	CE	Uncertain Significance	B	B	P	Likely Pathogenic
593G>A	Arg198His	EC2 domain	VUS	Likely Pathogenic	P	P	P	Pathogenic
602A>C	Gln201Pro	EC2 domain	VUS	Uncertain Significance	P	N/A	N/A	Uncertain Significance
617T>G	Phe206Cys	EC2 domain	CE	Uncertain Significance	P	B	P	Likely Pathogenic
769G>C	Val257Leu	EC3 domain	Population	Uncertain Significance	B	B	P	Likely Pathogenic
779T>G	Leu260Arg	EC3 domain	VUS	Uncertain Significance	P	B	N/A	Uncertain Significance
799G>A	Glu267Lys	EC3 domain	VUS	Uncertain Significance	P	N/A	N/A	Uncertain Significance
824A>G	Tyr275Cys	EC3 domain	VUS	Uncertain significance	P	N/A	N/A	Uncertain Significance

826T>C	Ser276Pro	EC3 domain	CE	Uncertain Significance	P	P	P	Likely Pathogenic
826T>G	Ser276Ala	EC3 domain	VUS	Uncertain Significance	P	B	B	Uncertain Significance
1019A>G	Asn340Ser	EC3 domain	CE	Likely Pathogenic	P	P	P	Pathogenic
1028C>A	Pro343Gln	EC3 domain	VUS	Uncertain Significance	B	P	B	Uncertain Significance
1123G>A	Asp375Asn	EC4 domain	CE	Uncertain Significance	P	P	P	Likely Pathogenic
1294A>G	Met432Val	EC4 domain	Population	Likely Benign	B	B	B	Benign
1322T>A	Val441Glu	EC4 domain	CE	Likely Pathogenic	P	P	P	Pathogenic
1321G>C	Val441Leu	EC4 domain	VUS	Uncertain Significance	B	B	B	Likely benign
1330A>G	Thr444Ala	EC4 domain	VUS	Uncertain Significance	B	B	B	Likely benign
1335C>G	Asp445Glu	EC4 domain	VUS	Uncertain Significance	P	B	N/A	Uncertain Significance
1335C>A	Asp445Glu	EC4 domain	VUS	Uncertain Significance	P	B	N/A	Likely Pathogenic
1469A>G	Tyr490Cys	EC5 domain	VUS	Likely Pathogenic	P	N/A	N/A	Likely Pathogenic
1469A>C	Tyr490Ser	EC5 domain	VUS	Likely Pathogenic	P	N/A	N/A	Likely Pathogenic
1495G>T	Asp499Tyr	EC5 domain	VUS	Uncertain Significance	P	B	B	Likely Benign
1618C>T	Leu540Phe	EC5 domain	Population	Likely Benign	B	B	B	Likely Benign
1671C>A	Asn557Lys	EC5 domain	VUS	Likely Pathogenic	P	P	P	Pathogenic
1671C>G	Asn557Lys	EC5 domain	CE	Uncertain	P	P	P	Pathogenic

				Significance				
1672G>C	Asp558His	EC5 domain	VUS	Uncertain Significance	P	N/A	N/A	Uncertain Significance
1672G>A	Asp558Asn	EC5 domain	VUS	Uncertain Significance	P	P	P	Likely Pathogenic
1682C>G	Pro561Arg	EC5 domain	CE	Likely Pathogenic	P	P	N/A	Likely Pathogenic
1700C>T	Pro567Leu	EC6 domain	CE	Uncertain Significance	B	P	P	Likely Pathogenic
1700C>G	Pro567Arg	EC6 domain	VUS	Likely Pathogenic	B	P	P	Pathogenic
1810A>G	Thr604Pro	EC6 domain	VUS	Uncertain Significance	P	N/A	N/A	Uncertain Significance
1877C>T	Thr626Ile	EC6 domain	VUS	Uncertain significance	B	N/A	N/A	Likely Benign
1919T>G	Leu640Arg	EC6 domain	VUS	Uncertain Significance	P	N/A	N/A	Likely Pathogenic

Table 1: *In silico* assessment of 43 *PCDH19* missense (45 DNA) variants (continuous)

cDNA	**Protein**	**Location**	**Classes of Variants**	**Automated InterVar (ACMG/AMP guideline)**	**Integrated (Combined) *In Silico* Prediction Outcome** **P≥3**	**Protein structural Modelling** **P≥3**	***In Vitro* Functional Assays** **P≥3**	**Final Prediction Outcome: ACMG/AMP guideline + Our study**
1960G>C	Ala654Pro	EC6 domain	VUS	Uncertain Significance	P	P	N/A	Uncertain Significance
2735G>A	Ser912Asn	CM1 domain	VUS	Uncertain Significance	B	N/A	N/A	Uncertain Significance
2873G>A	Arg958Gln	CM2 domain	Population	Likely benign	B	N/A	B	Likely benign
2938C>T	Arg980Cys	COOH-end	Population	Likely benign	B	N/A	N/A	Likely benign

3020A>C	Asp1007Ala	COOH-end	VUS	Uncertain significance	B	N/A	N/A	Uncertain Significance
3158G>A	Arg1053Gln	COOH-end	VUS	Uncertain significance	B	N/A	N/A	Uncertain Significance
3280C>G	Leu1094Val	COOH-end	Population	Likely benign	B	N/A	N/A	Likely benign
3319C>G	Arg1107Gly	COOH-end	Population	Uncertain Significance	B	N/A	P	Uncertain Significance
3400A>C	Asn1134His	COOH-end	Population	Likely Benign	B	N/A	B	Benign
3415G>A	Gly1139Ser	COOH-end	Population	Uncertain Significance	B	N/A	N/A	Likely Benign