Received: 25 February 2019 | Revised: 23 May 2019 | Accepted: 21 June 2019

DOI: 10.1002/humu.23849

SPECIAL ARTICLE

Human Mutation HGVS HUMAN GENOME VARIATION SOCIETY WILEY

Assessing the performance of in silico methods for predicting the pathogenicity of variants in the gene CHEK2, among Hispanic females with breast cancer

Alin Voskanian[1] | Panagiotis Katsonis[2] | Olivier Lichtarge[2,3] | Vikas Pejaver[4,5] | Predrag Radivojac[6] | Sean D. Mooney[4] | Emidio Capriotti[7] | Yana Bromberg[8,9,10] | Yanran Wang[8] | Max Miller[8] | Pier Luigi Martelli[11] | Castrense Savojardo[11] | Giulia Babbi[11] | Rita Casadio[11] | Yue Cao[12] | Yuanfei Sun[12] | Yang Shen[12] | Aditi Garg[13] | Debnath Pal[13] | Yao Yu[14] | Chad D. Huff[14] | Sean V. Tavtigian[15] | Erin Young[15] | Susan L. Neuhausen[16] | Elad Ziv[17] | Lipika R. Pal[18] | Gaia Andreoletti[19] | Steven E. Brenner[19] | Maricel G. Kann[1]

[1]Department of Biological Sciences, University of Maryland, Baltimore, Maryland

[2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas

[3]Department of Pharmacology, Computational and Integrative Biomedical Research Center, Baylor College of Medicine, Houston, Texas

[4]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, Washington

[5]The eScience Institute, University of Washington, Seattle, Washington

[6]Khoury College of Computer and Information Sciences, Northeastern University, Boston, Massachusetts

[7]BioFoID Unit, Department of Pharmacy and Biotechnology (FaBiT), University of Bologna, Bologna, Italy

[8]Department of Biochemistry and Microbiology, Rutgers University, New Brunswick, New Jersey

[9]Department of Genetics, Rutgers University, New Brunswick, New Jersey

[10]Institute for Advanced Study, Technical University of Munich, Garching, Germany

[11]Biocomputing Group, BiGeA/Giorgio Prodi Interdepartmental Center for Cancer Research, University of Bologna, Bologna, Italy

[12]Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas

[13]Department of Computational and Data Sciences, Indian Institute of Science, Bengaluru, India

[14]Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, Texas

[15]Huntsman Cancer Institute, University of Utah School of Medicine, Salt Lake City, Utah

[16]Department of Population Sciences, Beckman Research Institute of City of Hope, Duarte, California

[17]Division of General Internal Medicine, Department of Medicine, Institute of Human Genetics, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, California

[18]Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, Maryland

[19]Department of Plant and Microbial Biology, University of California, Berkeley, California

Correspondence
Maricel G. Kann, University of Maryland, Baltimore County, Department of Biological Sciences, 1000 Hilltop Circle, Baltimore, MD 21250.
Email: mkann@umbc.edu

Funding information
NIH, Grant/Award Numbers: R35GM124952, U01GM115486, U41 HG007346, R13 HG006650, R01CA CA184585, GM079656, GM066099; Andrew Sabin

Abstract

The availability of disease-specific genomic data is critical for developing new computational methods that predict the pathogenicity of human variants and advance the field of precision medicine. However, the lack of gold standards to properly train and benchmark such methods is one of the greatest challenges in the field. In response to this challenge, the scientific community is invited to participate in the Critical Assessment for Genome Interpretation (CAGI), where unpublished disease variants are available for classification by in silico methods. As part of the CAGI-5

Family Foundation Fellowships, Grant/Award Number: R01GM104390

challenge, we evaluated the performance of 18 submissions and three additional methods in predicting the pathogenicity of single nucleotide variants (SNVs) in checkpoint kinase 2 (CHEK2) for cases of breast cancer in Hispanic females. As part of the assessment, the efficacy of the analysis method and the setup of the challenge were also considered. The results indicated that though the challenge could benefit from additional participant data, the combined generalized linear model analysis and odds of pathogenicity analysis provided a framework to evaluate the methods submitted for SNV pathogenicity identification and for comparison to other available methods. The outcome of this challenge and the approaches used can help guide further advancements in identifying SNV-disease relationships.

KEYWORDS
breast cancer, CAGI, CHEK2, Hispanic women, SNV

1 | INTRODUCTION

Checkpoint kinase 2 (*CHEK2*) has been demonstrated to be an effector kinase in the DNA damage checkpoint pathway, thus variations in its sequence can lead to detrimental outcomes in downstream DNA replication processes (Chaturvedi et al., 1999). *CHEK2* alterations by single nucleotide variants (SNVs) are linked to DNA damage that is implicated in cases of breast cancer (BC; Apostolou & Papasotiriou, 2017). BC is the most commonly diagnosed cancer among female Hispanics in the United States and the leading cause of cancer related deaths (Lynce et al., 2016). By identifying pathogenic variants could lead to early diagnosis and/or prevention thus reducing mortality.

New methodologies to estimate the impact of variants on disease conditions are crucial for the advance of personal genome interpretation studies (Capriotti, Nehrt, Kann, & Bromberg, 2012; Peterson, Doughty, & Kann, 2013). These methods can be broadly divided into two categories: data-driven machine learning algorithms or algorithms trained with expert knowledge (Peterson et al., 2013). Unfortunately, the results from all these methods are not presented in a consistent way, mostly due to lack of benchmarks to assess the different methodologies (MacArthur et al., 2014; Pejaver, Mooney, & Radivojac, 2017).

Critical Assessment of Genome Interpretation (CAGI) is a community effort created to address the challenges around the benchmarking of computational methods for predicting phenotypic impacts of genomic variation (https://genomeinterpretation.org/). The purpose of the CAGI challenges is to provide a standardized framework for comparison across these different methodologies to classify human variants.The fifth CAGI experiment involved 14 challenges, one of which was the *CHEK2* challenge. The *CHEK2* challenge prompted the submitters to assign a *p* (case) value to a list of provided *CHEK2* variants to indicate whether they were protective or not. The *p* (case) values are numeric values between 0 and 1. A value of .5 would indicate that the variation is neutral (equal in both populations) while a $p < .5$ would be indicative of a variant that is protective. The CHEK2 protein reference used for this challenge can be found by the NCBI reference sequence ID: NP_001005735.1.

This challenge received 18 entries from eight different groups which were assessed in addition to three publicly available methodologies that differed from the ones used for the submissions: SIFT (Sim et al., 2012), Align-GVGD (Mathe et al., 2006; Tavtigian et al., 2006), and Blocks Substitution Matrix (BLOSUM) 62 (Jones, Taylor, & Thornton, 1992; Schwarz & Dayoff, 1979). We assessed these methods in two ways: (a) by providing a comparison on the performance of each method for each SNV using a generalized linear model (GLM); and (b) by calculating pathogenicity odds for each method. From our analysis, we showed that some methods performed well in both analysis measures, but also that each analysis method provided a different insight into the method's performance. Our GLM *p* values showed that Group 5.1 performed best followed by the submissions of Group 3. When evaluated on a subgroup, excluding variants with low number of cases/controls, Group 4.1 did best. However, if we considered both the odds of pathogenicity and significance of *p* values Group 3.1 performed well in both assessment methods.

2 | MATERIALS AND METHODS

2.1 | Data source

The participant *CHEK2* data were provided by Dr. Elad Ziv and Dr. Susan Neuhausen, from an R01-funded project of self-identified Hispanic women with BC who previously had tested negative for carrying a *BRCA1* or *BRCA2* (*BRCA*) variations. Further inclusion criteria were: age <51 years at BC diagnosis, bilateral BC, breast and ovarian cancer, odds ratio (OR) age at diagnosis between ages 51 and 70 years with a family history of BC in two or more first or second-degree relatives diagnosed at <70 years of age. Participants had been previously consented and enrolled in

center-specific Institutional Review Board approved protocols from three high-risk registry studies including the City of Hope Clinical Cancer Genomics Community Research Network (MacDonald, Blazer, & Weitzel, 2010), the University of California San Francisco Clinical Genetics and Prevention Program, and the University of Southern California Norris Comprehensive Cancer Center clinical genetics program. The data providers included *CHEK2* data from 1,078 Hispanics with familial BC who met inclusion criteria and 312 Hispanic controls from Southern California. The data providers also included *CHEK2* data from 887 participants from the Multiethnic Cohort (MEC) without BC (approximately half had diabetes) who were self-described Hispanics and had undergone whole exome sequencing at the Broad Sequencing Center. These controls are a subset of ExAC Hispanic samples. The data set of 53 *CHEK2* variants provided for the *CHEK2* challenge is described below and was made available to the community without the information about the unpublished case/control study findings for those sets.

2.2 | *CHEK2* variant data set

The data set included 43 exonic variants and 10 untranslated region variants. Of the variants, 34 were nonsynonymous SNVs, four were synonymous SNVs, one was a stop-gain SNV, two were nonframeshift deletions, and two were frameshift deletions. Submitters only predicted on the 34 exonic nonsynonymous SNVs. A summary of this subset can be seen in Figure 1. One of the nonsynonymous SNVs was misreported and did not match the amino acid in the reference (NP_001005735.1). For the purposes of the assessment, the misreported variant was adjusted to its correct initial amino acid.

2.3 | Reference methods

There are numerous methods available for variation analysis (Peterson et al., 2013; Tavtigian, Greenblatt, Lesueur, & Byrnes, 2008), some of which were used by the submitters. For the purpose of comparison, we selected the simplest approach a scoring function for amino acid variations regardless of the sequence context, BLOSUM62 (Henikoff & Henikoff, 1992; Jones et al., 1992), and two more sophisticated approaches that include position specific information and amino acid's physicochemical properties into the analysis, SIFT (Sim et al., 2012) and Align-GVGD (Mathe et al., 2006; Tavtigian et al., 2006), respectively.

2.3.1 | SIFT

SIFT (Sim et al., 2012) is an algorithm that predicts the probability of an amino acid variation affecting protein function. The algorithm was accessed online, https://sift.bii.a-star.edu.sg/www/SIFT_seq_submit2.html. The input files were the *CHEK2* sequence (NP_001005725.1) in FASTA format and the list of the variants as provided by CAGI. All other parameters were kept as defaults from December 2019. The output from this was a prediction of whether the variant would affect the protein function or not with a score and a confidence indicator.

2.3.2 | Align-GVGD

Align-GVGD was run using the default parameters as presented at http://agvgd.hci.utah.edu/agvgd_input.php. The Align-GVGD program (Mathe et al., 2006; Tavtigian et al., 2006) combines an alignment with amino acid physico-chemical characteristics to calculate the range of variation present at each position in the alignment (GV) and the distance of missense substitutions from that range of variation (GD).

Grades are assigned to each variation to provide an empirical mapping from GV-GD to genetic risk. For this analysis SNVs in C0 were considered benign, C15-C25 were considered indeterminate and C35-C65 were considered pathogenic.

2.3.3 | BLOSUM62

BLOSUM (Henikoff & Henikoff, 1992; Jones et al., 1992) was used to score the amino acid substitutions. Each substitution is given a score indicating the similarity or difference for each variant. BLOSUM62 (Henikoff & Henikoff, 1992; Jones et al., 1992) scores range from −4 to 11. A positive score indicates identity or common substitutions, a score of 0 is a commonly observed neutral substitution and a negative score indicates a rarely observed substitution, for this purpose assumed to be nonprotective. The more negative the BLOSUM62 substitution score is, the more likely it is that the variant results in a functionally significant, pathogenic, alteration of the protein.

2.4 | Submissions

For this challenge, CAGI received 18 submissions from eight different groups. The participating methods relied on evolutionary information, structural information, and machine learning approaches, providing predictions for all the 34 nonsynonymous SNVs. Only one method did not report a result for the misreported position, however, we estimated that this would not have significantly altered the assessment outcomes. Below is a brief description of each the methodologies that participated in this challenge provided by each submission group. The authors responsible for each submission are also provided as well as a link to the algorithms used, when available.

2.4.1 | Methodologies used by Group 1 (Olivier Lichtarge, Panagiotis Katsonis)

Available for nonprofit use at http://mammoth.bcm.tmc.edu/EvolutionaryAction

The Evolutionary Action (EA) measures the fitness effect of coding variations analytically from protein evolution data (Katsonis & Lichtarge, 2014), in the scale of 0 (benign) to 100 (pathogenic). The EA scores were used to calculate the *p* (case) that a CHEK2 variant is

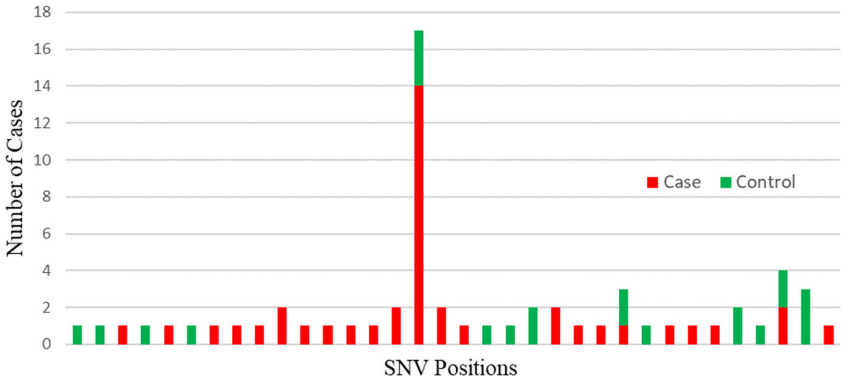

FIGURE 1 Distribution of patients in the case-control category per SNV position. SNV, single nucleotide variant

seen in cases rather than in controls. Ideally, benign variants are expected to be seen either in cases or in controls, while pathogenic variants are expected to be seen only in cases. Therefore, the EA scores were linearly transformed between 0.5 and 1, using: $p = 0.5 + EA/200$.

The input of EA was 60 homologous sequences of the human CHEK2 sequence (NP_009125). The homologous CHEK2 sequences were obtained by a standard protein BLAST (Altschul et al., 1997), using the UniRef100 database (Suzek et al., 2015). The sequences were selected to represent different evolutionary depths according to the sequence similarity of each homologous protein to human CHEK2 (the most distant homologous sequence was found in Caenorhabditis elegans). The selected sequences were aligned by using MUSCLE (Edgar, 2004). EA estimated the fitness effect of variations by using an equation that states the phenotype change ($\Delta\varphi$) equals to the product of the sensitivity of the mutated site to genotype changes ($\partial\varphi/\partial\gamma$) and the magnitude of the genotype change ($\Delta\gamma$). The sensitivity of the mutated site $\partial\varphi/\partial\gamma$ was approximated with the Evolutionary Trace algorithm which ranks the relative evolutionary importance of sequence positions in a family of aligned homologs (Lichtarge, Bourne, & Cohen, 1996) and the genotype change $\Delta\gamma$ was approximated with inverse amino acid substitution log-odds. The computed fitness change ($\Delta\varphi$), or EA score, has been shown to correlate with experimental loss of function, clinical association, morbidity, and mortality (Katsonis & Lichtarge, 2014; Neskey et al., 2015).

2.4.2 | Methodologies used by Group 2 (Vikas Pejaver, Predrag Radivojac, and Sean D. Mooney)

Available at http://mutpred.mutdb.org/#qform

MutPred2 (Pejaver, Mooney et al., 2017), an algorithm for the prediction of pathogenicity of missense variations (Pejaver, Urresti et al., 2017) was run on this data set to obtain scores between 0 and 1 representing the *p* (case). MutPred2 was run in the mode that uses gene-level homology count features. A score of 0 indicates a benign variation and a score of 1 indicates a pathogenic variation. The scores of MutPred2 approximate posterior probabilities (of pathogenicity, given sequence features). For each variant, this score was assumed to be equivalent to the probability that a variant is found in a case individual. Therefore, no further transformation of scores was performed.

2.4.3 | Methodologies used by Group 3 (Emidio Capriotti)

Various methods used available at http://biofold.org/resources.html

For the CHEK2 challenge, we predicted the presence of a given variant in cohorts of participants and controls using three methods and combinations of them. The predictions from PhD-SNP (Capriotti, Calabrese, & Casadio, 2006), PhD-SNPg (Capriotti & Fariselli, 2017) and SNPs&GO (Capriotti et al., 2013; Capriotti, Martelli, Fariselli, & Casadio, 2017) were used. These methods implement different machine learning algorithms using the protein, gene sequence-based and functional features as input. In particular, PhD-SNP and PhD-SNPg represent the simplest class of methods relying on sequence conservation scores calculated on protein and gene levels, respectively. The protein conservation score used in PhD-SNP were extracted from a BLAST (Altschul et al., 1997) search on the UniRef90 protein data set (Suzek, Huang, McGarvey, Mazumder, & Wu, 2007), while the PhyloP conservation (Pollard, Hubisz, Rosenbloom, & Siepel, 2010), calculated at nucleotide level, was included in the input features of PhD-SNPg. The PhyloP scores used, were available from the UCSC genome browser (https://genome.ucsc.edu/). A more sophisticated approach was used in SNPs&GO, which used functional information encoded by Gene Ontology (GO) terms as input. Specifically for this challenge, five sets of predictions were submitted: three of them considering separately the probabilistic output of PhD-SNP and PhD-SNPg, and the remaining ones used the average output of SNPs&GO with each of the other two methods.

2.4.4 | Methodologies used by Group 4 (Yana Bromberg, Yanran Wang, and Maximilian Miller)

Available at https://bromberglab.org/project/funtrp/ and https://bromberglab.org/project/snap/

Screen for Non-Acceptable Polymorphisms (SNAP; Bromberg & Rost, 2007), a neural network-based method for the prediction of the functional effects of nonsynonymous SNPs, was run on all missense

variations. In addition, fuNTRp was also run on these variations. (Miller, Vitale, Rost, & Bromberg, 2019)

The probability of each variation being more observed in the BC cohort was decided individually by assessing the reliability index (RI) of the SNAP prediction and the fuNTRp prediction. Specifically, a nonneutral SNAP variation at a toggle position was assigned a higher probability in the BC cohort and a higher *p* (case) value. A nonneutral SNAP variation at a rheostat or neutral position was assigned a *p* (case) closer to .5, and a neutral SNAP variation at a neutral position a *p* (case) of .5. A neutral SNAP variation at a rheostat or toggle position was assigned a score around .6 or .4. The allele frequencies of each variation in the general population were checked using ExAC, and if the variation had a relatively high variation (e.g. pI200T, MAF = 0.004), the *p* (case) was moved towards the healthy cohort ($p = 0.3$). If the variation was relatively rare, the *p* (case) was moved towards the BC cohort ($p > .5$). The distance the *p* (case) was moved depends manually on the confidence of SNAP and the fuNTRp prediction.

2.4.5 | Methodologies used by Group 5 (Pier Luigi Martelli, Castrense Savojardo, Giulia Babbi, and Rita Casadio)

Available at http://snps.biofold.org/snps-and-go/snps-and-go.html

Submission 5.1

The predictions were based on SNPs&GO (Calabrese, Capriotti, Fariselli, Martelli, & Casadio, 2009). SNPs&GO is a method based on Support Vector Machines for the prediction of deleterious single amino acid polymorphisms (SAP) using protein functional annotation. The output of SNPs&GO returns the effect (disease-associated variant or neutral variant) associated with a RI that is a number scoring from 0 (unreliable) to 10 (reliable). For each protein variant, the predictor scored the probability to be associated with human diseases. For the calibration of the final scores, a probability from .5 to 1 following a scale that is proportional to the RI of each prediction was assigned. When a variant is predicted to be neutral with an RI of 8, the predicted $p = .5$, when a variant is predicted to be disease associated with an RI of 8 the predicted $p = 1$. The maximum RI was set to 8 because this is the maximum RI found in this set of predictions. The prediction of one of the variants was manually curated because it was not coherent with the protein sequence. UniProt accession O96017 reports in that position a different amino acid. Finally, all predictions were assigned an arbitrary standard deviation (*SD*) of 0.1.

Submission 5.2

The predictions are based on the disease Index matrix (Casadio, Vassura, Tiwari, Fariselli, & Luigi Martelli, 2011). The matrix associates SAP with a corresponding probability to be associated with diseases (*p*d). For the calibration of the final scores, the *p*d values were scaled: considering neutral (prediction = 0.5) any SAP with a $pd \leq .4$, and disease-related (prediction = 1) any SAP with a $pd \geq .8$. All predictions were assigned an arbitrary *SD* of 0.1.

2.4.6 | Methodologies used by Group 6 (Yue Cao, Yuanfei Sun, and Yang Shen)

The training data set and the source code is available at: https://github.com/Shen-Lab/WSR-PredictPofPathogenicity/

BRCA-interacting tumor suppressor genes variation data from StringDB at https://string-db.org/cgi/ were identified and downloaded along with those of BRCA1/2 from ClinVar at https://www.ncbi.nlm.nih.gov/clinvar/. In total, 2,026 variations of six tumor suppressors (CHEK2, BRCA1, BRCA2, BRIP1, RBBP8, and TP53) were collected. Using MutPred2, 15 features were extracted; together with a constant as the16th feature, used in linear regression with a tailored loss function (Cao et al., 2019) Specifically, to describe a penalty more in line with the real biological processes while reducing the complexity of the optimization, the loss function needs to be convex and first-order differentiable. To accommodate these two conditions, a parabola-shaped polynomial of degree six as the loss function was implemented. The data is divided into five folds with four folds for training and one fold for testing and performed cross-validation of four-folds for optimizing the regularization constant C.

2.4.7 | Methodologies used by Group 7 (Aditi Garg, Debnath Pal)

Link to coarse-grained molecular dynamic simulation: http://pallab.cds.iisc.ac.in/CGMM/

Link to protein functional similarity match algorithm: http://pallab.cds.iisc.ac.in/dynfunc/

Two templates from the PDB (3I6W, 2CN5) corresponding to CHEK2 (92–586 residues) were used to create a single model by multichain modeling using Modeller (Fiser & Sali, 2003). The most stable structure was used to further create mutant models by replacing the specific amino acids. Each was subjected to C α atom-based MD simulation for 1 microsecond with Coarse-Grained Molecular Mechanics force field (Bhadra & Pal, 2014) at 300 K in vacuum. Identical parameters were used for each simulation namely, steepest descent energy minimization (max. force $\leq 100\ kJ \cdot mol^{-1} \cdot nm^{-1}$). A single short sequence of 273–279 K simulated annealing in six steps was used within the 70 ps equilibration step before reference temperature coupling. Structures during unconstrained dynamics simulation were recorded every 100 ps time from which 11 frames at every 100 ns were used for finding flexible regions based on RMSF norm (Bhadra & Pal, 2014). The filtered wild type protein and the variant pair were sent for a similarity score calculation using the formula: Similarity score = a/b, where a is the number of flexible regions in mutated protein and wild type (b). For residue positions, 1–91, a secondary structure was predicted using YASPIN (Lin, Simossis, Taylor, & Heringa, 2005).

Variation located at a position with regular secondary structure was deemed as damaging, while others were benign; E → Q variation

was treated as neutral. ATP binding is central to kinase activity and we found one segment (412–421) near this site which was flexible and could affect the ATP binding, and consequently the CHEK2 biochemical function. This segment was used in all variants for comparison and it was deemed that higher the similarity in flexibility of the variant protein to the wild type, the lesser are its chances to be damaging. The predicted similarity scores were normalized using identical reported cases from two previous studies for CHEK2 variants being damaging or benign (Desrichard, Bidet, Uhrhammer, & Bignon, 2011; Le Calvez-Kelm et al., 2011; Table S1). The ranges were mapped as follows: <70 damaging, 70–80 neutral, >80 benign for scores, and standard values of >.5 damaging, .5 neutral, <.5 benign for corresponding probabilities.

2.4.8 | Methodologies used by Group 8 (Yao Yu, Chad D. Huff, Sean V. Tavtigian, and Erin Young)

Accessible at http://www.hufflab.org/software/vaast/

Submissions 8.1–8.4 estimated *p* (case) using case-control and variant prioritization information as input to the Variant Annotation, Analysis & Search Tool (VAAST; Hu et al., 2013; Hu et al., 2014; Yandell et al., 2011). In the standard VAAST model, allele frequencies and ORs are estimated from the case-control data to test for gene and variant associations. In this application, the objective is to predict the phenotype status of a variant carrier for an established susceptibility gene. To meet this objective, the case allele frequency was constrained according to published CHEK2 variant ORs from one or more literature sources. *p* (case) was then estimated with the likelihood ratio from the VAAST model using case allele frequency data from BC cases in TCGA (Cancer Genome Atlas Network, 2012) and control allele frequency data from ExAC (Lek et al., 2016) (for details see Supporting Information Methods). *SD*s and confidence intervals (CIs) for *p* (case) were calculated using a parametric bootstrap with 10,000 iterations. For Submissions 8.5–8.6, OR was used as an approximation for the relative risk (RR) of a variant, with *p* (case) equal to RR/(RR + 1). SDs and CIs for *p* (case) were estimated using Monte Carlo simulations of reported ORs with 10,000 iterations.

Submission 8.1

Submission 8.1 applied the VAAST model using variant effect size estimates from (Young et al., 2016) for both truncating/splice-junction variants and missense variants "Overlap of missense analysis programs"; criteria).

Submission 8.2

This version applied the VAAST model with effect size estimates from (Young et al., 2016) for both truncating/splice-junction variants and missense variants "Overlap of missense analysis programs"; criteria). Known pathogenic variant information from ClinVar (Landrum et al., 2018) were also included. For any pathogenic variant reported in ClinVar, it was assumed that the probability of the alternative model was 1.

Submission 8.3

This version applied the VAAST model with used Align-GVGD variant effect sizes (Le Calvez-Kelm et al., 2011) and truncating/splice-junction estimates (Young et al., 2016).

Submission 8.4

This version applied the VAAST model with fixed variant effect sizes (Cybulski et al., 2011; truncating/splice-junction OR 3.6, causal missense OR 1.5 based on I157T).

Submission 8.5

Previously published estimates of CHEK2 variant effect size in BC as a function of one or more variant prioritization scores were applied. This version used effect size estimates (Young et al., 2016) for both truncating/splice-junction variants and missense variants "Overlap of missense analysis programs."

Submission 8.6

Previously published estimates of CHEK2 variant effect size in BC as a function of one or more variant prioritization scores were applied, along with Align-GVGD variant effect sizes (Le Calvez-Kelm et al., 2011) and truncating/splice-junction estimates (Young et al., 2016).

2.5 | Methodologies for benchmarking

The data set of 34 *CHEK2* SNVs was used to evaluate the 18 submission methods described above and compared with three existing techniques, used as reference, in the following ways:

1) A GLM which treats both the submissions and real experimental distinctions of variations as a continuous scale. The case/control participant data provided for each variant was used to compare each submission using a GLM with a GLM function in R (R Core Team, 2013) to calculate *p* values for each variant. As input, we provided the vector representing correct choices, where *X* was the participant category (case, control) and *Y* was the position as well as the entry per submission in the same format. The result was the intercept, standard error, *t* score, and *p* value. In addition, we obtained results for a subset of ten positions with more than one participant in at least one of the case or control group.
2) Analysis of the odds of pathogenicity was performed for participant data, clustered by each of the methodologies, as well as with BLOSUM62, Align-GVGD, and SIFT, into three groups, namely pathogenic, indeterminate, and benign. This method treats the submissions and the real experimental variations distinctions as categorical predictions. This assessment methodology identified the number of SNVs that each method categorized in benign, indeterminate, or pathogenic. To normalize across all methods the range of .45–.55 of *p* (case) were deemed indeterminate, *p* (case) values below that range were deemed benign and above were deemed pathogenic. The analysis proceeded in three steps. First, the number of case and control participants for the SNVs that the submitter categorized in one of the three groups was used to an OR for each of the categories.

ORs were also estimated for protein truncating variants (4.61; not part of the CAGI competition, but the data were available to the analysis team) and for the overall set of rare missense substitutions (2.15). Second, we estimated a proportion of pathogenic variants for the benign, indeterminate, and pathogenic categories in each methodology. To do this, we used the total number of missense substitutions placed in the category along with the categorical OR, the OR for truncating variants (4.61) and a theoretical OR of 1.00 for "pure" set of benign variants. We then determined the ratio of variants with OR = 4.61 to variants with OR = 1.00 that best approximated the observed categorical OR. Third, we estimated the odds in favor of pathogenicity for the benign, indeterminate, and pathogenic categories in each methodology. To do this, we treated the estimated proportion of pathogenic variants in the overall set of rare missense substitutions (0.32) as a prior probability ($p1$) and the proportion estimated for each of the three categories as a posterior probability ($p2$). Odds path were then estimated as Odds = ($p2 \times [1-p1]$)/([1− $p2$] x $p1$). These odds in favor of pathogenicity were compared against the ACMG scale (Tavtigian et al., 2018), where scores below 0.53 indicate strong benign, 0.053–0.481: benign, 0.481–2.08: indeterminate, 2.08–4.33: pathogenic, 4.33–18.72 moderate pathogenic, and scores above 18.72 are indicative of strong pathogenic relations.

3 | RESULTS

Figure 2 is a heatmap showing the p (case) predictions submitted by each of the methods separated into how the SNVs were classified in the participant data. The submission p (case) predictions for all variants is available (Table S2). Submission 8.3 and 8.4 generally made predictions in the neutral range, represented by the yellow color, while Submission 6 and 2 had most samples in the benign ranges, green. Submission 1, 5.1, 5.2, and 8.5 predicted most positions as pathogenic.

Table 1 shows p values for the 18 submissions and the three reference methods, namely SIFT, BLOSUM62, and Align-GVGD as calculated using the GLM in R. The ten positions sub grouped are the ones with more than one participant difference between case or control categories, and at least two participants for that position thus giving us higher confidence in the assessment of those variant positions. The remaining 24 *CHEK2* positions included in the "all positions" set were present in a total of only one participant, case or control. Method 8.6 lacks values for p values in the ten positions as all their p (case) values were the same not allowing for this type of analysis. The remaining results from the GLM analysis can be found in Table S3.

Submissions 2, 3.1, 3.2, 3.3, and 5.1, along with Align-GVGD, had significant p values at a .05 threshold, when looking at all positions. However, when looking at the selected ten positions, submissions 5.1 and 4 had p values under .05 in addition to SIFT. The p values were higher than .05 for all the other methods.

On the basis of the odds of pathogenicity, while the reference method Align-GVGD accurately classified SNVs in all three categories, the performance of the submitted groups varied in each of the categories. Figure 3 shows the odds of pathogenicity per submitter. It is distributed among categories (benign, indeterminate, and pathogenic) based on the submitter's provided predictions and then graphed according to the ACMG CI of pathogenicity. The size of the marker correlates to the number of positions that were attributed to that category.

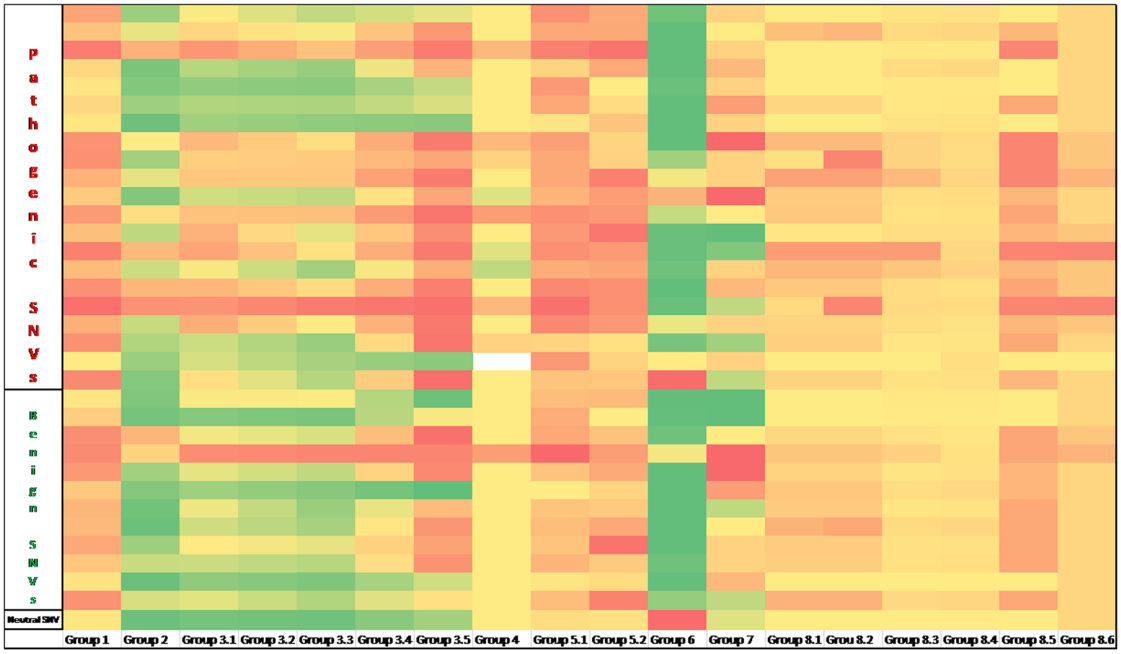

FIGURE 2 Representation of submission p (case). Green indicates values close to 0 (benign) and red indicates values close to 1 (pathogenic)

TABLE 1 The *p* values were calculated using the GLM function in R

Method	*p* value	
	All positions	Ten positions
SIFT	.60	.02
Align-GVGD	.03	.09
BLOSUM62	.43	1.28
Sub. 1	.16	.19
Sub. 2	.031	.061
Sub. 3.1	.016	.061
Sub. 3.2	.023	.065
Sub. 3.3	.039	.072
Sub. 3.4	.051	.15
Sub. 3.5	.11	.29
Sub. 4	.17	.037
Sub. 5.1	.002	.0035
Sub. 5.2	.24	.40
Sub. 6	.84	.86
Sub. 7	.67	.42
Sub. 8.1	.56	.06
Sub. 8.2	1.56	.12
Sub. 8.3	.31	1.54
Sub. 8.4	0.62	1.34
Sub. 8.5	0.25	.57
Sub. 8.6	0.31	N/A

Note: Method 8.6 lacks results for the subset of ten positions because there was not significant variation in their *p* (case) predictions per SNV.

Four submission groups performed accurately in two categories: 2 (benign, pathogenic) 3.1 (benign, pathogenic), 7 (benign, indeterminate), and 8.3 (indeterminate, pathogenic). In contrast, Groups 3.4 (pathogenic), 4 (pathogenic), 5.2 (indeterminate), 6 (indeterminate), 8.4 (indeterminate), 8.5 (indeterminate), and BLOSUM62 (benign) all had one of their three categories within the ACMG pathogenicity range used as a guideline here.

4 | DISCUSSION

The two methods of analysis, generalized linear modeling to assess in terms of *p* values and the odds of pathogenicity calculations, have their advantages but also several disadvantages. Both methods are impacted by the low number of datapoints in the original data set provided (Figure 1) and the unequal distribution of case and control instances. To account for some of this impact, we chose a subset of ten positions for the *p* value analysis. These ten positions all had more than one case that led to their classification as pathogenic or benign, giving us higher confidence on the statistical analysis. This subset of positions provided a framework where we no longer had zero participants in one category and just one in the other, but the data set was skewed, with a larger number of positions represented in case than in the control participant population.

Our results for the *p* value analysis in all positions and the subset of ten positions showed inconsistency in the performance of the methods. Five of the submitted methods (5.1, 3.1, 3.2, 2, and 3.3) and Align-GVGD had significant *p* values when analyzed across all SNVs while only three submissions (5.1 and 4) and SIFT performed well across the ten positions. These results confirm that methods which categorize more positions in the pathogenic category perform better since both in the overall sample and particularly in the ten extracted positions, more BC participants were present. This further illustrates the benefits of prediction methods favoring pathogenic predictions. Although, without a similar benchmarking data set that is more heavily benign, we could not assess whether this was just an artifact of the methods or an accurate bias towards the correct SNVs classification.

The reference methods, SIFT, Align-GVGD, and BLOSUM62, had some limitations in their evaluations. While BLOSUM62 and Align-GVGD have three categories, SIFT operates on a binary scale; this can explain why SIFT outperformed in the *p* values, among the ten positions, which with the exception of one were all confidently placed in case or control categories. BLOSUM62 as expected, being a simpler scoring, did very poorly.

To further understand the distribution of the *p* (case) submitted by each method, we had chosen three representative positions for each base (pathogenic, neutral, and protective; Figure 4). Thirteen groups predicted the pathogenic position correctly while only three called the protective variation correctly and nine called the neutral position correctly. The incorrect predictions generally fell in the pathogenic range, this may show some bias in how the methods evaluated the SNVs pathogenicity. This disbalance was also reflected

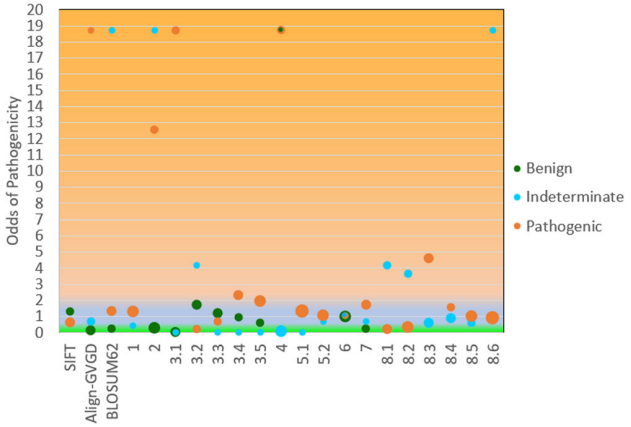

FIGURE 3 Summary of odds of pathogenicity results for all the submissions and reference methods used in the assessment. Dot size is proportional to the number of positions in each of the benign (green), indeterminate (cyan), and pathogenic (red) categories. The graph background color shows the ACMG odds of pathogenicity range use as a guide (green—benign, blue—indeterminate, orange—pathogenic)

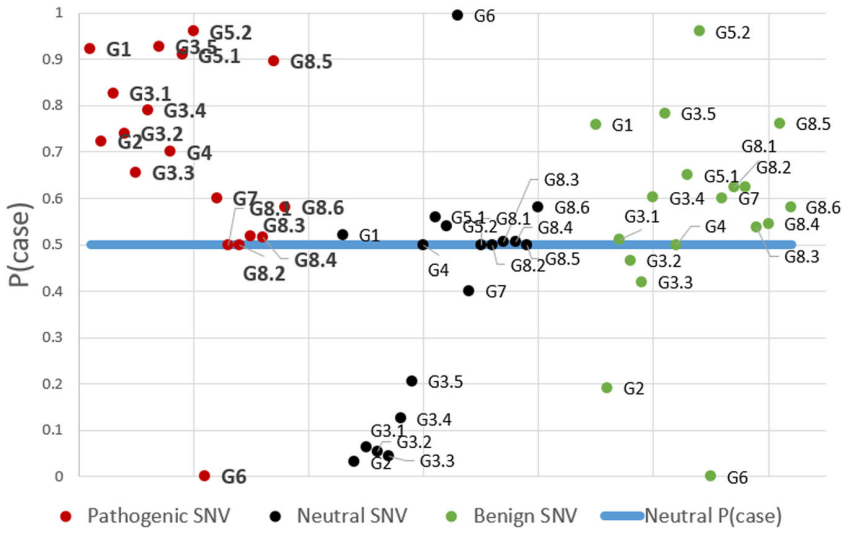

FIGURE 4 Representative positions for cancerous, neutral, and protective variants. Red dots represent a cancerous position, black dots represent a neutral position, and green dots represent protective variants. The blue line indicates the .5 neutral *p* value. SNV, single nucleotide variant

in the data set provided. Based on the *p* (case) predictions for these three positions, Group 4 performed the most accurately.

We performed an alternative assessment of the submissions by analyzing the odds of pathogenicity, which focused on the overall number of variants in each of the three categories instead of the performance per variant as in the *p* value assessment.

As shown in our results (Figure 3), Groups 3.1 and 8.3 performances were more closely correlated with the correct ACMG CI of pathogenicity. Even though our previous *p* value analysis indicated that Group 5.1 performed better, because they placed more positions in the pathogenic category, they performed worst in this analysis. The lack of range in some of the submitted method predictions, which is shown in the heatmap (Figure 2) is also clearly visible in Figure 3. Thus, the groups that more heavily favored one side of the scale were more scrutinized in this method of analysis. Also, many methods categorized the benign as indeterminate (range of 0.483–2.08). However, this could be an artifact of the preprocessing for this analysis. Due to the lack of consistent *SD*s across methods, we considered the arbitrary range of 0.45–0.55 to fall into the indeterminate category.

The pathogenicity odds results indicated that Align-GVGD was the best performer among the reference methods. However, this discrepancy may be due to SIFT relying on a binary classification and Align-GVGD having results that were more consistent with the challenge framework.

Regardless, Align-GVGD had a *p* value under .05 and performed well in the odds of pathogenicity analysis as did submissions by Groups 2 and 3.1, which utilized PhD-SNP, SNP&GO (Calabrese et al., 2009) and MutPred2 (Pejaver, Mooney et al., 2017).

5 | CONCLUSIONS

The analysis methods used each had different advantages. While the GLM analysis provided an assessment of the method's performance by specific positions and the classification it was given, the odds of pathogenicity assessed the overall classification of pathogenic, benign, and indeterminate positions. With relevance to the reference methods, Align-GVGD performed well overall, but not when estimating *p* values with the subset of 10 positions. From our assessment on this data set, which presented a strong bias towards pathogenic SNVs, it was difficult to extrapolate whether results would be different if the data set represented pathogenic and benign groups equally. We also concluded that it would be beneficial to perform this challenge and analysis with a larger, less biased data set. The comparison of these two results gave us a better idea of the correctness of the submissions, the appropriateness of the analysis method and the structure of the challenge.

ACKNOWLEDGMENT

The CAGI experiment coordination is supported by NIH U41 HG007346 and the CAGI conference by NIH R13 HG006650.

ORCID

Alin Voskanian http://orcid.org/0000-0003-2000-9574
Panagiotis Katsonis http://orcid.org/0000-0002-7172-1644
Vikas Pejaver http://orcid.org/0000-0002-1943-0284
Predrag Radivojac http://orcid.org/0000-0002-6769-0793
Sean D. Mooney http://orcid.org/0000-0003-2654-0833
Max Miller http://orcid.org/0000-0002-1335-9499
Yang Shen http://orcid.org/0000-0002-1703-7796
Debnath Pal http://orcid.org/0000-0002-3591-5978
Sean V. Tavtigian http://orcid.org/0000-0002-7543-8221
Gaia Andreoletti http://orcid.org/0000-0002-0452-0009
Steven Brenner http://orcid.org/0000-0001-7559-6185
Maricel G. Kann http://orcid.org/0000-0003-0116-3883

REFERENCES

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*, *25*(17), 3389–3402.

Apostolou, P., & Papasotiriou, I. (2017). Current perspectives on CHEK2 mutations in breast cancer. *Breast Cancer (Dove Med Press)*, *9*, 331–335. https://doi.org/10.2147/BCTT.S111394

Bhadra, P., & Pal, D. (2014). De novo inference of protein function from coarse-grained dynamics. *Proteins*, *82*(10), 2443–2454. https://doi.org/10.1002/prot.24609

Bromberg, Y., & Rost, B. (2007). SNAP: Predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Research*, *35*(11), 3823–3835. https://doi.org/10.1093/nar/gkm238

Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L., & Casadio, R. (2009). Functional annotations improve the predictive score of human disease-related mutations in proteins. *Human Mutation*, *30*(8), 1237–1244. https://doi.org/10.1002/humu.21047

Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature, 490(7418), 61–70. 10.1038/nature11412

Cao, Y., Sun, Y., Karimi, M., Chen, H., Moronfoye, O., & Shen, Y. (2019). Predicting pathogenicity of missense variants with weakly supervised regression. *Human Mutation*, https://doi.org/10.1101/545913

Capriotti, E., Calabrese, R., & Casadio, R. (2006). Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*, *22*(22), 2729–2734. https://doi.org/10.1093/bioinformatics/btl423

Capriotti, E., Calabrese, R., Fariselli, P., Martelli, P. L., Altman, R. B., & Casadio, R. (2013). WS-SNPs&GO: A web server for predicting the deleterious effect of human protein variants using functional annotation. *BMC Genomics*, *14*(Suppl 3), S6. https://doi.org/10.1186/1471-2164-14-S3-S6

Capriotti, E., & Fariselli, P. (2017). PhD-SNPg: A webserver and lightweight tool for scoring single nucleotide variants. *Nucleic Acids Research*, *45*(W1), W247–W252. https://doi.org/10.1093/nar/gkx369

Capriotti, E., Martelli, P. L., Fariselli, P., & Casadio, R. (2017). Blind prediction of deleterious amino acid variations with SNPs&GO. *Human Mutation*, *38*(9), 1064–1071. https://doi.org/10.1002/humu.23179

Capriotti, E., Nehrt, N. L., Kann, M. G., & Bromberg, Y. (2012). Bioinformatics for personal genome interpretation. *Briefings in Bioinformatics*, *13*(4), 495–512. https://doi.org/10.1093/bib/bbr070

Casadio, R., Vassura, M., Tiwari, S., Fariselli, P., & Luigi Martelli, P. (2011). Correlating disease-related mutations to their effect on protein stability: A large-scale analysis of the human proteome. *Human Mutation*, *32*(10), 1161–1170. https://doi.org/10.1002/humu.21555

Chaturvedi, P., Eng, W. K., Zhu, Y., Mattern, M. R., Mishra, R., Hurle, M. R., & Zhou, B. B. (1999). Mammalian Chk2 is a downstream effector of the ATM-dependent DNA damage checkpoint pathway. *Oncogene*, *18*(28), 4047–4054. https://doi.org/10.1038/sj.onc.1202925

Cybulski, C., Wokołorczyk, D., Jakubowska, A., Huzarski, T., Byrski, T., Gronwald, J., & Lubiński, J. (2011). Risk of breast cancer in women with a CHEK2 mutation with and without a family history of breast cancer. *Journal of Clinical Oncology*, *29*(28), 3747–3752. https://doi.org/10.1200/JCO.2010.34.0778

Desrichard, A., Bidet, Y., Uhrhammer, N., & Bignon, Y. J. (2011). CHEK2 contribution to hereditary breast cancer in non-BRCA families. *Breast Cancer Research*, *13*(6), R119. https://doi.org/10.1186/bcr3062

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Fiser, A., & Sali, A. (2003). Modeller: Generation and refinement of homology-based protein structure models. *Methods in Enzymology*, *374*, 461–491. https://doi.org/10.1016/S0076-6879(03)74020-8

Henikoff, S., & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(22), 10915–10919.

Hu, H., Huff, C. D., Moore, B., Flygare, S., Reese, M. G., & Yandell, M. (2013). VAAST 2.0: Improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution matrix. *Genetic Epidemiology*, *37*(6), 622–634. https://doi.org/10.1002/gepi.21743

Hu, H., Roach, J. C., Coon, H., Guthery, S. L., Voelkerding, K. V., Margraf, R. L., & Huff, C. D. (2014). A unified test of linkage analysis and rare-variant association for analysis of pedigree sequence data. *Nature Biotechnology*, *32*(7), 663–669. https://doi.org/10.1038/nbt.2895

Jones, D. T., Taylor, W. R., & Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*, *8*(3), 275–282.

Katsonis, P., & Lichtarge, O. (2014). A formal perturbation equation between genotype and phenotype determines the Evolutionary Action of protein-coding variations on fitness. *Genome Research*, *24*(12), 2050–2058. https://doi.org/10.1101/gr.176214.114

Landrum, M. J., Lee, J. M., Benson, M., Brown, G. R., Chao, C., Chitipiralla, S., & Maglott, D. R. (2018). ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Research*, *46*(D1), D1062–D1067. https://doi.org/10.1093/nar/gkx1153

Le Calvez-Kelm, F., Lesueur, F., Damiola, F., Vallée, M., Voegele, C., Babikyan, D., & Registry, B. C. F. (2011). Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: Results from a breast cancer family registry case-control mutation-screening study. *Breast Cancer Research*, *13*(1), R6. https://doi.org/10.1186/bcr2810

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

Lichtarge, O., Bourne, H., & Cohen, F. (1996). An evolutionary trace method defines binding surfaces common to protein families. *Journal of Molecular Biology*, *257*(2), 342–358.

Lin, K., Simossis, V. A., Taylor, W. R., & Heringa, J. (2005). A simple and fast secondary structure prediction method using hidden neural networks. *Bioinformatics*, *21*(2), 152–159. https://doi.org/10.1093/bioinformatics/bth487

Lynce, F., Graves, K. D., Jandorf, L., Ricker, C., Castro, E., Moreno, L., & Vadaparampil, S. T. (2016). Genomic disparities in breast cancer among Latinas. *Cancer Control*, *23*(4), 359–372. https://doi.org/10.1177/107327481602300407

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. https://doi.org/10.1038/nature13127

MacDonald, D. J., Blazer, K. R., & Weitzel, J. N. (2010). Extending comprehensive cancer center expertise in clinical cancer genetics and genomics to diverse communities: The power of partnership. *Journal of the National Comprehensive Cancer Network: JNCCN*, *8*(5), 615–624.

Mathe, E., Olivier, M., Kato, S., Ishioka, C., Hainaut, P., & Tavtigian, S. V. (2006). Computational approaches for predicting the biological effect of p53 missense mutations: A comparison of three sequence analysis based methods. *Nucleic Acids Research*, *34*(5), 1317–1325. https://doi.org/10.1093/nar/gkj518

Miller, M., Vitale, D., Rost, B., & Bromberg, Y. (2019). fuNTRp: Identifying protein positions for variation driven functional tuning. Bioinformatics. https://doi.org/10.1101/578757

Neskey, D. M., Osman, A. A., Ow, T. J., Katsonis, P., McDonald, T., Hicks, S. C., & Lichtarge, O. (2015). Evolutionary action score of TP53 identifies high-risk mutations associated with decreased survival and increased distant metastases in head and neck cancer. *Cancer Research*, *75*(7), 1527–1536. https://doi.org/10.1158/0008-5472.can-14-2735

Pejaver, V., Mooney, S. D., & Radivojac, P. (2017). Missense variant pathogenicity predictors generalize well across a range of function-specific prediction challenges. *Human Mutation*, *38*(9), 1092–1108. https://doi.org/10.1002/humu.23258

Pejaver, V., Urresti, J., Lugo-Martinez, J., Pagel, K., Lin, G. N., Nam, H., … Radivojac, P. (2017). MutPred2: Inferring the molecular and phenotypic impact of amino acid variants.

Peterson, T. A., Doughty, E., & Kann, M. G. (2013). Towards precision medicine: Advances in computational approaches for the analysis of human variants. *Journal of Molecular Biology*, *425*(21), 4047–4063. https://doi.org/10.1016/j.jmb.2013.08.008

Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., & Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Research*, *20*(1), 110–121. https://doi.org/10.1101/gr.097857.109

R Core Team (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Schwarz, R., & Dayoff, M. (1979). Matrices for detecting distant relationships. In Dayhoff, M. (Ed.), *Atlas of Protein Sequences*. Silver Spring, MD: National Biomedical Research Foundation.

Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res*, *40*(Web Server issue), W452–W457. https://doi.org/10.1093/nar/gks539

Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., & Wu, C. H. (2007). UniRef: Comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*, *23*(10), 1282–1288. https://doi.org/10.1093/bioinformatics/btm098

Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., & Wu, C. H. UniProt Consortium (2015). UniRef clusters: A comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics*, *31*(6), 926–932. https://doi.org/10.1093/bioinformatics/btu739

Tavtigian, S. V., Deffenbaugh, A. M., Yin, L., Judkins, T., Scholl, T., Samollow, P. B., & Thomas, A. (2006). Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *Journal of Medical Genetics*, *43*(4), 295–305. https://doi.org/10.1136/jmg.2005.033878

Tavtigian, S. V., Greenblatt, M. S., Harrison, S. M., Nussbaum, R. L., Prabhu, S. A., & Boucher, K. M. C. S. V. I. W. G (ClinGen SVI) (2018). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genetics in Medicine*, *20*(9), 1054–1060. https://doi.org/10.1038/gim.2017.210

Tavtigian, S. V., Greenblatt, M. S., Lesueur, F., & Byrnes, G. B. IARC Unclassified Genetic Variants Working Group (2008). In silico analysis of missense substitutions using sequence-alignment based methods. *Human Mutation*, *29*(11), 1327–1336. https://doi.org/10.1002/humu.20892

Yandell, M., Huff, C., Hu, H., Singleton, M., Moore, B., Xing, J., & Reese, M. G. (2011). A probabilistic disease-gene finder for personal genomes. *Genome Research*, *21*(9), 1529–1542. https://doi.org/10.1101/gr.123158.111

Young, E. L., Feng, B. J., Stark, A. W., Damiola, F., Durand, G., Forey, N., Francy, T. C., … Tavtiqian, S. V. Breast Cancer Family Registry (2016). Multigene testing of moderate-risk genes: Be mindful of the missense. *Journal of Medical Genetics*, *53*(6), 366–376. https://doi.org/10.1136/jmedgenet-2015-103398

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Voskanian A, Katsonis P, Lichtarge O, et al. Assessing the performance of in silico methods for predicting the pathogenicity of variants in the gene CHEK2, among Hispanic females with breast cancer. *Human Mutation*. 2019;40:1612–1622. https://doi.org/10.1002/humu.23849