

RESEARCH ARTICLE

Comprehensive analysis of the functional impact of single nucleotide variants of human *CHEK2*Claire E. McCarthy-Leo¹, George S. Brush², Roger Pique-Regi^{1,3}, Francesca Luca^{1,3a}, Michael A. Tainsky^{1,2}, Russell L. Finley, Jr.^{1*}

1 Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, Detroit, Michigan, United States of America, **2** Department of Oncology, Molecular Therapeutics Program, Barbara Ann Karmanos Cancer Institute, Wayne State University School of Medicine, Detroit, Michigan, United States of America, **3** Department of Obstetrics and Gynecology, Wayne State University School of Medicine, Detroit, Michigan, United States of America

✉ Current address: Department of Human Genetics, University of Chicago, Chicago, Illinois, United States of America

* rfinley@wayne.edu



OPEN ACCESS

Citation: McCarthy-Leo CE, Brush GS, Pique-Regi R, Luca F, Tainsky MA, Finley RL, Jr (2024) Comprehensive analysis of the functional impact of single nucleotide variants of human *CHEK2*. PLoS Genet 20(8): e1011375. <https://doi.org/10.1371/journal.pgen.1011375>

Editor: Haico van Attikum, Leiden University Medical Center, NETHERLANDS, KINGDOM OF THE

Received: January 7, 2024

Accepted: July 25, 2024

Published: August 15, 2024

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1011375>

Copyright: © 2024 McCarthy-Leo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are

Abstract

Loss of function mutations in the checkpoint kinase gene *CHEK2* are associated with increased risk of breast and other cancers. Most of the 3,188 unique amino acid changes that can result from non-synonymous single nucleotide variants (SNVs) of *CHEK2*, however, have not been tested for their impact on the function of the *CHEK2*-encoded protein (CHK2). One successful approach to testing the function of variants has been to test for their ability to complement mutations in the yeast ortholog of *CHEK2*, *RAD53*. This approach has been used to provide functional information on over 100 *CHEK2* SNVs and the results align with functional assays in human cells and known pathogenicity. Here we tested all but two of the 4,887 possible SNVs in the *CHEK2* open reading frame for their ability to complement *RAD53* mutants using a high throughput technique of deep mutational scanning (DMS). Among the non-synonymous changes, 770 were damaging to protein function while 2,417 were tolerated. The results correlate well with previous structure and function data and provide a first or additional functional assay for all the variants of uncertain significance identified in clinical databases. Combined, this approach can be used to help predict the pathogenicity of *CHEK2* variants of uncertain significance that are found in susceptibility screening and could be applied to other cancer risk genes.

Author summary

Variations in gene sequences account for many of the phenotypic differences between individuals, including different susceptibilities to diseases like cancer. Individuals who inherit specific variants in the *CHEK2* gene, for example, have increased risk of developing breast cancer and other cancers. These cancer-associated variants all abolish the normal activity of the *CHEK2*-encoded protein, a protein kinase called Chk2. Most of the *CHEK2*

within the paper and its [Supporting Information](#) files. Raw data used for producing all graphs and figures are in Supplemental Tables 1, 2, and 3.

Funding: This work was supported by a grant from the Wayne State University Center for Molecular Medicine and Genetics (to RLF), the Barbara and Fred Erb Endowed Chair in Cancer Genetics to MAT, and Rumble Fellowship from the Graduate School of Wayne State University (CML). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

variants found in individuals, however, have an unknown effect on the function of the Chk2 protein, and therefore an unknown potential role in cancer. This is because it is cumbersome to individually make and test each newly identified variant to see whether it affects protein function. Here we use an efficient version of a technique called deep mutational scanning to test nearly all possible *CHEK2* single nucleotide variants for their impact on Chk2 protein function. Analyses of the resulting data indicate that it will be useful for developing models to predict the cancer risk in individuals who inherit any *CHEK2* variants. This study also demonstrates a viable efficient approach to study variants of other genes, including other cancer related genes.

Introduction

Cell cycle checkpoint kinase 2 (CHK2) is a conserved protein involved in cell cycle control and DNA damage responses. First discovered in 1998 as the mammalian ortholog of Rad53 in *Saccharomyces cerevisiae*, CHK2 has established roles in DNA double strand break repair, DNA damage-induced apoptosis and cell cycle arrest [1]. CHK2 is a 543 amino acid nuclear serine/threonine protein kinase that contains three distinct functional domains; an SQ/TQ cluster domain (SCD), a forkhead-associated (FHA) domain and a kinase domain. When DNA damage occurs, activated ATM phosphorylates CHK2 at T68 [2–4]. This phosphorylation induces dimerization of CHK2 monomers through the binding of the phosphorylated SCD of one CHK2 monomer with the FHA domain of another, initiating autophosphorylation within the kinase domain (S260, T420), the activation loop (T383, T387), and the C-terminus (S516) [5–7]. Activated CHK2 monomers spread the DNA damage signal by phosphorylating downstream protein targets to halt the cell cycle, recruit various proteins to the site of DNA damage, and initiate apoptosis when necessary [1,8–10].

CHEK2 is a moderately penetrant gene that has been studied for its association with cancer predisposition [11]. Women with the heterozygous founder mutation *CHEK2* c.1100delC, which leads to a protein truncation in the kinase domain, have a 2–3-fold increased risk of developing breast cancer [12–14]. This pathogenic variant has also been associated with increased risk for male breast cancer, gastric, testicular, prostate, and thyroid cancers [15–20]. Additional *CHEK2* variants that impair protein function have been identified in the germline of individuals with cancer and are associated with the development of cancer [21–27]. Standards for the clinical classification of sequence variants established by the American College of Medical Genetics and Genomics (ACMG) use multiple pieces of evidence (family history, allele frequency, segregation and functional data, etc.) to classify sequence variants as either pathogenic (associated with increased cancer risk) or benign [28]. Most variants detected in susceptibility screening, however, lack supporting evidence by statistically significant population data to classify them as either pathogenic or benign, resulting in their classification as variants of uncertain significance (VUS) [28]. One solution to this problem is to test individual variants using assays for protein function to determine whether each variant is damaging to protein function. Functional analysis can provide supporting evidence for clinical classification of rare variants that otherwise could not be supported by population and family data [11,28].

Several different functional assays have been used to test individual and small sets of CHK2 missense variants identified in patients, including tests of protein stability, kinase activity toward various substrates, and activation by DNA damaging agents [22,23,29–32]. Another method to assay the function of CHK2 variants is to test their ability to complement a yeast *RAD53* loss-of-function mutant (*rad53*). Human *CHEK2* was cloned and first tested for its ability to complement yeast *rad53* in a complementation growth assay in 1998 [1]. This study

also demonstrated that the CHK2 kinase defective mutant, D347A, failed to complement *rad53*. Yeast *rad53* complementation has been used as a method for protein function assessment on several other *CHEK2* pathogenic and likely-pathogenic variants. For example, the *CHEK2* founder mutation, c.1100delC was tested for complementation of yeast *rad53* mutants and no complementation was observed [24]. The same study also showed that a *CHEK2* missense variant leading to S428F within the kinase domain does not complement, while a missense variant leading to P85L in the SCD has wildtype-like activity. Additionally, *rad53* complementation assays using the DNA damaging agent, methyl methanesulfonate (MMS) have been performed to evaluate the functional impact of over 100 germline missense variants detected in familial breast cancer patients or selected from the ClinVar database [33–35]. Over 2,700 *CHEK2* missense variants, however, have not been tested in a functional assay, including most of the VUS in the ClinVar database.

To test the ability of all possible *CHEK2* coding SNVs to complement a *rad53* mutant, we conducted a version of deep mutational scanning (DMS) [36,37]. This technique, also known as a multiplex assay of variant effects (MAVE) [38], has become a valuable tool to evaluate the impact of genetic variants on protein function and has been used in several studies to characterize human protein variants based on their function in yeast [39–43]. To carry out a DMS screen, we made a comprehensive library of *CHEK2* coding variants using error-prone PCR, screened the library *en masse* for *rad53* complementation, then used next generation sequencing (NGS) to identify complementors in the selected yeast. This resulted in testing 4,885 of the 4,887 possible *CHEK2* SNVs, including all but one of the 3,913 possible codon changes (stop, synonymous, or non-synonymous). This large-scale functional assay of *CHK2* variants evaluated the impact of almost every possible SNV in the open reading frame of *CHEK2*, providing data to aid in the classification of *CHEK2* VUS.

Results

A Screen for Variants That Affect CHK2 Function

To construct a library of all possible *CHEK2* SNVs, we used error-prone PCR (EP-PCR) on the *CHEK2* open reading frame (ORF) and subcloned the mutagenized ORF into a yeast expression vector (Methods). Subsequent next generation sequencing of the library after introduction into yeast revealed an average SNV frequency of 0.86 (SD = 0.46) per 1,000 bp, or 1.4 SNV per *CHEK2* ORF. The distribution of different variants is very similar to that reported for other EP-PCR studies [44]; all transitions and transversions are represented, though T>G and G>C are the rarest and G>A are the most frequent (S1 Fig). The frequency of variant detection outside the mutagenized region, but within the sequencing amplicon was substantially lower (0.097 per 1000 nt, SD = 0.108) and followed no significant nucleotide preference. The sequenced library included all but two of the 4,887 possible SNVs in the 543 bp *CHEK2* ORF.

To screen for *CHEK2* variants that retain function we selected yeast transformants that were able to grow in the absence of *RAD53* and determined the frequencies of the remaining SNVs by NGS (Methods). For each individual SNV, we compared the average frequency before selection (*i.e.*, in the library) to the average frequency after selection and calculated a *rad53* complementation score (RCS): \log_2 of the ratio of frequencies after/before selection. Lower ratios or more negative RCSs were interpreted to indicate that the SNVs were depleted after selection as expected for variants that are detrimental to *CHK2* function. Globally, 4,885 of the 4,887 possible SNVs in the *CHEK2* ORF could be compared in this way; one SNV was missing from the library, and another was not detected in any of the four independent replicate screens (S1 Table). The SNVs that were evaluated represented all but one of the 3,913 possible unique coding outcomes, including 195 stop codons, 530 synonymous changes, and 3,187 non-synonymous changes (S2 Table). The average RCS for all variants was -0.94 (range -7.30–5.20).

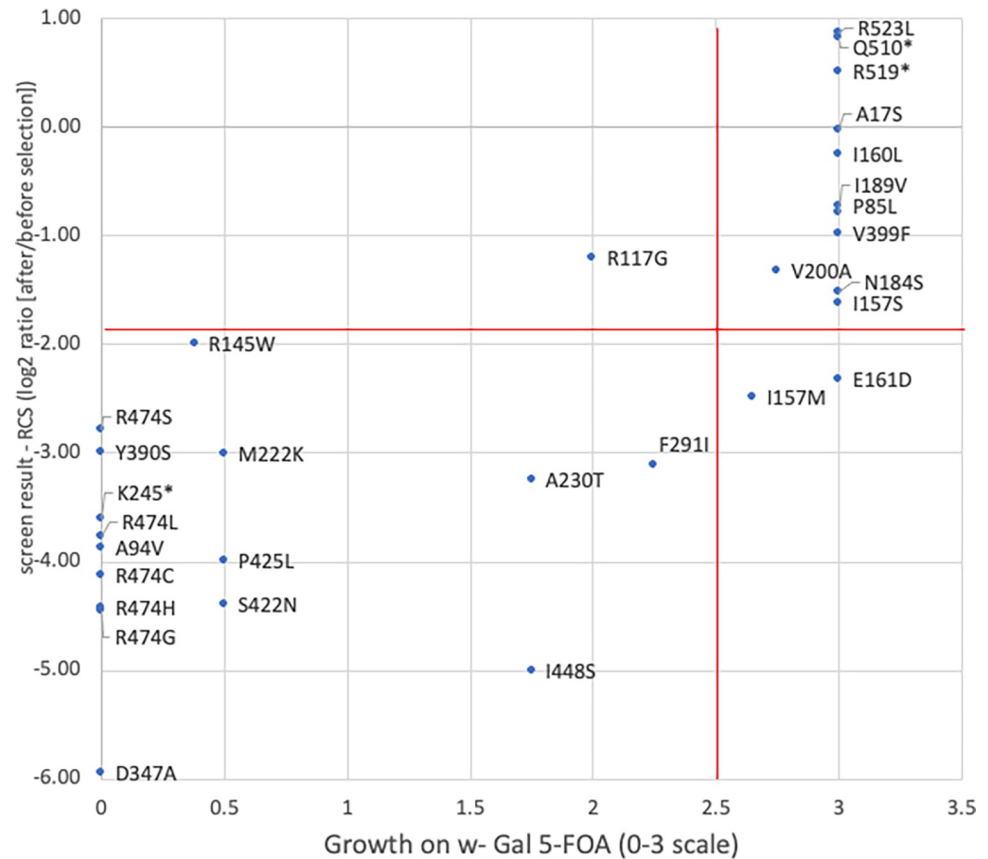


Fig 1. Screen results for individually tested *CHEK2* variants. Yeast *rad53* cells containing the *pURA3::RAD53* plasmid and expressing the indicated individual *CHEK2* variants were plated on media containing 5-FOA to select for loss of *pURA3::RAD53*. Growth is plotted in the X axis on a 0–3 scale where 0 is no growth (equivalent to vector only) and 3 is maximum growth (equivalent to wild-type *CHEK2* growth). Growth of 2.5 or less (vertical red line) was considered defective. Screen depletion values of each variant are plotted on the Y axis. Subsequent analyses indicate that most variants that are known to be functionally defective had depletion scores below -1.9 (horizontal red line).

<https://doi.org/10.1371/journal.pgen.1011375.g001>

To begin to test the assertion that loss of *CHEK2* function correlates with screen depletion, we constructed 30 variants and tested them individually for *rad53* complementation. We found that the 9 variants that did not fully complement *rad53* in individual assays were more depleted in the screen (ave. RCS -3.64, SD = 1.12) than were the 11 variants that fully complemented *rad53* in individual assays (ave. RCS -0.76, SD = 1.10) (Fig 1). These results suggest that depletion in the screen can be used as a predictor of variant function.

Correlation of *rad53* Complementation Screen Results with *CHK2* Function

To see how well the screen results correlate with previous annotations of pathogenicity, we looked at SNVs in ClinVar [35]. ClinVar currently has information for 1,418 *CHEK2* SNVs, including 98 that are annotated as being pathogenic or likely pathogenic, 1,268 that are variants of uncertain significance (VUS), and 58 noted as having conflicting information regarding pathogenicity; ClinVar has no information on 2,237 of the non-synonymous SNVs that we evaluated in our screen. Consistent with the predictive power of our screen, the pathogenic variants were significantly more depleted (mean RCS -3.15, stdev = 1.49) than the VUS (mean

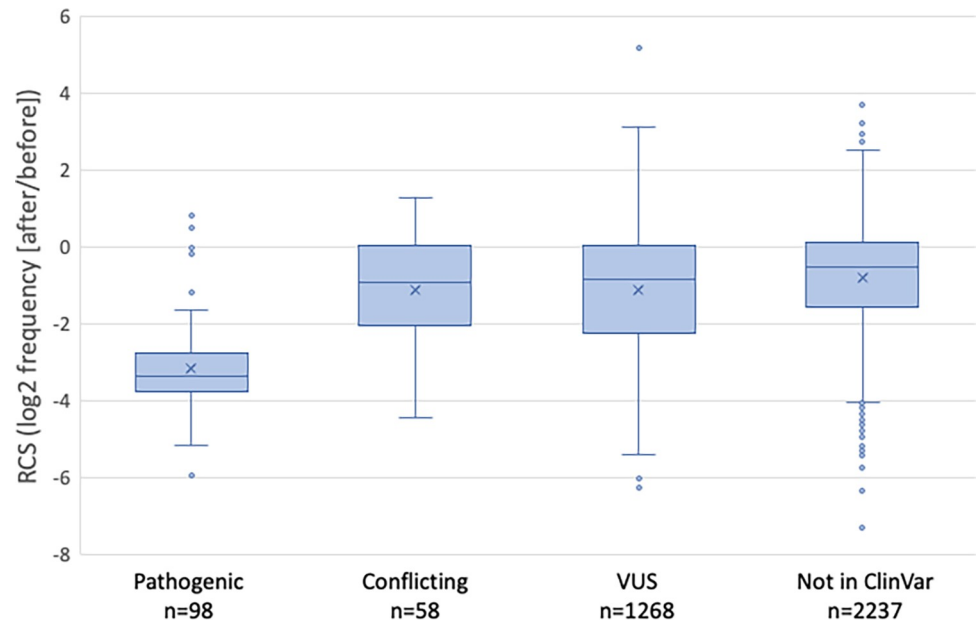


Fig 2. *rad53* complementation screen scores (RCS) correlate with ClinVar annotations and expected variant effects. Boxplots show the RCS for SNVs annotated in ClinVar as pathogenic or likely pathogenic (Pathogenic), with conflicting information about pathogenicity (Conflicting), and variants of uncertain significance (VUS). Also shown are the screen results for 2237 non-synonymous variants that are not listed in ClinVar. The differences between the means (marked with an X) of the pathogenic group and any other group are significant (P values $< 10^{-14}$).

<https://doi.org/10.1371/journal.pgen.1011375.g002>

RCS -1.10, stdev = 1.54) or the variants with conflicting annotations (Fig 2). The VUSs as well as the SNVs that are not in ClinVar had a wider distribution of depletion values in the screen, consistent with the likelihood that only a fraction of these may be functionally defective.

Most (92/98) of the *CHEK2* SNVs that are considered pathogenic or likely pathogenic in ClinVar create premature nonsense (stop) codons in the ORF. To further validate our screen, we compared the results for SNVs that generated all 195 stop codons to SNVs that result in synonymous or non-synonymous codon changes (Fig 3). As expected, stop codons variants were significantly depleted in the *rad53* complementation screen, while the synonymous variants were not. Curiously, this analysis also showed that several stop codon variants were outliers in the screen, with relatively modest levels of depletion. To take a closer look at these, we plotted the depletion scores for all stop codons along the primary axis of the 543 residue CHK2 protein (Fig 4). This revealed that depleted stop codons were evenly distributed through most of the ORF, yet most of the outliers resided beyond codon 500. The average depletion score for the stops in the first 500 codons (-3.26, stdev = 0.91) was significantly different from that of the last 43 codons (0.42, stdev = 0.84; $p = 3.2 \times 10^{-23}$), which mimicked the synonymous SNVs (Fig 3). Using individual clones, we confirmed that variants (p.Gln510stop) and (p.Arg519stop) were indistinguishable from wild-type in *rad53* complementation (Fig 1). Combined, these results strongly suggest that premature termination beyond residue 500 does not adversely affect the ability of *CHEK2* to complement *rad53* (discussed further below).

Classification of Non-Synonymous SNVs as functionally Damaging or Tolerated

Having established that the results from the *rad53* complementation screen correlate well with known function or pathogenicity, we used the screen data to functionally classify the 3,187

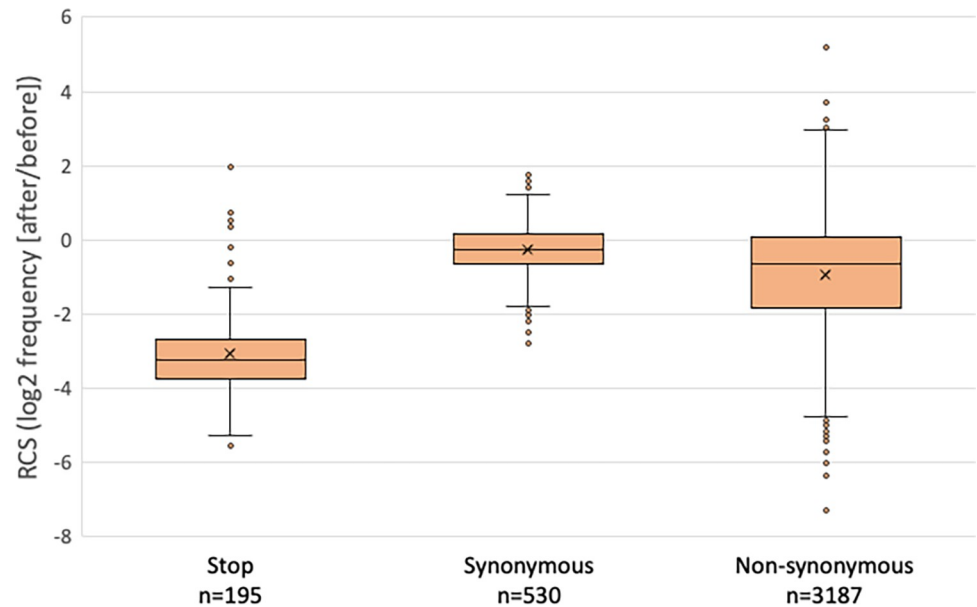


Fig 3. *rad53* complementation screen scores (RCS) correlate with expected variant effects. RCS for variants that result in the 195 possible stop codons, 530 synonymous, and 3188 non-synonymous changes. For this analysis, the SNV frequencies were combined for all SNVs that led to the same protein outcome (stop, synonymous, and each non-synonymous amino acid change). The differences between the means (marked with an X) of the stop codon group and any other group are significant (P values $< 10^{-80}$).

<https://doi.org/10.1371/journal.pgen.1011375.g003>

unique amino acids changes that result from non-synonymous SNVs. Here we refer to variants that appear to reduce *CHK2* function in the yeast assay as ‘damaging’ and those that are functionally indistinguishable from wild type as ‘tolerated’. We tested various machine learning algorithms optimized for accuracy using as training data the 181 stop variants in the first 500 codons, presumed to be damaging, and the 587 synonymous variants also in the first 500 codons, presumed to be tolerated (see [Methods](#)). We found that the only data feature that was useful in predicting function was the RCS depletion score (log₂ frequency [after/before selection]). Moreover, a logistic regression model with an RCS cutoff of near -1.9 achieved the optimal accuracy, correctly classifying 174/181 stop codons as functionally damaging and 480/487 synonymous variants as tolerated, for an overall prediction accuracy of 97.9% (654/668) ([S2 Fig](#)). For the synonymous variants beyond residue 500, which were not used in the training data, the classifier correctly labeled 43/43 as tolerated. In all, we functionally classified 3,579 non-synonymous SNVs resulting in 3,187 unique amino acid changes, including 770 amino acid changes labeled as damaging and 2,417 amino acid changes labeled as tolerated. These results are summarized in an expandable variant effect map ([S3 Fig](#)) and are deposited into MAVdb [[45](#)] (accession urn:mavedb:00001203).

Several functional studies of *CHEK2* variants have included an “intermediate” category purported to represent variant function that is intermediate between wild-type (tolerated) and complete loss of function (damaging). To see if our *rad53* complementation score could be used to predict intermediate function, we looked at 443 variants that have been tested in another functional assay that includes an intermediate category. There was no correlation between the RCS and the intermediate category of the other functional assays ([S4 Fig](#)), demonstrating that the variants classified as intermediate by other assays could not serve as gold standard training data to classify intermediate function based on the RCS. It is worth noting that the intermediate category appears to be the least reliable based the lack of agreement among

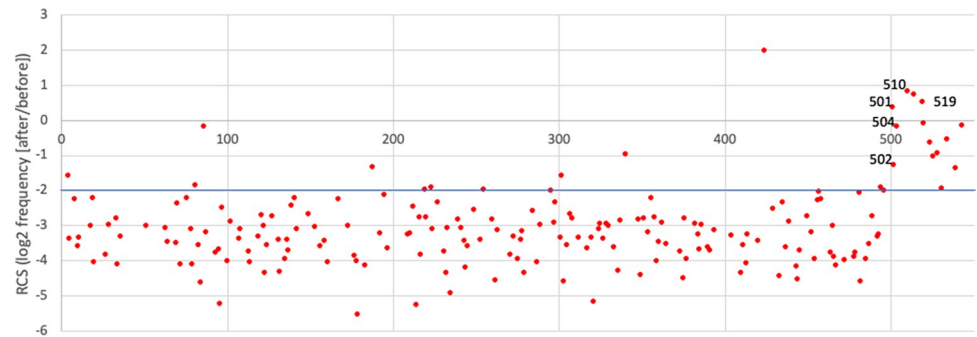


Fig 4. Stop variants. Screen scores for the 195 possible stop codons plotted along the *CHK2* protein primary axis from residue 1 (left) to residue 543 (right). All of the stop codons beyond residue 500 have screen scores > -2.0 (line highlighted for emphasis).

<https://doi.org/10.1371/journal.pgen.1011375.g004>

different functional studies (discussed in detail below). For example, of the 108 variants scored as intermediate in a least one of the other 14 functional assays and with at least one other assay to compare with (not including our assay), 94% (101/108) of the intermediate calls disagree with another assay.

Functional consequences of missense variants

The positions of putative damaging and tolerated variants predicted from the screen correlate well with previous knowledge about the structure and function of *CHK2*. We looked at the frequency of damaging and tolerated variants at each position across the length of the protein, or the average RCS score of all variants at each position, and several patterns emerged (Figs 5, and S5). The number of different amino acid changes tested at each residue varies from 4 to 7, distributed uniformly across the protein. While damaging variants are found throughout the protein, they are rarest at the N- and C-terminal ends. Residues 1–69 and 506–543 are thought to be disordered regions, both by prediction algorithms [46] and by structural analyses [47]. The fraction of protein variants that are damaging in these two disordered regions is 7.2% (29/405) and 10.6% (23/217), respectively, while the fraction of variants that are damaging across the entire protein is 24.2% (770/3187). The average RCS for all variants in these two disordered regions also shows that most variants are tolerated (Figs 5C and S5). The amino terminal disordered region also includes the SQ/TQ cluster domain (SCD) from residue 19 to 69, which has several serines (S) or threonines (T), and seven instances of S or T followed by a glutamine (Q). This region includes several SQ or TQ residues that are phosphorylated in response to DNA damage, including the main ATM target, T68 [2]. Phosphorylation of T68 promotes dimerization interactions between the SCD of one molecule and the FHA domain of another molecule, which in turn leads to autophosphorylation and activation of kinase activity [2–4]. While the SQ and TQ motifs are highly conserved, T68 specifically is not conserved in the *CHK2* orthologs in *S. cerevisiae* (Rad53), *C. elegans*, *Drosophila*, or *S. pombe*, though these proteins do have other SQ and TQ in the region. In our screen, all 5 amino acid changes at residue T68 were tolerated, consistent with the previous finding that T68A was tolerated in another yeast assay [34]. Moreover, all variants in the other 6 TQ/SQ residues were also individually tolerated in our screen. Combined, these results suggest that phosphorylation of T68 or any other single TQ or SQ residue in the SCD of *CHK2* is not required for *rad53* complementation in yeast. This is consistent with studies showing that the Rad53 SQ/TQ residues are not required to promote growth, while any one of the first four SQ residues in Rad53 is sufficient for DNA-damage-induced activation [48].

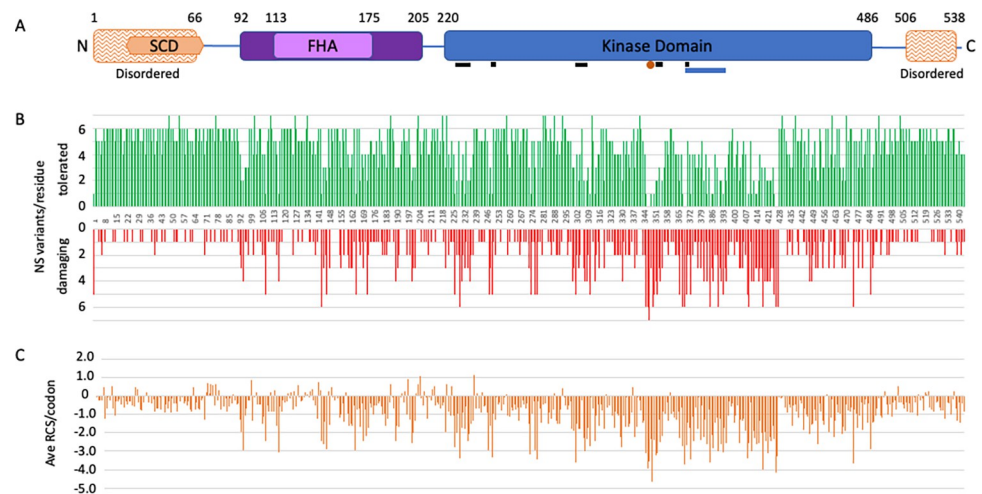


Fig 5. (A) Primary domain structure of the 543-residue CHK2 protein. Residues 1–66 and 506–538 are predicted to be intrinsically disordered regions [46]. The N-terminal disordered region overlaps with the SQ/TQ Cluster Domain (SCD) from residue 19 to 69, which contains seven SQ or TQ sites [2]. The forkhead-associated (FHA) domain extends from residue 113 to 175, according to sequence alignments [62]; a larger region from 92 to 205 has been characterized structurally as the FHA domain [47,63]. The conserved kinase domain extends from residue 220–486 [49,50]. Also shown are conserved ATP binding pocket residues (black bars under kinase domain), the proton acceptor residue D347 (orange dot), and the activation loop from residue 368–394 (blue bar under kinase domain). **(B)** Frequency of damaging or tolerated protein variants at each residue. The numbers of non-synonymous (NS) variants at each residue are plotted along the CHK2 primary axis as in (A) for tolerated variants (upper) and damaging variants (lower). This reveals individual residues and regions where variants appear to be infrequently tolerated. **(C)** The average RCS score for all variants at each codon is plotted along the CHK2 primary axis as in (A).

<https://doi.org/10.1371/journal.pgen.1011375.g005>

The first region with a high frequency of damaging variants was the FHA domain (aa92–205), with 21.5% (146/672) of the variants damaging; 51% (34/66) of the variants in FHA residues that are 100% conserved are damaging. Not surprisingly, the kinase domain (residues 220–486) has the highest fraction of protein variants that are damaging, at 34.6% (559/1562) (Figs 5 and 6, and S5). Furthermore, the frequency of damaging variants within the kinase domain is higher in amino acids and clusters that have conserved functions (Fig 5). For example, in the conserved ATP-binding pocket (Figs 5 and 6) [49], 52% (44/84) of the variants are damaging. Among the variants in the 46 kinase domain residues that are invariant from yeast to human, 69.1% (193/265) are damaging. No variants were tolerated in the proton acceptor residue (D347), while only 2 variants were tolerated in the two other conserved catalytic residues, K249 and D368 [50]; the two tolerated variants, K249R and D368H, have marginally low RCS of -1.55 and -1.78, respectively. In the rest of the activation loop or T loop (residues 368–394), most of the variants are damaging, and no variants were tolerated in residues F369, G370, C385, Y390. Four out of five variants were damaging in each of the autophosphorylation sites, T383 and T387. In the two residues that contact pT383 and pT387 in the dimer, R346 and K373, only one variant was found to be tolerated, the conservative variant K373R. A second region within the kinase domain, from V408 to F427, had a preponderance (70%) of damaging variants and low RCS (ave. -2.47), visible in Fig 5.

The damaging classification correlated well with previous annotations

Previous functional assays for 162 *CHEK2* missense variants generated by SNVs were summarized in [51] (S3 Table). These assays include complementation of yeast *rad53* mutants either in the presence or absence of DNA damaging agents [24,32–34], phosphorylation of various CHK2 substrates *in vitro* or in mammalian cells [23,26,29–31,52,53], DNA damage responses

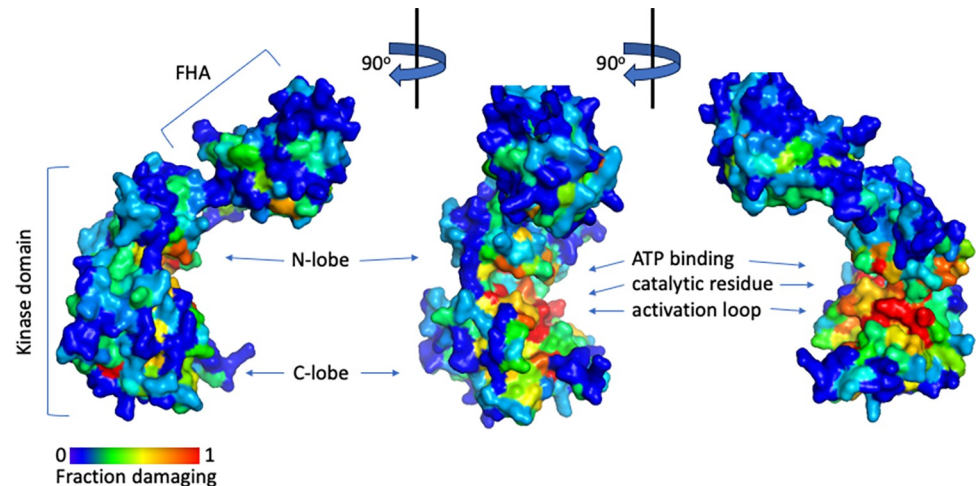


Fig 6. Space-filled peptide backbone structure of CHK2 from three angles with backbone residues colored according to the fraction of variants at that residue that are damaging. The Protein Database structure (3I6U) is from residues 89 to 501, including the FHA domain and the kinase domain with its conserved N-lobe and C-lobe forming an active site cleft between them; the structure lacks the unstructured N- and C-termini [47]. The active site region containing the ATP binding pocket, catalytic residue, and activation loop (partially shown) contains many residues unable to tolerate variants (yellow to red).

<https://doi.org/10.1371/journal.pgen.1011375.g006>

in cells [53,54], and protein stability [30,31]. The results of these various assays have been used to functionally classify each variant as either damaging to the protein's function, tolerated (sometimes referred to as "functional"), or with some intermediate level of function that is less than that of wild-type. The various assays do not always agree on the results for any given variant. The disagreements may be due in part to the fact that the different assays are probing different aspects of CHK2 function; in some cases, however, disagreements between assays may be due to varying levels of miss-classification in each approach. Nevertheless, our screen classification agrees well with the other functional studies. For example, of the 53 previously tested variants that we classified as damaging, only 5 were classified as tolerated by another study, and in all cases it was just one other study (see below). None of our damaging calls were contradicted by more than one study.

Roeb *et al.* [33] and Delimitsou *et al.* [34], assayed *rad53* complementation in the presence of DNA damaging agents to test 26 and 119 *CHEK2* missense SNVs, respectively, and functionally classified each as either damaging, tolerated, or intermediate. While these two assays were very similar to each other, the results agreed on only 7 of the 17 variants that were tested in common. In 6 of the 10 differences, one or the other study classified the variant as intermediate, a classification that we did not attempt to make and that may be more prone to error. Our data agree with Delimitsou *et al.* [34], for 89/119 (75%) of the variants; half of the disagreements were classified as intermediate by Delimitsou *et al.* [34]. In contrast, our data agree with Roeb *et al.* [33], for only 9/26 (35%) of the variants, and again nearly half of the disagreements were classified as intermediate by Roeb *et al.* [33]. There were only three variants where our classification (tolerated) was different from that of both other studies (damaging): R117G, G167R, and Y424H. Interestingly, when we tested R117G individually, it supported less than wild-type growth even though that difference was not detected in our large-scale screen (Fig 1).

Ten additional studies summarized by Boonen *et al.* [51] evaluated a total of 75 SNV variants in functional assays. Our functional classification agreed with 75% of the 132 individual functional assays; again, over half of the differences included assays that found an intermediate

phenotype. For the 71 variants that we found to be damaging, all but one other assay that was performed also found to be damaging, and the one exception found the variant to be intermediate. Boonen *et al.* [51] suggested a counting system to settle differences between assays. At that time, 48 variants had been tested in more than one assay. Among these 48, our damaging calls agree with the consensus for all but one variant where the consensus was intermediate. Our tolerated calls agree with most of the consensus calls and half of the disagreements are with a consensus call of intermediate. I157T is a well-studied example of a variant that we classified as tolerated but where there is disagreement among assays: eight other assays also called I157T tolerated, while one [26] called it intermediate and one [33] called it damaging. Only seven of our calls, all tolerated, differed from the consensus of multiple assays. For two of those, adding our tolerated call to the calculation of a consensus would change the consensus to tolerated. Thus, other assays provide strong evidence that only 4 of our calls are wrong: our tolerated designation of R117G, D162G, G167R, and S428F is countered by multiple other assays that found them to be damaging.

Another recent large study by Stolarova *et al.* [22], functionally characterized 427 *CHEK2* VUS selected by the international ENIGMA (Evidence-based Network for the Interpretation of Germline Mutant Alleles) consortium for variants with an association to an increased risk of breast cancer. That study used two parallel functional assays that quantified the phosphorylation of the KRAB-associated protein (KAP1) at S473 and the autophosphorylation of CHK2 at S516 to characterize 427 missense SNVs as functionally wild-type, intermediate, or impaired. The two assays were concordant for 79.1% of the variants. When compared to our results, the phosphorylation of KAP1 and autophosphorylation of CHK2 experiments agreed in 322/427 (75.4%) and 318/427 (74.5%) of the variant classifications, respectively. This analysis does not include those classified as functionally intermediate, for which there is substantial disagreement even within the Stolarova *et al.* [22] assays. For example, while the KAP1 and *CHEK2* phosphorylation assays agreed on 6 intermediate calls, they disagreed on the remaining 71 variants where one or the other assay called intermediate function: these account for 79% (71/90) of the disagreements between those two assays. Compared with the concordant KAP1 and CHK2 phosphorylation data, our functional classifications were consistent in 337/427 (78.9%) of the variants.

Functional assays and variant effect predictors

Several algorithms have been developed to predict the function or pathogenicity of protein variants in the absence of functional data. We compared our classifications to three variant effect predictors (VEPs) that either combine various algorithms (REVEL) [55] or various structural, functional, or conservation data (HELIX [23] and AlphaMissense [56]) to predict altered protein function and therefore pathogenicity. Each of these VEPs score variants on a scale of 0 to 1 reflecting the probability of the variant being pathogenic or damaging to protein function. The REVEL, HELIX and AlphaMissense scores correlated with the *rad53* complementation scores from our screen (Fig S6 and S3 Table). The average REVEL, HELIX, and AlphaMissense scores for variants we found to be damaging were 0.566, 0.659, and 0.73, respectively, while the average scores for the tolerated variants were 0.304, 0.309, and 0.36, respectively. REVEL correctly predicted 461/770 (59%) of the functionally damaging variants while HELIX correctly predicted 532/770 (69%). These results underscore the importance of experimental verification of function predictions and the variability between these prediction algorithms.

While it remains to be determined whether computational predictions like AlphaMissense are more accurate at assigning function to variants than are experimental functional assays [57], the predicted classifications can still be used as a benchmark to compare the various

functional assays with each other. As we observed when comparing between functional assays, the ability of AlphaMissense to accurately predict the ‘intermediate’ phenotype was very poor (often limited to 11–15% accuracy). Thus, we only compared the ability of AlphaMissense to predict damaging or tolerated variants in the functional assays and did not consider variants that were classified as intermediate by AlphaMissense or experimentally. AlphaMissense was able to accurately predict 71.1% of the variants that we found to be damaging or tolerated. Similarly, AlphaMissense accurately predicted from 71–81% of damaging or tolerated variants in other, smaller datasets (S7 Fig and S3 Table). Interestingly, AlphaMissense appears to be particularly accurate at predicting damaging variants. For example, AlphaMissense accurately predicted all 82 variants that were classified as damaging by more than one functional assay. In contrast, AlphaMissense accurately predicted only 73.5% of the 174 variants that were classified as tolerated in more than one functional assay. These results also support the use of multiple functional assays to accurately classify the damaging variants.

Discussion

The fact that several known loss-of-function *CHK2* variants are associated with increased cancer risk leads to the expectation that any *CHK2* loss-of-function variant would also contribute to cancer risk. This forms the rationale for determining the functional consequences of *CHK2* variants, either experimentally with functional assays or computationally with VEPs. Applying functional data, therefore, can contribute as evidence to the clinical classification system put forth by the ACMG, even for rare variants that have no segregation data [28]. The high throughput approach presented here provided a functional assay for most of the *CHK2* variants that can be generated by SNVs. The screen evaluated the functional impact of 4,885 of the 4,887 possible SNVs in the *CHEK2* ORF, which included all but one of the 3,913 possible unique protein-coding changes. This included functional data for the 1,268 VUS in ClinVar along with an additional 2,237 non-synonymous variants. The data also provide additional supporting evidence for the 58 variants that have a ClinVar clinical classification as “conflicting interpretations of pathogenicity”, and for 443 missense variants that have been tested in at least one other functional assay.

A problem that has emerged in using computational or experimental data to determine variant function is that the various assays or prediction methods do not always agree on the classification. There are several possible explanations for these disagreements. For some variants, the explanation lies in how different studies report variants and whether the classification systems are based on two (tolerated or damaging) or three (including intermediate) functional classes. For example, of the five variants annotated in ClinVar as “benign/likely benign”, our screen classified one (I448S) as functionally damaging. In contrast, four other functional assays all found this variant to be intermediate (meaning somewhat damaging), while the VEP algorithms REVEL, HELIX, and AlphaMissense gave this SNV a score of 0.142 (tolerated), 0.348 (tolerated), and 0.546 (ambiguous), respectively. Interestingly, while our screen was incapable of discriminating damaging from intermediate function, a test of the I448S individual clone did show intermediate levels of *rad53* growth complementation (Fig 1). Another example is I157T, the variant most studied across *CHK2* functional screens (12 total assays; S3 Table). Our screen as well as six others found this variant to be tolerated; however, three assays functionally classified it as intermediate, and one classified it as damaging. In agreement with the majority of the functional screens regarding I157T, the three VEP scores (0.316, 0.167, and 0.229) all result in a tolerated classification. I448S and I157T may be examples of variants that have a truly intermediate effect on function, which different assays interpret differently based on their different sensitivities and classification systems.

Disagreements among individual functional screens may be due to the differences in the model organism or in the assays themselves. Different functional assays may assess distinct aspects of CHK2 function. For example, our assay measured the ability of CHK2 variants to promote growth in the absence of Rad53 and DNA damaging agents, while other assays in yeast measured the ability of CHK2 variants to overcome DNA damage in a strain that does not require *Rad53* or CHK2 for growth. Thus, our screen may miss damaging variants in positions in CHK2 that are not required for *rad53* growth complementation, but instead may be required for activity in response to DNA damaging agents. It is also possible that differences may arise for variants that affect CHK2 activity toward one subset of its targets and not another. Furthermore, while *Saccharomyces cerevisiae* has been used to successfully assess human genetic variants it does have its limitations. Some ORF variants may impact transcript splicing, protein binding sites, thermostability, or subcellular localization differently in yeast cells compared to mammalian cells and *in situ*. For example, our results suggest that beyond residue 500, premature termination does not negatively impact *CHEK2*'s ability to complement *rad53*. However, residues 515–523 contain a nuclear localization signal (NLS) that is required for nuclear localization in human cells [58], and this was confirmed by the ENIGMA study showing that truncations at position 519 (p.R519X) or 523 (p.R523fsX), or NLS missense variants p.Arg521Gln and p.Arg521Trp, did not localize to the nucleus in human cells. p.Arg521Trp was tolerated in our screen and in another yeast assay [34]. Combined, these results suggest that the C-terminal NLS is required for nuclear localization in human cells, but not in yeast cells. Yeast cells may be capable of using one of the other two consensus NLS in *CHEK2*, even though they are not required in human cells [58]. Given the lack of requirement for the C-terminal 43 amino acids for function in our yeast assay, it is puzzling that 25 of the 249 missense variants in that region were found to be damaging, well beyond the expected false positive rate in our assay. An explanation for this apparent discrepancy may require additional functional assays for these variants in both human cells and yeast.

A common solution to the problem of discordance between functional assays is to seek consensus, as suggested by Boonen *et al.* for CHK2 functional assays [51]. In this approach, functional classifications confirmed by more than one assay or VEP are taken to be higher quality. Such an approach is supported by the finding that concordant data from multiple assays has significantly more overlap with independent VEP predictions than data from a single assay. The value of a second, confirming assay is further supported by an analysis of ENIGMA data and two functional assays (KAP1 phosphorylation and CHK2 phosphorylation) carried out by Stolarova *et al.* [22], which shows that CHK2 variants confirmed to be damaging in more assays are associated with increasing risk for breast cancer. For example, the 140 variants that either of their functional assays classified as damaging, even if the two assays were discordant, showed a combined increased risk for breast cancer (OR 2.29; 95% CI 1.97–2.66), while the 91-variant subset of these that were confirmed to be damaging in our screen showed a marginally increased risk (OR 2.95; 95% CI 2.36–3.69), and the risk associated with 74 variants confirmed to be damaging in all three assays was even higher (OR 3.05; 2.39–3.90) (S3 Table and [22]). These results support the continued application of functional assays to establish a high confidence clinical classification of variant effects. Our study provides an additional piece of functional data that may be combined with other evidence for the classification of *CHEK2* variants.

Material and methods

Plasmids

pRS314-GU-HR is a *TRP1*, CEN/ARS yeast plasmid with the *GAL1* promoter driving expression of inserts. pRS314-GU-HR was made from pRS314-GU (a gift from Phil Heiter), a derivative of pRS314 [59] in which the *GAL1* promoter was inserted at one end of the multiple

cloning site and the *URA3* promoter at the other end in the opposite orientation to serve as a terminator. The oligonucleotides 314GU-A (CGATTTGACTGTATCGCCGGAATTCGGGC CCGGATCCCTGCAGCCAAGCTAATTCCGGT) and 314GU-B (GATCACCGGAATTAGC TTGGCTGCAGGGATCCGGGCCCGAATTCGGGCGATACAGTCAAAT), containing 5RT and 3RT flanking and EcoRI and BamHI cloning sites were annealed and ligated into the ClaI-BamHI interval of pRS314-GU to generate pRS314-GU-HR. The *CHEK2* open reading frame corresponding to transcript NM_007194, encoding *CHEK2* isoform a, within a Gateway entry clone was changed by site-directed mutagenesis to eliminate the BamHI with a silent single nucleotide change (ORF T1410->C), to make pENTR-CHEK2-BamHI. pENTR-CHEK2-BamHI was then used as a PCR template with primers that added an EcoRI site immediately adjacent to the ATG and a stop codon and BamHI site immediately after the last codon. The resulting EcoRI-BamHI fragment was cloned into EcoRI/BamHI cut pRS314-GU-HR to make pRS314-GU-HR-CHEK2-BamHI. Site-directed mutagenesis was carried out on pENTRY-CHEK2-BamHI to create the 30 variants shown in Fig 1. We subsequently learned that the absence of a stop codon in the original pENTRY-CHEK2 diminished the ability of *CHEK2* to complement yeast *RAD53* mutants when cloned into a yeast vector. Thus, we converted all of the pENTRY-CHEK2-BamHI variants to versions with the natural stop codon of *CHEK2* as follows: We PCR amplified wild type *CHEK2* and each variant from pENTRY-CHEK2-BamHI with primers that added an attB1 site to the 5' end and a stop codon and attB2 to the 3' end, then used the PCR product in a BP reaction into pDONR221. The resulting pENTRY-CHEK2-stop-BamHI variants were then cloned into the yeast vector pAG414GAL (a gift from Susan Lindquist; Addgene plasmid # 14143) using the LR reaction. All *CHEK2* clones were verified by sequencing.

RAD53 Complementation assays

For complementation of the *rad53* deletion, we used yeast strain YGB272 (*MATa rad53::HIS3 pRAD53::URA3*), which was derived from a *MATa/MATα RAD53/rad53::HIS3 SML1/sml1-1* heterozygote (W303 background) transformed with pJA92 encoding *RAD53* and *URA3* [60] (diploid and plasmid kindly provided by Dr. Stephen Elledge). To test individual *CHEK2* variants for *rad53* complementation, YGB272 was transformed with pRS314-GU-HR or pAG414-GAL containing wild-type *CHEK2* or variants and transformants were selected on synthetic complete media lacking tryptophan (-trp) and containing 2% glucose (Glu). YGB272 requires *RAD53* for growth and will not grow on 5-fluoroorotic acid (5-FOA), which is toxic to strains containing *URA3* [61]. Previous studies have shown that expression of wild-type human *CHEK2* in such a strain will allow it to lose the *pRAD53::URA3* plasmid and to grow on 5-FOA [1]. Individual transformants were picked and grown in 2 ml of -trp liquid media supplemented with 2% galactose and 1% raffinose (Gal/Raf) at 30 degrees with shaking to saturation. Cells were then serially diluted in sterile water and 5 μ L of each dilution was plated to -trp Gal/Raf plates supplemented with 1mg/ml 5-FOA (Gold Biotechnology). After 4 days at 30 degrees, the spots were scored for growth as follows: 0 for no growth, 1 for very weak growth, 2 for moderate growth, and 3 for complete growth. Strains with the vector only always scored 0 while strains expressing wild-type *CHK2* always scored 3.

CHEK2 Variant Library

The library was constructed by first PCR amplifying the *CHEK2* ORF from pENTR-CHEK2-BamHI with primers (HsCHEK2_5RT and HsCHEK2stop_3RT) that added an EcoRI site immediately adjacent to the ATG and a stop codon and BamHI site immediately after the last codon, respectively, using high-fidelity PCR conditions (Platinum Super-Fi polymerase,

Invitrogen). The fragment was gel purified and used as template in error-prone PCR (EP-PCR) using the GeneMorph II kit and Mutazyme II polymerase (Agilent Technologies) and the primers (5RT 3RT). Ideal EP-PCR conditions were established in pilot assays with *CHEK2* and other ORFs in both complementation assays and yeast two-hybrid assays and by sequencing to maximize the number of clones with a mutation while minimizing the number of mutations per clone. The final EP-PCR conditions used to make the *CHEK2* library were: five separate 50 μ L reactions with 0.8 mM dNTP mix, 0.25 μ M each primer, 100 ng template DNA, and 0.017 units/ μ L of Mutazyme II and these were treated for 27 cycles of 95 degrees C for 30 seconds, 57 degrees for 30 seconds, and 72 degrees for 2 minutes. The EP-PCR products were purified through a GeneJET PCR purification column (ThermoFisher), cut with hi-fidelity EcoRI and BamHI (New England Biolabs), and gel purified. The fragments were then ligated into pRS314-GU-HR linearized with EcoRI and BamHI. Bacterial strain OmniMax (ThermoFisher) was electroporated with the ligations and plated onto 22cmx22cm LB amp plates. Plasmid DNA was prepared from 22,400 transformants.

Library screening

The pRS314-GU-HR-CHEK2 EP-PCR library DNA was used to transform strain YGB272 and 62,000 trp⁺ colonies were obtained on -trp Glu plates. The colonies were pooled and frozen in aliquots containing 2×10^7 colony forming units (cfu)/100 μ L (after the freeze/thaw cycle). Two aliquots (LibA and LibB) were used to prepare DNA for sequencing. Complementation screens were carried out in quadruplicate (screen S1 to S4), each by inoculating 100 ml -trp Gal/Raf with 100 μ L of a library aliquot and growing at 30 degrees with shaking for 24 hours to induce expression of the *CHEK2* ORF; based on OD600 this resulted in a 21-fold amplification of cells. 4×10^7 cfu were plated on -trp Gal/Raf supplemented with 1mg/ml 5-fluoroorotic acid (5-FOA) (Gold Biotechnology) and were grown at 30 degrees for 96 hours. In the quadruplicate screens, 28,000, 26,600, 26,600, and 30,400, 5-FOA-resistant colonies, respectively, were collected and frozen aliquots were made containing 5×10^6 cfu/100 μ L in -trp Gal/Raf (after the freeze/thaw cycle). Aliquots of the unselected library transformants (LibA and LibB) and the four 5-FOA-resistant yeast (S1 to S4) were thawed and used to prepare crude plasmid DNA using zymolyase and Qiagen miniprep columns. To create amplicons for next generation sequencing, the mutagenized inserts were amplified with vector primer sequences from 46 bp upstream of the *CHEK2* ATG and 144 bp downstream of the *CHEK2* stop codon.

Sequencing libraries were constructed with the Nextera XT DNA Library Prep kit with Nextera XT Indexes (Illumina), pooled, and sequenced on a NextSeq 500 to obtain 76 bp paired-end reads. Sequence data was processed within Geneious using BBDuk to trim adapters, filter out quality scores below 30 and sequences less than 60 bp, then the Geneious algorithm was used to map reads to the *CHEK2* amplicon and to detect variants. The number of mapped reads for each library ranged from 2.5 to 5.6 million. Subsequent analyses focused on the single nucleotide variants (SNVs). All SNVs in the *CHEK2* ORF were represented in the library sequences, except one (CDS position 1129 A->C, p.Ile374Leu). For each SNV and each library we calculated the frequency as the number of reads with that SNV divided by the coverage at that nucleotide position. For every SNV present in the library, we calculated the average of frequencies in the two aliquots before selection (LibA and LibB) and the average of frequencies across the four screens, S1-S4 after selection. We then calculated a *RAD53* Complementation Score (RCS) as \log_2 (average frequency after/average frequency before selection). 669 codon changes (synonymous, non-synonymous, or stop) were represented by more than one (2–4) SNVs. For these we calculated the average frequencies and depletion scores across all SNVs leading to the same change.

Supporting information

S1 Fig. Nucleotide variant frequencies. Variant frequencies are shown as the average of variant counts at each position divided by the sequencing coverage at that position x 1000. All frequencies were normalized to the most frequent wild-type nucleotide in the fragment being sequenced. Fragments include the 1629 bp mutagenized *CHEK2* ORF (ORF) and two regions of the sequenced amplicon that were not mutagenized, combining 46 bp upstream of the *CHEK2* ATG and 144 bp downstream of the *CHEK2* stop codon (no-EPCR). For each, the results are shown for the library after introduction into yeast (Library), and after selection for *rad53* complementation (Selected).

(TIFF)

S2 Fig. Logistic regression model to classify non-synonymous variants as functionally damaging or tolerated. (A) Training data, model parameters, and output applied to the *CHEK2* ORF. (B) Receiver operating characteristic (ROC) curve for the final model as applied to the training data. (C) Confusion matrix showing the numbers of training data (gained stops and synonymous SNVs) correctly classified (true positives, green) and incorrectly classified (false positives, red) by the model. Note that false positives in one class are equivalent to false negatives in the other class.

(TIFF)

S3 Fig. A variant effect map summarizing the effect on function of all *CHK2* protein variants from this study. Residue positions and original amino acids are shown along the bottom while variants are listed at the left. Variants are colored according to their probability of being damaging (1.0 = red, 0.0 = blue), based on a logistic regression model of the screen depletion scores or RCS.

(TIFF)

S4 Fig. Lack of correlation between the *rad53* complementation score and the intermediate functional classification of other functional assays. In both graphs the *rad53* complementation score (RCS) score is on the X axis, and as a reminder a red vertical line shows the cutoff between damaging (< -1.9) and tolerated (> -1.9). On the y axis is the average score of other functional assays, where 1.0 is damaging, 0 is tolerated, and anything in between is intermediate (data in S3 Table). In both graphs, the variants with intermediate function in other assays are spread uniformly across the RCS scale. This is the case for the 443 variants that have been tested in at least one other functional assay (left graph) and for the subset of variants tested in the KAP1 kinase assay (by Stolarova et al. [22]).

(TIFF)

S5 Fig. Frequencies of variants and damaging variants across the *CHK2* protein.

(TIFF)

S6 Fig. Comparison of functional classifications from this study to the prediction algorithms, REVEL [55] (A) and HELIX [23] (B), and Alphasense [56] (C). In each, variants are scored on a scale of 0 to 1 (Y axis), with scores ≥ 0.5 predicting pathogenic (damaging) and scores < 0.5 predicting benign (tolerated). The average REVEL, HELIX, and Alphasense scores (x in the boxplot) for variants we found to be damaging was 0.566, 0.659, and 0.73 respectively, while the average scores for the tolerated variants was 0.304, 0.309, and 0.36 respectively.

(TIFF)

S7 Fig. Accuracy of AlphaMissense at predicting results of functional assays. Overall accuracy: the fraction damaging, tolerated, and ambiguous or intermediate (if the assays included

such a designation) that were accurately predicted by AlphaMissense. The “Accuracy of damaging or tolerated only” ignores all variants designated as ambiguous or intermediate by AlphaMissense or the functional assay. The last row includes only variants tested in more than one assay, where all assays agree that they are tolerated or damaging.
(TIFF)

S1 Table. *rad53* screen data for all *CHEK2* ORF SNVs
(XLSX)

S2 Table. Damaging and tolerated CHK2 protein variants.
(XLSX)

S3 Table. CHK2 variants and results from functional assays and VEPs.
(XLSX)

Acknowledgments

We thank Adnan Alazizi and the Genomics Services Center in the Center for Molecular Medicine and Genetics, Wayne State University School of Medicine for help with sequencing. We also thank Jeffery Tseng for adding our custom values to the CHK2 PDB file.

Author Contributions

Conceptualization: George S. Brush, Michael A. Tainsky, Russell L. Finley, Jr.

Formal analysis: Claire E. McCarthy-Leo, Russell L. Finley, Jr.

Investigation: Claire E. McCarthy-Leo, George S. Brush, Russell L. Finley, Jr.

Methodology: Russell L. Finley, Jr.

Resources: Roger Pique-Regi, Francesca Luca.

Supervision: Michael A. Tainsky, Russell L. Finley, Jr.

Writing – original draft: Claire E. McCarthy-Leo, Russell L. Finley, Jr.

Writing – review & editing: George S. Brush, Francesca Luca, Michael A. Tainsky.

References

1. Matsuoka S, Huang M, Elledge SJ. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. *Science*. 1998; 282(5395):1893–7. Epub 1998/12/04. <https://doi.org/10.1126/science.282.5395.1893> PMID: 9836640.
2. Matsuoka S, Rotman G, Ogawa A, Shiloh Y, Tamai K, Elledge SJ. Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro. *Proc Natl Acad Sci U S A*. 2000; 97(19):10389–94. Epub 2000/09/06. <https://doi.org/10.1073/pnas.190030497> PMID: 10973490; PubMed Central PMCID: PMC27034.
3. Melchionna R, Chen XB, Blasina A, McGowan CH. Threonine 68 is required for radiation-induced phosphorylation and activation of Cds1. *Nat Cell Biol*. 2000; 2(10):762–5. <https://doi.org/10.1038/35036406> PMID: 11025670.
4. Ahn JY, Schwarz JK, Piwnica-Worms H, Canman CE. Threonine 68 phosphorylation by ataxia telangiectasia mutated is required for efficient activation of Chk2 in response to ionizing radiation. *Cancer Res*. 2000; 60(21):5934–6. Epub 2000/11/21. PMID: 11085506.
5. Gabant G, Lorphelin A, Nozerand N, Marchetti C, Bellanger L, Dedieu A, et al. Autophosphorylated residues involved in the regulation of human chk2 in vitro. *J Mol Biol*. 2008; 380(3):489–503. Epub 20080501. <https://doi.org/10.1016/j.jmb.2008.04.053> PMID: 18538787.
6. Schwarz JK, Lovly CM, Piwnica-Worms H. Regulation of the Chk2 protein kinase by oligomerization-mediated cis- and trans-phosphorylation. *Mol Cancer Res*. 2003; 1(8):598–609. PMID: 12805407.

7. Wu X, Chen J. Autophosphorylation of checkpoint kinase 2 at serine 516 is required for radiation-induced apoptosis. *J Biol Chem.* 2003; 278(38):36163–8. Epub 20030709. <https://doi.org/10.1074/jbc.M303795200> PMID: 12855706.
8. Bartek J, Falck J, Lukas J. CHK2 kinase—a busy messenger. *Nature reviews Molecular cell biology.* 2001; 2(12):877–86. <https://doi.org/10.1038/35103059> PMID: 11733767.
9. Hirao A, Kong YY, Matsuoka S, Wakeham A, Ruland J, Yoshida H, et al. DNA damage-induced activation of p53 by the checkpoint kinase Chk2. *Science.* 2000; 287(5459):1824–7. Epub 2000/03/10. <https://doi.org/10.1126/science.287.5459.1824> PMID: 10710310.
10. Zannini L, Delia D, Buscemi G. CHK2 kinase in the DNA damage response and beyond. *J Mol Cell Biol.* 2014; 6(6):442–57. Epub 2014/11/19. <https://doi.org/10.1093/jmcb/mju045> PMID: 25404613; PubMed Central PMCID: PMC4296918.
11. Hanson H, Astiazaran-Symonds E, Amendola LM, Balmana J, Foulkes WD, James P, et al. Management of individuals with germline pathogenic/likely pathogenic variants in CHEK2: A clinical practice resource of the American College of Medical Genetics and Genomics (ACMG). *Genet Med.* 2023; 25(10):100870. Epub 20230725. <https://doi.org/10.1016/j.gim.2023.100870> PMID: 37490054; PubMed Central PMCID: PMC10623578.
12. Consortium CBCC-C. CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. *American journal of human genetics.* 2004; 74(6):1175–82. Epub 20040430. <https://doi.org/10.1086/421251> PMID: 15122511; PubMed Central PMCID: PMC1182081.
13. Weischer M, Bojesen SE, Ellervik C, Tybjaerg-Hansen A, Nordestgaard BG. CHEK2*1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls. *J Clin Oncol.* 2008; 26(4):542–8. Epub 20080102. <https://doi.org/10.1200/JCO.2007.12.5922> PMID: 18172190.
14. Weischer M, Nordestgaard BG, Pharoah P, Bolla MK, Nevanlinna H, Van't Veer LJ, et al. CHEK2*1100delC heterozygosity in women with breast cancer associated with early death, breast cancer-specific death, and increased risk of a second breast cancer. *J Clin Oncol.* 2012; 30(35):4308–16. Epub 20121029. <https://doi.org/10.1200/JCO.2012.42.7336> PMID: 23109706; PubMed Central PMCID: PMC3515767.
15. Wasielewski M, den Bakker MA, van den Ouweland A, Meijer-van Gelder ME, Portengen H, Klijn JG, et al. CHEK2 1100delC and male breast cancer in the Netherlands. *Breast Cancer Res Treat.* 2009; 116(2):397–400. Epub 20080831. <https://doi.org/10.1007/s10549-008-0162-7> PMID: 18759107.
16. Teodorczyk U, Cybulski C, Wokolorczyk D, Jakubowska A, Starzynska T, Lawniczak M, et al. The risk of gastric cancer in carriers of CHEK2 mutations. *Fam Cancer.* 2013; 12(3):473–8. <https://doi.org/10.1007/s10689-012-9599-2> PMID: 23296741.
17. AlDubayan SH, Pyle LC, Gamulin M, Kulis T, Moore ND, Taylor-Weiner A, et al. Association of Inherited Pathogenic Variants in Checkpoint Kinase 2 (CHEK2) With Susceptibility to Testicular Germ Cell Tumors. *JAMA Oncol.* 2019; 5(4):514–22. <https://doi.org/10.1001/jamaoncol.2018.6477> PMID: 30676620; PubMed Central PMCID: PMC6459214.
18. Cybulski C, Wokolorczyk D, Huzarski T, Byrski T, Gronwald J, Gorski B, et al. A large germline deletion in the Chek2 kinase gene is associated with an increased risk of prostate cancer. *J Med Genet.* 2006; 43(11):863–6. <https://doi.org/10.1136/jmg.2006.044974> PMID: 17085682; PubMed Central PMCID: PMC2563179.
19. Siolak M, Cybulski C, Gasior-Periczak D, Kowalik A, Kozak-Klonowska B, Kowalska A, et al. CHEK2 mutations and the risk of papillary thyroid cancer. *Int J Cancer.* 2015; 137(3):548–52. Epub 20150128. <https://doi.org/10.1002/ijc.29426> PMID: 25583358.
20. Wu Y, Yu H, Zheng SL, Na R, Mamawala M, Landis T, et al. A comprehensive evaluation of CHEK2 germline mutations in men with prostate cancer. *Prostate.* 2018; 78(8):607–15. Epub 20180309. <https://doi.org/10.1002/pros.23505> PMID: 29520813.
21. Wagener R, Walter C, Auer F, Alzoubi D, Hauer J, Fischer U, et al. The CHK2 kinase is recurrently mutated and functionally impaired in the germline of pediatric cancer patients. *Int J Cancer.* 2023; 152(7):1388–98. Epub 20221227. <https://doi.org/10.1002/ijc.34390> PMID: 36468172.
22. Stolarova L, Kleiblova P, Zemankova P, Stastna B, Janatova M, Soukupova J, et al. ENIGMA CHEK2-gether Project: A Comprehensive Study Identifies Functionally Impaired CHEK2 Germline Missense Variants Associated with Increased Breast Cancer Risk. *Clin Cancer Res.* 2023:OF1–OF14. Epub 20230713. <https://doi.org/10.1158/1078-0432.CCR-23-0212> PMID: 37449874.
23. Boonen R, Wiegant WW, Celosse N, Vroling B, Heijl S, Kote-Jarai Z, et al. Functional Analysis Identifies Damaging CHEK2 Missense Variants Associated with Increased Cancer Risk. *Cancer Res.* 2022; 82(4):615–31. <https://doi.org/10.1158/0008-5472.CAN-21-1845> PMID: 34903604.

24. Shaag A, Walsh T, Renbaum P, Kirchoff T, Nafa K, Shiovitz S, et al. Functional and genomic approaches reveal an ancient CHEK2 allele associated with breast cancer in the Ashkenazi Jewish population. *Hum Mol Genet.* 2005; 14(4):555–63. Epub 2005/01/15. <https://doi.org/10.1093/hmg/ddi052> PMID: 15649950.
25. Laitman Y, Kaufman B, Lahad EL, Papa MZ, Friedman E. Germline CHEK2 mutations in Jewish Ashkenazi women at high risk for breast cancer. *Isr Med Assoc J.* 2007; 9(11):791–6. PMID: 18085035.
26. Kleiblova P, Stolarova L, Krizova K, Lhota F, Hojny J, Zemankova P, et al. Identification of deleterious germline CHEK2 mutations and their association with breast and ovarian cancer. *Int J Cancer.* 2019; 145(7):1782–97. Epub 20190520. <https://doi.org/10.1002/ijc.32385> PMID: 31050813.
27. Dong X, Wang L, Taniguchi K, Wang X, Cunningham JM, McDonnell SK, et al. Mutations in CHEK2 associated with prostate cancer risk. *American journal of human genetics.* 2003; 72(2):270–80. Epub 20030117. <https://doi.org/10.1086/346094> PMID: 12533788; PubMed Central PMCID: PMC379222.
28. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015; 17(5):405–24. Epub 20150305. <https://doi.org/10.1038/gim.2015.30> PMID: 25741868; PubMed Central PMCID: PMC4544753.
29. Wu X, Webster SR, Chen J. Characterization of tumor-associated Chk2 mutations. *J Biol Chem.* 2001; 276(4):2971–4. Epub 20001026. <https://doi.org/10.1074/jbc.M009727200> PMID: 11053450.
30. Lee SB, Kim SH, Bell DW, Wahrer DC, Schiripo TA, Jorczak MM, et al. Destabilization of CHK2 by a missense mutation associated with Li-Fraumeni Syndrome. *Cancer Res.* 2001; 61(22):8062–7. Epub 2001/11/24. PMID: 11719428.
31. Bell DW, Kim SH, Godwin AK, Schiripo TA, Harris PL, Haserlat SM, et al. Genetic and functional analysis of CHEK2 (CHK2) variants in multiethnic cohorts. *Int J Cancer.* 2007; 121(12):2661–7. <https://doi.org/10.1002/ijc.23026> PMID: 17721994; PubMed Central PMCID: PMC3090684.
32. Tischkowitz MD, Yilmaz A, Chen LQ, Karyadi DM, Novak D, Kirchoff T, et al. Identification and characterization of novel SNPs in CHEK2 in Ashkenazi Jewish men with prostate cancer. *Cancer Lett.* 2008; 270(1):173–80. Epub 20080620. <https://doi.org/10.1016/j.canlet.2008.05.006> PMID: 18571837; PubMed Central PMCID: PMC2969172.
33. Roeb W, Higgins J, King MC. Response to DNA damage of CHEK2 missense mutations in familial breast cancer. *Hum Mol Genet.* 2012; 21(12):2738–44. <https://doi.org/10.1093/hmg/dds101> PMID: 22419737; PubMed Central PMCID: PMC3363333.
34. Delimitsou A, Fostira F, Kalfakakou D, Apostolou P, Konstantopoulou I, Kroupis C, et al. Functional characterization of CHEK2 variants in a *Saccharomyces cerevisiae* system. *Hum Mutat.* 2019; 40(5):631–48. Epub 20190309. <https://doi.org/10.1002/humu.23728> PMID: 30851065.
35. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: improvements to accessing data. *Nucleic Acids Res.* 2020; 48(D1):D835–D44. <https://doi.org/10.1093/nar/gkz972> PMID: 31777943; PubMed Central PMCID: PMC6943040.
36. Fowler DM, Fields S. Deep mutational scanning: a new style of protein science. *Nat Methods.* 2014; 11(8):801–7. Epub 2014/07/31. <https://doi.org/10.1038/nmeth.3027> PMID: 25075907; PubMed Central PMCID: PMC4410700.
37. Fowler DM, Araya CL, Fleishman SJ, Kellogg EH, Stephany JJ, Baker D, et al. High-resolution mapping of protein sequence-function relationships. *Nat Methods.* 2010; 7(9):741–6. Epub 2010/08/17. <https://doi.org/10.1038/nmeth.1492> PMID: 20711194; PubMed Central PMCID: PMC2938879.
38. Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G, et al. Variant Interpretation: Functional Assays to the Rescue. *American journal of human genetics.* 2017; 101(3):315–25. <https://doi.org/10.1016/j.ajhg.2017.07.014> PMID: 28886340; PubMed Central PMCID: PMC5590843.
39. Sun S, Weile J, Verby M, Wu Y, Wang Y, Cote AG, et al. A proactive genotype-to-patient-phenotype map for cystathionine beta-synthase. *Genome Med.* 2020; 12(1):13. Epub 20200130. <https://doi.org/10.1186/s13073-020-0711-1> PMID: 32000841; PubMed Central PMCID: PMC6993387.
40. Weile J, Sun S, Cote AG, Knapp J, Verby M, Mellor JC, et al. A framework for exhaustively mapping functional missense variants. *Molecular systems biology.* 2017; 13(12):957. Epub 2017/12/23. <https://doi.org/10.15252/msb.20177908> PMID: 29269382; PubMed Central PMCID: PMC5740498.
41. Sun S, Yang F, Tan G, Costanzo M, Oughtred R, Hirschman J, et al. An extended set of yeast-based functional assays accurately identifies human disease mutations. *Genome research.* 2016; 26(5):670–80. Epub 20160314. <https://doi.org/10.1101/gr.192526.115> PMID: 26975778; PubMed Central PMCID: PMC4864455.
42. Gersing S, Schulze TK, Cagiada M, Stein A, Roth FP, Lindorff-Larsen K, et al. Characterizing glucokinase variant mechanisms using a multiplexed abundance assay. *Genome biology.* 2024; 25(1):98. Epub 20240416. <https://doi.org/10.1186/s13059-024-03238-2> PMID: 38627865.

43. Newberry RW, Leong JT, Chow ED, Kampmann M, DeGrado WF. Deep mutational scanning reveals the structural basis for alpha-synuclein activity. *Nat Chem Biol.* 2020; 16(6):653–9. Epub 20200309. <https://doi.org/10.1038/s41589-020-0480-6> PMID: 32152544; PubMed Central PMCID: PMC7339969.
44. Rasila TS, Pajunen MI, Savilahti H. Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment. *Anal Biochem.* 2009; 388(1):71–80. Epub 2009/05/21. <https://doi.org/10.1016/j.ab.2009.02.008> PMID: 19454214.
45. Esposito D, Weile J, Shendure J, Starita LM, Papenfuss AT, Roth FP, et al. MaveDB: an open-source platform to distribute and interpret data from multiplexed assays of variant effect. *Genome biology.* 2019; 20(1):223. Epub 20191104. <https://doi.org/10.1186/s13059-019-1845-6> PMID: 31679514; PubMed Central PMCID: PMC6827219.
46. Necci M, Piovesan D, Dosztanyi Z, Tosatto SCE. MobiDB-lite: fast and highly specific consensus prediction of intrinsic disorder in proteins. *Bioinformatics.* 2017; 33(9):1402–4. <https://doi.org/10.1093/bioinformatics/btx015> PMID: 28453683.
47. Cai Z, Chehab NH, Pavletich NP. Structure and activation mechanism of the CHK2 DNA damage checkpoint kinase. *Mol Cell.* 2009; 35(6):818–29. <https://doi.org/10.1016/j.molcel.2009.09.007> PMID: 19782031.
48. Lee SJ, Schwartz MF, Duong JK, Stern DF. Rad53 phosphorylation site clusters are important for Rad53 regulation and signaling. *Mol Cell Biol.* 2003; 23(17):6300–14. <https://doi.org/10.1128/MCB.23.17.6300-6314.2003> PMID: 12917350; PubMed Central PMCID: PMC180918.
49. Oliver AW, Paul A, Boxall KJ, Barrie SE, Aherne GW, Garrett MD, et al. Trans-activation of the DNA-damage signalling protein kinase Chk2 by T-loop exchange. *EMBO J.* 2006; 25(13):3179–90. Epub 2006/06/24. <https://doi.org/10.1038/sj.emboj.7601209> PMID: 16794575; PubMed Central PMCID: PMC1500991.
50. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S. The protein kinase complement of the human genome. *Science.* 2002; 298(5600):1912–34. <https://doi.org/10.1126/science.1075762> PMID: 12471243.
51. Boonen R, Vreeswijk MPG, van Attikum H. CHEK2 variants: linking functional impact to cancer risk. *Trends Cancer.* 2022; 8(9):759–70. Epub 20220525. <https://doi.org/10.1016/j.trecan.2022.04.009> PMID: 35643632.
52. Chrisanthar R, Knappskog S, Lokkevik E, Anker G, Ostenstad B, Lundgren S, et al. CHEK2 mutations affecting kinase activity together with mutations in TP53 indicate a functional pathway associated with resistance to epirubicin in primary breast cancer. *PLoS One.* 2008; 3(8):e3062. Epub 20080826. <https://doi.org/10.1371/journal.pone.0003062> PMID: 18725978; PubMed Central PMCID: PMC2518116.
53. Falck J, Mailand N, Syljuasen RG, Bartek J, Lukas J. The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. *Nature.* 2001; 410(6830):842–7. Epub 2001/04/12. <https://doi.org/10.1038/35071124> PMID: 11298456.
54. Wang N, Ding H, Liu C, Li X, Wei L, Yu J, et al. A novel recurrent CHEK2 Y390C mutation identified in high-risk Chinese breast cancer patients impairs its activity and is associated with increased breast cancer risk. *Oncogene.* 2015; 34(40):5198–205. Epub 20150126. <https://doi.org/10.1038/onc.2014.443> PMID: 25619829.
55. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *American journal of human genetics.* 2016; 99(4):877–85. Epub 20160922. <https://doi.org/10.1016/j.ajhg.2016.08.016> PMID: 27666373; PubMed Central PMCID: PMC5065685.
56. Cheng J, Novati G, Pan J, Bycroft C, Zemgulyte A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science.* 2023; 381(6664):eadg7492. Epub 20230922. <https://doi.org/10.1126/science.adg7492> PMID: 37733863.
57. Livesey BJ, Marsh JA. Updated benchmarking of variant effect predictors using deep mutational scanning. *Molecular systems biology.* 2023; 19(8):e11474. Epub 20230613. <https://doi.org/10.15252/msb.202211474> PMID: 37310135; PubMed Central PMCID: PMC10407742.
58. Zannini L, Lecis D, Lisanti S, Benetti R, Buscemi G, Schneider C, et al. Karyopherin-alpha2 protein interacts with Chk2 and contributes to its nuclear import. *J Biol Chem.* 2003; 278(43):42346–51. Epub 20030809. <https://doi.org/10.1074/jbc.M303304200> PMID: 12909615.
59. Sikorski RS, Hieter P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics.* 1989; 122(1):19–27. <https://doi.org/10.1093/genetics/122.1.19> PMID: 2659436; PubMed Central PMCID: PMC1203683.
60. Allen JB, Zhou Z, Siede W, Friedberg EC, Elledge SJ. The SAD1/RAD53 protein kinase controls multiple checkpoints and DNA damage-induced transcription in yeast. *Genes Dev.* 1994; 8(20):2401–15. <https://doi.org/10.1101/gad.8.20.2401> PMID: 7958905.

61. Boeke JD, Trueheart J, Natsoulis G, Fink GR. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. *Methods Enzymol.* 1987; 154:164–75. [https://doi.org/10.1016/0076-6879\(87\)54076-9](https://doi.org/10.1016/0076-6879(87)54076-9) PMID: 3323810.
62. Hofmann K, Bucher P. The FHA domain: a putative nuclear signalling domain found in protein kinases and transcription factors. *Trends Biochem Sci.* 1995; 20(9):347–9. [https://doi.org/10.1016/s0968-0004\(00\)89072-6](https://doi.org/10.1016/s0968-0004(00)89072-6) PMID: 7482699.
63. Li J, Williams BL, Haire LF, Goldberg M, Wilker E, Durocher D, et al. Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2. *Mol Cell.* 2002; 9(5):1045–54. [https://doi.org/10.1016/s1097-2765\(02\)00527-0](https://doi.org/10.1016/s1097-2765(02)00527-0) PMID: 12049740.