# Functional analysis of genetic variants in PALB2 and CHEK2: linking functional impact with cancer risk

Boonen, R.A.C.M.

# CHAPTER 6

6

## Functional analysis identifies damaging *CHEK2* missense variants associated with increased cancer risk

Rick A.C.M. Boonen, Wouter W. Wiegant, Nandi Celosse, Bas Vroling, Stephan Heijl, Zsofia Kote-Jarai, Martina Mijuskovic, Simona Cristea, Nienke Solleveld-Westerink, Tom van Wezel, Niko Beerenwinkel, Rosalind Eeles, Peter Devilee, Maaike P.G. Vreeswijk, Giancarlo Marra and Haico van Attikum

Published in *Cancer Research*
(PMID: 34903604)

**ABSTRACT**

Heterozygous carriers of germline loss-of-function variants in the tumor suppressor gene checkpoint kinase 2 (*CHEK2)* are at an increased risk for developing breast and other cancers. While truncating variants in *CHEK2* are known to be pathogenic, the interpretation of missense variants of uncertain significance (VUS) is challenging. Consequently, many VUS remain unclassified both functionally and clinically. Here we describe a mouse embryonic stem (mES) cell-based system to quantitatively determine the functional impact of 50 missense VUS in human *CHEK2*. By assessing the activity of human CHK2 to phosphorylate one of its main targets, Kap1, in *Chek2* knockout mES cells, 31 missense VUS in *CHEK2* impaired protein function to a similar extent as truncating variants, and 9 *CHEK2* missense VUS resulted in intermediate functional defects. Mechanistically, most VUS impaired CHK2 kinase function by causing protein instability or by impairing activation through (auto)phosphorylation. Quantitative results showed that the degree of CHK2 kinase dysfunction correlates with an increased risk for breast cancer. Both damaging *CHEK2* variants as a group (OR 2,23; 95% CI 1,62-3,07; $p$<0,0001) and intermediate variants (OR 1,63; 95% CI 1,21-2,20; $p$=0,0014) were associated with an increased breast cancer risk, while functional variants did not show this association (OR 1,13; 95% CI 0,87-1,46; $p$=0,378). Finally, a damaging VUS in *CHEK2*, c.486A>G/p.D162G, was also identified, which co-segregated with familial prostate cancer. Altogether, these functional assays efficiently and reliably identified VUS in *CHEK2* that associate with cancer.

**STATEMENT OF SIGNIFICANCE**

Quantitative assessment of the functional consequences of *CHEK2* variants of uncertain significance identifies damaging variants associated with increased cancer risk, which may aid in the clinical management of patients and carriers.

**KEYWORDS**

*CHEK2* gene; CHK2 protein; Variant of Uncertain Significance (VUS); Functional assays; Kap1 phosphorylation; Breast and prostate cancer; Cancer risk

## INTRODUCTION

The importance of genome stability for preventing breast and other cancers is evident from the increased cancer risk that results from inherited loss-of-function (LOF) variants in DNA damage repair genes such as *BRCA1/2* and *PALB2*, as well as in genes that control genome integrity checkpoints. The checkpoint kinase 2 (*CHEK2*) gene is a well-known example, which encodes the serine-threonine kinase CHK2 protein that becomes activated in response to DNA damage, and regulates cell cycle progression and apoptosis (1,2). The CHK2 protein is therefore believed to act as a tumor suppressor by delaying cell cycle progression to allow time for DNA repair, or by eliminating genomically unstable cells through induction of cell death (3). In 2002, association analysis of the truncating *CHEK2* c.1100delC/p.T367Mfs variant indeed revealed that it confers a moderate risk of breast cancer (4,5). Meanwhile, other studies have also shown that carriers of such LOF variants in the *CHEK2* gene are at a significantly increased risk for developing breast cancer (OR ~2,5) (6,7), as well as several other cancers such as prostate cancer (8-10). These studies firmly established that *CHEK2* is a low to moderate penetrance cancer susceptibility gene.



The growing body of evidence that associates *CHEK2* with breast cancer has led to increased genetic testing of *CHEK2,* and as a consequence to the identification of more (rare) genetic variants in this gene for which clinical significance is unknown (11-15). In fact, 1332 variants of uncertain significance (VUS) in *CHEK2* have currently been reported in ClinVar (16) (as of October 2021), most of which (i.e., 1139) are missense variants. For many of these missense variants the impact on protein function and the associated cancer risk remain to be elucidated. Assessment of pathogenicity of these VUS in a moderate risk gene such as *CHEK2* is mostly dependent on family history of cancer. To overcome this limitation, quantitative methods are required that can determine the functional impact of VUS in *CHEK2* and establish their relationship with cancer risk.

The CHK2 protein, which is expressed throughout the cell cycle, consists of 543 amino acids, and possesses three characteristic domains: an N-terminal SQ/TQ cluster domain (residues 19-69), a fork head-associated (FHA) domain (residues 92-205), and a serine/threonine kinase domain (residues 212-501). A nuclear localization signal (NLS) is located at the C-terminus of CHK2 (residues 515-522) (17). Activation of CHK2 kinase activity occurs specifically in response to DNA damage and is a multistep process initiated by ATM-mediated phosphorylation of several SQ/TQ sites, particularly p.T68, in its N-terminal regulatory domain (1,2). This promotes homodimerization and intermolecular autophosphorylation of CHK2 on p.T383 and p.T387 within the T-loop region (residues 366-406) (18), and on p.S516 within the NLS, collectively leading to efficient kinase activation and the subsequent phosphorylation of target proteins (19,20). The spectrum of known CHK2 targets includes proteins involved in cell cycle control (i.e., CDC25A and CDC25C

phosphatases), regulation of cell death (i.e., p53) (1,2,21), and DNA damage repair (i.e., BRCA1 and KAP1) (22-24). Following DNA damage, CHK2 phosphorylates KAP1 specifically at p.S473. This modification attenuates KAP1 binding to heterochromatin protein 1 family proteins, leading to relaxation of the damaged heterochromatin and promoting DNA damage repair (24-28).

In an effort to interpret *CHEK2* VUS, several studies assessed their functional consequences (29-36). The largest set of *CHEK2* variants to date was analyzed by Delimitsou and colleagues (34). They employed a yeast-based functional assay that assesses the ability of yeast strains expressing different *CHEK2* variants to resume proliferation and cell growth following repair of DNA damage induced by methyl methanesulfonate (MMS) (31,32). Other recent studies also assessed the ability of CHK2 variants to phosphorylate downstream targets such as CDC25C, BRCA1 and KAP1 (29,30,35). Although these studies have assayed >130 patient-derived *CHEK2* variants and identified numerous damaging missense variants, results were often discordant and the relationship with risk of breast and other cancers remained unclear. Consequently, there is a need to further improve the functional analysis of missense variants in *CHEK2,* and develop assays that can link the functional impact of such variants to cancer risk.

Here, we developed a mouse embryonic stem (mES) cell-based assay for the functional analysis of VUS in *CHEK2*. The assay allows a semi high-throughput analysis of variants in human *CHEK2* cDNA in *Chek2* knockout mES cells, using CHK2-mediated Kap1 p.S473 phosphorylation as a quantitative readout. Using this approach, we identified 31 *CHEK2* missense VUS to impair protein function to a similar extent as *CHEK2* truncating variants, while 9 missense VUS showed intermediate functional defects. Our results further indicate that at least two mechanisms are at play by which VUS in *CHEK2* impair protein function: loss of protein stability and defective (auto)phosphorylation/activation. Importantly, the degree of CHK2 kinase dysfunction observed for *CHEK2* missense variants highly correlates with increased breast cancer risk.

## RESULTS

### A cell-based functional assay for *CHEK2* variants

To assess the functional impact of *CHEK2* variants, we developed a mES cell-based system that allows for the semi high-throughput testing of variants in human *CHEK2.* To this end, we employed our mES cells carrying the well-established DR-GFP reporter for homologous recombination (HR) at the *Pim1* locus, and the recombination-mediated cassette exchange (RMCE) system at the *Rosa26* locus (38,43). CRISPR/Cas9-mediated genome-editing was

used to knockout (KO) mouse *Chek2* in these cells (Fig. 1a, Supplementary Fig. S1a-c) (38,43). Given that BRCA1, a crucial player in HR, becomes phosphorylated by CHK2, and given that this event promotes the dispersion of BRCA1 from DNA breaks (46), we assessed whether KO of *Chek2* affects the efficiency of HR in the DR-GFP reporter. Analysis of one heterozygous and two homozygous *Chek2*$^{KO}$ clones revealed that HR remained unaffected in these cells (Supplementary Fig. S1d), suggesting that loss of *Chek2* does not affect HR.

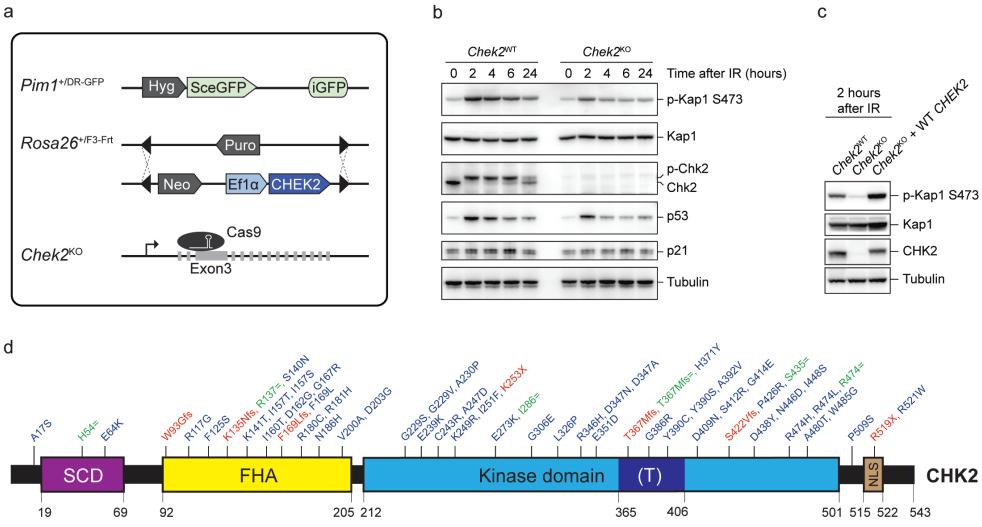



**Figure 1.** Generation of a cDNA-based complementation system for the functional analysis of human *CHEK2* variants. **a** Schematic representation of the mES cell- and cDNA-based complementation system for functional analysis. The DR-GFP reporter and Recombination-Mediated Cassette Exchange system (RMCE) have been stably integrated at the *Pim1* and *Rosa26* loci, respectively. Endogenous mouse *Chek2* was targeted with CRISPR/Cas9 using a gRNA against exon 3. **b** Western blot analysis of the indicated proteins from unirradiated and IR-exposed (10Gy) *Chek2*$^{WT}$ and *Chek2*$^{KO}$ mES cells. Tubulin was used as a loading control. **c** Western blot analysis of the indicated proteins from IR-exposed (10Gy) *Chek2*$^{WT}$, *Chek2*$^{KO}$, and *Chek2*$^{KO}$ mES cells complemented with human *CHEK2* cDNA. Tubulin was used as a loading control. **d** Schematic representation of the CHK2 protein with variant positions indicated and categorized as either synonymous (green), truncating (red) and missense VUS (blue). The amino acid numbers are shown to demarcate CHK2's evolutionarily conserved functional domains. (T) refers to the T-loop or activation segment.

CHK2 is known for its role in p53-mediated cell cycle control and apoptosis, as well as DNA damage repair in heterochromatin (1,2,21-24). Although we did not detect major changes in the cell cycle profile of *Chek2*$^{KO}$ cells when compared to wild type cells (Supplementary Fig. S1e), we did observe a slight, though not significant growth advantage for the *Chek2*$^{KO}$ cells over the course of 5 days (Supplementary Fig. S1f). In agreement with previous studies

(47,48), this growth advantage became more pronounced after DNA break induction by the radiomimetic agent phleomycin (Supplementary Fig. S1g). Moreover, p53 protein levels were moderately reduced in these cells after exposure to ionizing radiation (IR, 10Gy) (Fig. 1b). Accordingly, the expression of p53 target genes was also reduced, as evidenced by reduced p21 and Mdm2 transcript and p21 protein levels (Fig. 1b, Supplementary Fig. S1h). Most evidently, however, we observed that Kap1 phosphorylation at p.S473, which is required for DNA repair in heterochromatin (24), was strongly impaired in *Chek2*$^{KO}$ cells after IR (Fig. 1b).

We decided to exploit the strong impact of *Chek2* loss on Kap1 p.S473 phosphorylation as a read-out for the functional analysis of human *CHEK2* variants. To this end, we stably integrated human wild type *CHEK2* cDNA by RMCE in *Chek2*$^{KO}$ mES cells (Fig. 1a). Prior to examining CHK2 kinase activity, we pooled all the neomycin-resistant clones with stably integrated *CHEK2* cDNA (Fig. 1a), to average out any clonal variability in *CHEK2* expression. We found that the defect in IR-induced Kap1 p.S473 phosphorylation in *Chek2*$^{KO}$ cells was efficiently rescued following expression of human *CHEK2* (Fig. 1c). Strikingly, human CHK2 appeared to phosphorylate mouse Kap1 even more efficiently when compared to endogenous mouse Chk2, while their expression levels were comparable (Fig. 1c). Thus, we established a cDNA-based complementation system for the functional analysis of human *CHEK2* genetic variants using Kap1 p.S473 phosphorylation as a read-out.

**Validation of a cell-based functional assay for *CHEK2* variants**

To validate our system, we selected 7 truncating and 6 synonymous *CHEK2* variants for functional analysis (Fig. 1d). Sequence-verified constructs were introduced by RMCE into the *Chek2*$^{KO}$ mES cells and their ability to phosphorylate Kap1 at p.S473 after IR was assessed by western blot analysis. As expected, in *Chek2*$^{KO}$ cells complemented with an empty vector or a truncating *CHEK2* variant, phosphorylation of Kap1 p.S473 was strongly impaired at both 2 and 6 hours after IR (Fig. 2). The exception to this was the nonsense variant p.R519X which moderately impacted Kap1 p.S473 phosphorylation at 2 hours after IR (Fig. 2), even though it was classified as likely pathogenic in ClinVar. p.R519X leads to a truncated CHK2 protein that lacks part of its NLS domain (Fig. 1d; amino acids 515-522). Possibly, residual nuclear localization of this variant is sufficient to induce partial Kap1 p.S473 phosphorylation after IR, suggesting it acts as a hypomorphic variant. In contrast to truncating *CHEK2* variants, cells that expressed synonymous variants showed phospho-Kap1 p.S473 levels comparable to cells expressing wild type *CHEK2* (Fig. 2). Neither the expression of different *CHEK2* variants, nor the exposure to IR affected overall Kap1 protein levels, suggesting that CHK2 activity does not affect Kap1 stability or expression (Fig. 2).

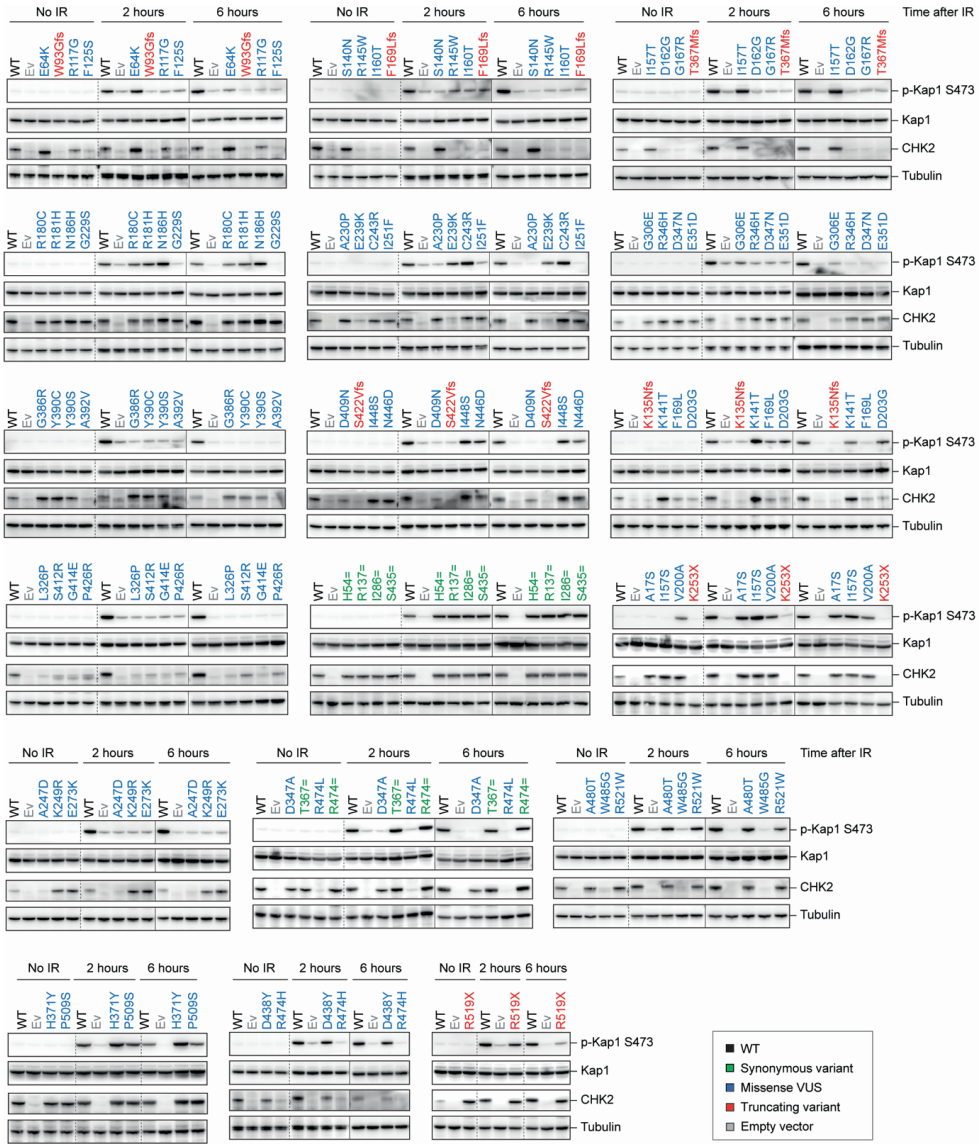



**Figure 2.** Human *CHEK2* variants and their effect on CHK2 expression and kinase activity toward Kap1 p.S473. Western blot analysis of the indicated proteins from *Chek2*$^{KO}$ mES cells expressing wild type (WT, black) human untagged CHK2, empty vector (Ev, grey), or the indicated untagged CHK2 variants in untreated conditions (no IR) or at 2 or 6 hours after IR exposure (10Gy). WT and Ev served as controls on each blot and variants are categorized by color as either synonymous (green), truncating (red) and missense VUS (blue). Tubulin was used as a loading control. Dashed lines represent a marking of different set of samples on the same blot, whereas continuous lines are used to mark different sets of samples from distinct and separately exposed blots.

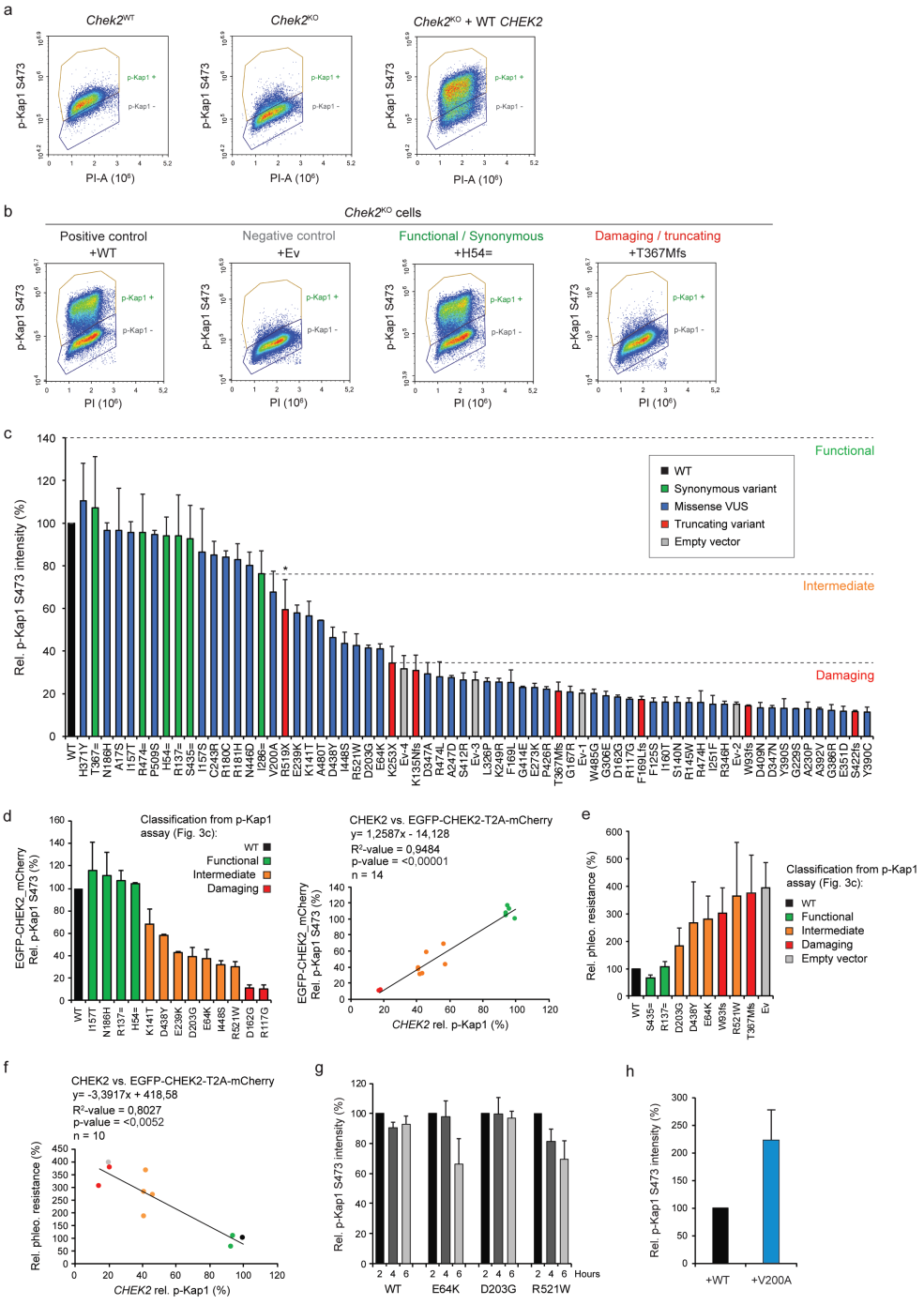

a
Chek2WT
Chek2KO
Chek2KO + WT CHEK2
p-Kap1 S473
PI-A (10^6)
p-Kap1 +
p-Kap1 -
b
Chek2KO cells
Positive control
+WT
Negative control
+Ev
Functional / Synonymous
+H54=
Damaging / truncating
+T367Mfs
PI (10^6)
c
Rel. p-Kap1 S473 intensity (%)
Functional
Intermediate
Damaging
WT
Synonymous variant
Missense VUS
Truncating variant
Empty vector
d
EGFP-CHEK2_mCherry Rel. p-Kap1 S473 (%)
Classification from p-Kap1 assay (Fig. 3c):
WT
Functional
Intermediate
Damaging
CHEK2 vs. EGFP-CHEK2-T2A-mCherry
y= 1,2587x - 14,128
R2-value = 0,9484
p-value = <0,00001
n = 14
CHEK2 rel. p-Kap1 (%)
e
Rel. phleo. resistance (%)
Classification from p-Kap1 assay (Fig. 3c):
WT
Functional
Intermediate
Damaging
Empty vector
f
CHEK2 vs. EGFP-CHEK2-T2A-mCherry
y= -3,3917x + 418,58
R2-value = 0,8027
p-value = <0,0052
n = 10
Rel. phleo. resistance (%)
CHEK2 rel. p-Kap1 (%)
g
Rel. p-Kap1 S473 intensity (%)
Hours
WT
E64K
D203G
R521W
h
Rel. p-Kap1 S473 intensity (%)
+WT
+V200A

**Figure 3.** Human *CHEK2* variants and their effect on CHK2's kinase activity toward Kap1 p.S473. **a** Quantitative FACS-based analysis of Kap1 p.S473 phosphorylation in *Chek2*$^{WT}$, *Chek2*$^{KO}$, and *Chek2*$^{KO}$ mES cells complemented with human untagged *CHEK2* cDNA at 2 hours after IR exposure (10Gy). **b** Quantitative FACS-based analysis of Kap1 p.S473 phosphorylation in *Chek2*$^{KO}$ mES cells complemented with the indicated untagged constructs at 2 hours after IR exposure (10Gy). **c** Quantification of FACS measurements of Kap1 p.S473 phosphorylation in *Chek2*$^{KO}$ mES cells expressing wild type (WT, black) human untagged CHK2, empty vector (Ev, grey), or the indicated untagged CHK2 variants (blue and red) at 2 hours after IR exposure (10Gy). Data represent mean percentages ± SEM of the average phospho-Kap1 p.S473 intensity observed in the 'p-Kap1 +' gate as shown in b from 2 independent experiments. Data are relative to WT, which was set to 100%. Ev1-4 refer to four independent Ev controls that were included. Dashed lines indicate functional thresholds based on the synonymous or truncating variant with the lowest or highest Kap1 p.S473 phosphorylation level, respectively. The asterisk marks p.R519X, which acted as a hypomorphic variant and was therefore not used for thresholding. **d** Quantification of FACS measurements (left) of Kap1 p.S473 phosphorylation in *Chek2*$^{KO}$ mES cells complemented with EGFP-CHEK2-T2A-mCherry, with or without a *CHEK2* variant, at 2 hours after IR exposure (10Gy). Data represent mean percentages ± SEM of the average phospho-Kap1 p.S473 intensity observed after gating for mCherry-positive cells from 2 independent experiments. Data are relative to WT, which was set to 100%. Scatter plot (right) shows the correlation between phospho-Kap1 p.S473 intensities measured in *Chek2*$^{KO}$ mES cells expressing untagged *CHEK2* or EGFP-tagged *CHEK2 (*from stably integrated EGFP-CHEK2-T2A-mCherry). Conditions are colored as indicated based on functional classification using untagged *CHEK2* cDNA as shown in c. **e** Phleomycin sensitivity assay using *Chek2*$^{KO}$ mES cells complemented with the indicated untagged CHK2 constructs or empty vector (Ev). Cells were exposed to 2,5 μM of phleomycin for two days. Cell viability was measured after one additional day of incubation in drug-free medium using FACS (using only forward and sideways scatter). Data represent the mean percentage ± SEM of viability relative to untreated cells from 3 independent experiments. **f** Scatter plot showing the correlation between phospho-Kap1 p.S473 intensities and the relative resistance to 2,5 μM phleomycin as measured in e in *Chek2*$^{KO}$ mES cells expressing untagged CHK2 variants. **g** Quantification of FACS measurements of Kap1 p.S473 phosphorylation in *Chek2*$^{KO}$ mES cells expressing wild type (WT) untagged *CHEK2* or three selected variants at the indicated times after 10Gy of IR. For each condition, data are plotted relative to the 2 hours timepoint, which was set to 100%. **h** Quantification of FACS measurements of Kap1 p.S473 phosphorylation in *Chek2*$^{KO}$ mES cells expressing wild type (WT, black) untagged CHK2, or untagged CHK2 carrying the p.V200A variant (blue) at 2 hours after IR exposure (10Gy). Data from 2 independent experiments are represented as in c.



### A quantitative cell-based functional assay for *CHEK2* variants

Complementary western blot analysis, which is at best semi-quantitative in our setup, we next aimed for a more quantitative approach. To this end, we used fluorescence-activated cell sorting (FACS) to determine the levels of phospho-Kap1 p.S473. Consistent with results from western blot analysis (Fig. 1c), we observed a strong reduction in the phospho-Kap1 p.S473 signals in *Chek2*$^{KO}$ cells 2 hours after IR (Fig. 3a). Surprisingly, we also observed substantial Kap1 p.S473 phosphorylation in unirradiated *Chek2*$^{KO}$ cells, albeit this was most likely restricted to M-phase cells and disappeared after IR exposure (Supplementary Fig. S2). Complementation of *Chek2*$^{KO}$ cells with wild type human *CHEK2* cDNA rescued the defect in IR-induced Kap1 p.S473 phosphorylation and even led to higher phospho-Kap1 p.S473 signals when compared to that in *Chek2* wild type cells (Fig. 3a). This effect was also seen for

the 6 synonymous *CHEK2* variants (Fig. 3b, Supplementary Fig. S3, 4a). In contrast, complementation with the empty vector or the truncating variants (except the hypomorphic variant p.R519X), resulted in a complete absence of cells that were positive for phospho-Kap1 p.S473 (Fig. 3b, Supplementary Fig. S3, 4a). Thus, the quantitative results obtained using a FACS-based approach fully corroborated the results obtained by western blot analysis.

Notably, our FACS-based analysis showed a large population of cells that is negative for phospho-Kap1 p.S473, even after expression of wild type *CHEK2* or a synonymous variant (Fig. 3a, b, Supplementary Fig. S3, 4a). Stable introduction of a construct that carries a T2A sequence for co-expression of EGFP-CHEK2 and mCherry (EGFP-CHEK2-T2A-mCherry) showed that there is both a GFP/mCherry-positive as well as GFP/mCherry-negative population of cells (Supplementary Fig. S5a). These data suggest that a large portion of cells lose *CHEK2* expression after stable integration. Importantly the GFP/mCherry-negative population of cells was clearly phospho-Kap1 p.S473-negative, even following exposure to IR (Supplementary Fig. S5a). We therefore excluded this population from our analysis and quantified the mean intensity of phospho-Kap1 p.S473 (Fig. 3c) only for cells that were positively gated for phospho-Kap1 p.S473 (Fig. 3a, b). As expected, this showed that synonymous variants exhibited kinase activity comparable to that of wild type CHK2 (i.e., a reduction of <24%), whereas the truncating *CHEK2* variants (except the hypomorphic variant p.R519X) caused a major reduction in kinase activity of >69%. Thus, our cell-based system can classify functional/synonymous and damaging/truncating *CHEK2* variants based on their effect on Kap1 p.S473 phosphorylation.

**Functional analysis of *CHEK2* missense VUS**

Having established a quantitative cell-based functional assay for *CHEK2* variants, we next examined the effect of 50 missense VUS. The majority of these VUS were identified using a multigene panel analysis of a large case-control association study performed by the BRIDGES consortium and Breast Cancer Association Consortium (BCAC) (7). Importantly, for all 50 missense VUS, the contribution with respect to cancer risk is largely unclear and insights into their functionality may aid in their clinical classification. Following their expression in *Chek2*$^{KO}$ cells using the non-tagged *CHEK2* cDNA, we found that 31 VUS strongly impaired CHK2 kinase activity toward Kap1 p.S473, comparable to that observed for *CHEK2* truncating variants and the empty vector conditions (Fig. 3c, Supplementary Fig. S3, 4a). Importantly, p.R519X was not used to set the threshold for damaging variants as it distinguished itself from the other truncating variants by acting as a hypomorphic variant (Fig. 2, Fig. 3c, Supplementary Fig. S4a). In addition to p.R519X, 9 *CHEK2* missense VUS similarly exhibited intermediate functional defects (Fig. 3c, Supplementary Fig. S3, 4a). The remaining 10 *CHEK2* missense VUS did not impact CHK2's functionality (Fig. 3c, Supplementary Fig. S3, 4a). These results

were in agreement with those from the western blot analysis (Fig. 2). However, correlation analysis showed that especially among the functional and intermediate *CHEK2* variants, western blot analysis is inefficient in discriminating functional differences ($R^2$ = 0,71; $p$<0,0001) (Supplementary Fig. S4b). Thus, the FACS-based phospho-Kap1 p.S473 analysis allows for a quantitative and therefore more accurate functional classification of *CHEK2* variants.

We noticed that with FACS analysis, differentiating the positive phospho-Kap1 p.S473 population from the negative population was difficult for cells that expressed *CHEK2* VUS with intermediate function (p.E64K, p.K141T, p.D203G, p.E239K, p.D438Y, p.I448S, p.A480T and p.R521W). We therefore repeated the FACS-based quantification of phospho-Kap1 p.S473 for several missense variants (4 functional, 7 intermediate, and 2 damaging variants) following co-expression of EGFP-*CHEK2* and mCherry (EGFP-CHEK2-T2A-mCherry). Following selection of GFP/mCherry-positive cells, the effects of these variants on Kap1 p.S473 phosphorylation fully corroborated those obtained with cells expressing non-tagged *CHEK2* (i.e., $R^2$ = 0.95), as all intermediate variants displayed intermediate effects on kinase activity (Fig. 3d, Supplementary Fig. S5b).



As Kap1 represents only one of the many targets of CHK2, an important question was whether the functional defects with regards to Kap1 p.S473 phosphorylation also translate to other functions of CHK2. To address this, we used a more general readout, i.e., cell growth after DNA damage induction, which is likely regulated by CHK2's activity on multiple downstream targets. For this, we assessed the impact of two benign (p.R137= and p.S435=), two pathogenic (p.W93fs and p.T367Mfs) and four intermediate *CHEK2* variants (p.E64K, p.D203G, p.D438Y and p.R521W) on cell survival after phleomycin treatment (Fig. 3e). Their impact on cell survival correlated well with phospho-Kap1 p.S473 levels as measured by FACS ($R^2$ = 0,80; $p$=0,0052) (Fig. 3f). However, the growth effects for intermediate variants were variable among replicate experiments, whereas the effects observed for the benign and pathogenic variants were reproducible. These data suggest that our FACS-based assay is a robust and reliable approach for the functional classification of *CHEK2* variants and that phosphorylation of Kap1 p.S473 is a suitable readout to assess the general impact of variants on CHK2 function.

**Several variants alter the kinetics of CHK2**

The analysis of phospho-Kap1 p.S473 levels in unirradiated cells, and 2 or 6 hours after IR, showed that two *CHEK2* missense VUS (p.E64K and p.R521W) were unable to maintain phosphorylation of Kap1 at p.S473 at the later timepoint (Fig. 2). To confirm this, we expressed these VUS in *Chek2*$^{KO}$ cells using the non-tagged *CHEK2* cDNA and assessed phospho-Kap1 p.S473 levels by FACS at 2, 4 and 6 hours after IR (Supplementary Fig. S6). Quantification of the average intensity of phospho-Kap1 p.S473 showed that for wild type *CHEK2*, the signal

intensity only slightly decreases in time compared to that at 2 hours after IR (Fig. 3g, Supplementary Fig. S6). Similarly, for p.D203G, which we identified as a variant with intermediate functional impact (Fig. 3c), we observed that phospho-Kap1 p.S473 levels are maintained in time (Fig. 3g, Supplementary Fig. S6), even though overall phospho-Kap1 p.S473 levels at 2h after IR were lower than in cells expressing wild type *CHEK2*. For both p.E64K and p.R521W, however, the phospho-Kap1 p.S473 levels were strongly reduced at 6 hours after IR (Fig. 3g, Supplementary Fig. S6). Additionally, we observed that the truncating *CHEK2* variant p.R519X resulted in the same kinetic defect as p.R521W (Fig. 2). Functional classification of such variants is therefore strongly dependent on the timepoint after IR at which CHK2 activity is measured. This may also explain why previous reports using different approaches classified p.E64K and p.R521W as either neutral or damaging, rather than intermediate (Supplementary Fig. S7a, b) (34,35). In addition, we found that one variant (i.e., p.V200A) displayed unregulated CHK2 activity in the absence of DNA damage induction (Fig. 2). Analysis of phospho-Kap1 p.S473 levels by FACS analysis confirmed this functional effect (Fig. 3h, Supplementary Fig. S4c). In conclusion, p.E64K, p.V200A, p.R519X and p.R521W alter the kinetics of CHK2 activity, implicating a mechanism for aberrant protein function that has not been previously reported for *CHEK2* genetic variants.

**Correlation between computational predictions and functionality of variants**

With the rapid accumulation of identified VUS in cancer associated genes (49,50), computational tools can aid in the clinical interpretation of such variants (51). We therefore compared the quantitative outcome of our functional assays for *CHEK2* missense variants (Fig. 3c) with the predictions from twelve algorithms: Helix, PolyPhen (hvar), PolyPhen (hdiv), VEST4, REVEL, PrimateAI, CADD, Provean, Deogen2, MVP, SIFT and FATHMM (Fig. 4a, b, Supplementary Fig. S8). Interestingly, Helix (52) outperformed all other tools (Fig. 4a). This tool is a missense variant effect predictor built on an extensive resource of protein data, in which protein structures, together with high-quality structure-based multiple sequence alignments (MSAs) for the complete structural space, are combined with full length sequence-based MSAs for the human proteome. Furthermore, Helix was trained on a large set of well-annotated variants using a strict training regime where circularity is actively avoided (53). When comparing the predictions from Helix to our functional data, we observed a significant correlation ($R^2$ = 0,66; $p$<0,00001) (Fig. 4b). Such a correlation was also observed for the functional data from Delimitsou et al. (34) ($R^2$ = 0,48; $p$<0,0001), but not for those from Kleiblova et al. (35) ($R^2$ = 0,31; $p$=0,13) (Fig. 4b). For the *CHEK2* VUS in our study, both versions of PolyPhen (hvar and hdiv) also appeared to predict functional effects relatively well ($R^2$ = 0,52 and 0,44, respectively), but the effects of intermediate *CHEK2* VUS, as well as of

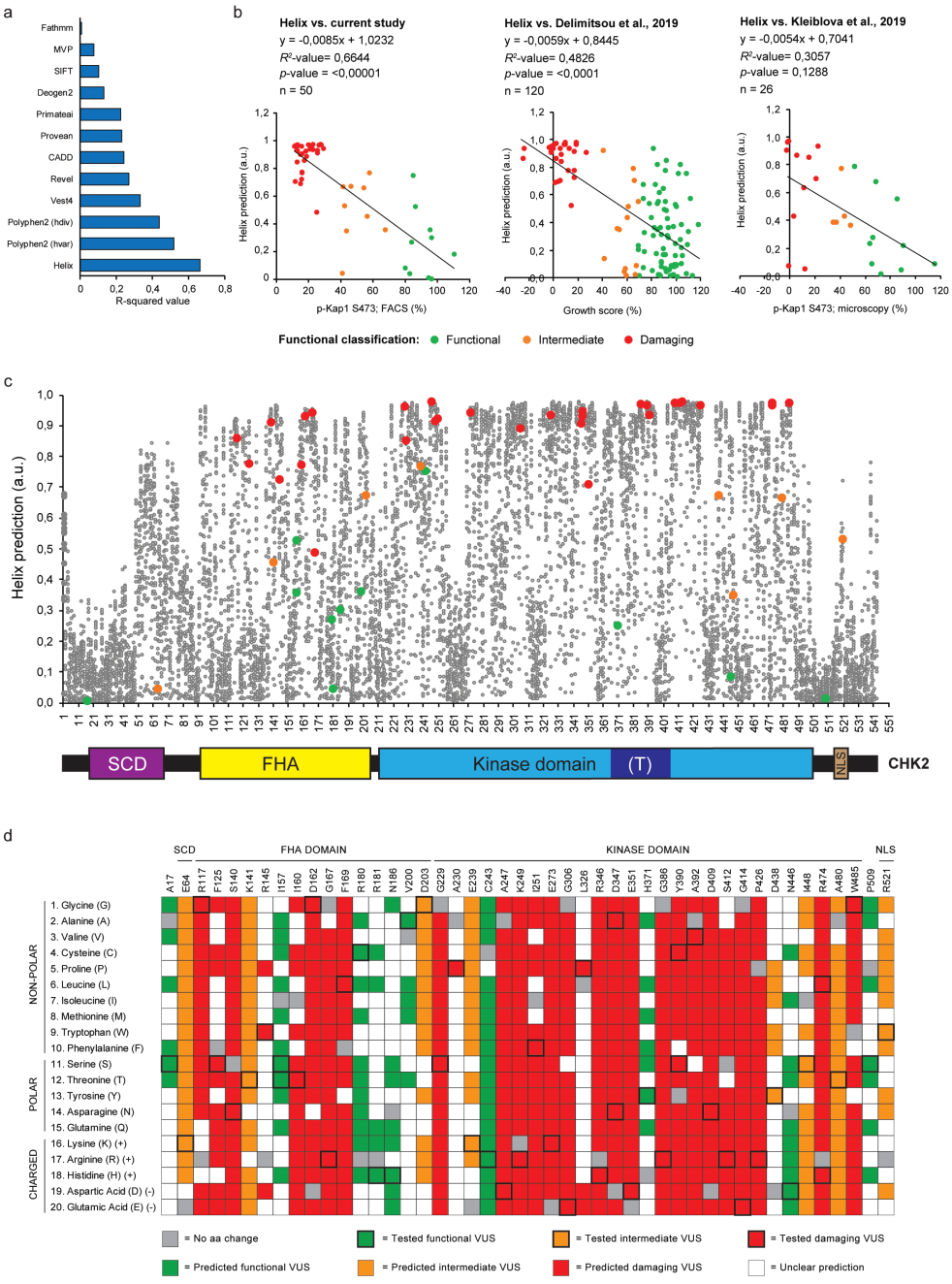



**Figure 4.** Correlation between computational predictions and functionality of *CHEK2* variants. **a** Bar plot showing the $R^2$-correlation values between the FACS-based analysis of Kap1 p.S473 phosphorylation as shown in Fig. 3c and computational predictions from twelve different prediction algorithms. **b** Scatter plot showing the correlation between Helix-based *in silico* predictions and results from functional assays

presented in our study (Fig. 3c), or those from Delimitsou et al. 2019 (34) and Kleiblova et al. 2019 (35). Datapoints are colored based on functional classification (green, functional; orange, intermediate; red, damaging). Helix provides predictions for pathogenicity ranging from 0-1, with values close to 1 representing pathogenic predictions. **c** En masse prediction plot from Helix for all possible missense changes in human *CHEK2*. Schematic representation of the CHK2 protein and its functional domains demarcated by the amino acid numbers at the X-axis of the plot. **d** Heatmap showing predictions from Helix combined with functional data for CHK2 amino acid changes that were analyzed in Fig. 3c (outlined in bold). For functional variants indicated in green (with bold outline), amino acid changes with a similar (+0.05) or lower prediction from Helix are also indicated in green. For intermediate variants indicated in orange (with bold outline), amino acid changes with a similar (-0.05) or higher prediction from Helix are also indicated in orange. For damaging variants indicated in red (with bold outline), amino acid changes with a similar (-0.05) or higher prediction from Helix are also indicated in red. For each amino acid position, amino acid changes with a similar color code are expected to result in similar functional effects. Squares in grey and white represent changes into the original amino acid or variant changes for which predictions are unclear, respectively.

several functional VUS, were overestimated (Supplementary Fig. S8). Importantly, other tools, particularly REVEL, Provean, Deogen and FATHMM, underestimated the effect of several damaging variants in *CHEK2* (Supplementary Fig. S8). Together, these findings highlight the potential of Helix with regards to interpretation of missense variants in *CHEK2*.

To better understand the functional effects of missense variants throughout the entire CHK2 protein, we next visualised the predictions from Helix for all possible missense alterations in *CHEK2* (Fig. 4c, Supplementary Table S1). Interestingly, many missense changes were predicted to exhibit damaging effects. This may be due to the relatively small size of the CHK2 protein (62 kDa, 543 amino acids), in which unfavourable missense substitutions (based on amino acid characteristics) may be more prone to affect function than in larger proteins. Furthermore, we used the predictions from Helix to examine the functional effects of alternative amino acid changes for each *CHEK2* missense VUS in this study (Fig. 4d). This suggested that several conserved CHK2 amino acid residues (e.g., p.S140, p.G229, p.A247, p.K249, p.E273, p.R346, p.D347, p.E351, p.G386, p.D409, p.G414, p.P426 and p.R474) are critical for kinase function. Not surprisingly, this included the p.S140 autophosphorylation site that regulates CHK2 dimerization (54), p.E273 which is important for ATP hydrolysis (55,56), and the catalytic residue p.D347A (55). Thus, Helix is a powerful tool to predict the impact of missense alterations in *CHEK2* and can highlight regions and specific residues that are crucial for protein function.

### *CHEK2* VUS affect protein function through distinct mechanisms

Our western blot analysis showed that many *CHEK2* missense variants result in reduced protein levels (Fig. 2). To further assess their effect on protein stability, we selected 30 VUS and introduced these in our EGFP-CHEK2-T2A-mCherry construct. Following RMCE in

*Chek2*$^{KO}$ mES cells, steady-state abundance of CHK2 protein variants was measured based on GFP fluorescence in mCherry-positive cells, ruling out transcriptional effects on EGFP-CHEK2 expression. The GFP signal for the two synonymous *CHEK2* variants (p.H54= and p.R137=), as well as that for several other functional, intermediate and damaging VUS (e.g., p.E64K, p.K141T, p.I157T, p.N186H, p.E273K, p.G306E, p.G386R, p.I448S, p.R521W), was comparable to wild type *CHEK2* (Fig. 5a). However, all variants that displayed clearly reduced CHK2 protein levels on western blot (Fig. 2), also exhibited strongly reduced GFP signals (i.e., <65%) (Fig. 5a). Overall, we identified 18 *CHEK2* VUS that exhibit major effects on CHK2 protein stability, thereby hampering CHK2 kinase function.

Several damaging variants (e.g., p.E273K and p.G386R) did not affect CHK2 protein stability, yet impaired IR-induced Kap1 p.S473 phosphorylation (Fig. 2, Fig. 5a). We therefore questioned whether these variants affect CHK2 kinase activation. Autophosphorylation of CHK2 is essential for its activation and occurs, amongst others, on residues p.T383 and p.T387 in the T-loop region located within the kinase domain (Fig. 1d) (19,20). Consistent with a role for ATM in CHK2 activation (20,57), exposure of cells to ATM inhibitor completely abolished IR-induced autophosphorylation of CHK2 on p.T383 (Fig. 5b). Subsequently, we examined the effect of 7 intermediate and 13 damaging *CHEK2* variants, which did not affect CHK2 protein stability (with exception of p.D203G and p.D438Y), on CHK2 p.T383 phosphorylation (Fig. 5c, Table 1). Most of these *CHEK2* variants reduced (n=8) or completely abolished autophosphorylation (n=8). Surprisingly, 5 *CHEK2* variants (i.e., p.I251F, p.E273K, p.Y390C, p.Y390S, and particularly p.E351D) that did not grossly impact CHK2 p.T383 autophosphorylation, still impaired kinase activity toward Kap1 p.S474 (Fig. 5c), possibly by impacting ATP binding/hydrolysis. Thus, our results suggest that the damaging effect of *CHEK2* variants is a consequence of protein instability, impaired kinase activation, or perhaps reduced ATP binding/hydrolysis.

**Association of CHK2 functional defects with breast cancer risk**

Having determined the functional impact of VUS in *CHEK2*, we next investigated whether the observed impact correlates with increased breast cancer risk. For this, we considered all 30 population-based BCAC studies, which were combined in a case-control association study performed by the BRIDGES consortium (48826 breast cancer cases and 50703 controls) (7). Due to the low allele frequency of most *CHEK2* variants, we were only able to identify two variants, c.190G>A/p.E64K (OR 1,78; 95% CI 1,14-2,77; *p*=0.0112) and c.349A>G/p.R117G (OR 2,22; 95% CI 1,34-3,68; *p*=0,0020) (Table 1), that associate with significantly increased breast cancer risk and for which the population-based ORs had a relatively narrow CI. p.E64K had an intermediate functional impact, whereas p.R117G was damaging (Fig. 3c), suggesting that the degree of functional impact correlates with the breast cancer risk level.

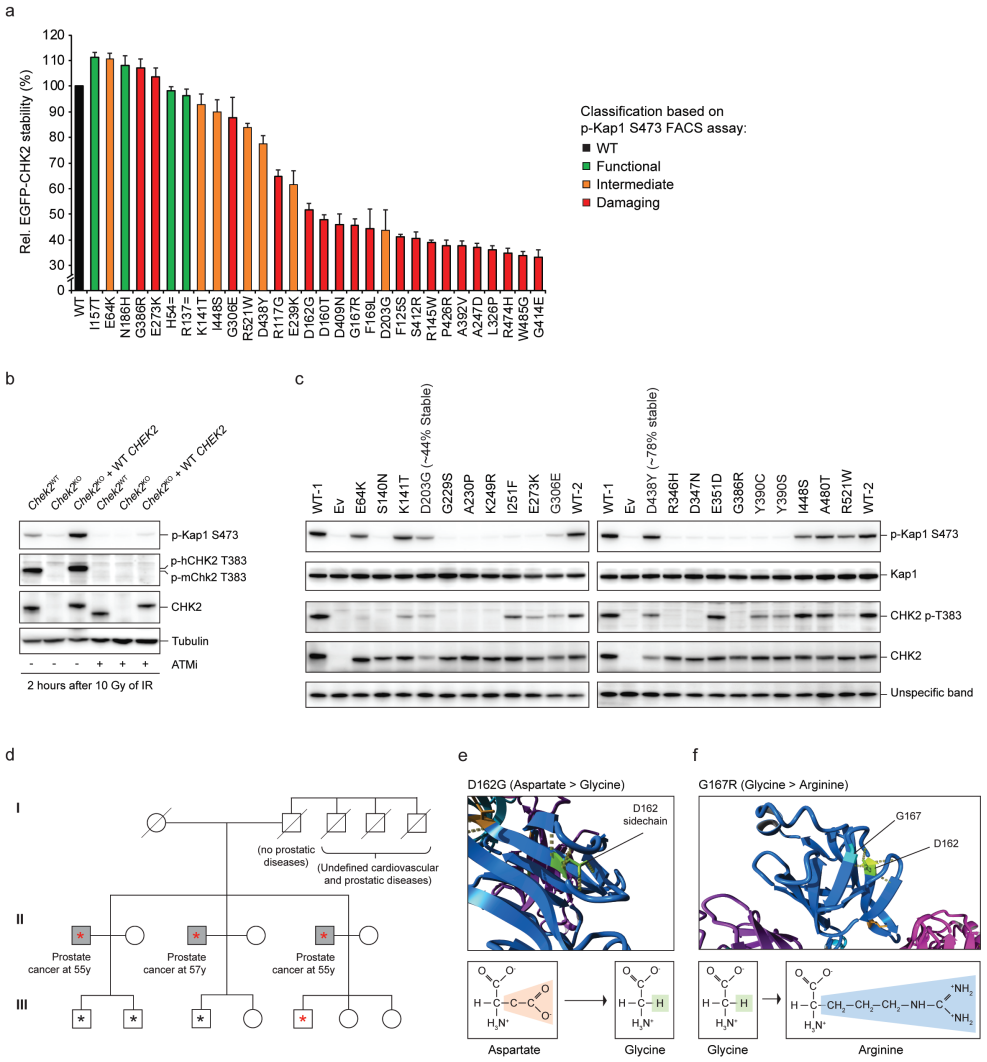



**Figure 5.** Analysis of pathogenic mechanisms of *CHEK2* VUS and the association of two VUS with prostate cancer. **a** Quantification of FACS measurements of the average EGFP intensity in *Chek2*$^{KO}$ mES cells complemented with EGFP-CHEK2-T2A-mCherry, with or without the indicated *CHEK2* variants. EGFP intensities were measured in mCherry-positive gated cells. Data represent mean percentages ± SEM for 3 independent measurements and are relative to WT which is set at 100%. **b** Western blot analysis of the indicated proteins from IR-exposed (10Gy) *Chek2*$^{WT}$, *Chek2*$^{KO}$, and *Chek2*$^{KO}$ mES cells complemented with human *CHEK2* cDNA that were left untreated or treated with ATM inhibitor (ATMi). Tubulin was used as a loading control. **c** Western blot analysis of the indicated proteins from IR-exposed (10Gy) *Chek2*$^{KO}$ mES cells complemented with human *CHEK2* cDNA without or with a *CHEK2* variant that displayed intermediate or damaging effects in Fig. 3c. An unspecific band produced by the anti-CHK2 antibody was used as a loading control. **d** Pedigree of the family with the *CHEK2* c.485A>G/p.D162G variant. Three male siblings carrying *CHEK2* c.485A>G/p.D162G developed prostate cancer in their fifties (grey squares). Circles indicate females and squares indicate males. The asterisks indicate family members whose blood cell DNA was subjected to exome

sequencing. The red asterisks indicate members carrying the *CHEK2* c.485A>G/p.D162G variant. **e-f** Partial structures (top) of the CHK2 FHA domain showing the effect of two CHK2 variants exhibiting protein instability as shown in a. Formulas and changes for the indicated amino acids are shown (bottom).

Under the assumption that variants with a similar impact on CHK2 functionality confer the same level of cancer risk, we performed a burden-type association analysis (Table 2). Accordingly, we defined three groups of *CHEK2* VUS based on their impact on CHK2 function (i.e., functional, intermediate or damaging) and established the joint frequencies of the individual variants within the same group in both cases and controls. The two variants mentioned above (p.E64K and p.R117G) were excluded from these groups as they were already associated with a significant breast cancer risk (Table 1). This analysis revealed that functional *CHEK2* VUS as a group (n=6, excl. p.I157T and p.R180C for which carrier frequencies were not available) (Fig. 3c), are not associated with an increased risk for breast cancer (OR 1,13; 95% CI 0,87-1.46; *p*=0,3773) (Table 2). However, *CHEK2* VUS that exhibited an intermediate functional effect (n=7, excl. p.E64K) (Fig. 3c) were associated with a significantly increased risk for breast cancer (OR 1,52; 95% CI 1.01-2,28; *p*=0.0448) (Table 2). Importantly, damaging *CHEK2* VUS (n=27, excl. p.R117G) (Fig. 3c), were associated with an even higher risk than intermediate variants (OR 2,23; 95% CI 1,48-3,38; *p*<0.0001) (Table 2). In addition to population-based ORs, cancer risks described in Table 1 and 2 were also calculated based on all 44 BCAC studies (combination of 30 population-based and 14 family-based studies) (7). Although this generally resulted in slightly higher risk estimations for most *CHEK2* variants or variant groups, a similar correlation between functional impact of variants and cancer risk was observed (Table 1, Table 2). These results suggest that our quantitative functional assay can identify pathogenic *CHEK2* variants.

**Association of the *CHEK2* c.485A>G/p.D162G variant with prostate cancer**

Functional defects caused by *CHEK2* VUS are not only associated with an increased risk of developing breast cancer, but have also been linked to other cancers, including prostate cancer (9,10). We therefore examined three male siblings from a family that all presented with prostate cancer >10 years earlier than the average age of onset for sporadic prostate cancer. This revealed that they were all heterozygous for the germline *CHEK2* c.485A>G/p.D162G allele (Fig. 5d), which was characterized as a damaging variant in this study (Fig. 3c). Similarly, the closely located *CHEK2* VUS p.G167R had also been linked to prostate cancer (10). Our results showed that both variants lead to protein instability (Fig. 5a, Supplementary Fig. S5b), rendering CHK2 non-functional (Fig. 2, Fig. 3c). Consistently, using the crystal structure of CHK2 (PDB - 3I6U) (55), in silico modeling of CHK2 p.D162G and p.G167R showed that these

substitutions are extremely unfavorable for correct folding of the region they locate to, as they lead to loss of two hydrogen bonds (Fig. 5e-f). Interestingly, analysis of prostate tumor DNA of two of the three siblings carrying the *CHEK2* c.485A>G/p.D162G variant showed no evidence for loss of heterozygosity (LOH) (Table 3), resembling observations made for the well-known *CHEK2* c.1100delC/p.T367Mfs allele in breast cancer (Table 3) (58). These results suggest that LOH for individuals carrying a monoallelic damaging *CHEK2* variant may not be a prerequisite for cancer development, although we cannot rule out that promotor methylation silenced expression of the intact allele, thereby mimicking LOH (59). The findings on *CHEK2* c.485A>G/p.D162G suggest that our functional analysis can also identify pathogenic VUS in *CHEK2* that associate with prostate cancer.

**Table 1.** Complete list of human *CHEK2* variants analyzed in this study.

| Protein change | pKap1 (%) | Classification | Helix | Stability (%) | p.T383 phos. | Nr. cases | Nr. controls | Odds Ratio (95% CI), p-value (all studies) | Odds Ratio (95% CI), p-value (population-based studies) |
|---|---|---|---|---|---|---|---|---|---|
| p.A17S | 96,5 | Functional | 0,00 | n/a | n/a | n/a | n/a | n/a | n/a |
| p.H54= | 94,10 | Functional | n/a | 98,10 | n/a | n/a | n/a | n/a | n/a |
| p.E64K | 41,09 | Intermediate | 0,04 | 110,76 | Absent | 53 | 31 | 1,77 (1,16-2,69), p=0,008 | 1,78 (1,14-2,77), p=0,011 |
| p.W93Gfs | 14,31 | Damaging | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| p.R117G | 17,42 | Damaging | 0,86 | 64,72 | n/a | 47 | 22 | 2,93 (1,82-4,73), p<0,0001 | 2,22 (1,34-3,68), p=0,002 |
| p.F125S | 15,99 | Damaging | 0,77 | 41,03 | n/a | 0 | 1 | n/a | n/a |
| p.K135Nfs | 30,89 | Damaging | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| p.R137= | 94,09 | Functional | n/a | 96,42 | n/a | n/a | n/a | n/a | n/a |
| p.S140N | 15,90 | Damaging | 0,91 | n/a | Absent | 1 | 0 | n/a | n/a |
| p.K141T | 56,40 | Intermediate | 0,45 | 92,77 | Intermediate | 1 | 0 | n/a | n/a |
| p.R145W | 15,82 | Damaging | 0,72 | 38,83 | n/a | 10 | 9 | 1,96 (0,89-4,32), p=0,093 | 1,15 (0,47-2,84), p=0,756 |
| p.I157S | 86,39 | Functional | 0,53 | n/a | n/a | 1 | 0 | n/a | n/a |
| p.I157T | 95,55 | Functional | 0,36 | 111,22 | n/a | n/a | n/a | n/a | n/a |
| p.I160T | 15,95 | Damaging | 0,77 | 48,00 | n/a | 0 | 1 | n/a | n/a |
| p.D162G | 18,40 | Damaging | 0,93 | 51,69 | n/a | | | n/a | n/a |
| p.G167R | 20,76 | Damaging | 0,94 | 45,72 | n/a | 8 | 3 | 5,01 (1,47-17,10), p=0,010 | 2,77 (0,73-10,44), p=0,133 |
| p.F169Lfs | 17,11 | Damaging | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| p.F169L | 25,19 | Damaging | 0,49 | 44,35 | n/a | 1 | 2 | 3,09 (0,64-14,9), p=0,159 | 0,52 (0,05-5,73), p=0,593 |
| p.R180C | 84,08 | Functional | 0,27 | n/a | n/a | n/a | n/a | n/a | n/a |
| p.R181H | 82,93 | Functional | 0,04 | n/a | n/a | 33 | 22 | 1,16 (0,69-1,93), p=0,578 | 1,56 (0,91-2,67), p=0,108 |
| p.N186H | 96,50 | Functional | 0,30 | 108,21 | n/a | 17 | 14 | 1,59 (0,85-2,99), p=0,149 | 1,26 (0,62-2,56), p=0,5206 |
| p.V200A | 67,63 | Intermediate | 0,36 | n/a | n/a | 0 | 1 | n/a | n/a |
| p.D203G | 41,43 | Intermediate | 36,00 | 43,79 | Intermediate | 4 | 0 | n/a | n/a |
| p.G229S | 12,94 | Damaging | 0,96 | n/a | Absent | 0 | 1 | n/a | n/a |
| p.A230P | 12,90 | Damaging | 0,85 | n/a | Absent | n/a | n/a | n/a | n/a |
| p.E239K | 57,85 | Intermediate | 0,77 | 61,58 | n/a | 12 | 7 | 2,27 (0,95-5,44), p=0,065 | 1,78 (0,70-4,52), p=0,226 |
| p.C243R | 85,04 | Functional | 0,75 | n/a | n/a | 4 | 8 | 0,44 (0,13-1,47), p=0,183 | 0,52 (0,16-1,72), p=0,285 |
| p.A247D | 27,50 | Damaging | 0,97 | 36,99 | n/a | 1 | 0 | n/a | n/a |

**Table 1.** Continued

| Protein change | pKap1 (%) | Classification | Helix | Stability (%) | p.T383 phos. | Nr. cases | Nr. controls | Odds Ratio (95% CI), p-value (all studies) | Odds Ratio (95% CI), p-value (population-based studies) |
|---|---|---|---|---|---|---|---|---|---|
| p.K249R | 25,48 | Damaging | 0,91 | n/a | Absent | n/a | n/a | n/a | n/a |
| p.I251F | 15,11 | Damaging | 0,69 | n/a | Intermediate | 3 | 0 | n/a | n/a |
| p.K253X | 34,35 | Damaging | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| p.E273K | 22,77 | Damaging | 0,94 | 103,64 | Intermediate | n/a | n/a | n/a | n/a |
| p.I286= | 76,33 | Functional | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| p.G306E | 18,84 | Damaging | 0,89 | 87,77 | Intermediate | 1 | 0 | n/a | n/a |
| p.L326P | 25,49 | Damaging | 0,93 | 35,98 | n/a | 1 | 0 | n/a | n/a |
| p.R346H | 15,03 | Damaging | 0,90 | n/a | Absent | 4 | 2 | 2,65 (0,54-13,14), p=0,232 | 2,08 (0,38-11,34), p=0,399 |
| p.D347N | 13,23 | Damaging | 0,93 | n/a | Absent | 4 | 3 | 0,88 (0,22-3,54), p=0,861 | 1,38 (0,31-6,19), p=0,6701 |
| p.D347A | 29,31 | Damaging | 0,94 | n/a | n/a | n/a | n/a | n/a | n/a |
| p.E351D | 11,72 | Damaging | 0,71 | n/a | Normal | 7 | 2 | 4,42 (0,97-20,18), p=0,055 | 3,63 (0,76-17,50), p=0,108 |
| p.T367Mfs | 21,13 | Damaging | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| p.T367= | 107,13 | Functional | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| p.H371Y | 110,43 | Functional | 0,25 | n/a | n/a | 37 | 38 | 0,78 (0,52-1,17), p=0,225 | 1,01 (0,64-1,59), p=0,962 |
| p.G386R | 12,09 | Damaging | 0,97 | 107,07 | Absent | 1 | 0 | n/a | n/a |
| p.Y390C | 11,27 | Damaging | 0,96 | n/a | Intermediate | n/a | n/a | n/a | n/a |
| p.Y390S | 13,12 | Damaging | 0,96 | n/a | Intermediate | 2 | 3 | 0,88 (0,18-4,38), p=0,880 | 0,69 (0,12-4,14), p=0,687 |
| p.A392V | 12,72 | Damaging | 0,93 | 37,55 | n/a | 12 | 4 | 3,32 (1,10-9,99), p=0,033 | 3,12 (1,00-9,66), p=0,0491 |
| p.D409N | 13,24 | Damaging | 0,97 | 46,02 | n/a | 1 | 1 | 0,88 (0,06-14,14), p=0,931 | 1,04 (0,06-16,60), p=0,979 |
| p.S412R | 26,39 | Damaging | 0,97 | 40,53 | n/a | 1 | 0 | n/a | n/a |
| p.G414E | 22,86 | Damaging | 0,97 | 33,33 | n/a | 1 | 0 | n/a | n/a |
| p.S422Vfs | 11,54 | Damaging | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| p.P426R | 22,05 | Damaging | 0,96 | 37,67 | n/a | 1 | 0 | n/a | n/a |
| p.S435= | 92,62 | Functional | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| p.D438Y | 46,39 | Intermediate | 0,67 | 77,56 | Intermediate | 26 | 27 | 1,24 (0,76-2,04), p=0,385 | 1,00 (0,58-1,71), p=0,999 |
| p.N446D | 80,18 | Functional | 0,08 | n/a | n/a | 4 | 3 | 1,47 (0,35-6,17), p=0,596 | 1,38 (0,31-6,19), p=0,670 |
| p.I448S | 43,56 | Intermediate | 0,35 | 90,08 | Normal | 1 | 3 | 0,88 (0,18-4,38), p=0,880 | 0,35 (0,04-3,33), p=0,358 |
| p.R474H | 15,73 | Damaging | 0,96 | 34,92 | n/a | 11 | 0 | n/a | n/a |
| p.R474L | 27,86 | Damaging | 0,97 | n/a | n/a | n/a | n/a | n/a | n/a |
| p.R474= | 95,54 | Functional | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| p.A480T | 54,42 | Intermediate | 0,66 | n/a | Intermediate | 4 | 0 | n/a | n/a |
| p.W485G | 20,05 | Damaging | 0,97 | 33,98 | n/a | 0 | 1 | n/a | n/a |
| p.P509S | 94,60 | Functional | 0,01 | n/a | n/a | 21 | 22 | 0,88 (0,50-1,58), p=0,676 | 0,99 (0,55-1,81), p=0,977 |
| p.R519X | 59,34 | Intermediate | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| p.R521W | 42,51 | Intermediate | 0,53 | 83,94 | Intermediate | 9 | 2 | 2,48 (0,89-6,87), p=0,082 | 4,67 (1,01-21,63), p=0,049 |

All variants are indicated at the protein level in the protein change column, where missense variants are indicated in blue, synonymous variants in green and truncating variants in red. Nucleotide annotations for each variant are available in the published manuscript, where nucleotide numbering reflects Human Genome Variation Society (HGVS) nomenclature and cDNA number +1 corresponds to the A of the ATG translation initiation codon in the reference sequence (*CHEK2* NM_007194.4). The initiation codon is codon 1. For each variant, results for three functional readouts (i.e., Kap1 p.S473 phosphorylation, EGFP-CHK2 stability and CHK2 p.T383 phosphorylation), Helix-based predictions, population-based case-control frequencies and odds ratios are shown. Functional classification is based on the phospho-

Kap1 FACS assay (Fig. 2c) and population-based case-control frequencies and odds ratios are based on a study from the BRIDGES consortium in collaboration with the BCAC (7).

**Table 2.** Burden-type cancer risk association analysis for human *CHEK2* variants.

| Variant group based on function | Aa change | Cases | Controls | Odds Ratio (95% CI), p-value (population-based studies) | Odds Ratio (95% CI), p-value (all studies) |
|---|---|---|---|---|---|
| Functional VUS | p.I157S | 1 | 0 | 1,13 (0,87-1,46), p = 0,378 | 0,97 (0,76-1,23), p = 0,7943 |
| | p.R181H | 33 | 22 | | |
| | p.N186H | 17 | 14 | | |
| | p.V200A | 0 | 1 | | |
| | p.C243R | 4 | 8 | | |
| | p.H371Y | 37 | 38 | | |
| | p.N446D | 4 | 3 | | |
| | p.P509S | 21 | 22 | | |
| Intermediate VUS | p.E64K | 53 | 31 | 1,63 (1,21-2,20), p = 0,0014 | 1,79 (1,36-2,36), p < 0,0001 |
| | p.K141T | 1 | 0 | | |
| | p.D203G | 4 | 0 | | |
| | p.E239K | 12 | 7 | | |
| | p.D438Y | 26 | 27 | | |
| | p.I448S | 1 | 3 | | |
| | p.A480T | 4 | 0 | | |
| | p.R521W | 9 | 2 | | |
| Intermediate VUS (excl. p.E64K) | | | | 1,52 (1,01-2,28), p = 0,0448 | 1,81 (1,25-2,62), p = 0,0016 |
| Damaging VUS | p.R117G | 47 | 22 | 2,23 (1,62-3,07), p < 0,0001 | 3,03 (2,25-4,08), p < 0,0001 |
| | p.F125S | 0 | 1 | | |
| | p.S140N | 1 | 0 | | |
| | p.R145W | 10 | 9 | | |
| | p.I160T | 0 | 1 | | |
| | p.G167R | 8 | 3 | | |
| | p.F169L | 1 | 2 | | |
| | p.G229S | 0 | 1 | | |
| | p.A230P | n/a | n/a | | |
| | p.A247D | 1 | 0 | | |
| | p.K249R | n/a | n/a | | |
| | p.I251F | 3 | 0 | | |
| | p.E273K | n/a | n/a | | |
| | p.G306E | 1 | 0 | | |
| | p.L326P | 1 | 0 | | |
| | p.R346H | 4 | 2 | | |
| | p.D347N | 4 | 3 | | |
| | p.E351D | 7 | 2 | | |
| | p.G386R | 1 | 0 | | |
| | p.Y390C | n/a | n/a | | |
| | p.Y390S | 2 | 3 | | |
| | p.A392V | 12 | 4 | | |
| | p.D409N | 1 | 1 | | |
| | p.S412R | 1 | 0 | | |
| | p.G414E | 1 | 0 | | |
| | p.P426R | 1 | 0 | | |
| | p.R474H | 11 | 0 | | |
| | p.W485G | 0 | 1 | | |
| Damaging VUS (excl. p.R117G) | | | | 2,23 (1,48-3,38), p < 0,0001 | 3,09 (2,11-4,53), p < 0,0001 |

Variants with similar impact of CHK2 functionality were grouped (Fig. 2c). Only missense variants for which case-control frequencies from population- or family-based studies have been reported were included (7). The case-control frequencies reflect those of the population-based studies alone. The analysis was also performed for groups of *CHEK2* variants without p.E64K or p.R117G, for which the carrier frequencies are high.

**Table 3.** No LOH in *CHEK2* c.485A>G/p.D162G or c.1100delC/p.T367Mfs carriers.

| *CHEK2* variant carriers | Tissue type | VAF c.485A>G | VAF c.1100delC |
|---|---|---|---|
| c.485A>G/p.D162G carrier 1 (brother 1) | Tumor tissue | 0,521 | |
| | Control tissue | 0,482 | |
| c.485A>G/p.D162G carrier 2 (brother 2) | Tumor tissue | 0,526 | |
| | Control tissue | 0,476 | |
| c.1100delC/p.T367Mfs carrier 1 | Tumor tissue | | 0,538 |
| | Control tissue | | 0,485 |
| c.1100delC/p.T367Mfs carrier 2 | Tumor tissue | | 0,466 |
| | Control tissue | | N/A |

VAF refers to variant allele frequency.

## DISCUSSION

We developed a mES cell-based system that allows for the quantitative functional classification of genetic variants in the *CHEK2* gene that associate with breast and prostate cancer. Of the 50 *CHEK2* missense VUS tested in this study, 9 variants (18%) had an intermediate impact on CHK2 function, while 31 (62%) were damaging (Table 1, Fig. 3c). Importantly, 23 *CHEK2* missense VUS constitute variants that have, to our knowledge, not been functionally characterized in previous studies (29-35,60). At least 18 of the intermediate and damaging VUS in our study (>50%) exhibited defects in protein stability (Fig. 5a), which is a common pathogenic mechanism originating from missense variants (38,61). Moreover, at least 11 VUS (22%) showed reduced or complete lack of autophosphorylation on p.T383 (Fig. 5c), explaining the impaired kinase activity for most of these VUS (19,62). For 5 damaging VUS (i.e., p.I251F, p.E273K, p.E351D, p.Y390C and p.Y390S) considerable levels of autophosphorylation were observed, while kinase activity towards Kap1 was lacking. As these VUS mostly localize to the ATP-binding pocket of CHK2, they likely impair the ability of CHK2 to bind or hydrolyze ATP, the latter of which has already been reported for p.E273K (55,56). Thus, we examined numerous *CHEK2* missense VUS for which we quantified functional effects (i.e., kinase activity) and assessed pathogenic mechanisms of action. Correlation between our quantitative results and breast cancer risk further demonstrated that our functional assay can identify pathogenic missense variants in *CHEK2*.

Our results are generally in line with two recent studies describing functional analysis of *CHEK2* missense variants (34,35). Kleiblova et al. employed both an in vitro kinase assay

and a RPE1 *CHEK2*$^{KO}$ cell-based system for functional classification of *CHEK2* variants (35). For most overlapping variants, our results are consistent with their functional assessment (Supplemental Fig. S7a). Although further research is required to explain the differences observed for three variants (i.e., p.I157T, p.R346H and p.D438Y), differences in the functional classification of p.E64K may be explained by its kinetic effect on Kap1 phosphorylation (Fig. 3g, Supplemental Fig. S7a). That is, we based our 'intermediate' functional classification on phospho-Kap1 p.S473 levels observed at 2 hours after IR, whereas Kleiblova et al. based their 'damaging' classification on the KAP1 phosphorylation levels observed at 4 hours after IR in the RPE1 cell-based assay. On the other hand, Delimitsou et al. employed a yeast *rad53* mutant cell-based system for functional characterization of human *CHEK2* variants (34), whose results were also highly consistent with those from our study (Supplemental Fig. S7b). However, all *CHEK2* variants (with the exception of p.E64K) that we classified as intermediate and Delimitsou et al. as neutral (Supplemental Fig. S7b), are variants that impaired protein stability in our assays (i.e., p.D203G, p.E239K, p.D438Y and p.R521W) (Fig. 5a). Possibly, several intermediate effects are not picked up efficiently in the yeast assays as yeast cells grow at 30°C rather than at 37°C, which may reduce the thermodynamic instability of proteins. Thus, while the outcome of the different functional analysis of *CHEK2* variants are generally consistent, discrepancies for some variants remain, complicating their classification and calling for further analysis.

The Helix algorithm predicted functionality of *CHEK2* missense variants more accurately than several other algorithms did (Fig. 4b). Therefore, the en masse Helix predictions (Fig. 4c, Supplementary Table S1) may aid in the classification of missense variants in *CHEK2* for which functional outcomes were inconsistent (e.g., p.L174V) (35), or for which functional analysis have yet to be performed. In support of the remarkable performance of Helix, in both our study and that of Delimitsou et al. (34), no variants predicted to be benign by Helix were found to be damaging (Fig. 4b). Although computational predictions should be handled with care, discrepancies with Helix may also highlight variants that require further validation of their functional impact, thereby aiding in the classification of *CHEK2* variants.

The BRIDGES consortium in collaboration with the BCAC, showed that rare *CHEK2* missense VUS in aggregate associate with a low, yet significant risk for breast cancer (OR 1.42; 95% CI, 1.28 to 1.58; p<0,0001) (7). However, a major challenge is to discriminate which VUS associate with cancer risk and which do not. Our study addressed this challenge and showed that the degree of CHK2 dysfunction, for numerous *CHEK2* missense VUS, correlates with increased breast cancer risk (Table 1, Table 2, Fig. 3c). Furthermore, the OR for the damaging *CHEK2* VUS in aggregate (OR 2,23; 95% CI 1,48-3,38; $p$<0.0001), as well as that for the damaging VUS c.349A>G/p.R117G alone (OR 2,22; 95% CI 1,34-3,68; $p$=0,0020), compared well to the population-based ORs for c.1100delC/p.T367Mfs (OR 2,66; 95% CI 2,27-

3,11; *p*<0,0001) and that of all other *CHEK2* truncating variants in aggregate (OR 2,13; 95% CI 1,60-2,84; *p*<0,0001) (6,7). The OR for the intermediate *CHEK2* variant c.190G>A/p.E64K (OR 1,78; 95% CI 1,14-2,77; *p*=0.0112) associated with significantly increased breast cancer risk comparable to that calculated for its functional classification group (OR 1,52; 95% CI 1.01-2,28; *p*=0.0448). These results strongly suggest that intermediate *CHEK2* VUS associate with significantly increased breast cancer risk and that damaging *CHEK2* VUS likely associate with a similar risk for breast cancer as truncating *CHEK2* variants.

*E*ffects of *CHEK2* variants on splicing could not be examined since we employed human *CHEK2* cDNA-based complementation assays. However, in silico splice site prediction analysis was performed using four different algorithms (Splice Site Finder-like, MaxEntScan, GeneSplicer, NNSplice) in Alamut (http://www.interactivebiosoftware.com/). For most VUS, an effect on RNA splicing was unlikely, except for five variants (p.A17S, p.I157S, p.I160T, p.D162G, p.F169L, p.G229S and p.A230P) for which these algorithms predicted the introduction of weak acceptor or donor recognition sites in the corresponding exons (Supplementary Table S2). Consistently, the recently developed deep learning-based SpliceAI tool (63) predicted no major splice effects for the *CHEK2* missense VUS examined in this study, except for (i.e., p.V200A and p.G229S) for which the loss or introduction of a splice acceptor site was predicted with low to moderate confidence (Supplementary Table S2). The path to clinical implementation of functional analysis, in line with ACMG guidelines (64), involves having a well-calibrated assay. Even though we note that the slight difference in homology between mouse and human *CHEK2* (82% identical and 88% similar in protein sequence) could affect the functional analysis presented in this study, we believe that our quantitative data and the correlation with breast cancer risk supports the robustness and validity of our functional assay for *CHEK2*, and thus its value as clinical diagnostic tool.



## MATERIALS AND METHODS

### Cell lines and cell culture conditions

129/Ola E14 IB10 mES cells (37) were cultured on gelatine-coated dishes in 50% 2i ES medium of which 500 ml contains 1) 250 mL Knockout Dulbecco's modified Eagle's medium (Gibco, 21710-025) supplemented with 2,5 ml 100 mM sodium pyruvate (Gibco 11360-039), 2,5 ml 100x non-essential amino acids (Gibco 11140-035) and 25 ml Fetal Calf Serum (FCS); 2) 125 ml DMEM/F2 HEPES supplemented with 1,25 ml 100x N2 Supplement (Gibco 17502-048), 85 µl 7.5% BSA (Gibco # 15260-037) and 500 µL 0,1M β-MeOH; and 3) 125 ml NEUROBASAL medium (Gibco, 21103-049) supplemented with 2,5 mL 50x B27 Supplement (Gibco # 17504-044), 1,25 ml 200 mM L-glutamine (Gibco 25030-024) and 500 µL 0,1M β-

MeOH. The total 500 ml is supplemented with 5 ml 5000 units/ml penicillin/streptomycin (Gibco 15070063), 5 ml $10^5$ units/ml LIF (Millipore ESG1107), 250 µL 0,1M β-MeOH, 250 µL 3mM CHIR (Axon Medchem 1386) and 250 µL 1 mM PD (Axon Medchem 1408).

**Generation of *Chek2*$^{KO}$ mES cells with DR-GFP and RMCE**

mES cells carrying the DR-GFP reporter and RMCE system at the *Pim1* and *Rosa26* locus, respectively, were generated previously (38). Using these mES cells, *Chek2*$^{KO}$ cells were generated by transfecting1 µg of pSpCas9(BB)-2A-GFP (*pX458*) (39), encoding Cas9, GFP and a gRNA that targets exon 3 of mouse *Chek2* (5'-ACTGTGTTAACGACAACTAC-3'). GFP-positive cells were FACS-sorted and seeded. Individual clones were examined by TIDE (https://tide.nki.nl) and western blot analysis for loss of Chk2 expression.

**Selection of human *CHEK2* variants**

Seven previously reported *CHEK2* truncating variants were included as negative controls (16,40). Six synonymous variants, which have not yet been observed in carriers were selected based on their position throughout the *CHEK2* protein and were included as positive controls. Truncating and missense *CHEK2* VUS were selected based on one or more of the following criteria: 1) identification in the case-control association study performed by the BRIDGES consortium in collaboration with the BCAC (7) or prostate cancer family members reported in this study, 2) clinical classification in ClinVar (16), 3) position in the CHK2 protein sequence, 4) computational predictions from Helix and 5) presence/absence in previous functional studies (34,35).

**Cloning and generation of human *CHEK2* variants**

Vector pBudCE4.1 (ThermoFisher, V53220) was modified by adding two *Pac*I restriction sites as previously described (38). Human HA-tagged *CHEK2* cDNA (NM_007194.4) was subcloned from pBabe-HA-CHK2 (41) using the *Bsr*GI and *Xho*I restriction sites into pBudCE4.1-*Pac*I using the *Bsr*GI-compatible *Acc*65I restriction site and *Xho*I restriction site. pBabe-HA-CHK2 was a gift from Stephen Elledge (Addgene plasmid #41901). An Ef1α-*CHEK2*-containing fragment from pBudCE4.1-*Pac*I-CHEK2 was then cloned into the RMCE vector (pRNA 251-MCS RMCE) (TaconicArtemis GmbH) using the *Pac*I restriction sites in both vectors. *CHEK2* variants were introduced by site-directed mutagenesis (SDM) using the Quick-Change Lightning protocol (Agilent Technologies). All SDM primers are shown in Supplementary Table S3. Constructs were verified by Sanger sequencing and used for mES cell-based assays.

The RMCE vector carrying EGFP-CHEK2-T2A-mCherry was generated as follows. The RMCE vector carrying *CHEK2* was digested with *Eco*RI. EGFP was PCR amplified from

an EGFP-carrying construct (pcDNA-FRT-TO-puro-EGFP) with the following primers: forward primer 5'-CCCAGTGTGGTGGTACGTAGATGGTGAGCAAGGGCGAGG-3' and reverse primer 5'-TATGGGTAAGCCATGAATTCCTTGTACAGCTCGTCCATGCCG-3'. Gibson assembly was then performed to generate the RMCE vector carrying EGFP-CHEK2. Next, three different fragments were PCR amplified: *CHEK2* (forward primer 5'-ACGAGCTGTACAAGGAATTCATGTCTCGGGAGTCGGATGT-3' and reverse primer 5'-AGCAGACTTCCTCTGCCCTCCAACACAGCAGCACACACAGC-3') and the hGH sequence (forward primer 5'-TGGACGAGCTGTACAAGTGAACTCCGTGGTTTGAACACTCTAG-3' and reverse primer 5'-GCATAACTAGTGTCACGCGTCATATGGCCGGCCTATTTAAATAAGC-3') from the RMCE vector carrying EGFP-CHEK2, and T2A-mCherry (forward primer 5'-GAGGGCAGAGGAAGTCTGCTAAC-3' and reverse primer 5'-TCACTTGTACAGCTCGTCCATGC-3') from a T2A-mCherry carrying construct (pX459-Cas9-T2A-mCherry). The RMCE vector carrying EGFP-CHEK2 was then digested with *Eco*RI and *Mlu*I after which the plasmid backbone (lacking *CHEK2*) was gel extracted. By employing Gibson assembly, the three PCR fragments were cloned into the backbone to generate the RMCE vector carrying EGFP-CHEK2-T2A-mCherry. The construct was verified by Sanger sequencing and used to generate *CHEK2* variants and perform mES cell-based assays.



**Western blot analysis**

$2x10^6$ *Chek2*$^{KO}$ mES cells carrying the DR-GFP reporter and RMCE system were subjected to RMCE by co-transfecting 1 µg FlpO expression vector (pCAGGs-FlpO-IRES-puro) (42) with 1 µg RMCE exchange vector. Neomycin-resistant cells from ~500 resistant clones were pooled and expanded as previously described (38). For various conditions, protein levels for mouse Chk2, human CHK2, human phospho-CHK2 p.T383, mouse Kap1, mouse phospho-Kap1 p.S473, mouse p53, mouse p21 and mouse tubulin were examined by protein extraction and western blot analysis. Briefly, samples were generated by taking up ~$1,5x10^6$ cells in 75 µl Laemmli buffer and boiling them at 95°C for 5 minutes. Samples were incubated with 0,2 µl benzonase (Merck Millipore 70746, 250 U/µl) for 20 minutes at room temperature and then loaded for gel electrophoresis followed by immunoblotting. Primary antibodies used were: mouse monoclonal antibody against mouse/human CHK2 (1:1000; BD Biosciences 611571), rabbit polyclonal antibody against mouse/human phospho-CHK2 p.T383 (1:1500; Abcam 59408), rabbit polyclonal antibody against mouse/human Kap1 (1:10000; Abcam 10484), mouse monoclonal antibody against mouse/human phospho-Kap1 p.S473 (1:2000; Biolegend 654102), mouse monoclonal antibody against mouse/human p53 (1:1000; Cell Signaling 2524), rabbit polyclonal antibody against mouse/human p21 *(Cdkn1a)* (1:800; Santa Cruz sc-397) and mouse monoclonal antibody against $\alpha$-tubulin (1:5000, Sigma, T6199). Peroxidase-AffiniPure goat polyclonal anti-rabbit (1:5000; Jackson laboratories 111-035-003) and affinity

isolated goat polyclonal anti-mouse (1:5000; Dako P0447) were used as secondary antibodies. SuperSignal West Femto Maximum Sensitivity Substrate (ThermoFisher Scientific 34095) and ECL Prime Western Blotting Detection Reagents (Merck RPN2232) were used for development of blots on the Amersham Imager 600 (GE Healthcare Life Sciences).

**HR Reporter Assays**

1-2x$10^6$ *Chek2*$^{KO}$ mES cells carrying the DR-GFP reporter and RMCE system were subjected to HR assays by transfecting 1 µg of plasmid that co-expresses I-*Sce*I and mCherry (pCMV-Red-Isce, kind gift from Jos Jonkers) using Lipofectamine 2000 (ThermoFisher) (43). A co-transfection of 1 µg pCAGGs (44) with 0,05 µg of an mCherry expression vector was included as control. Two days after transfection, mCherry/GFP double-positive cells were scored using a Novocyte Flow Cytometer (ACEA Biosciences, Inc.).

**Phleomycin sensitivity assays**

For proliferation-based phleomycin sensitivity assays, mES cells were seeded in triplicate at 10000 cells per well of a 96-well plate. The next day, cells were treated with phleomycin (InvivoGen ant-ph-2p) for two days, after which the medium was refreshed, and cells were cultured for one more day in drug-free medium. Viable cells were subsequently counted using the Novocyte Flow Cytometer (ACEA Biosciences, Inc.).

**RT-qPCR analysis**

RNA was isolated from mES cells grown on 6-well plates using Trizol (ThermoFisher Scientific 15596026) as per the manufacturer's protocol. For each condition, 3 µg RNA was treated with RQ1 RNase-free DNase (Promega M6101) and cDNA was synthesized from 0,2 µg DNase-treated RNA using hexamer primers (ThermoFisher Scientific N8080127) and SuperScript™ IV Reverse Transcriptase (ThermoFisher Scientific 18090050) as per the manufacturer's protocols. RT-qPCRs were carried out using GoTaq qPCR Master mix (Promega A6002), a CFX384 Real-Time System (Bio-Rad) and the following qPCR primers directed at the mouse *Mdm2*, *p21 (Cdkn1a),* or the mouse control gene *Pim1*: Mdm2-exon11-Fw 5'-GTCTATCAGACAGGAGAAAGCGATACAG-3', Mdm2-exon12-Rv 5'-GTCCAGCATCTTTTGCAGTGTGATGGAAG-3'. p21-exon2-Fw 5'-GCTGTCTTGCACTCTGGTGTCTGAG-3', p21-exon3-Rv 5'-GACCAATCTGCGCTTGGAGTGATAG-3'. Pim1-exon4-Fw 5'-GCGGCGAAATCAAACTCATCGAC-3' and Pim1-exon5-Rv 5'-GTAGCGATGGTAGCGAATCCACTCTGG-3'.

**Flow cytometry (FACS) analysis**

As for western blot analysis, *Chek2*$^{KO}$ mES cells expressing human *CHEK2* variants were generated and expanded. For phospho-Kap1 p.S473 FACS-based assays, 1x10$^6$ mES cells were seeded on 60 mm dishes one day prior to exposure to 10 Gy of IR. Two, four or six hours after IR, cells were trypsinized and fixed in 5 ml 2% formaldehyde for 15 minutes. A volume of 2 ml 0,125 M glycine was added and cells were centrifuged for 5 minutes at 1500 rpm. Cells were then washed in PBS and fixed for a second time in 100% ice-cold methanol and incubated overnight at -20°C. After washing once in PBS, fixed cells were permeabilized for 15 minutes using 0,25% Triton X-100 in PBS, after which cells were stained in 200 µl PBS$^+$ (5 g/l BSA, 1,5 g/l glycine) with 1 µl mouse anti-phospho-Kap1 p.S473 (0,5 µg/µl, Biolegend 654102) for 3 hours at room temperature, with gentle resuspension every 30 minutes. Alexa-488 goat anti-mouse (1:200 in 200 µl PBS; ThermoFisher Scientific A-21424) was used as a secondary antibody, followed by a propidium iodide staining (25 µg/ml PI, RNaseA 0,1 mg/ml, 0.05% Triton X-100). Phospho-Kap1 p.S473 intensity was analysed using the Novocyte Flow Cytometer (ACEA Biosciences, Inc.). For FACS-based assays with mES cells expressing EGFP-CHEK2-T2A-mCherry, phospho-Kap1 p.S473 was stained with alexa-647 goat anti-mouse (1:200 in 200 µl PBS; ThermoFisher Scientific A-21235), propidium iodide staining was not performed and Phospho-Kap1 p.S473 intensity was measured after gating for mCherry- or GFP-positive cells using a Fortessa1 (BD Biosciences).



**Exome sequencing in prostate cancer family members**

Three brothers were diagnosed with prostate cancer >10 years earlier than the average age of onset of sporadic prostate cancer, suggesting that they might be carriers of a germline mutation responsible for predisposition to this type of cancer. Copy-number variations (deletions or amplifications) in blood cell DNA from these four brothers and their sons were not detected using the SNP6 microarray (Affymetrix). The Agilent SureSelect Human All Exon V5+UTRs protocol was used to carry out targeted enrichment of all exonic sequences from the total DNA material for each sample. Paired-end Illumina sequencing with 100 cycles was performed to minimize the ambiguities of read alignment to the reference genome. Two sequencing lanes resulted in an average of 20 million fragments per sample. All sequence fragments were aligned to the reference human genome (version hg19) using BWA mem (v. 0.7.10), after quality and TruSeq adapter trimming using Cutadapt (v.1.5). Sam files were manipulated using Samtools (v.1.1) and Picard tools (v.1.119) were used to run quality metrics (insert size, hybridization quality) and mark PCR duplicates. VerifyBamID (v. 1.1) was used to

estimate contamination. Samples were genotyped and variants jointly called using GATK (v. 3.5). For this purpose, padded targeted intervals were created based on Agilent targets. Annotation was performed using wAnnovar, Oncotator (v.1.8) and WGSA (Amazon EC2 cloud, AWS community instance: WGSA055-ubuntu-800G). Transcript annotation was taken from the Oncotator pipeline using the transcript list giving priority to known clinical protein changes (list downloaded in Feb 2016). GENCODE (Version 19 - July 2013 freeze, GRCh37 - Ensembl 74) was used as a reference transcript set. Unfiltered variants were jointly called over all samples. Filtering was performed based on genotyping quality. All variants that did not have a minimum read depth of 8 and genotype quality of 20 in all affected family members were removed. Finally, all variants with MAF >1% (based on ExAc European non-Finnish cohort, annotation from WGSA) were excluded. Variants classified as pathogenic by ClinVar were not discarded even if MAF was >1%. Analysis of the remaining variants showed that all three affected brothers, as well as one of their sons, carried the *CHEK2* allele rs587781652 harbouring the c.485A>G/p.D162G VUS.

**LOH assessment**

Tumor DNA was isolated from formalin-fixed paraffin-embedded (FFPE) tissue blocks either by taking three 0.6 mm tumor cores or by microdissection of tumor areas with at least 70% tumor cells (10 mm slides). Fully automated DNA isolation was performed using the Tissue Preparation System (Siemens Healthcare Diagnostics) as described previously (45). The Qubit dsDNA HS Assay Kit was used for DNA quantification according to the manufacturer's protocol (Qubit 2.0 Fluorometer, Invitrogen, Waltham, MA USA, cat. Q32851). Next-generation sequencing (NGS) was performed using 40 ng of tumor DNA per sample isolated from FFPE tissue blocks. The custom Ampliseq HDR15v1-panel (Thermo Fisher) was used for variant detection in *CHEK2*. LOH of *CHEK2* was determined by comparing the variant allele frequency (VAF) of heterozygous c.485A>G/p.D162G and c.1100delC/p.T367Mfs in tumor and normal tissue as described previously (45). LOH was considered present when the tumor cell percentage was >20% and the germline *CHEK2* variant allele frequency was >0.6. LOH was considered inconclusive when the tumor cell percentage was <20% or considered absent when the germline *CHEK2* variant VAF was <0.6.

**Ethics declaration**

Individuals of the prostate cancer family were identified and evaluated at the University Hospital Zurich. The study protocol was approved by the hospital's research ethics committee, and donors provided written consent to tissue collection, testing, and data publication. LOH assessment was performed at Leiden University Medical Center under protocols approved by hospital's local ethics committee.

**ACKNOWLEDGEMENTS**

The authors would like to thank Jos Jonkers and Peter Bouwman for providing the pTT5-Puro (RMCE acceptor cassette), pRNA-251-MCS-RMCE (RMCE exchange vector) and pCMV-Red-I-SceI constructs, Maria Jasin and Francis Stewart for sharing the DR-GFP reporter and FlpO constructs, respectively, Doug Easton, Jamie Allen, Leila Dorling and the BRIDGES consortium for providing *CHEK2* variants and case-control frequencies for OR calculations, Emanuele Valtorta and the Flowcytometry Core Facility of the LUMC for technical assistance, and Josef Jiricny for fruitful discussions. This project has received funding from the Giuliana and Giorgio Stefanini Foundation (G.M.), the European Union's Horizon 2020 research and innovation program BRIDGES under grant agreement 634935 (P.D., M.V., and H.v.A.), and the Dutch Cancer Society (KWF-7473; P.D. and H.v.A.).



**AUTHOR CONTRIBUTIONS**

R.B. generated *CHEK2* cDNA-containing RMCE constructs, introduced *CHEK2* variants therein using SDM, generated *Chek2*$^{KO}$ mES cells harboring the DR-GFP reporter and RMCE acceptor cassette, performed DR-GFP and FACS-based assays, as well as western blot analysis, PCR and DNA sequencing analysis and performed in silico splice predictions. W.W. performed FACS-based assays. N.C. generated and validated *Chek2*$^{KO}$ mES cells. B.V. performed in silico modeling and S.H. performed in silico predictions of *CHEK2* variants. Z.K-J., S.C., M.M., N.B. and R.E. performed NGS data analysis for the prostate cancer-affected family and performed bioinformatics analysis of the acquired data. G.M. characterized the prostate cancer family and exome sequenced blood DNA from its family members. N.S-W. and T.v.W. performed the LOH analysis. M.V., P.D., and H.v.A. conceived the project. H.v.A. supervised the project. R.B and H.v.A. wrote the paper.

## REFERENCES


1. Ahn J, Urist M, Prives C. The Chk2 protein kinase. DNA Repair (Amst) **2004**;3(8-9):1039-47 doi 10.1016/j.dnarep.2004.03.033.
2. Bartek J, Lukas J. Chk1 and Chk2 kinases in checkpoint control and cancer. Cancer cell **2003**;3(5):421-9 doi 10.1016/s1535-6108(03)00110-7.
3. Hirao A, Cheung A, Duncan G, Girard PM, Elia AJ, Wakeham A*, et al.* Chk2 is a tumor suppressor that regulates apoptosis in both an ataxia telangiectasia mutated (ATM)-dependent and an ATM-independent manner. Mol Cell Biol **2002**;22(18):6521-32 doi 10.1128/mcb.22.18.6521-6532.2002.
4. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R*, et al.* Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet **2002**;31(1):55-9 doi 10.1038/ng879.
5. Vahteristo P, Bartkova J, Eerola H, Syrjakoski K, Ojala S, Kilpivaara O*, et al.* A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. American journal of human genetics **2002**;71(2):432-8 doi 10.1086/341943.
6. Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J*, et al.* Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. JAMA Oncol **2017**;3(9):1190-6 doi 10.1001/jamaoncol.2017.0424.
7. Breast Cancer Association C, Dorling L, Carvalho S, Allen J, Gonzalez-Neira A, Luccarini C*, et al.* Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. The New England journal of medicine **2021** doi 10.1056/NEJMoa1913948.
8. Cybulski C, Gorski B, Huzarski T, Masojc B, Mierzejewski M, Debniak T*, et al.* CHEK2 is a multiorgan cancer susceptibility gene. American journal of human genetics **2004**;75(6):1131-5 doi 10.1086/426403.
9. Cybulski C, Huzarski T, Gorski B, Masojc B, Mierzejewski M, Debniak T*, et al.* A novel founder CHEK2 mutation is associated with increased prostate cancer risk. Cancer research **2004**;64(8):2677-9 doi 10.1158/0008-5472.can-04-0341.
10. 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 doi 10.1086/346094.
11. Le Calvez-Kelm F, Lesueur F, Damiola F, Vallee M, Voegele C, Babikyan D*, et al.* 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 Res **2011**;13(1):R6 doi 10.1186/bcr2810.
12. Dufault MR, Betz B, Wappenschmidt B, Hofmann W, Bandick K, Golla A*, et al.* Limited relevance of the CHEK2 gene in hereditary breast cancer. International journal of cancer **2004**;110(3):320-5 doi 10.1002/ijc.20073.

13. Ingvarsson S, Sigbjornsdottir BI, Huiping C, Hafsteinsdottir SH, Ragnarsson G, Barkardottir RB*, et al.* Mutation analysis of the CHK2 gene in breast carcinoma and other cancers. Breast Cancer Res **2002**;4(3):R4 doi 10.1186/bcr435.
14. Schutte M, Seal S, Barfoot R, Meijers-Heijboer H, Wasielewski M, Evans DG*, et al.* Variants in CHEK2 other than 1100delC do not make a major contribution to breast cancer susceptibility. American journal of human genetics **2003**;72(4):1023-8 doi 10.1086/373965.
15. Sodha N, Bullock S, Taylor R, Mitchell G, Guertl-Lackner B, Williams RD*, et al.* CHEK2 variants in susceptibility to breast cancer and evidence of retention of the wild type allele in tumours. British journal of cancer **2002**;87(12):1445-8 doi 10.1038/sj.bjc.6600637.
16. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM*, et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic acids research **2014**;42(Database issue):D980-5 doi 10.1093/nar/gkt1113.
17. Stolarova L, Kleiblova P, Janatova M, Soukupova J, Zemankova P, Macurek L*, et al.* CHEK2 Germline Variants in Cancer Predisposition: Stalemate Rather than Checkmate. Cells **2020**;9(12) doi 10.3390/cells9122675.
18. Guo X, Ward MD, Tiedebohl JB, Oden YM, Nyalwidhe JO, Semmes OJ. Interdependent phosphorylation within the kinase domain T-loop Regulates CHK2 activity. J Biol Chem **2010**;285(43):33348-57 doi 10.1074/jbc.M110.149609.
19. 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.
20. Ahn JY, Li X, Davis HL, Canman CE. Phosphorylation of threonine 68 promotes oligomerization and autophosphorylation of the Chk2 protein kinase via the forkhead-associated domain. J Biol Chem **2002**;277(22):19389-95 doi 10.1074/jbc.M200822200.
21. Kastan MB, Bartek J. Cell-cycle checkpoints and cancer. Nature **2004**;432(7015):316-23 doi 10.1038/nature03097.
22. 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 doi 10.1016/s1097-2765(02)00527-0.
23. Zhang J, Willers H, Feng Z, Ghosh JC, Kim S, Weaver DT*, et al.* Chk2 phosphorylation of BRCA1 regulates DNA double-strand break repair. Mol Cell Biol **2004**;24(2):708-18 doi 10.1128/mcb.24.2.708-718.2004.
24. Hu C, Zhang S, Gao X, Gao X, Xu X, Lv Y*, et al.* Roles of Kruppel-associated Box (KRAB)-associated Co-repressor KAP1 Ser-473 Phosphorylation in DNA Damage Response. J Biol Chem **2012**;287(23):18937-52 doi 10.1074/jbc.M111.313262.
25. Cann KL, Dellaire G. Heterochromatin and the DNA damage response: the need to relax. Biochem Cell Biol **2011**;89(1):45-60 doi 10.1139/O10-113.
26. Czerwinska P, Mazurek S, Wiznerowicz M. The complexity of TRIM28 contribution to cancer. J Biomed Sci **2017**;24(1):63 doi 10.1186/s12929-017-0374-4.
27. Bolderson E, Savage KI, Mahen R, Pisupati V, Graham ME, Richard DJ*, et al.* Kruppel-associated Box (KRAB)-associated co-repressor (KAP-1) Ser-473 phosphorylation regulates

heterochromatin protein 1beta (HP1-beta) mobilization and DNA repair in heterochromatin. J Biol Chem **2012**;287(33):28122-31 doi 10.1074/jbc.M112.368381.

28. Lemaitre C, Soutoglou E. Double strand break (DSB) repair in heterochromatin and heterochromatin proteins in DSB repair. DNA Repair (Amst) **2014**;19:163-8 doi 10.1016/j.dnarep.2014.03.015.
29. 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. International journal of cancer **2007**;121(12):2661-7 doi 10.1002/ijc.23026.
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 research **2001**;61(22):8062-7.
31. 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 doi 10.1093/hmg/dds101.
32. Tischkowitz MD, Yilmaz A, Chen LQ, Karyadi DM, Novak D, Kirchhoff 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 doi 10.1016/j.canlet.2008.05.006.
33. 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 doi 10.1038/onc.2014.443.
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 doi 10.1002/humu.23728.
35. 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. International journal of cancer **2019**;145(7):1782-97 doi 10.1002/ijc.32385.
36. Cuella-Martin R, Hayward SB, Fan X, Chen X, Huang JW, Taglialatela A*, et al.* Functional interrogation of DNA damage response variants with base editing screens. Cell **2021**;184(4):1081-97 e19 doi 10.1016/j.cell.2021.01.041.
37. Robanus-Maandag E, Dekker M, van der Valk M, Carrozza ML, Jeanny JC, Dannenberg JH*, et al.* p107 is a suppressor of retinoblastoma development in pRb-deficient mice. Genes & development **1998**;12(11):1599-609.
38. Boonen R, Rodrigue A, Stoepker C, Wiegant WW, Vroling B, Sharma M*, et al.* Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. Nature communications **2019**;10(1):5296 doi 10.1038/s41467-019-13194-2.
39. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc **2013**;8(11):2281-308 doi 10.1038/nprot.2013.143.
40. Susswein LR, Marshall ML, Nusbaum R, Vogel Postula KJ, Weissman SM, Yackowski L*, et al.* Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. Genet Med **2016**;18(8):823-32 doi 10.1038/gim.2015.166.

41. Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER, 3rd, Hurov KE, Luo J*, et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science **2007**;316(5828):1160-6 doi 10.1126/science.1140321.
42. Kranz A, Fu J, Duerschke K, Weidlich S, Naumann R, Stewart AF*, et al.* An improved Flp deleter mouse in C57Bl/6 based on Flpo recombinase. Genesis **2010**;48(8):512-20 doi 10.1002/dvg.20641.
43. Bouwman P, van der Gulden H, van der Heijden I, Drost R, Klijn CN, Prasetyanti P*, et al.* A high-throughput functional complementation assay for classification of BRCA1 missense variants. Cancer discovery **2013**;3(10):1142-55 doi 10.1158/2159-8290.CD-13-0094.
44. Niwa H, Yamamura K, Miyazaki J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene **1991**;108(2):193-9.
45. de Jonge MM, Auguste A, van Wijk LM, Schouten PC, Meijers M, Ter Haar NT*, et al.* Frequent Homologous Recombination Deficiency in High-grade Endometrial Carcinomas. Clinical cancer research : an official journal of the American Association for Cancer Research **2019**;25(3):1087-97 doi 10.1158/1078-0432.CCR-18-1443.
46. Parameswaran B, Chiang HC, Lu Y, Coates J, Deng CX, Baer R*, et al.* Damage-induced BRCA1 phosphorylation by Chk2 contributes to the timing of end resection. Cell Cycle **2015**;14(3):437-48 doi 10.4161/15384101.2014.972901.
47. Takai H, Naka K, Okada Y, Watanabe M, Harada N, Saito S*, et al.* Chk2-deficient mice exhibit radioresistance and defective p53-mediated transcription. EMBO J **2002**;21(19):5195-205 doi 10.1093/emboj/cdf506.
48. 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 doi 10.1126/science.287.5459.1824.
49. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T*, et al.* Analysis of protein-coding genetic variation in 60,706 humans. Nature **2016**;536(7616):285-91 doi 10.1038/nature19057.
50. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q*, et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv **2019**:531210 doi 10.1101/531210.
51. 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 doi 10.1038/gim.2015.30.
52. Vroling B, Heijl S. White paper: The Helix Pathogenicity Prediction Platform (https://arxiv.org/abs/2104.01033). **2021**.
53. Heijl S, Vroling B, Bergh Tvd, Joosten H. Mind the gap: preventing circularity in missense variant prediction (https://doi.org/10.1101/2020.05.06.080424). **2020**.

54. Li J, Taylor IA, Lloyd J, Clapperton JA, Howell S, MacMillan D*, et al.* Chk2 oligomerization studied by phosphopeptide ligation: implications for regulation and phosphodependent interactions. J Biol Chem **2008**;283(51):36019-30 doi 10.1074/jbc.M804075200.

55. Cai Z, Chehab NH, Pavletich NP. Structure and activation mechanism of the CHK2 DNA damage checkpoint kinase. Mol Cell **2009**;35(6):818-29 doi 10.1016/j.molcel.2009.09.007.

56. Lountos GT, Tropea JE, Zhang D, Jobson AG, Pommier Y, Shoemaker RH*, et al.* Crystal structure of checkpoint kinase 2 in complex with NSC 109555, a potent and selective inhibitor. Protein Sci **2009**;18(1):92-100 doi 10.1002/pro.16.

57. 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 research **2000**;60(21):5934-6.

58. Oldenburg RA, Kroeze-Jansema K, Kraan J, Morreau H, Klijn JG, Hoogerbrugge N*, et al.* The CHEK2*1100delC variant acts as a breast cancer risk modifier in non-BRCA1/BRCA2 multiple-case families. Cancer research **2003**;63(23):8153-7.

59. Sharifi MJ, Zaker F, Nasiri N, Yaghmaie M. Epigenetic changes in FOXO3 and CHEK2 genes and their correlation with clinicopathological findings in myelodysplastic syndromes. Hematol Oncol Stem Cell Ther **2020**;13(4):214-9 doi 10.1016/j.hemonc.2019.11.004.

60. 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 doi 10.1371/journal.pone.0003062.

61. Matreyek KA, Starita LM, Stephany JJ, Martin B, Chiasson MA, Gray VE*, et al.* Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat Genet **2018**;50(6):874-82 doi 10.1038/s41588-018-0122-z.

62. Lee CH, Chung JH. The hCds1 (Chk2)-FHA domain is essential for a chain of phosphorylation events on hCds1 that is induced by ionizing radiation. J Biol Chem **2001**;276(32):30537-41 doi 10.1074/jbc.M104414200.

63. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI*, et al.* Predicting Splicing from Primary Sequence with Deep Learning. Cell **2019**;176(3):535-48 e24 doi 10.1016/j.cell.2018.12.015.

64. Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD*, et al.* Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med **2019**;12(1):3 doi 10.1186/s13073-019-0690-2.

SUPPLEMENTARY INFORMATION

# Functional analysis identifies damaging *CHEK2* missense variants associated with increased cancer risk

Rick A.C.M. Boonen, Wouter W. Wiegant, Bas Vroling, Stephan Heijl, Zsofia Kote-Jarai, Martina Mijuskovic, Simona Cristea, Nandi Celosse, Nienke Solleveld-Westerink, Peter Devilee, Tom van Wezel, Maaike P.G. Vreeswijk, Niko Beerenwinkel, Rosalind Eeles, Giancarlo Marra and Haico van Attikum



The supplementary information contains*:

- Supplementary Tables S2-S3
- Supplementary Figures S1-S8
- Supplementary References

**Go to the online published manuscript to access Supplementary Table S1*

**Supplementary Table S1*.** Complete list of the predictions from Helix (version 4.2.0) for all possible missense amino acid changes in human *CHEK2*.

**Go to the online published manuscript to access Supplementary Table S1*

**Supplementary Table S2.** List of human *CHEK2* missense VUS analyzed in this study and their predicted splice effects using Alamut and SpliceAI.

| Genomic location (on Assembly GRCh37) | Protein change | SpliceAI score | | | | SpliceAI pre-mRNA position | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *Acceptor Loss* | *Donor Loss* | *Acceptor Gain* | *Donor Gain* | *Acceptor Loss* | *Donor Loss* | *Acceptor Gain* | *Donor Gain* |
| chr22_29130661_C_A | A17S | 0 | 0 | 0,01 | 0 | n/a | n/a | -12 bp | n/a |
| chr22_29130520_C_T | E64K | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29121326_T_C | R117G | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29121301_A_G | F125S | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29121256_C_T | S140Q | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29121253_T_G | K141T | 0 | 0 | 0 | 0,01 | n/a | n/a | n/a | -26 bp |
| chr22_29121242_G_A | R145W | 0 | 0,01 | 0 | 0 | n/a | -15 bp | n/a | n/a |
| chr22_29121087_A_G | I157T | 0 | 0 | 0,01 | 0 | n/a | n/a | 21 bp | n/a |
| chr22_29121087_A_C | I157S | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29121078_A_G | I160T | 0 | 0 | 0,02 | 0 | n/a | n/a | -3 bp | n/a |
| chr22_29121072_T_C | D162G | 0,05 | 0 | 0 | 0 | 36 bp | n/a | n/a | n/a |
| chr22_29121058_C_T | G167R | 0,06 | 0 | 0 | 0 | 50 bp | n/a | n/a | n/a |
| chr22_29121050_A_C | F169L | 0 | 0 | 0 | 0,08 | n/a | n/a | n/a | 0 bp |
| chr22_29121019_G_A | R180C | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29121015_C_T | R181H | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29121001_T_G | N186H | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29115467_A_G | V200A | 0,28 | 0 | 0,04 | 0 | -21 bp | n/a | 6 bp | n/a |
| chr22_29115458_T_C | D203G | 0,05 | 0 | 0 | 0 | -12 bp | n/a | n/a | n/a |
| chr22_29108004_C_T | G229S | 0 | 0 | 0,66 | 0 | n/a | n/a | -2 bp | n/a |
| chr22_29108001_C_G | A230P | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29107974_C_T | E239K | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29107962_A_G | C243R | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29107949_G_T | A247D | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29107943_T_C | K249R | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29107938_T_A | I251F | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29106023_C_T | E273K | 0,03 | 0,04 | 0 | 0 | 24 bp | -29 bp | n/a | n/a |
| chr22_29095917_C_T | G306E | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29095857_A_G | L326P | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |

**Supplementary Table S2.** Continued

| Genomic location (on Assembly GRCh37) | Protein change | SpliceAI score | | | | SpliceAI pre-mRNA position | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | *Acceptor Loss* | *Donor Loss* | *Acceptor Gain* | *Donor Gain* | *Acceptor Loss* | *Donor Loss* | *Acceptor Gain* | *Donor Gain* |
| chr22_29092947_C_T | R346H | 0,04 | 0 | 0 | 0 | 28 bp | n/a | n/a | n/a |
| chr22_29092945_C_T | D347N | 0,07 | 0 | 0 | 0 | 30 bp | n/a | n/a | n/a |
| chr22_29092944_T_G | D347A | 0,01 | 0 | 0 | 0 | 31 bp | n/a | n/a | n/a |
| chr22_29092931_C_A | E351D | 0,04 | 0,01 | 0 | 0 | 44 bp | -42 bp | n/a | n/a |
| chr22_29091846_G_A | H371Y | 0 | 0 | 0,02 | 0 | n/a | n/a | -9 bp | n/a |
| chr22_29091801_C_G | G386R | 0,01 | 0 | 0 | 0 | 36 bp | n/a | n/a | n/a |
| chr22_29091788_T_C | Y390C | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29091788_T_G | Y390S | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29091782_G_A | A392V | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29091732_C_T | D409N | 0 | 0,01 | 0 | 0 | n/a | -34 bp | n/a | n/a |
| chr22_29091721_A_T | S412R | 0 | 0 | 0 | 0,01 | n/a | n/a | n/a | -23 bp |
| chr22_29091716_C_T | G414E | 0 | 0 | 0 | 0,01 | n/a | n/a | n/a | -18 bp |
| chr22_29091213_G_C | P426R | 0,12 | 0 | 0,04 | 0 | -16 bp | n/a | 11 bp | n/a |
| chr22_29091178_C_A | D438Y | 0,05 | 0 | 0,02 | 0 | 19 bp | n/a | 46 bp | n/a |
| chr22_29091154_T_C | N446D | 0 | 0 | 0,01 | 0 | n/a | n/a | 43 bp | n/a |
| chr22_29091147_A_C | I448S | 0 | 0 | 0 | 0 | n/a | n/a | n/a | n/a |
| chr22_29090060_C_T | R474H | 0,01 | 0,01 | 0 | 0 | 45 bp | -40 bp | n/a | n/a |
| chr22_29090060_C_A | R474L | 0,01 | 0,01 | 0 | 0 | 45 bp | -40 bp | n/a | n/a |
| chr22_29090043_C_T | A480T | 0 | 0,01 | 0 | 0 | n/a | -23 bp | n/a | n/a |
| chr22_29090028_A_C | W485G | 0 | 0,01 | 0 | 0 | n/a | -8 bp | n/a | n/a |
| chr22_29085140_G_A | P509S | 0 | 0 | 0 | 0,07 | n/a | n/a | n/a | -5 bp |
| chr22_29083956_G_A | R521W | 0,01 | 0 | 0 | 0 | 18 bp | n/a | n/a | n/a |

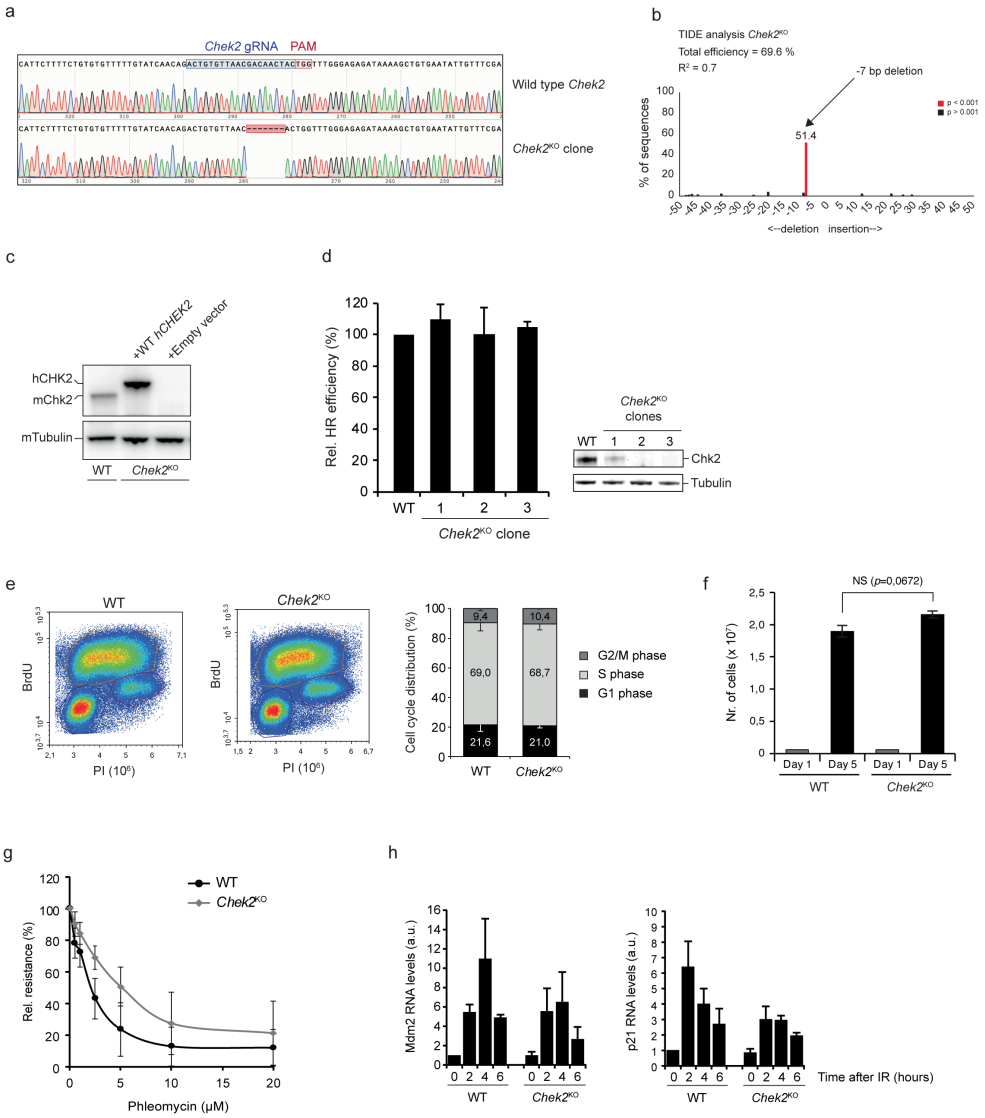
Only predictions from SpliceAI are shown in this table. Predictions using Alamut (i.e., from four algorithms; SpliceSiteFinder-like, MaxEntScan, GeneSplicer and NNSPLICE) are available in the online version of this table. SpliceAI scores range from 0-1 and can be interpreted as the probability that the variant affects splicing at any position within a window of +/- 50 bp. For each variant, SpliceAI looks within a window of +/- 50 bp to see how the variant affects the probabilities of different positions in the pre-mRNA being splice acceptors or donors. The numbers in the pre-mRNA position column represent the positions with the biggest change in probability within the window. Negative values are upstream (5') of the variant and positive values are downstream (3') of the variant. n/a; not applicable.

**Supplementary Table S3.** Complete list SDM primers for all human *CHEK2* variants analyzed in this study.

| Protein change | Forward SDM primer | Reverse SDM primer |
|---|---|---|
| p.A17S | 5'-agtctcatggcagcagttcctgttcacagcc-3' | 5'-ggctgtgaacaggaactgctgccatgagact-3' |
| p.H54= | 5'-ctccagccagtcctctcattccagctctg-3' | 5'-cagagctggaatgagaggactggctggag-3' |
| p.E64K | 5'-tctgggacactgagctccttaaagacagtgtcc-3' | 5'-ggacactgtctttaaggagctcagtgtcccaga-3' |
| p.W93Gfs | 5'-tacccctgccccc gggctcgattatg-3' | 5'-cataatcgagcccgggggcaggggta-3' |
| p.R117G | 5'-acaactactggtttgggggggacaaaagctgtgaa-3' | 5'-ttcacagctttgtccccccaaaccagtagttgt-3' |
| p.F125S | 5'-caaaagctgtgaatattgctctgatgaaccactgctg-3' | 5'-cagcagtggttcatcagagcaatattcacagcttttg-3' |
| p.K135Nfs | 5'-gaaccactgctgaaaagaacagataatacc gaacatacag-3' | 5'-ctgtatgttcggtattatctgttcttttcagcagtggttc-3' |
| p.R137= | 5'-ctgaaaagaacagataaataccgtacatacagcaagaaacactttcg-3' | 5'-cgaaagtgtttcttgctgtatgtacggtatttatctgttcttttcag-3' |
| p.S140N | 5'-aacagataaataccgaacatacaacaagaaacactttcggattttca-3' | 5'-tgaaaatccgaaagtgtttcttgttgtatgttcggtatttatctgtt-3' |
| p.K141T | 5'-gataaataccgaacatacagcacgaaacactttcggattttcagg-3' | 5'-cctgaaaatccgaaagtgtttcgtgctgtatgttcggtatttatc-3' |
| p.R145W | 5'-ccgaacatacagcaagaaacactttggattttcaggga-3' | 5'-tccctgaaaatccaaaagtgtttcttgctgtatgttcgg-3' |
| p.I157T | 5'-gggaagtgggtcctaaaaactcttacactgcatacatagaag-3' | 5'-cttctatgtatgcagtgtaagagtttttaggacccacttccc-3' |
| p.I157S | 5'-gggaagtgggtcctaaaaactcttacagtgcatacatagaag-3' | 5'-cttctatgtatgcactgtaagagtttttaggacccacttccc-3' |
| p.I160T | 5'-tcctaaaaactcttacattgcatacacagaagatcacagtggc-3' | 5'-gccactgtgatcttctgtgtatgcaatgtaagagtttttagga-3' |
| p.D162G | 5'-cattgcatacatagaaggtcacagtggcaatggaac-3' | 5'-gttccattgccactgtgaccttctatgtatgcaatg-3' |
| p.G167R | 5'-tgcatacatagaagatcacagtggcaatagaacctttgtaaataca-3' | 5'-tgtatttacaaaggttctattgccactgtgatcttctatgtatgca-3' |
| p.F169Lfs | 5'-acagtggcaatggaaccttgtaaatacagagcttgtag-3' | 5'-ctacaagctctgtatttacaaggttccattgccactgt-3' |
| p.F169L | 5'-acagtggcaatggaaccttggtaaatacagagcttgtag-3' | 5'-ctacaagctctgtatttaccaaggttccattgccactgt-3' |
| p.R180C | 5'-cagagcttgtagggaaaggaaaatgccgtcctttga-3' | 5'-tcaaaggacggcattttccttccctacaagctctg-3' |
| p.R181H | 5'-tgtagggaaaggaaaacgccatcctttgaataacaattctg-3' | 5'-cagaattgttattcaaaggatggcgttttccttccctaca-3' |
| p.N186H | 5'-acgccgtcctttgaataaccattctgaaattgcactgtc-3' | 5'-gacagtgcaatttcagaatggttattcaaaggacggcgt-3' |
| p.V200A | 5'-cactaagcagaaataaagtttttgccttttttgatctgactgtagatga-3' | 5'-tcatctacagtcagatcaaaaaaggcaaaaactttatttctgcttagtg-3' |
| p.D203G | 5'-gaaataaagtttttgtctttttttggtctgactgtagatgatcagtcag-3' | 5'-ctgactgatcatctacagtcagaccaaaaaagacaaaaactttatttc-3' |
| p.G229S | 5'-caaaaactcttggaagtagtgcctgtggagaggta-3' | 5'-tacctctccacaggcactacttccaagagtttttg-3' |
| p.A230P | 5'-aaaactcttggaagtggtccctgtggagaggtaaa-3' | 5'-tttacctctccacagggaccacttccaagagtttt-3' |
| p.E239K | 5'-gagaggtaaagctggctttcaagaggaaaacatgtaagaaa-3' | 5'-tttcttacatgttttcctcttgaaagccagctttacctctc-3' |
| p.C243R | 5'-tggctttcgagaggaaaacacgtaagaaagtagccataaag-3' | 5'-ctttatggctacttttcttacgtgttttcctctcgaaagcca-3' |
| p.A247D | 5'-gaggaaaacatgtaagaaagtagacataaagatcatcagcaaaagga-3' | 5'-tcctttgctgatgatctttatgtctactttcttacatgttttcctc-3' |
| p.K249R | 5'-aacatgtaagaaagtagccataaggatcatcagcaaaaggaagttt-3' | 5'-aaacttccttttgctgatgatccttatggctactttcttacatgtt-3' |
| p.I251F | 5'-gtaagaaagtagccataaagatcttcagcaaaaggaagtttgctatt-3' | 5'-aatagcaaacttcctttgctgaagatctttatggctactttcttac-3' |
| p.K253X | 5'-gtagccataaagatcatcagctaaaggaagtttgctattggtt-3' | 5'-aaccaatagcaaacttcctttagctgatgatctttatggctac-3' |
| p.E273K | 5'-gacccagctctcaatgttgaaacaaaaatagaaattttgaaaaagctaa-3' | 5'-ttagctttttcaaaatttctatttttgtttcaacattgagagctgggtc-3' |
| p.I286= | 5'-aattttgaaaaagctaaatcatcctgcatcataaagattaaaaactttttgatgca-3' | 5'-tgcatcaaaaaagtttttaatctttatgatgcaaggatgatttagctttttcaaaatt-3' |
| p.G306E | 5'-gttttggaattgatggaagagggagagctgtttgacaaa-3' | 5'-tttgtcaaacagctctccctcttccatcaattccaaaac-3' |
| p.L326P | 5'-aaagaagctacctgcaagccctatttttaccagatgctc-3' | 5'-gagcatctggtaaaaatagggcttgcaggtagcttcttt-3' |
| p.R346H | 5'-catgaaaacggtattatacaccatgacttaaagccagagaatgtt-3' | 5'-aacattctctggctttaagttacggtgtataatacctttcatg-3' |
| p.D347N | 5'-ccttcatgaaaacggtattatacaccgtaacttaaagccagaga-3' | 5'-tctctggctttaagttacggtgtataatacctttcatgaagg-3' |
| p.D347A | 5'-gaaaacggtattatacaccgtggcttaaagccagagaatgtttta-3' | 5'-taaaacattctctggctttaagccacggtgtataatacctttc-3' |
| p.E351D | 5'-caccgtgacttaaagccagataatgttttactgtcatctca-3' | 5'-tgagatgacagtaaaacattatctggctttaagtcacggtg-3' |
| p.T367Mfs | 5'-caagaagaggactgtcttataaagattatgattttgggcactc-3' | 5'-gagtgcccaaaatcataatctttataagacagtcctcttcttg-3' |
| p.T367= | 5'-gaggactgtcttataaagattacagattttgggcactccaag-3' | 5'-cttggagtgcccaaaatctgtaatctttataagacagtcctc-3' |
| p.H371Y | 5'-tataaagattactgattttgggtactccaagattttgggagagac-3' | 5'-gtctctcccaaaatcttggagtacccaaaatcagtaatctttata-3' |
| p.G386R | 5'-tctctcatgagaaccttatgtcgaacccccacctac-3' | 5'-gtaggtgggggttcgacataaggttctcatgagaga-3' |

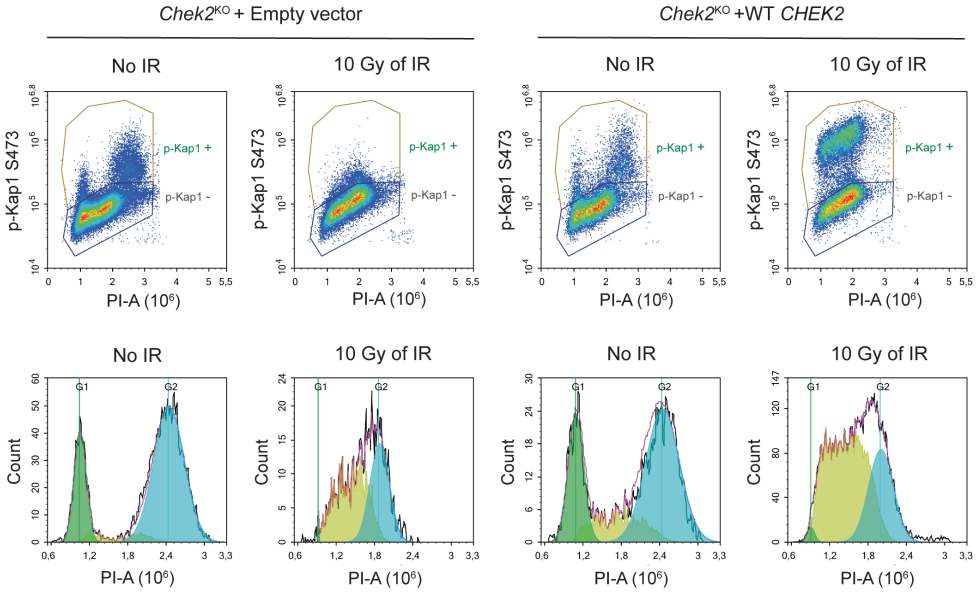
**Supplementary Table S3.** Continued

| Protein change | Forward SDM primer | Reverse SDM primer |
|---|---|---|
| p.Y390C | 5'-ggaacccccacctgcttggcgcctgaa-3' | 5'-ttcaggcgccaagcaggtgggggttcc-3' |
| p.Y390S | 5'-ggaacccccacctccttggcgcctgaa-3' | 5'-ttcaggcgccaaggaggtgggggttcc-3' |
| p.A392V | 5'-cccccacctacttggtgcctgaagttcttgt-3' | 5'-acaagaacttcaggcaccaagtaggtgggggg-3' |
| p.D409N | 5'-ggtataaccgtgctgtgaactgctggagtttagga-3' | 5'-tcctaaactccagcagttcacagcacggttatacc-3' |
| p.S412R | 5'-gtgctgtggactgctggagattaggagttattcttttta-3' | 5'-taaaaagaataactcctaatctccagcagtccacagcac-3' |
| p.G414E | 5'-tgtggactgctggagtttagaagttattcttttatctgcc-3' | 5'-ggcagataaaaagaataacttctaaactccagcagtccaca-3' |
| p.S422Vfs | 5'-ttaggagttattcttttatctgcctagtgggtatccacc-3' | 5'-ggtggatacccactaggcagataaaaagaataactcctaa-3' |
| p.P426R | 5'-ccttagtgggtatccacgtttctctgagcatagga-3' | 5'-tcctatgctcagagaaacgtggatacccactaagg-3' |
| p.S435= | 5'-cataggactcaagtgtctctgaaggatcagatcac-3' | 5'-gtgatctgatccttcagagacacttgagtcctatg-3' |
| p.D438Y | 5'-ctcaagtgtcactgaagtatcagatcaccagtgga-3' | 5'-tccactggtgatctgatacttcagtgacacttgag-3' |
| p.N446D | 5'-gatcaccagtggaaaatacgacttcattcctgaagtctg-3' | 5'-cagacttcaggaatgaagtcgtattttccactggtgatc-3' |
| p.I448S | 5'-atcaccagtggaaaatacaacttcagtcctgaagtctgg-3' | 5'-ccagacttcaggactgaagttgtattttccactggtgat-3' |
| p.R474H | 5'-tagtggatccaaaggcacattttacgacagaagaagc-3' | 5'-gcttcttctgtcgtaaaatgtgcctttggatccacta-3' |
| p.R474L | 5'-tagtggatccaaaggcacttttttacgacagaagaagc-3' | 5'-gcttcttctgtcgtaaaaagtgcctttggatccacta-3' |
| p.R474= | 5'-gtggatccaaaggcacgatttacgacagaagaagc-3" | 5'-gcttcttctgtcgtaaatcgtgcctttggatccac-3' |
| p.A480T | 5'-aggcacgttttacgacagaagaaaccttaagacaccc-3' | 5'-gggtgtcttaaggtttcttctgtcgtaaaacgtgcct-3' |
| p.W485G | 5'-gccttaagacacccggggcttcaggatgaag-3' | 5'-cttcatcctgaagccccgggtgtcttaaggc-3' |
| p.P509S | 5'-aaatgaatccacagctctatcccaggttctagccc-3' | 5'-gggctagaacctgggatagagctgtggattcattt-3' |
| p.R519X | 5'-ccagccttctactagttgaaagcggcccc-3' | 5'-ggggccgctttcaactagtagaaggctgg-3' |
| p.R521W | 5'-gccttctactagtcgaaagtggccccgtgaag-3' | 5'-cttcacggggccactttcgactagtagaaggc-3' |

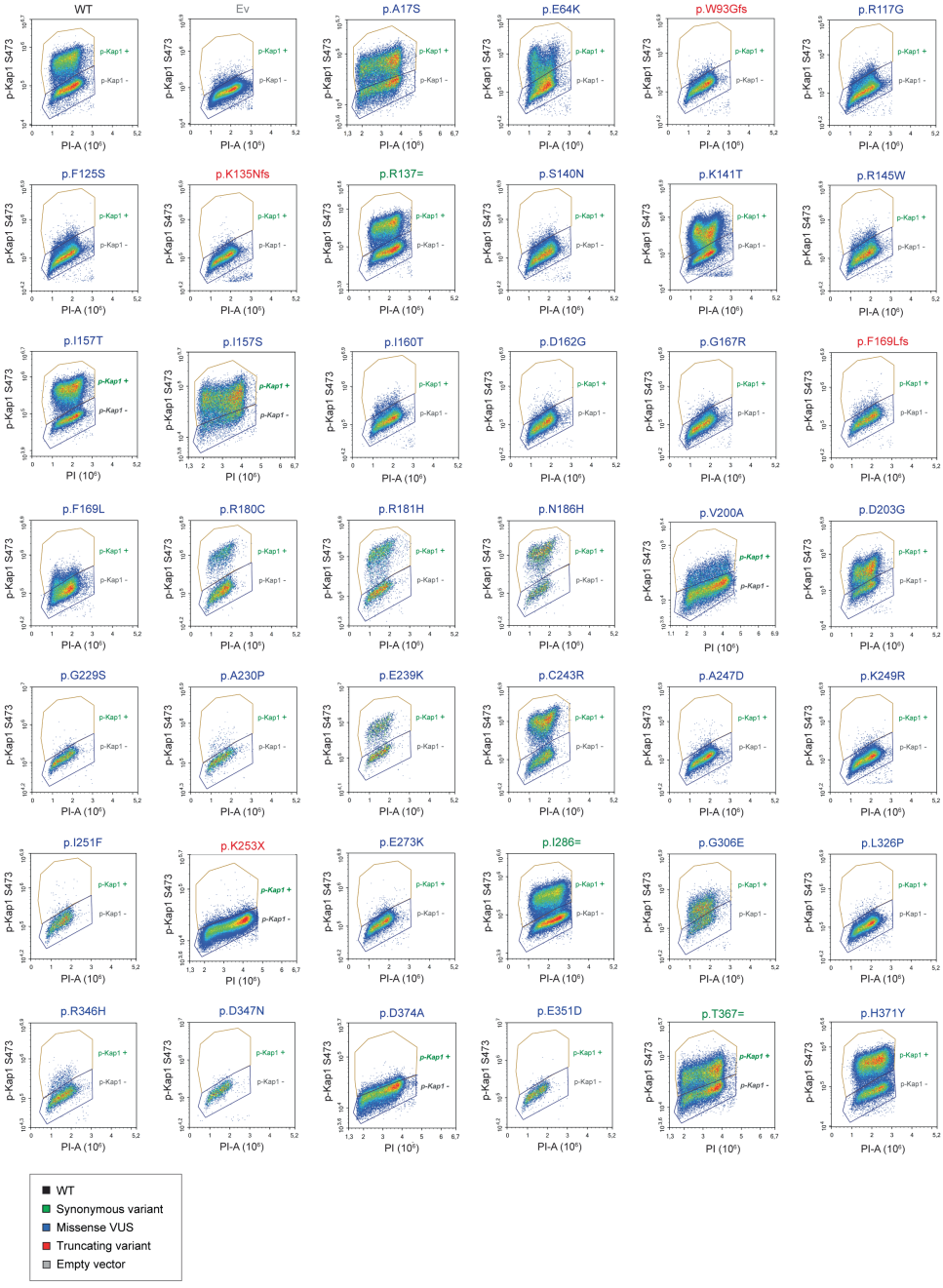



**Supplementary Figure S1.** Validation and functional analysis of *Chek2*$^{KO}$ mES cells. **a** Sequence alignment of a fragment of exon 3 of the *Chek2* gene showing a -7 bp deletion. **b** TIDE analysis confirming the -7 bp deletion in exon 3 of the *Chek2* gene. **c** Western blot analysis confirming the KO of mouse *Chek2* and subsequent complementation/expression of human *CHEK2* in mES cells. Tubulin was used as a loading control. **d** Analysis of the HR efficiency using the DR-GFP reporter in three additional *Chek2*$^{KO}$ clones (left) and western blot analysis confirming the heterozygous or homozygous KO (right). HR efficiency was examined after transient co-expression of I-SceI and mCherry. GFP expression was monitored by FACS. Data represent mean percentages (±SEM) of GFP-positive cells among the mCherry-positive cells relative to that for the wild type (WT), which was set to 100%, from two independent experiments. **e** FACS-based analysis of cell cycle profiles from *Chek2*$^{WT}$ and WT cells after BrdU and propidium iodide (PI) staining. A positive BrdU signal marks cells that are in S-phase.

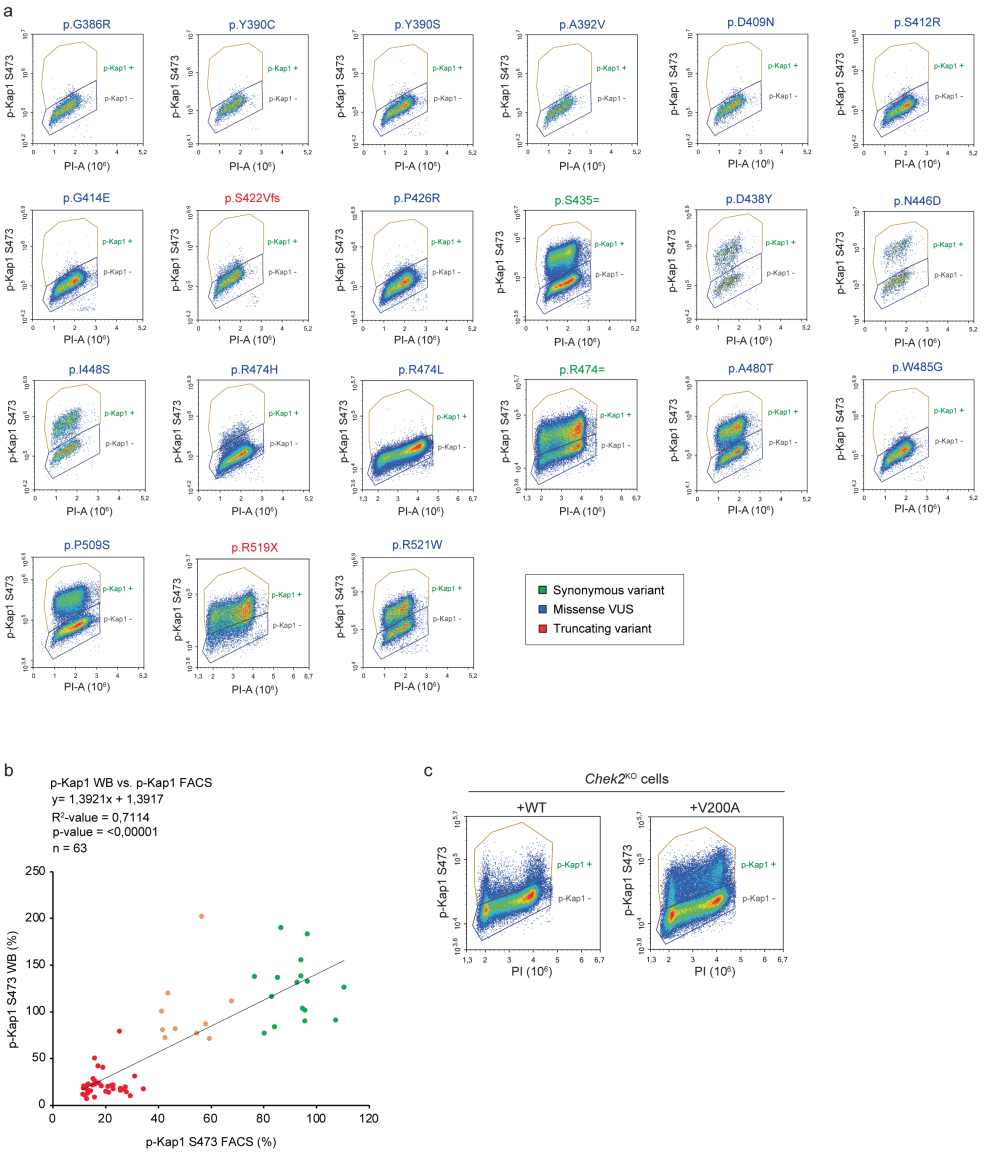
The bar graph represents the mean percentage of cell cycle phase distributions from 2 independent measurements. **f** Analysis of the proliferation rate of WT and *Chek2*$^{KO}$ mES cells. On day 1, 0.5x10$^{6}$ cells were seeded for both conditions and on day 5, cell growth was assessed by cell counting. **g** Phleomycin sensitivity assay using WT and *Chek2*$^{KO}$ mES cells. Cells were exposed to the indicated concentrations of phleomycin for two days. Cell viability was measured after one additional day of incubation in drug-free medium using FACS (using only forward and sideways scatter). Data represent the mean percentage of viability/resistance relative to untreated cells (± SEM) from 2 independent experiments. **f** RT-qPCR analysis of mouse Mdm2 (left) and p21 (right) transcripts in WT versus *Chek2*$^{KO}$ mES cells after the indicated timepoints after IR. Data represent the mean transcript levels (±SEM) from two independent RNA isolation experiments and are relative to the 0 hour timepoint, which was set to 1.




**Supplementary Figure S2.** Kap1 p.S473 phosphorylation in the absence or presence of DNA damage induction. FACS-based analysis, without or 2 hours after IR, of Kap1 p.S473 phosphorylation in *Chek2*$^{KO}$ mES cells complemented with WT *CHEK2* or an empty vector. Cell cycle profiles are shown in the bottom panels and confirm stalling of the cell cycle 2 hours after IR.

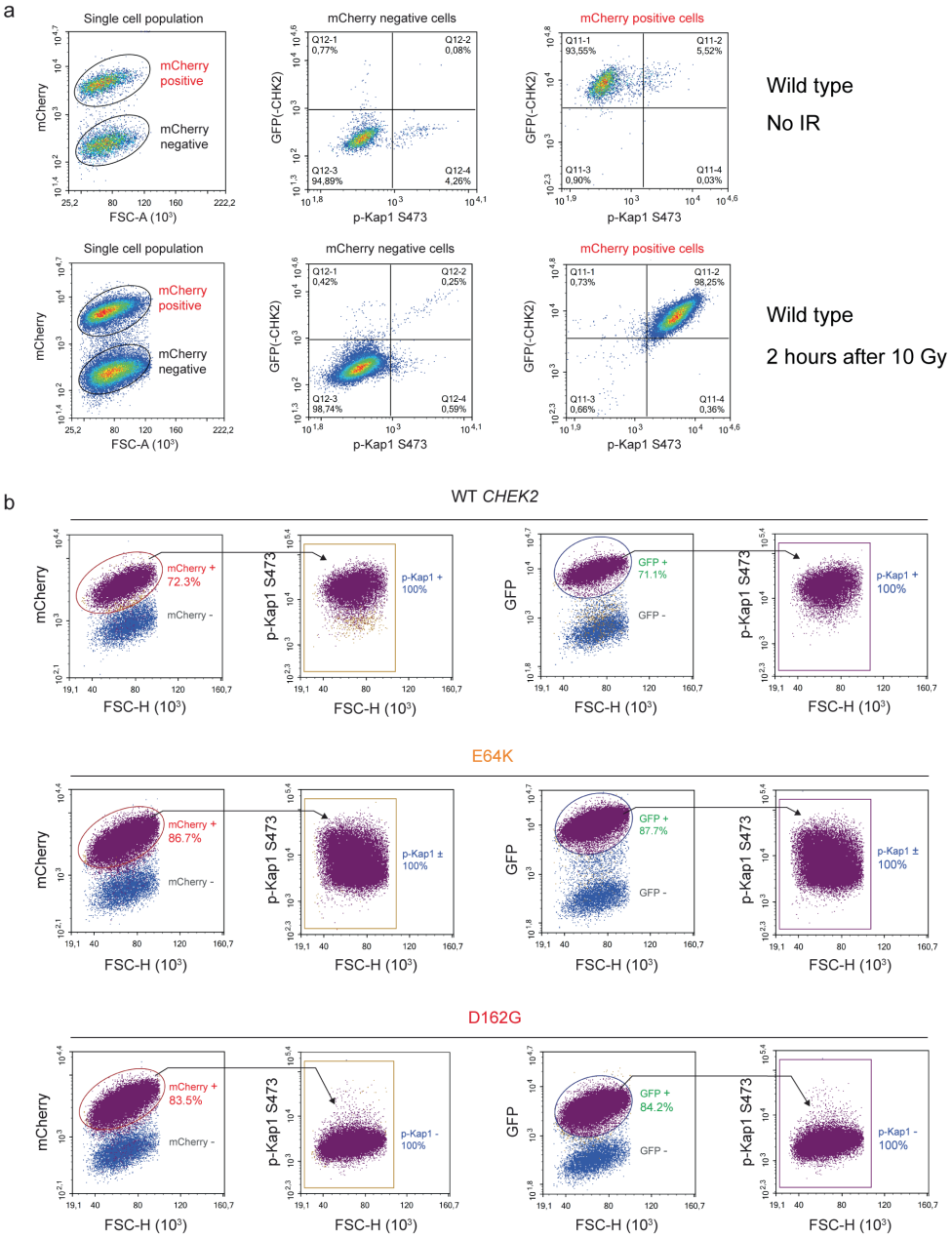



**Supplementary Figure S3.** Human *CHEK2* variants and their effect on CHK2's kinase activity toward Kap1 p.S473. Quantitative FACS-based analysis, 2 hours after IR, of Kap1 p.S473 phosphorylation in *Chek2*$^{KO}$ mES cells complemented with the indicated conditions. Variants/conditions are categorized by color as either wild type (WT, black), synonymous variant (green), truncating variant (red), VUS (blue), or empty vector (Ev, grey).



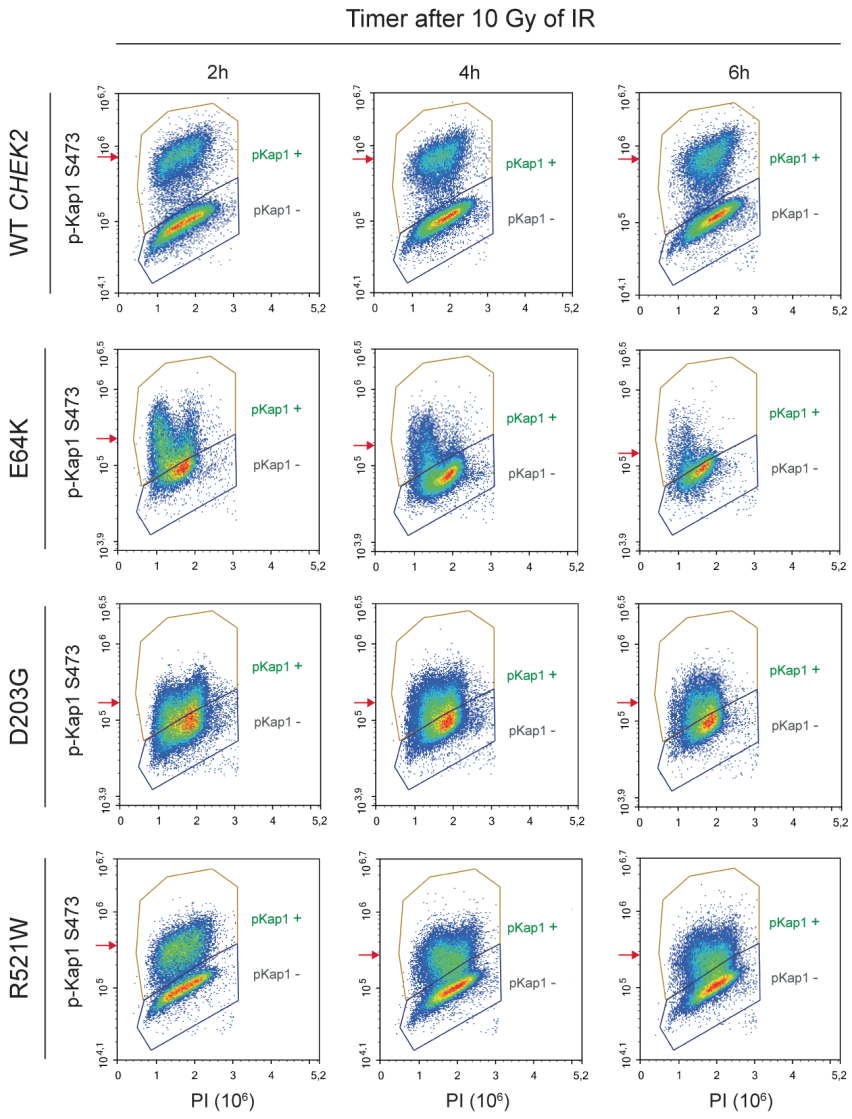

**Supplementary Figure S4.** Human *CHEK2* variants and their effect on CHK2's kinase activity toward Kap1 p.S473. **a** Quantitative FACS-based analysis, 2 hours after IR, of Kap1 p.S473 phosphorylation in *Chek2*$^{KO}$ mES cells complemented with the indicated conditions. Variants/conditions are categorized by color as either wild type (WT, black), synonymous variant (green), truncating variant (red), VUS (blue), or empty vector (Ev, grey). **b** Scatter plot showing the correlation between phospho-Kap1 p.S473 intensities at 2 hours after IR (10 Gy) in *Chek2*$^{KO}$ mES cells expressing untagged CHK2 measured by either FACS and western blot analysis. For quantification of western blots as shown in Fig. 2, phospho-Kap1 p.S473 levels were first normalized to the total Kap1 signals on each blot with its respective wild type and empty vector control (demarcated by the dashed and continuous lines). For each blot, the phospho-Kap1 p.S473 intensities for the *CHEK2* variants were calculated relative to that of wild type

*CHEK2,* which was set to a 100%. Datapoints representing *CHEK2* variants are categorized by color based on functional classification as shown in Fig. 3c (green is functional, orange is intermediate, red is damaging). **c** FACS-based analysis of Kap1 p.S473 phosphorylation at 2 hours after IR in *Chek2*$^{KO}$ mES cells complemented with wild type (WT) untagged CHK2 or untagged CHK2 carrying the p.V200A variant.




**Supplementary Figure S5.** Phospho-Kap1 p.S473 FACS-based analysis after gating for *CHEK2* expression. **a** FACS-based analysis, without or 2 hours after IR, of Kap1 p.S473 phosphorylation in mES cells complemented with EGFP-CHEK2. Left panels show gates for mCherry positive and negative cells, as mCherry is co-expressed from the same cDNA through to a T2A sequence. Middle panels show signals negative for EGFP and phospho-Kap1 p.S473 after gating for mCherry negative cells.

Right panels show positive signals for EGFP and phospho-Kap1 p.S473 after gating for mCherry positive cells. **b** FACS-based analysis, 2 hours after IR, of Kap1 p.S473 phosphorylation in mES cells complemented with EGFP-CHEK2. Results for three conditions (WT, functional; p.E64K, intermediate; p.D162G, damaging) are shown and are quantified in Fig. 2d. Kap1 p.S473 phosphorylation can be quantified after gating for the mCherry-positive signal (left 2 panels) or GFP-positive signal (right 2 panels).

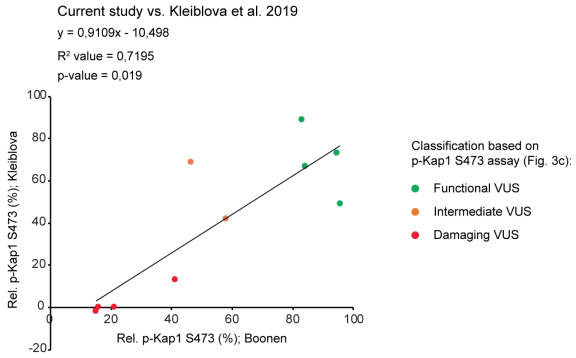



**Supplementary Figure S6.** CHK2 kinase activity in time after DNA damage induction. Quantitative FACS-based analysis of Kap1 p.S473 phosphorylation in *Chek2*$^{KO}$ mES cells complemented with the indicated conditions. Cells were fixed and measured at the indicated times after IR. The red arrows indicate the mean phospho-Kap1 S473 intensity.

a

10 overlapping variants

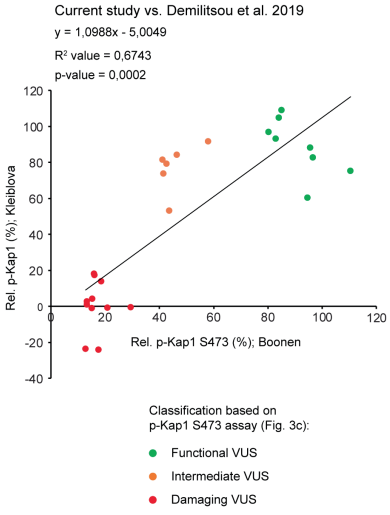
| *CHEK2* variants | | Current study | Kleiblova et al., 2019 | |
|---|---|---|---|---|
| Nt position | Aa position | p-Kap1 intensity (%) | p-Kap1 microscopy (%) | In vitro |
| c.190G>A | p.E64K | 41 | 13 | neutral |
| c.470 T>C | p.I157T | 96 | 49 | neutral |
| c.538C>T | p.R180C | 84 | 67 | neutral |
| c.542G>A | p.R181H | 83 | 89 | neutral |
| c.715G>A | p.E239K | 58 | 42 | intermediate |
| c.1037G>A | p.R346H | 15 | -2 | neutral |
| c.1100delC | p.T367Mfs | 21 | 0 | damaging |
| c.1312G>T | p.D438Y | 46 | 69 | intermediate |
| c.1421G>A | p.R474H | 16 | 0 | damaging |
| c.1525C>T | p.P509S | 95 | 73 | neutral |




b

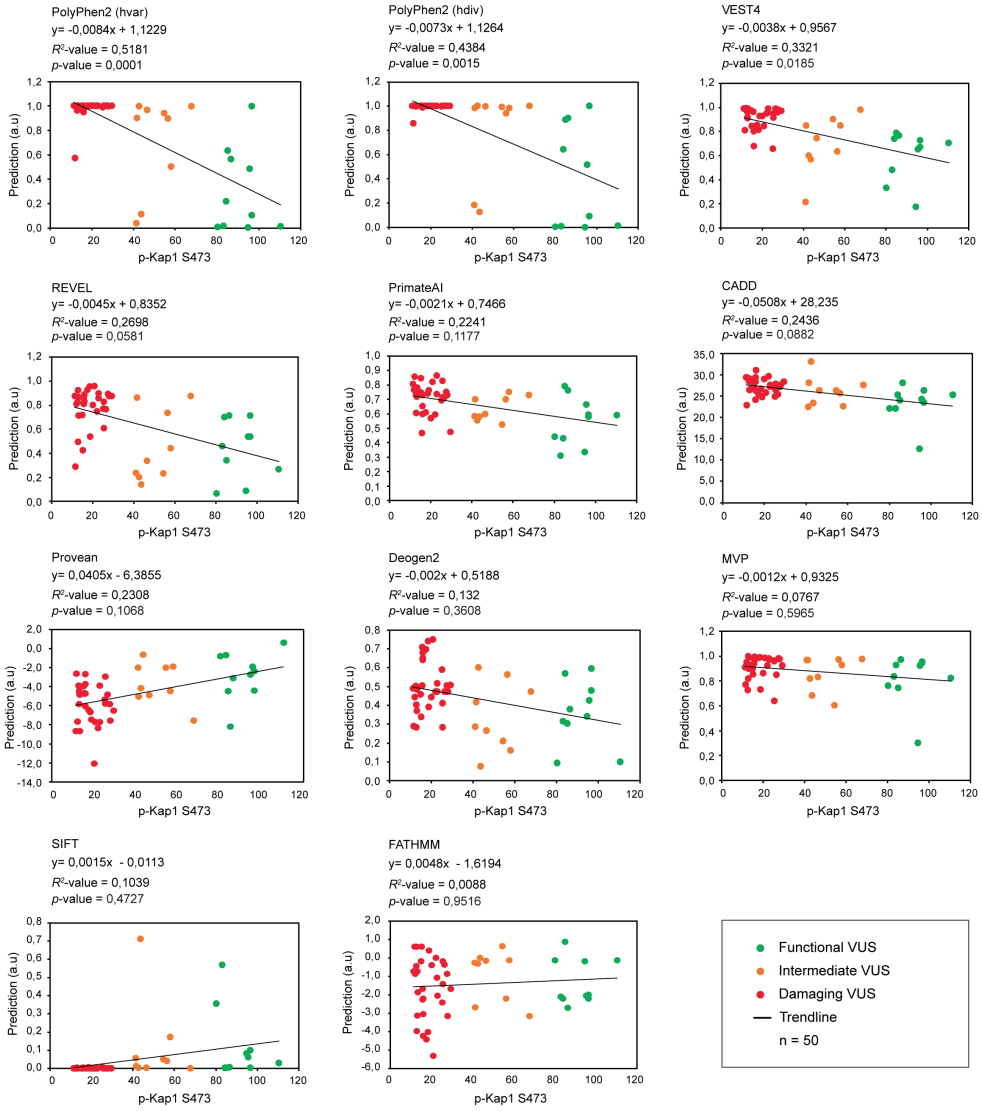
25 overlapping variants

| *CHEK2* variants | | Current study | Delimitsou et al., 2019 |
|---|---|---|---|
| Nt position | Aa position | p-Kap1 intensity (%) | Growth score (%) |
| c.190G>A | p.E64K | 41,09 | 81,42 |
| c.349A>G | p.R117G | 17,42 | -24,08 |
| c.433C>T | p.R145W | 15,82 | 18,04 |
| c.470 T>C | p.I157T | 95,55 | 88,27 |
| c.479T>C | p.I160T | 15,95 | 17,35 |
| c.485A>G | p.D162G | 18,40 | 13,99 |
| c.499G>A | p.G167R | 20,76 | -0,86 |
| c.538C>T | p.R180C | 84,08 | 104,74 |
| c.542G>A | p.R181H | 82,93 | 93,09 |
| c.556A>C | p.N186H | 96,50 | 82,67 |
| c.608A>G | p.D203G | 41,43 | 73,85 |
| c.715G>A | p.E239K | 57,85 | 91,78 |
| c.727T>C | p.C243R | 85,04 | 109,11 |
| c.751A>T | p.I251F | 15,11 | 4,28 |
| c.1037G>A | p.R346H | 15,03 | -0,98 |
| c.1039G>A | p.D347N | 13,23 | 0,89 |
| c.1040A>C | p.D347A | 29,31 | -0,53 |
| c.1111C>T | p.H371Y | 110,43 | 75,29 |
| c.1169A>C | p.Y390S | 13,12 | 2,67 |
| c.1175C>T | p.A392V | 12,72 | -23,49 |
| c.1312G>T | p.D438Y | 46,39 | 84,19 |
| c.1336A>G | p.N446D | 80,18 | 96,85 |
| c.1343T>G | p.I448S | 43,56 | 53,26 |
| c.1525C>T | p.P509S | 94,60 | 60,26 |
| c.1561C>T | p.R521W | 42,51 | 79,35 |




**Supplementary Figure S7.** Correlation between two previously published studies and our current functional analysis of human *CHEK2* missense variants. **a** Table (top) and scatter plot (bottom) showing the correlation between our current phospho-Kap1 p.S473 FACS-based readout as shown in Fig. 2c

and two distinct and indicated functional assays from Kleiblova et al. (1). The scatter plot shows the correlation between our data and the semi-quantitative microscopy-based phospho-Kap1 p.S473 quantification. In the scatter plot, datapoints are colored based on our current functional classification (green is functional, orange is intermediate, red is damaging). **b** Table (left) and scatter plot (right) showing the correlation between our current phospho-Kap1 p.S473 FACS-based readout as shown in Fig. 2c and a yeast-based functional classification from Demilitsou et al. (2). The scatter plot shows the correlation between our data and the yeast-based growth scores. In the scatter plot, datapoints are colored as in a.




**Supplementary Figure S8.** Correlation between in silico predictions and the outcome of functional assays for missense variants in human *CHEK2*. **a** Scatter plots showing the correlation between the indicated in silico predictions algorithms and results from the FACS-based assay examining Kap1 p.S473 phosphorylation as shown in Fig. 2c. Datapoints are colored based on their functional classification (green is functional, orange is intermediate, red is damaging).

## SUPPLEMENTARY REFERENCES

1. Kleiblova, P., Stolarova, L., Krizova, K., Lhota, F., Hojny, J., Zemankova, P., Havranek, O., Vocka, M., Cerna, M., Lhotova, K., et al. (2019). Identification of deleterious germline CHEK2 mutations and their association with breast and ovarian cancer. International journal of cancer 145, 1782-1797.
2. Delimitsou, A., Fostira, F., Kalfakakou, D., Apostolou, P., Konstantopoulou, I., Kroupis, C., Papavassiliou, A.G., Kleibl, Z., Stratikos, E., Voutsinas, G.E., et al. (2019). Functional characterization of CHEK2 variants in a Saccharomyces cerevisiae system. Hum Mutat 40, 631-648.