Comprehensive analysis and ACMG-based classification of *CHEK2* variants in Spanish hereditary cancer patients

Gardenia Vargas-Parra[1], Jesus del Valle[2], Paula Rofes[3], Mireia Gausachs[1], Agostina Stradella[3], Jose Marcos Moreno-Cabrera[4], Angela Velasco[5], Eva Tornero[2], Mireia Menéndez[5], Xavier Muñoz[6], Silvia Iglesias[7], Adriana Lopez-Doriga[8], Daniel Azuara[2], Olga Campos[9], Raquel Cuesta[2], Esther Darder[10], Rafael de Cid[4], Sara Gonzalez[11], Alex Teulé[5], Matilde Navarro[7], Joan Brunet[12], Gabriel Capellá[13], Marta Pineda[14], Lídia Feliubadaló[9], and Conxi Lázaro[15]

[1]Bellvitge Institute for Biomedical Research
[2]Institut Català d'Oncologia
[3]Institut Catala d' Oncologia
[4]The Institute for Health Science Research Germans Trias i Pujol (IGTP) - PMPPC
[5]Institut Català d'Oncologia-IDIBELL
[6]Molecular Epidemiology Group, Translational Research Laboratory, Catalan Institute of Oncology-IDIBELL
[7]Catalan Institute of Oncology, IDIBELL
[8]Institut d'Investigacions Biomèdiques de Bellvitge (IDIBELL)
[9]Catalan Institute of Oncology (ICO-IDIBELL)
[10]Programa de Consell Genètic en Càncer
[11]Instituto Catala d'Oncologia
[12]Institut d'Investigació Biomèdica de Girona (IdIBG), Institut Català d'Oncologia, Hospital Josep Trueta
[13]ICO-IDIBELL
[14]Institut Català d'Oncologia - IDIBELL
[15]Catalan Institute of Oncology (ICO)

May 26, 2020

Abstract

Background: *CHEK2* variants are associated with intermediate breast cancer risk among other cancers. We aimed to comprehensively describe *CHEK2* variants in a Spanish hereditary cancer (HC) cohort and adjust American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) guidelines for their classification. Methods: First, three *CHEK2* frequent variants were screened in a retrospective Hereditary Breast and Ovarian Cancer cohort of 516 patients. After, the whole *CHEK2* coding region was analyzed by next-generation sequencing in 1,848 prospective patients with HC suspicion. We refined ACMGAMP criteria and applied different combinatorial rules to classify *CHEK2* variants and define risk alleles. Results: We identified 10 *CHEK2* null variants, 6 missense variants with discordant interpretation in ClinVar database, and 35 additional variants of unknown significance. Twelve variants were classified as (likely)-pathogenic; 2 can also be considered "established risk-alleles" and one as "likely risk-allele". The prevalence of (likely)-pathogenic variants in the HC cohort was 0.8% (1.3% in breast cancer patients and 1.0% in hereditary non-polyposis colorectal cancer patients). Conclusions: Here we provide ACMG adjustment guidelines to classify *CHEK2* variants. We hope that this work would be useful for variant

classification of other genes with low effect variants

CONFLICT OF INTEREST

Authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS STATEMENT

Conceptualization and design: Lázaro, Vargas-Parra, del Valle, Gausachs. Data curation: Rofes, Gausachs, Stradella, Velasco, Tornero, Menéndez, Muñoz, Iglesias, López-Doriga, Azuara, Campos, Cuesta, Darder, Teulé, Navarro. Formal analysis and interpretation of data: Vargas-Parra, del Valle, Pineda, Feliubadaló, Moreno-Cabrera, del Cid, Lázaro. Funding acquisition: Lázaro, Brunet, Capellá, Pineda, Feliubadaló. Investigation: Vargas-Parra, del Valle, Gausachs, Rofes, Gausachs, Stradella. Methodology: Vargas-Parra, del Valle, Gausachs, Lázaro. Project administration: Lázaro. Resources: del Cid, González. Software: Moreno-Cabrera , López-Doriga. Supervision: Lázaro. Validation: Pineda, Feliubadaló. Visualization: Vargas-Parra, del Valle, Rofes. Drafting of manuscript: Vargas-Parra, del Valle, Rofes, Lázaro. Critical revision: Lázaro, Brunet, Capellá, Pineda, Feliubadaló.

DATA AVAILABILITY STATEMENT

The data files that support the findings of this study are available upon request.

INTRODUCTION

Extensive efforts to standardize variant classification criteria in highly penetrant genes have been made by different groups such as the joint consensus of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) (C. S. Richards et al., 2008; S. Richards et al., 2015), ENIGMA consortium for*BRCA1/2* genes (Spurdle et al., 2012) or InSiGHT variant interpretation group for MMR genes (Plon et al., 2008). However, there is still an important work to be done in moderate or low-penetrance genes (Katona et al., 2018) since multigene panels for hereditary cancer (HC) include them. A recent publication proposed a framework for classification of variants in low-penetrance genes, in which a variant could be classified as established risk allele (ERA) if it has been assessed in case-control studies of good design and data quality, demonstrated to be cancer-related and determined through robust meta-analysis (Senol-Cosar et al., 2019).

In the present work we have focused in *CHEK2* (checkpoint kinase 2; MIM# 604373), which is a tumor suppressor gene associated with different forms of HC, such as breast cancer (BC), colorectal cancer (CRC) and Li-Fraumeni syndrome (Bell et al., 1999; Meijers-Heijboer et al., 2002), among others. *CHEK2* is included in most of the in-house and commercial HC panels (Easton et al., 2015). *CHEK2*mRNA has a total length of 1844-bp distributed in 15 exons, is located at chromosome 22q12.1 and encodes for a human protein of 543-aa, analogue of the yeast checkpoint kinases Cds1 and Rad53 (Matsuoka et al., 2000). CHK2 protein is a kinase involved in several cellular processes, including the control of mitosis and meiosis progression, and plays an important role in the DNA-damage signaling network (Bartek, Falck, & Lukas, 2001; Zannini, Delia, & Buscemi, 2014). ATM activates CHK2 in response to DNA damage. Once activated, CHK2 is capable of phosphorylating many substrates involved in DNA repair, cell cycle regulation, p53 signaling and apoptosis (Zannini et al., 2014).

A few *CHEK2* variants have been described as recurrent or founder variants in some populations. The most well-known *CHEK2* variant is c.1100delC and it is primarily present in individuals of Northern and Eastern European descent; it results in a premature stop codon within exon 10, impairing the kinase ability of the enzyme (Wu, Webster, & Chen, 2001). A meta-analysis of 44,777 patients and 42,997 controls established a BC odds ratio (OR) of 2.26 for *CHEK2* c.1100delC carriers (Schmidt et al., 2016). Another frameshift founder mutation, the deletion of exons 9 and 10, is considered to double BC risk (Cybulski et al., 2007). The missense variant c.470T>C, p(.Ile157Thr) is described to confer a lower risk compared to the two previous ones (OR of 1.58 and 1.67 for BC and CRC, respectively) (Han, Guo, & Liu, 2013). According to a study of

13,087 BC cases and 5,488 controls, the OR for 73 *CHEK2* rare missense variants was of 1.36 (95% CI, 0.99-1.87) and of 1.51 (95% CI, 1.02-2.24) if considering only variants in functional domains (Decker et al., 2017; Han et al., 2013). Furthermore, in a recent study of 1,355 BC cases, the OR for *CHEK2* missense variants varied between 3.79 and 5.9 (95% CI, 1.86-7.12 and 2.38-14.78) when compared to ExAC and FLOSSIES controls, respectively (Fostira et al., 2020).

The challenge of *CHEK2* variant classification is reflected in numerous discrepancies in ClinVar classification (Decker et al., 2017), to the point of being recognized as the gene with more conflicting interpretations in HC diagnosis (Balmaña et al., 2016). Moreover, there is a current controversy whether to use *CHEK2* missense variants at clinical level. For instance, the National Comprehensive Cancer Network's BC management recommendations for *CHEK2* carriers only apply to carriers of truncating variants. In the same line, the UK Cancer Genetics Group decided not to take into account non-truncating variants in the clinical routine until a precise utility is stated for missense variants (Taylor et al., 2018).

Here, we present our effort to characterize the *CHEK2* mutational spectrum in Spanish HC patients which has resulted in the need to refine ACMG-AMP guidelines for this gene.

MATERIAL AND METHODS

Patients and control cohort

A total of 2,346 HC suspected patients were screened in two phases, first 516 cases were screened for c.1100delC, exon 9-10 deletion and c.470T>C recurrent variants and after 1,848 HC patients and 194 healthy controls were analyzed by mutligene panel testing (see Figure 1 and Supplementary Material). Additionally, 1,501 control samples were genotyped for the c.320-5T>A variant. Written informed consent was obtained from all patients and the study protocol was approved by the Ethics Committee of IDIBELL (PR278/19).

***CHEK2* variant annotation and collection of variant information.** Variant annotation was performed using NM_007194.3 for *CHEK2* gene (coding region and +/-20bp of intronic region). All variants identified were submitted to Alamut Software Suite v2.15.0 (Interactive Biosoftware, Rouen, France) to retrieve population frequency and *in silico* prediction data. Variant classification in ClinVar as well as literature review were collected.

Criteria used to assess pathogenicity

PVS1 and PVS1_strong were considered met according to Tayoun decision tree (Abou Tayoun et al., 2018; ClinGen-TP53_Expert_Panel, 2019), PS3 was weighted when a functional defect was found in at least 2 independent studies in absence of discordant results. PS4 was weighted for variants with an odds ratio (OR) >5.0 in case-control studies, PS4_moderate for low-moderate penetrant genes if the OR was between 1.5 and 5, with a p-value <0.01 as long as the phenotype was in accordance with the described for the gene. PM1, if the variant affected a highly conserved amino acid located in the FHA and/or kinase domain. PM2 was weighted when the variant was absent or in less than 1 out of 100,000 alleles in gnomAD v2.1.1 from "all" non-cancer population dataset; if present in [?] 2 individuals within any sub-population, it should be present in <1 out of 50,000 alleles in that subpopulation. Since some *CHEK2* variants in spite of being frequent in the population, the associated risk is significant, PM2_supporting was applied if the variant was present in [?] 1 out of 20,000 alleles in gnomAD v2.1.1 dataset (Karczewski et al., 2019). PP3 was weighted if the *in silico* predictors suggested a splicing alteration (reduction of [?]20% in Alamut score) and/or protein function alteration according to Varsome genome interpreter (Kopanos et al., 2018). Variant classification was performed using different combination of rules according to classical ACMG-AMP guidelines (S. Richards et al., 2015), ClinGen-*TP53* suggested modifications to ACMG (ClinGen-TP53_Expert_Panel, 2019) and to ACMG-Bayesian modelling (Tavtigian et al., 2018) (Table 1).

Risk allele categorization was ascertained when possible as previously described (Senol-Cosar et al., 2019) (Table S2). Accordingly, ERA classification was given to variants reported in multiple association studies or to those determined by robust meta-analysis; likely risk allele (LRA) was assigned if either the variant

showed association in at least 2 independent studies, had been reported in a large study of high quality or in multiple studies with almost complete concordance .

RESULTS

Nature and distribution of variants and clinical classification

After *CHEK2* mutational analysis of 2,346 cases with suspicion of HC and discarding benign variants, we identified 51 different variants. Sixteen of which corresponded to variants expected to produce loss of function proteins or missense variants with conflicting interpretation in the literature (Table 1, Figure 2, pedigrees in Figure S1). The remaining 35 variants were clearly variants of unknown significance (VUS) (Table S1). The control group carried one conflicting interpretation missense variant and one VUS (Table 1 and Table S1).

To apply ACMG-AMP guidelines we split them based on the presence or absence of PVS1 (criterion for a predicted loss of function variant; Table 2).

Variants meeting PVS1 criterion

Nonsense and frameshift variants

Only one patient was carrier of the recurrent *CHEK2* c.1100delC mutation, p.(Thr367Metfs*15) (1 out of 2,346, 0.04%). Given the great amount of data related to *CHEK2* c.1100delC, this variant meets PS3 (well established functional studies) and PS4 (higher prevalence in affected individuals versus controls), besides PVS1**.** However, PS4 was assigned with moderate strength (PS4_moderate), since OR>5.0 for a moderately penetrant gene cannot be achieved. The combination of these rules classified this variant as pathogenic (P) in any combination of rules framework, and since it is well-studied and frequent in some populations, it was classified as established risk allele (ERA) within the Senol-Cosar framework (Table 1 and Table S2). c.1368dupA, p.(Glu457Argfs*33) variant meets PS3 and PM2_supporting, being classified as P in all frameworks. c.715G>T, p.(Glu239*) variant meets PM2, therefore was classified as likely pathogenic (LP) using ACMG and ClinGen-*TP53* frameworks. According to Tavtigian's Bayes model (Tavtigian et al., 2018), it gathers enough evidence to be classified as P. Variants c.279G>A, p.(Trp93*) and c.591delA, p.(Val198Phefs*7) were weighted PM2_supporting. The sum of PVS1 and a supporting criterion is not enough to classify a variant as LP/P using ACMG guidelines (S. Richards et al., 2015). However, application of Bayesian modelling of this combination of rules gives a posterior probability of 0.988, resulting in its classification as LP according to Tavtigian's (S. Richards et al., 2015; Tavtigian et al., 2018) as well as following ClinGen-*TP53* modifications (ClinGen-TP53_Expert_Panel, 2019).

Canonical Splice Site variants

PVS1 was weighted for splicing variants predicted to produce an exon skipping with a subsequent frameshift. PM2 was weighted for c.593-1G>T and c.792+2T>C. Neither of them received PP3 to avoid redundancy with PVS1, remaining as LP according to ACMG and ClinGen-*TP53* frameworks. Notwithstanding, the combination of these rules in the Bayes model gives a posterior probability of 0.994, allowing its classification as P (Tavtigian et al., 2018). c.792+2T>C was reported in a previous study from our group (Feliubadalo et al., 2017), it produces a partial retention of intron 6, decreasing the expression of wildtype. It is classified as LP by ClinVar.

Copy number variants

The whole *CHEK2* deletion was weighted as PVS1 Stand-alone, as proposed for full gene deletions of known haploinsufficiency (Abou Tayoun et al., 2018), being classified P by all frameworks. Deletion of exons 3 and 4 occurs in-frame and produces the loss of the entire critical FHA domain, for this reason PVS1 was weighted. Together with PM2_supporting it would be a VUS with traditional ACMG combination rules but would be classified as LP following ClinGen-*TP53* as well as using Tavtigian's calculations. Deletion of exon 2 removes the first methionine and deletes 45 amino acids of the FHA domain, essential for CHK2 protein function, therefore, PVS1 was applied as "strong". Together with PM2_supporting, it did not reach LP/P classification in any framework.

Variants not meeting PVS1

We found 6 missense variants with discordant classifications of pathogenicity in ClinVar (Table 1) in 13 unrelated patients. In addition, one of the healthy (non-cancer) controls carried the *CHEK2* c.349A>G, p.(Arg117Gly) variant. To better interpret missense variants, a comprehensive review of previous functional studies was done, main results are summarized in Table S3 and Table S4. In a further effort to improve variant classification, after classical ACMG we also followed the allele risk criteria reported recently (Senol-Cosar et al., 2019). For this, we searched for association studies of our *CHEK2* variants (Table S2).

*CHEK2*c.190G>A, p.(Glu64Lys) is located in a weakly conserved amino acid in the SQ/TQ cluster domain (SCD). It is predicted deleterious by *in silico* analysis. It shows a partially reduced phosphorylation by ATM at the Thr68 residue, as well as partially reduced auto-phosphorylation and Cdc25C phosphorylation. It affects KAP1 phosphorylation and has discrepant results about DNA damage response (Table S3). Furthermore, there are no high-quality case-control studies. Therefore, this variant only meets PP3 criterion, remaining as VUS (Table 1). Variant c.349A>G, p.(Arg117Gly) affects a highly conserved amino acid (class C65 according to GVGD) in the FHA domain. It is predicted deleterious by *in silico* analysis. It does not affect phosphorylation by ATM nor oligomerization, but affects all the rest of the studied protein functions (Table S3). This variant accomplished PS3, PS4_moderate, PM1 and PP3 criteria, being classified as LP by all frameworks. It has been studied in a large high-quality case-control study, reporting a BC OR of 2.26 (95% CI, 1.29-3.95) (Table 1), therefore it could be considered as LRA within the Senol-Cosar framework (Table S2). Variant c.433C>T, p.(Arg145Trp) is located in a moderately conserved amino acid of the FHA domain. It is predicted deleterious by *in silico* . It reduces CHK2 expression and stability. In functional assays, it has been consistently reported to impair kinase and DNA repair activity. Evidences for classification includes PS3, PM1 and PP3, being classified as LP by all frameworks. Variant c.470T>C, p.(Ile157Thr) lies in a weakly conserved amino acid of the FHA domain. It is predicted deleterious by *in silico* analysis. It has been widely studied, nevertheless the functional assays reported to date show discordant results (Table S3). The reported OR in the biggest *CHEK2*meta-analysis was 1.58 (95% CI, 1.42 - 1.75), therefore PS4_moderate was applied, application of PP3 was not enough to classify this variant as LP/P. However, following recommendations from Senol-Cosar et al, it would be an ERA due to the existence of multiple case-control studies (Senol-Cosar et al., 2019). Variant c.499G>A, p.(Gly167Arg) is located in a highly conserved amino acid of the FHA domain. It is predicted deleterious by *in silico* analysis. Although there are only 2 functional studies, they both reported an impaired DNA repair activity in yeast assays (Table S3). PS3, PM1, PM2_supporting and PP3 were assigned, being classified as LP by all frameworks. Variant c.1427C>T, p.(Thr476Met) lies in a moderately conserved amino acid of the kinase domain. It is predicted deleterious by *in silico* analysis. Functional assessment of KAP1 phosphorylation results deleterious *in vitro* and good enough *in vivo* . Furthermore, SOX phosphorylation was reported equal to that of the pathogenic c.1100delC variant. Assays on DNA repair activity have found it damaging or with intermediate activity (Table S3). Due to these discordant functional assay results PS3 was not weighted. Classification remained as VUS since c.1427C>T only accomplished PP3.

Variants of unknown significance

Thirty-five unique VUS (with less than 2 LP/P interpretations in ClinVar) were encountered in our cases (Table S1). We aimed to perform RNA analysis in 3 of these, due to *in silico* prediction results (c.320-5T>A and c.1376-8T>C) or to the nature of the variant (duplication of exons 3 and 4). Lymphocytes for RNA analysis were available from one carrier of the duplication of exons 3 and 4, for several samples with c.320-5T>A and were unattainable from c.1376-8T>C carriers. RNA analysis showed that the duplication of exons 3 and 4 occurs in tandem and produces 30% of aberrant transcript containing an in-frame insertion of 273 bp (Figure 3). This affects the region that codifies for the FHA domain, unfortunately there were no polymorphisms in the region to perform quantitative analysis. This variant remains as VUS following all guidelines. Regarding c.320-5T>A variant, *in silico*programs predicted a reduction in the recognition of the splicing acceptor site of exon 3. cDNA analysis in two carriers showed the generation of an aberrant transcript, consisting in an in-frame deletion of exons 3 and 4 (Figure 4), as previously reported (Kraus et

al., 2017). The amount of abnormal transcript seemed greater than 20%, although the absence of exonic polymorphisms prevented an accurate quantification. Of note, the frequency of c.320-5T>A is 0.12% in gnomAD (NFE) and of 1.35% (25 out of 1,848) in our HC cohort. In order to understand the differences in frequency in our population with relation to international databases, we screened 1,501 control samples. *CHEK2* c.320-5T>A had a frequency of 0.8% (12 out of 1,501) in our controls, not a statistically significant difference, preventing it to be considered as risk allele.

CHEK2 variants in the different HC groups

Applying the Bayesian combination of rules by clinical suspicion subgroups of the HC cohort, *CHEK2* LP/P variants were identified in 1.3% of HBC cases (n=9), in 0.5% HBOC cases (n=1), 1% of the hereditary non-polyposis colorectal cancer patients (HNPCC, n=3) and in one patient from the minority cancer group (0.5%), who had two kidney tumors, pheochromocytoma and prostate cancer.

Among the 10 families with HBC/HBOC, 2 proband females had two variants in *CHEK2* . One female, with BC at 42, was a compound heterozygous of a whole *CHEK2* deletion and variant c.499G>A. The other patient with bilateral BC at 35 carried two *CHEK2* missense variants (c.433C>T and c.470T>C) *in trans.* Both cases were previously reported by our group (Stradella et al., 2018). In addition, a third proband diagnosed of BC at age 49 carried the *CHEK2* c.349A>G and a pathogenic variant in *ERCC3* , so she could be considered a multilocus inherited neoplasia allele syndrome (MINAS) patient. Interestingly, the three HNPCC patients with *CHEK2* LP/P variants developed CRC at a young age (22, 25 and 44) and their tumors were MMR proficient.

DISCUSSION

We have made an effort to classify variants in the low-moderate penetrance *CHEK2* gene. For that, we analyzed the whole coding region of *CHEK2* in a large HC cohort, performed in-depth literature review and have defined specific cut-offs for ACMG criteria to allow classification of variants with low effect. Furthermore, we applied different combinatorial rules that enabled us to compare classification rates. Concluding that the Bayesian model is the most optimal framework to classify variants to a greater extent.

From our experience in variant classification and after a comprehensive literature review, we propose two adaptations of the ACMG criteria. Regarding PS4 we propose to score PS4_moderate for low-moderate penetrant genes if an OR is given between 1.5 and 5, with a p value of <0.01, when the phenotype is in accordance with the previously described. In relation to PM2 evidence, in our laboratory we use an extremely conservative approach and assign PM2 only if the variant is absent or present in less than 1 out of 100,000 alleles in gnomAD (0.001% of maximum frequency) for high penetrant genes. However, we propose to assign PM2_supporting when the variant is [?]1 out of 20,000 alleles.

Variants meeting PVS1 criterion tend to be easier to classify as LP/P. For instance, the founder mutation c.1100delC is the most studied*CHEK2* mutation and it has a prevalence of 0.26% in NFE population. *CHEK2* c.1100delC has a moderate penetrance (Meijers-Heijboer et al., 2002; Oldenburg et al., 2003), conferring an increased BC risk for overall population (OR= 2.89, 95% CI, 2.63–3.16) (Liang et al., 2018) and for carriers with familial BC (OR= 3.21, 95% CI, 2.41-4.29) (Liang et al., 2018). It has been reported absent in Spanish population (Bellosillo et al., 2005), or with frequencies of 0.93% in Basque population, 0.36% in Galician population and 0.3% in a study of *BRCA* -negative HBC Basque and Catalan families (Fachal, Santamarina, Blanco, Carracedo, & Vega, 2013; Gutierrez-Enriquez, Balmana, Baiget, & Diez, 2008; Martinez-Bouzas et al., 2007). In our larger cohort, only one case was identified (0.08%, 1 out of 1,251 BC affected cases), confirming its low prevalence in our population. Moreover, in a recent study analyzing 15 truncating *CHEK2*variants in 213 patients and 29 control carriers, the BC risk OR was 3.11 (95% CI, 2.15-4.69) (Decker et al., 2017). Here we identified 10 proband carriers of truncating variants, 8 of which developed the first tumor before the age of 50, consistent with previous findings of early cancer development in carriers of truncated variants (Decker et al., 2017; Han et al., 2013). Nonetheless, the median age at first cancer diagnosis in our study was not very different amongst carriers of truncating and missense LP/P variants, being 42 (range 25-65) and 40 (range 22-51) years, respectively. Bilateral BC has been mainly

reported in c.1100delC carriers (M Kriege & J M Collee, 2014), and truncating variants in this gene have been associated to other non-breast second primary tumor diagnosis in a study using multigene panel testing (Fostira et al., 2020). In our cohort, 4 cases with two or multiple cancers were carriers of truncating variants and only one was carrier of a LP missense, confirming a higher aggressiveness of truncating variants over missense variants.

Conflicting results are common for missense hypomorphic variants and represent one of the biggest challenges we faced for *CHEK2*variant classification due to the lack of more sensitive functional assays and the use of different controls, complicating replication and therefore bypassing PS3 application. The c.470T>C founder mutation conveys a moderate susceptibility for overall cancer (OR= 1.39; p<0.00001) and for BC only (OR= 1.58; p<0.00001) in a large meta-analysis (Han et al., 2013). Its pathogenicity has been established for ovary cystadenomas in young Polish carriers (OR = 2.6; p=0.006) (Szymanska-Pasternak et al., 2006) and is associated to a 2-fold risk of non-Hodgkin lymphoma, colon, kidney, thyroid and prostate cancers (Cybulski et al., 2004). We found it in a male patient diagnosed of testicular cancer at 25 years. Interestingly, in a recent study of 448 Croatian testicular cancer patients it was found in 5.1% of them, resulting in an OR of 3.93 (95% CI, 1.53-9.95) even when its population frequency is of 1-2% (AlDubayan et al., 2019). Of note, when applying ACMG-AMP guidelines, c.470T>C remains as VUS even applying PS4_moderate. To our knowledge, c.470T>C is the most studied *CHEK2* missense variant, but as shown in Table S3, it has conflictive interpretations of pathogenicity at almost all functional studies, therefore PS3 was ruled out, remaining as VUS in the ACMG context. However, we were able to classify it as ERA according to the risk allele-based classification (Senol-Cosar et al., 2019). Of note, this variant is classified as LP by GeneDx, and as P by Ambry, Color and Invitae diagnostic laboratories (Table S3), which could convey errors in clinical interpretation. PS3 was also not possible to apply for 2 other missense variants: c.190G>A and c.1427C>T.*CHEK2* c.190G>A is a fairly frequent variant found in 0.03% of NFE by gnomAD, with partial reduction of Thr68 phosphorylation, auto-phosphorylation and Cdc25C phosphorylation, but DNA repair assays in yeast are discordant (Table S3). Variant c.1427C>T is another relatively frequent variant present in 0.05% of NFE (gnomAD). It has been reported to affect DNA damage response in yeast at intermediate-high level. In addition, it shows reduced SOX phosphorylation almost equally to c.1100delC. However,*in vivo* and *in vitro* studies of KAP1 phosphorylation from the same group showed discordant results of pathogenicity (Table S3). As noted in Table S2, lack of robust association studies and meta-analysis of these variants hampered the possibility of applying risk allele-based classification. Both remained as VUS in any classification framework, although are classified as LP by at least 2 different reputable sources (Table S3).

To summarize, we describe here a comprehensive *CHEK2* mutational analysis in a large Spanish cohort of HC patients, providing full data of the actual prevalence of *CHEK2* pathogenic variants in our population. The frequency of LP/P variants in the HBC suspected cases in the whole gene analysis was 1.3% (9 out of 689), similar to the reported by Couch et al (Couch et al., 2017) in a study of 58,798 BC patients, in which they found 1.41% of truncating variants and 2.22% of LP/P *CHEK2* missense variants. Interestingly, 3 young CRC cases carried an LP/P *CHEK2* variant and none of them had any additional pathogenic variant in our NGS panel analysis. By this means,*CHEK2* represents the most frequently mutated gene after MMR genes in our hereditary non-polyposis CRC (HNPCC) cohort. *CHEK2*c.1100delC was reported in 6 out of 234 HNPCC families from Poland (Meijers-Heijboer et al., 2003). In their study, 3 of them also carried germline MMR P variants. In addition, c.470T>C has been found in familial CRC. To our knowledge this is the largest Spanish dataset presenting the sequencing of the whole *CHEK2* coding region together with the first attempt to apply ACMG-AMP guidelines for this gene. We detailed different strategies that can be helpful to classify VUS using different frameworks with the aim of being of help not only for the curation of *CHEK2* variants but also for other genes. We hope our work serves as a starting point to better tune ACMG criteria in the case of low-penetrance and low effect size variants associated with disease risk.

ACKNOWLEDGEMENTS

We thank CERCA Program / Generalitat de Catalunya for their institutional support. We wish also to

thank all the members of the ICO Hereditary Cancer Program.

REFERENCES

Abou Tayoun, A. N., Pesaran, T., DiStefano, M. T., Oza, A., Rehm, H. L., Biesecker, L. G., . . . SVI), C. S. V. I. W. G. C. (2018). Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. *Hum Mutat, 39* (11), 1517-1524. doi:10.1002/humu.23626

AlDubayan, S. H., Pyle, L. C., Gamulin, M., Kulis, T., Moore, N. D., Taylor-Weiner, A., . . . Lessel, D. (2019). Association of Inherited Pathogenic Variants in Checkpoint Kinase 2 (CHEK2) With Susceptibility to Testicular Germ Cell Tumors. *JAMA Oncol* . doi:10.1001/jamaoncol.2018.6477

Balmana, J., Digiovanni, L., Gaddam, P., Walsh, M. F., Joseph, V., Stadler, Z. K., . . . Domchek, S. M. (2016). Conflicting Interpretation of Genetic Variants and Cancer Risk by Commercial Laboratories as Assessed by the Prospective Registry of Multiplex Testing. *J Clin Oncol, 34* (34), 4071-4078. doi:10.1200/JCO.2016.68.4316

Bartek, J., Falck, J., & Lukas, J. (2001). CHK2 kinase–a busy messenger. *Nat Rev Mol Cell Biol, 2* (12), 877-886.

Bell, D. W., Varley, J. M., Szydlo, T. E., Kang, D. H., Wahrer, D. C., Shannon, K. E., . . . Haber, D. A. (1999). Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. *Science, 286* (5449), 2528-2531.

Bellosillo, B., Tusquets, I., Longaron, R., Perez-Lezaun, A., Bellet, M., Fabregat, X., . . . Sole, F. (2005). Absence of CHEK2 mutations in Spanish families with hereditary breast cancer. *Cancer Genet Cytogenet, 161* (1), 93-95. doi:10.1016/j.cancergencyto.2005.01.016

ClinGen-TP53_Expert_Panel. (2019, August 6, 2020). TP53 Rule Specifications for the ACMG/AMP

Variant Curation Guidelines. Retrieved from*https://www.clinicalgenome.org/affiliation/50013*

Couch, F. J., Shimelis, H., Hu, C., Hart, S. N., Polley, E. C., Na, J., . . . Dolinsky, J. S. (2017). Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. *JAMA Oncol, 3* (9), 1190-1196. doi:10.1001/jamaoncol.2017.0424

Cybulski, C., Gorski, B., Huzarski, T., Masojc, B., Mierzejewski, M., Debniak, T., . . . Lubinski, J. (2004). CHEK2 is a multiorgan cancer susceptibility gene. *Am J Hum Genet, 75* (6), 1131-1135.

Cybulski, C., Wokolorczyk, D., Huzarski, T., Byrski, T., Gronwald, J., Gorski, B., . . . Lubinski, J. (2007). A deletion in CHEK2 of 5,395 bp predisposes to breast cancer in Poland. *Breast Cancer Res Treat, 102* (1), 119-122.

Decker, B., Allen, J., Luccarini, C., Pooley, K. A., Shah, M., Bolla, M. K., . . . Easton, D. F. (2017). Rare, protein-truncating variants in.*J Med Genet, 54* (11), 732-741. doi:10.1136/jmedgenet-2017-104588

Easton, D. F., Pharoah, P. D. P., Antoniou, A. C., Tischkowitz, M., Tavtigian, S. V., Nathanson, K. L., . . . Foulkes, W. D. (2015). Gene-panel sequencing and the prediction of breast-cancer risk.*The New England journal of medicine, 372* (23), 2243-2257. doi:10.1056/NEJMsr1501341

Fachal, L., Santamarina, M., Blanco, A., Carracedo, A., & Vega, A. (2013). CHEK2 c.1100delC mutation among non-BRCA1/2 Spanish hereditary breast cancer families. *Clin Transl Oncol, 15* (2), 164-165. doi:10.1007/s12094-012-0967-z

Feliubadalo, L., Tonda, R., Gausachs, M., Trotta, J. R., Castellanos, E., Lopez-Doriga, A., . . . Lazaro, C. (2017). Benchmarking of Whole Exome Sequencing and Ad Hoc Designed Panels for Genetic Testing of Hereditary Cancer. *Sci Rep, 7* , 37984. doi:10.1038/srep37984

Fostira, F., Kostantopoulou, I., Apostolou, P., Papamentzelopoulou, M. S., Papadimitriou, C., Faliakou, E., . . . Yannoukakos, D. (2020). One in three highly selected Greek patients with breast cancer carries a

loss-of-function variant in a cancer susceptibility gene. *J Med Genet, 57* (1), 53-61. doi:10.1136/jmedgenet-2019-106189

Gutierrez-Enriquez, S., Balmana, J., Baiget, M., & Diez, O. (2008). Detection of the CHEK2 1100delC mutation by MLPA BRCA1/2 analysis: a worthwhile strategy for its clinical applicability in 1100delC low-frequency populations? *Breast Cancer Res Treat, 107* (3), 455-457. doi:10.1007/s10549-007-9555-2

Han, F. F., Guo, C. L., & Liu, L. H. (2013). The effect of CHEK2 variant I157T on cancer susceptibility: evidence from a meta-analysis.*DNA Cell Biol, 32* (6), 329-335. doi:10.1089/dna.2013.1970

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alfoldi, J., Wang, Q., . . . MacArthur, D. G. (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* , 531210. doi:10.1101/531210

Katona, B. W., Yurgelun, M. B., Garber, J. E., Offit, K., Domchek, S. M., Robson, M. E., & Stadler, Z. K. (2018). A counseling framework for moderate-penetrance colorectal cancer susceptibility genes. *Genet Med, 20* (11), 1324-1327. doi:10.1038/gim.2018.12

Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Albarca Aguilera, M., Meyer, R., & Massouras, A. (2018). VarSome: the human genomic variant search engine. *Bioinformatics, 35* (11), 1978-1980. doi:10.1093/bioinformatics/bty897

Kraus, C., Hoyer, J., Vasileiou, G., Wunderle, M., Lux, M. P., Fasching, P. A., . . . Reis, A. (2017). Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2. *Int J Cancer, 140* (1), 95-102. doi:10.1002/ijc.30428

Liang, M., Zhang, Y., Sun, C., Rizeq, F. K., Min, M., Shi, T., & Sun, Y. (2018). Association Between CHEK2*1100delC and Breast Cancer: A Systematic Review and Meta-Analysis. *Mol Diagn Ther* . doi:10.1007/s40291-018-0344-x

M Kriege, A. H., A Jager, P E A Huijts, E M Berns, A M Sieuwerts, M E Meijer-van Gelder,, & J M Collee, P. D., M J Hooning, J W M Martens and C Seynaeve. (2014). Survival and contralateral breast cancer in

CHEK2 1100delC breast cancer patients:

impact of adjuvant chemotherapy. *British Journal of Cancer, 111* , 1004–1013. doi:10.1038/bjc.2014.306

Martinez-Bouzas, C., Beristain, E., Guerra, I., Gorostiaga, J., Mendizabal, J. L., De-Pablo, J. L., . . . Tejada, M. I. (2007). CHEK2 1100delC is present in familial breast cancer cases of the Basque Country. *Breast Cancer Res Treat, 103* (1), 111-113. doi:10.1007/s10549-006-9351-4

Matsuoka, S., Rotman, G., Ogawa, A., Shiloh, Y., Tamai, K., & Elledge, S. J. (2000). Ataxia telangiectasia-mutated phosphorylates Chk2 in vivo and in vitro. *Proc Natl Acad Sci U S A, 97* (19), 10389-10394.

Meijers-Heijboer, H., van den Ouweland, A., Klijn, J., Wasielewski, M., de Snoo, A., Oldenburg, R., . . . Stratton, M. R. (2002). Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet, 31* (1), 55-59.

Meijers-Heijboer, H., Wijnen, J., Vasen, H., Wasielewski, M., Wagner, A., Hollestelle, A., . . . Schutte, M. (2003). The CHEK2 1100delC mutation identifies families with a hereditary breast and colorectal cancer phenotype. *Am J Hum Genet, 72* (5), 1308-1314.

Oldenburg, R. A., Kroeze-Jansema, K., Kraan, J., Morreau, H., Klijn, J. G., Hoogerbrugge, N., . . . Devilee, P. (2003). The CHEK2*1100delC variant acts as a breast cancer risk modifier in non-BRCA1/BRCA2 multiple-case families. *Cancer Res, 63* (23), 8153-8157.

Plon, S. E., Cooper, H. P., Parks, B., Dhar, S. U., Kelly, P. A., Weinberg, A. D., . . . Hilsenbeck, S. (2008). Genetic testing and cancer risk management recommendations by physicians for at-risk relatives.*Genet Med, 13* (2), 148-154.

Richards, C. S., Bale, S., Bellissimo, D. B., Das, S., Grody, W. W., Hegde, M. R., . . . Ward, B. E. (2008). ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007. *Genetics in medicine : official journal of the American College of Medical Genetics, 10* (4), 294-300. doi:10.1097/GIM.0b013e31816b5cae

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., . . . Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med, 17* (5), 405-424. doi:10.1038/gim.2015.30

Schmidt, M. K., Hogervorst, F., van Hien, R., Cornelissen, S., Broeks, A., Adank, M. A., . . . Easton, D. F. (2016). Age- and Tumor Subtype-Specific Breast Cancer Risk Estimates for CHEK2*1100delC Carriers. *J Clin Oncol, 34* (23), 2750-2760. doi:10.1200/JCO.2016.66.5844

Senol-Cosar, O., Schmidt, R. J., Qian, E., Hoskinson, D., Mason-Suares, H., Funke, B., & Lebo, M. S. (2019). Considerations for clinical curation, classification, and reporting of low-penetrance and low effect size variants associated with disease risk. *Genet Med, 21* (12), 2765-2773. doi:10.1038/s41436-019-0560-8

Spurdle, A. B., Healey, S., Devereau, A., Hogervorst, F. B. L., Monteiro, A. N. A., Nathanson, K. L., . . . Enigma. (2012). ENIGMA–evidence-based network for the interpretation of germline mutant alleles: an international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. *Human mutation, 33* (1), 2-7. doi:10.1002/humu.21628

Stradella, A., Del Valle, J., Rofes, P., Feliubadalo, L., Grau Garces, E., Velasco, A., . . . Lazaro, C. (2018). Does multilocus inherited neoplasia alleles syndrome have severe clinical expression? *J Med Genet* . doi:10.1136/jmedgenet-2018-105700

Szymanska-Pasternak, J., Szymanska, A., Medrek, K., Imyanitov, E. N., Cybulski, C., Gorski, B., . . . Lubinski, J. (2006). CHEK2 variants predispose to benign, borderline and low-grade invasive ovarian tumors.*Gynecol Oncol, 102* (3), 429-431. doi:10.1016/j.ygyno.2006.05.040

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

Taylor, A., Brady, A. F., Frayling, I. M., Hanson, H., Tischkowitz, M., Turnbull, C., . . . (UK-CGG), U. C. G. G. (2018). Consensus for genes to be included on cancer panel tests offered by UK genetics services: guidelines of the UK Cancer Genetics Group. *J Med Genet, 55* (6), 372-377. doi:10.1136/jmedgenet-2017-105188

Wu, X., Webster, S. R., & Chen, J. (2001). Characterization of tumor-associated Chk2 mutations. *J Biol Chem, 276* (4), 2971-2974.

Zannini, L., Delia, D., & Buscemi, G. (2014). CHK2 kinase in the DNA damage response and beyond. *J Mol Cell Biol, 6* (6), 442-457. doi:10.1093/jmcb/mju045

TABLES

FIGURE LEGENDS

Figure 1. Diagram of the study

BC, breast cancer; HBOC, hereditary breast and ovarian cancer; PBL, peripheral blood lymphocytes; CSCE, conformation-sensitive capillary electrophoresis; MLPA, multiplex, ligation-dependent probe amplification; NGS, next generation sequencing; P, pathogenic; LP, likely pathogenic; VUS, variant of unknown significance; ERA, established risk allele.

Figure 2. Schematic representation of *CHEK2* variants found in our cohort

Color code: dark red: pathogenic; red: likely pathogenic; pink: established risk allele; yellow: variant of uncertain significance. Shape code: diamond: nonsense variants; triangle: frameshift variants; square: splicing variants; circle: missense variants; star: copy number variants. Solid horizontal lines correspond to a copy number variant, each found in 1 index case.

Figure 3. mRNA analysis of *CHEK2* E3-E4dup

Top, schematic representation of *CHEK2* E3-E4dup. cDNA amplification showed a double band, one corresponding to the full-length transcript (708 bp) and the other to the transcript carrying the duplication (981 bp), as shown in the electropherogram on the bottom left. Bottom right, agarose gel of a carrier and 2 controls with (P+) and without puromycin (P-).

Figure 4. mRNA analysis of *CHEK2* c.320-5T>A

Top, schematic representation of *CHEK2* c.320-5T>A splicing effect. cDNA amplification showed a double band, one corresponding to the full-length transcript (860 bp) and the other to the transcript lacking exons 3 and 4 (587 bp), as shown in the electropherogram on the bottom left. Bottom right, agarose gel of a carrier and 2 controls with (P+) and without puromycin (P-).

Figure S1 a-c: Pedigrees from families carrying 16 *CHEK2*variants discussed

Filled symbol, cancer confirmed by pathologist report; partially filled symbol, cancer referred by relative; arrow, index case. Cosegregation results are indicated with the name of the variant if present and WT for non-carriers. Current ages and ages at death, when available, are indicated on the top-left corner of each individual's symbol. BC, breast cancer; BlC, bladder cancer; BrC, brain cancer; CRC, colorectal cancer; EC, endometrial cancer GC, gastric cancer; HFN, head/face/neck cancer; KC, kidney cancer; LC, lung cancer; Leu, leukemia; LiC, Liver cancer; Lym, Lymphoma; OC, ovarian cancer; PC, pancreas cancer; Para, parathyroid cancer; PCC, pheochromocytoma; PrC, prostate cancer; SC, skin cancer; SA, sebaceous adenoma; SAR, sarcoma; T, thyroid cancer; TeC, testicular cancer.