RESEARCH ARTICLE

Mutations of the CHEK2 gene in patients with cancer and their presence in the Latin American population [version 1; peer review: 3 approved with reservations]

Sandra Guauque-Olarte ¹, Ana-Lucia Rivera-Herrera², Laura Cifuentes-C ¹

¹GIOD Group, Faculty of Dentistry, Universidad Cooperativa de Colombia, Pasto, Colombia

²Human Molecular Genetics Lab, Biology Department, Universidad del Valle, Cali, Colombia

V1 First published: 29 Nov 2016, 5:2791
<https://doi.org/10.12688/f1000research.9932.1>
 Latest published: 29 Nov 2016, 5:2791
<https://doi.org/10.12688/f1000research.9932.1>

Abstract

Background: *CHEK2* (Checkpoint Kinase 2) encodes CHK2, a serine/threonine kinase involved in maintaining the G1/S and G2/M checkpoints and repair of double-strand DNA breaks via homologous recombination. Functions of CHK2 include the prevention of damaged cells from going through the cell cycle or proliferating and the maintenance of chromosomal stability. *CHEK2* mutations have been reported in a variety of cancers including glioblastoma, ovarian, prostate, colorectal, gastric, thyroid, and lung cancer in studies performed mainly in White populations. The most studied mutation in *CHEK2* is c.1100delC, which was associated with increased risk of breast cancer. The objective of this study was to compile mutations in *CHEK2* identified in cancer genomics studies in different populations and especially in Latin American individuals.

Methods: A revision of cancer genomics data repositories and a profound literature review of Latin American studies was performed.

Results: Mutations with predicted high impact in *CHEK2* were reported in studies from Australia, Japan, United States, among other countries. The TCGA cancer types with most mutations in *CHEK2* were breast, colorectal, and non-small cell lung cancer. The most common mutation found was E321* in three patients with uterine cancer. In Latin American individuals nine mutations were found in melanoma, lymphoma, and head and neck cohorts from TCGA and ICGC. Latin American studies have been restricted to breast and colorectal cancer and only two mutations out of four that have been interrogated in this population were identified, namely c.1100delC and c.349A>G.

Conclusions: This study presents a compilation of mutations in *CHEK2* with high impact in different cancer types in White, Hispanic and other populations. We also show the necessity of screening *CHEK2* mutations in Latin American in cancer types different than breast and colorectal.

Open Peer Review

Approval Status

	1	2	3
version 1			
29 Nov 2016	view	view	view

1. **Claire Palles**, NIHR Comprehensive Biomedical Research Centre, Oxford, UK
2. **Muhammad Usman Rashid**, Department of Basic Sciences Research, Lahore, Pakistan
3. **Ewa Grzybowska**, Centre of Oncology-MSK memorial Institute, Gliwice Branch, Wybrzeze Armii Krajowej 15, 44-101, Poland

Any reports and responses or comments on the article can be found at the end of the article.

Keywords

CHEK2 , CHK2 , cancer , Latin America , databases , mutations ,
CHEK2*1100delC , genomics

Corresponding author: Laura Cifuentes-C (lauracifuentes@gmail.com)

Competing interests: No competing interests were disclosed.

Grant information: LCC received funding by CONADI-Universidad Cooperativa de Colombia (Grant ID1450).

Copyright: © 2016 Guauque-Olarte S *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Guauque-Olarte S, Rivera-Herrera AL and Cifuentes-C L. **Mutations of the CHEK2 gene in patients with cancer and their presence in the Latin American population [version 1; peer review: 3 approved with reservations]** F1000Research 2016, 5:2791 <https://doi.org/10.12688/f1000research.9932.1>

First published: 29 Nov 2016, 5:2791 <https://doi.org/10.12688/f1000research.9932.1>

Introduction

CHEK2 (Checkpoint Kinase 2) (OMIM +604373) encodes CHK2 a serine/threonine kinase that is the human homolog of *Saccharomyces cerevisiae* RAD53 and *Schizosaccharomyces pombe* CDS1¹. In mammalian cells, ATM activates CHK2 in response to ionizing radiation through phosphorylation. This leads to a variety of cellular responses, such as cell cycle checkpoint activation², where CHK2 is involved in maintaining the G1/S and G2/M checkpoints by phosphorylation of CDC25A, CDC25C and p53³ and in the repair of double-strand DNA breaks via homologous recombination (HR) through phosphorylation of BRCA1⁴ and BRCA2⁵. CHK2 is also involved in the induction of p53-dependent apoptosis through phosphorylation of p53 on Ser20⁶, and, in a p53-independent manner, via phosphorylation of PML and E2F1³. These responses prevent damaged cells from going through the cell cycle or proliferating. CHK2 also plays an important role during mitosis by maintaining chromosomal stability⁷.

CHEK2 c.1000delC, a truncating mutation in exon 10 that abolishes kinase activity of the protein, was the first mutation being reported for this gene and was found in a woman with breast cancer and family history of Li-Fraumeni syndrome-2⁸. The role of this mutation in breast cancer was confirmed by Meijers-Heijboer *et al.*⁹ and in several other studies^{10–22}. Based on these studies, *CHEK2* has been proposed as a moderate penetrance breast cancer susceptibility gene⁹ and mutations in this gene are associated with almost a 3-fold increase in the risk of breast cancer in women and a 10-fold increase in the risk of breast cancer in men²³.

Given the role of *CHEK2* in maintaining genomic stability and the fact that the *CHEK2* protein is expressed in a wide range of tissues, it was not surprising that alterations in this protein were found in other cancers, including glioblastoma, ovarian, prostate, colorectal, gastric, thyroid, and lung cancer^{18,24–28}. The studies in *CHEK2* included individuals mainly from the United States and Europe while Latin American individuals were underrepresented. In order to infer the role of the *CHEK2* gene in the cancer etiology in the Latin American population we compiled mutations in the *CHEK2* gene registered in genomics data repositories and the literature, that had been reported in this population.

Methods

Search of cancer genomics data repositories and the GWAS catalog

Mutations in *CHEK2* were identified in The Exome Aggregation Consortium (ExAC, RRID:SCR_004068, <http://exac.broadinstitute.org/>)²⁹ browser, the Cancer Genome Atlas (TCGA, RRID:SCR_003193)³⁰ data sets extracted from the cBioPortal for Cancer Genomics (RRID:SCR_014555, www.cbioportal.org/)³¹, and The International Cancer Genome Consortium (ICGC) (<http://icgc.org/>)³². From the GWAS catalog (RRID:SCR_012745, <https://www.ebi.ac.uk/gwas/>)³³ a list of SNPs mapped to *CHEK2* and associated with a disease was also downloaded. Data obtained from cell line studies was not included.

ICGC, the cBioportal and ExAc use prediction tools to assess functional impact of non-synonymous (SO term: missense_variant) somatic mutations on protein coding genes. ICGC uses FatHMM

(<http://fathmm.biocompute.org.uk/>)³⁴, Mutation Assessor (RRID:SCR_005762)³⁵ and SIFT (RID:SCR_012813)³⁶ to compute functional impact scores and assign impact categories (High, Medium, Low and Unknown). The cBioPortal uses Mutation Assessor and reports the same impact categories. We used those functional impact categories to filter the mutations and extract possible pathogenic mutations by selecting only high and medium impact mutations and nonsense alterations. The percentage of mutations in *CHEK2* per cancer study and the percentage of cases altered per cancer type was also calculated. The filter used for the ExAC information was based on the annotation of possible damaging and deleterious mutations made by two *in silico* tools: Polyphen2 (RID:SCR_013200)³⁷ and SIFT³⁶. The assessment of stop gained, splice site disrupting and frameshift variants was made through Loss of Function Transcript Effect Estimator (LOFTEE), a plugin of the Ensembl Variant Effect Predictor (VEP) (RRID SCR_007931)³⁸. The Latino annotation was examined in the databases that reported ethnicity data; this search was done before filtering the datasets, with the purpose to report all genetic alterations found in Latin American populations.

The plots were generated with R version 3.3.1 (RRID:SCR_001905)³⁹.

Literature review of Latin American studies

In order to include all the studies identifying *CHEK2* gene mutations in Latin America, a deep search of literature was conducted by using the terms “CHEK2”, “CHEK2 Latin America”, and “CHEK2 cancer” in electronic academic literature search engines. PUBMED (RRID:SCR_004846) was the relevant database used followed by Google Scholar (RRID:SCR_008878). References of the retrieved articles were also screened for relevant studies. This search strategy was performed iteratively up to and including 10 October 2016.

Results

The complete list of mutations in *CHEK2* reported in the cBioPortal and ICGC, before applying filters, are available in **Dataset 1** and **Dataset 2**, respectively.

Dataset 1. A complete list of mutations, before applying filters, in *CHEK2* reported in the cBioPortal

<http://dx.doi.org/10.5256/f1000research.9932.d142129>

Dataset 2. A complete list of mutations, before applying filters, in *CHEK2* reported in the ICGC

<http://dx.doi.org/10.5256/f1000research.9932.d142130>

CHEK2 mutations in the data genomics repositories

cBioPortal. The available data sets consisted of 147 studies that included only cancer samples. Mutations in *CHEK2* were reported in 39 out of the 147 studies. Before applying filters, cholangiocarcinoma (8.6%), uterine carcinosarcoma (7.0%), and colorectal adenocarcinoma (6.9%) were the types of cancer that showed the higher number of cases (**Figure 1**); meanwhile, breast, colorectal and non-small cell lung cancer (NSCLC) had more mutations in *CHEK2* than other cancer types (**Figure 2**).

Figure 1. Percentage of cases with mutations in *CHEK2* per cancer study. The X axis shows the type of cancer in which at least one case has a mutation in *CHEK2*, the Y axis indicates the percentage of cases per study that have mutations in *CHEK2* (source: cBioPortal).

Figure 2. Percentage of mutations in *CHEK2* per cancer study. The X axis shows the type of cancer in which at least one mutation in *CHEK2* was identified, the Y axis indicates the percentage of mutations in *CHEK2* per cancer type (source: cBioPortal). n unique mutations = 159. Synonymous mutations are not included in the cBioPortal database.

Using the Mutation Assessor from cBioPortal, we filtered out mutations labeled to have neutral and low impact. In **Table 1** we are reporting the mutations with high and medium impact and also nonsense mutations and frameshifts. **Table 1** shows the 78 mutations that remained after the filtering process, 38 of which were classified as with high impact. 51.2% of mutations were missense mutations, 20.5% were frameshift mutations, 19.2% were nonsense mutations and 9% were in splice sites. The type of cancer with most mutations (13/78) was breast cancer, followed by uterine, lung, and colorectal cancer. The rest of cancer types had six or less mutations. The most frequent mutation was E321* reported in three patients with uterine cancer.

Before filtering the mutations found in the cBioPortal we identified Latino individuals with the ethnicity data obtained from the TCGA clinical data available at the NCI's Genomic Data Commons portal (GDC, RRID:SCR_014514, <https://gdc-portal.nci.nih.gov/>) (**Table 2**). Two patients with three mutations in the gene were found. One of the samples was a Latino patient from the head and neck squamous cell carcinoma cohort (HNSC); this patient carries the neutral variant K373E. Because this is a neutral variant it was not included in **Table 2**. The second Latino patient was part of the diffuse large B-cell lymphoma (DLBC) cohort; this patient carries a frameshift and a nonsense mutation.

Table 1. Mutations in *CHEK2* identified in TCGA studies after filtering out neutral and low impact mutations.

AA change	No of patients	Type	Ethnicity	Type of cancer
A247T	1	Missense	white_not hispanic or latino	Uterine cancer
A392V	2	Missense	white_not reported	Lung cancer
			No data availableNo data available	Melanoma
A480T	1	Missense	white_not hispanic or latino	Glioma and glioblastoma
A540Cfs*9	1	FS del	No data availableNo data available	Colorectal cancer
A98Mfs*13	1	FS ins	white_hispanic or latino	Lymphoid neoplasm diffuse large B-cell lymphoma [DLBC]
D208Ifs*9	1	FS del	white_not hispanic or latino	Lung cancer
E122*	1	Nonsense	No data available	Liver cancer
E275*	1	Nonsense	not reported_hispanic or latino	Uterine cancer
			No data availableNo data available	Colorectal cancer
E321*	3	Nonsense	native hawaiian or other pacific islander_not hispanic or latino	Uterine cancer
			asian_not hispanic or latino	
E351D	2	Missense	No data available	Colorectal cancer
			asian_not hispanic or latino	Uterine cancer
E394Kfs*20	1	FS del	white_not hispanic or latino	Stomach cancer
E457Rfs*33	1	FS ins	white_not hispanic or latino	Breast cancer
E478*	1	Nonsense	white_not hispanic or latino	Thymoma
F103L	2	Missense	white_not reported	Esophageal cancer
			No data available	Renal cancer
F144L	1	Missense	not reported_not reported	Colorectal cancer
F202Lfs*3	1	FS del	not reported_not reported	Stomach cancer

AA change	No of patients	Type	Ethnicity	Type of cancer
F310V	1	Missense	No data available	Breast cancer
G232R	1	Missense	No data available	Esophageal cancer
G259Wfs*13	1	FS ins	No data available	Breast cancer
G306W	1	Missense	white_not hispanic or latino	Cervical cancer
G342V	1	Missense	No data available	Prostate cancer
G386V	1	Missense	white_not hispanic or latino	Lung cancer
H143N	1	Missense	white_not hispanic or latino	Renal cancer
H143Q	1	Missense	asian_not reported	Thyroid cancer
H282D	1	Missense	white_not hispanic or latino	Lung cancer
H339Y	1	Missense	white_not hispanic or latino	Bladder cancer
H345L	1	Missense	white_not reported	Esophageal cancer
H54Pfs*6	1	FS del	No data available	Colorectal cancer
I160M	1	Missense	No data available	Ewing sarcoma
I276S	1	Missense	white_not reported	Uterine cancer
I364T	1	Missense	No data available	Breast cancer
I419Yfs*4	1	FS ins	No data available	Colorectal cancer
K520Afs*4	1	FS del	No data available	Renal cancer
L226F	2	Missense	No data available	Head and Neck cancer
			white_not hispanic or latino	Sarcoma
L338R	1	Missense	white_not reported	Uterine cancer
L354V	1	Missense	white_not reported	Uterine cancer
L466Ffs*3	1	FS del	not reported_not reported	Lung cancer
N166S	1	Missense	white_not hispanic or latino	Head and Neck cancer
N185Y	1	Missense	No data available	Lung cancer
N290Tfs*14	1	FS del	white_not hispanic or latino	Liver cancer
N316Efs*2	1	FS ins	No data available	Colorectal cancer
P182H	1	Missense	No data available	Stomach cancer
P182S	1	Missense	No data available	Colorectal cancer
Q100*	1	Nonsense	white_hispanic or latino	Lymphoid neoplasm diffuse large B-cell lymphoma [DLBC]
Q29*	1	Nonsense	No data available	Breast cancer
Q36*	1	Nonsense	No data available	Colorectal cancer
R117G	1	Missense	black or african american_not hispanic or latino	Lung cancer
R318C	1	Missense	No data available	Colorectal cancer
R346C	1	Missense	white_not reported	Breast cancer

AA change	No of patients	Type	Ethnicity	Type of cancer
R346G	1	Missense	white_not hispanic or latino	Breast cancer
R346H	2	Missense	No data available	Breast cancer
			white_not reported	Ovarian cancer
R346S	1	Missense	No data available	Head and Neck cancer
R474H	1	Missense	white_not hispanic or latino	Stomach cancer
R95*	1	Nonsense	white_not hispanic or latino	Adrenocortical carcinoma
S210*	1	Nonsense	white_not reported	Uterine cancer
S223*	1	Nonsense	No data available	Breast cancer
S356*	1	Nonsense	white_not hispanic or latino	Bladder cancer
S372F	1	Missense	white_not hispanic or latino	Head and neck cancer
S456*	1	Nonsense	No data available	Pleural mesothelioma
S52P	1	Missense	No data available	Lung cancer
S55P	1	Missense	not reported_not reported	Liver cancer
T225Lfs*10	1	FS del	No data available	Colorectal cancer
T323Lfs*14	1	FS del	white_not hispanic or latino	Head and Neck cancer
T367Mfs*15	1	FS del	not reported_not reported	Breast cancer
V9A	1	Splice	white_not hispanic or latino	Uterine cancer
W97*	1	Nonsense	No data available	Breast cancer
W485*	1	Nonsense	No data available	Breast cancer
W93L	1	Missense	No data available	Lung cancer
X107_splice	1	Splice	No data available	Esophageal cancer
X198_splice	1	Splice	white_not hispanic or latino	Stomach cancer
X365_splice	1	Splice	white_not hispanic or latino	Lung cancer
X366_splice	1	Splice	asian_not hispanic or latino	Stomach cancer
X420_splice	2	Splice	No data available	Bladder cancer
			asian_not hispanic or latino	Uterine cancer
X514_splice	1	Splice	white_not hispanic or latino	Uterine cancer
Y123C	1	Missense	No data available	Breast cancer
Y404C	1	Missense	not reported_not reported	Lung cancer
Y337D	1	Missense	white_not hispanic or latino	Lung cancer
Y390*	1	Nonsense	white_not hispanic or latino	Thyroid cancer

Table 2. Mutations found in three cancer genomics data repositories for Latin American populations.

Database	Genomic DNA change	AA change*	Effect	cancer type	Population	Frequency in the sample
ICGC	22:29121093	S155F	Missense	Melanoma	Brazil	1/70 (1.43%)
	22:29138126		5 UTR			1/70 (1.43%)
	22:29091835	I117M	Missense			1/70 (1.43%)
	22:29138096		5 UTR			1/70 (1.43%)
TCGA	22:29130418	A98Mfs*13	Frameshift	Diffuse large B-cell lymphoma [DLBC]	Latino	0,42%
	22:29130412	Q100*	Nonsense	Diffuse large B-cell lymphoma [DLBC]	Latino	0,42%
	22:29091840	K373E	Missense	Head and neck cancer	Latino	3,10%
ExAC	22:29090018	c.1590+2T>G	splice donor	NA	Latino	8,81E-02
	22:29090030	p.Pro527Leu	missense	NA	Latino	0.0004404
	22:29090054	p.Thr519Met	missense	NA	Latino	8,81E-02
	22:29091225	p.Ser465Asn	missense	NA	Latino	8,99E-02
	22:29091768	p.Val440PhefsTer17	frameshift	NA	Latino	8,66E-02
	22:29091856	p.Thr410MetfsTer15*	frameshift	NA	Latino	0.0001733
	22:29092960	p.Gly385Ser	missense	NA	Latino	8,76E-02
	22:29099495	p.Glu345Asp	missense	NA	Latino	0.0008772
	22:29107942	p.Lys292Asn	missense	NA	Latino	8,64E-02
	22:29107982	p.Leu279Pro	missense	NA	Latino	0.003112
	22:29107994	p.Glu56Ter	stop gained	NA	Latino	8,65E-02
	22:29115401	p.Met265AsnfsTer24	frameshift	NA	Latino	0.0004003
	22:29115403	p.Ile264Met	missense	NA	Latino	0.0001938
	22:29120978	p.Leu236LysfsTer4	frameshift	NA	Latino	8,65E-02
	22:29121000	p.Asn229Ser	missense	NA	Latino	0.0001729
	22:29121078	p.Ile203Arg	missense	NA	Latino	8,64E-02
	22:29121229	c.573+2T>G	splice donor	NA	Latino	8.64e-05
	22:29121242	p.Arg188Trp	missense	NA	Latino	8,64E-02
	22:29121326	p.Arg160Gly	missense	NA	Latino	8,65E-02
	22:29130427	p.Arg95Ter	stop gained	NA	Latino	8,67E-02
	22:29130430	p.Ala94Ser	missense	NA	Latino	8,67E-02
	22:29130431	p.Trp93Ter	stop gained	NA	Latino	8,67E-02
	22:29130576	p.Thr45Met	missense	NA	Latino	8,65E-02

*The nomenclature used for the mutation annotation is as follow: ICGC (ENST00000328354), ExAC (NP_665861) and TCGA (NP_009125).

ICGC

A total of 279 mutations including up- and down-stream mutations were reported in 185 donors. From this number, seven mutations are predicted to have high impact (Table 3). For the Latin American population in ICGC, the Brazilian melanoma study (SKCA-BR) reported four mutations inside the gene, one of them with high impact (Table 2 and Table 3).

ExAC browser

A total of 742 mutations for the CHEK2 gene were reported in this database and 132 of them were present in the Latino population before filters (Dataset 3). After applying the filter of possibly damaging and deleterious alterations, 23 mutations in the Latino population were left. In this group the mutation p.Leu279Pro was the most frequent (0.003112). CHEK2 c.1100delC (p.Thr410MetfsTer15*), the most interrogated mutation in CHEK2, was found in two samples (Table 2).

Dataset 3. Mutations in CHEK2 identified in Latino American samples before applying filters (source:ExAC)

<http://dx.doi.org/10.5256/f1000research.9932.d142131>

GWAS catalog

Mutations rs132390-C and rs17879961-A mapped to or near CHEK2 were associated in European populations with breast and lung cancer, respectively. Mutations rs4822983-T and rs2239815-T were associated with esophageal squamous cell carcinoma in

individuals with Han Chinese ancestry. In addition, in a Han Chinese cohort of esophageal and gastric cancer the mutation rs738722-T was also associated with those cancers (Dataset 4).

Dataset 4. Variants reported in CHEK2 that have been associated with cancer according to data in the GWAS catalog

<http://dx.doi.org/10.5256/f1000research.9932.d142132>

All of these variants were found in the cBioPortal or ICGC data.

CHEK2 mutations in Latinos reported in the literature

In total, we found nine studies in which mutations in CHEK2 were evaluated in Latino populations. Two of these studies were international and included Latin American cancer patients^{10,22} and the other six studies were country-based. The country in which most studies have been performed was Brazil with four studies⁴⁰⁻⁴³. In Argentina⁴⁴, Chile⁴⁵, and Mexico⁴⁶ one study per country was identified. In eight out of the nine studies, the presence of variants in CHEK2 was interrogated in breast cancer patients. Only one study used samples of patients with hereditary breast and colorectal cancer. The mutation most frequently evaluated in these investigations was c.1100delC (in six studies); while other two studies^{42,44} interrogated the other two most frequent mutations in the CHEK2 gene (c.470T>C and c.444+IG>A) in addition to c.1100delC. Additionally, Chaudury *et al.* performed a complete sequencing of the gene and found a different mutation, c.478A>G (p.Arg160Gly)⁴⁶. Table 4 shows the Latin American studies that reported the presence of mutations in CHEK2 mutations and their frequency.

Table 3. Mutations in CHEK2 (5' to 3'UTR) with high impact in the ICGC portal excluding TCGA data.

Genomic DNA change	Functional impact	Consequences*	Donors affected
22:29099515C>T	high	CHEK2 D229N, D47N, D205N, D39N, D75N, D296N, D339N	MELA-AU:1/183
22:29099522A>C	high	CHEK2 D202E, D36E, D336E, D293E, D226E, D72E, D44E 3 UTR: CHEK2 Exon: CHEK2 Intron: CHEK2	MELA-AU:1/183
22:29106038G>A	high	CHEK2 L268F, L177F, L311F, L47F, L11F, L19F, L201F 3 UTR: CHEK2 Exon: CHEK2 Intron: CHEK2	MELA-AU:1/183
22:29121093G>A	high	Missense: CHEK2 S155F, S186F, S198F, S165F	SKCA-BR:1/70
22:29121242G>A	high	Missense: CHEK2 R188W, R176W, R145W, R155W Start Gained: CHEK2 3 UTR: CHEK2 Exon: CHEK2 Intron: CHEK2	BRCA-EU:1/560 ESAD-UK:1/203
22:29130564G>A	high	CHEK2 S59F, S49F 5 UTR: CHEK2 Exon: CHEK2 Intron: CHEK2	LINC-JP:1/244
22:29130703G>A	high	Missense: CHEK2 R13W, R3W Start Gained: CHEK2 Exon: CHEK2 Intron: CHEK2	BRCA-EU:1/560 BRCA-FR:1/64

*Depending of transcript. All mutations are single base substitutions. MELA-AU: melanoma, Australia. BRCA-EU: breast ER+ and HER-cancer, European Union. ESAD-UK: esophageal adenocarcinoma, United Kingdom. SKCA-BR: skin adenocarcinoma, Brazil. LINC-JP: liver cancer, Japan. BRCA-FR: breast cancer, France.

Table 4. Mutations in *CHEK2* reported in the literature for Latin American populations.

Position	dbSNP	Nucleotide change	AA change	Effect	Population	Cancer type	Carriers n (frequency of carriers, %)		Source	Reference
							Case	Control		
22:29091857	rs555607708	c.1100delC	p.Thr367Metfs	Stop codon	Latin-American	Breast cancer	1/362 (0.0027)	0/384 (0%)	Blood	Bell <i>et al.</i> (2007)
					Brazil	Breast cancer	1/155 (0.7%)	0/377 (0%)	Blood	Zhang <i>et al.</i> (2008)
					Brazil	Hereditary Breast and Colorectal cancer	1/59 (1.7%)	0	Blood	Abud <i>et al.</i> (2012)
					Brazil	Breast cancer predisposition syndromes (BCPS)	1/7 families (14.3%)	0	Blood	Palmero <i>et al.</i> (2016)
22:29121326	rs28909982	c.478A>G	p.Arg160Gly	Missense	Mexico	Breast cancer	2/92 (2.17%)	0	Blood	Chaudury <i>et al.</i> (2013)

Dataset 5. Number of individuals per cancer study and ethnicity in the TCGA cohort

<http://dx.doi.org/10.5256/f1000research.9932.d142133>

Only studies in which at least one mutation in *CHEK2* was found were included.

Discussion

A search in cancer genomics data repositories and the literature was performed to identify mutations in *CHEK2* in different cancer types, with specific emphasis on mutations found in Latino American populations. The database with the most number of mutations reported in *CHEK2* for Latino populations was ExAC with 132 mutations, followed by ICGC with four mutations, and TCGA with three mutations. After filtering 30 mutations with high and medium impact according to the databases functional impact categories were kept: seventeen missense, eight ‘stop gain’ mutations, one frameshift mutation, two mutations in the 5’UTR, and two mutations in splice donor sites of *CHEK2*. These mutations included the most analyzed mutation of *CHEK2*, c.1100delC (p.Thr367Metfs) (Table 2).

Worldwide, according to our findings in the ICGC and TCGA databases, *CHEK2* mutations were reported in 23 cancer types, while in the Latin American population *CHEK2* mutations were only found in head and neck cancer, lymphoma and melanoma. In this context, it is important to highlight, that Latino populations have been underrepresented in other worldwide studies. As shown in Dataset 4, the cohorts of TCGA are biased toward the inclusion of white individuals and individuals from other ethnicities are underrepresented. The same was observed in ICGC in which only a Latin American cohort from Brazil was available for our analysis. Regarding the data found in our literature review, *CHEK2* has only been studied in the Latin American population in breast and colorectal cancer.

In the ExAC repository, the mutations c.1100delC and c.478A>G were found two times and one time, respectively, in the Latino population (Dataset 3). In TCGA, c.1100delC was found in a patient with breast cancer but information about its ethnicity was not available (Table 1). Up to now, only nine studies evaluating mutations in *CHEK2* have been performed in Latin America and only six of them found mutations in the gene, five studies found the c.1100delC mutation and one found the c.478A>G (p.Arg160Gly)^{10,22,40,43,46}. Two mutations, c.1100delC and c.478A>G, were classified in the ClinVar archive (<https://www.ncbi.nlm.nih.gov/clinvar/>) as pathogenic and likely pathogenic, respectively. These mutations are the only ones in common with the mutations found in genomics data repositories.

Although c.1100delC is the *CHEK2* mutation most evaluated in the Latin American population, it should be noted that its frequency, seen from literature reports and data repositories, is rather low. Because the highest frequency of this mutation is found in populations from the Northern and Western Europe, c.1100delC is proposed as an allele with population gradient, which originated

in these populations and its frequency decreases as you get to the southern regions of Europe (Basque Country, Spain, and Italy)⁴⁷. Taking into account the European genetic component of Latin American populations, it is expected that if the frequency of c.1100delC is low in the Spanish population, in our mixed populations the frequency would be even lower.

Because cancer types other than breast and colorectal cancer, such as uterine, lung, bladder and head and neck cancer, presented mutations in *CHEK2* in several populations, it is relevant to focus the search for mutations in these types of cancer in the Latin American populations. Additionally, the interrogation of *CHEK2* mutations in the Latin American population has been focused mainly on the c.1100delC mutation, but the data obtained from the ExAC database showed that in Latin American samples there are 23 germline mutations (Table 2) that could generate cancer susceptibility. It would therefore be important to examine the frequencies of these mutations in the Latin American population and its association with the development of cancer.

This study has limitations; for example, information about race and ethnicity was not available for at least 28 studies in the cBioPortal, and consequently some Latinos may be hidden in those studies. Thus, the small number of Latinos included in the genomics data repositories could be a reason why we have found a small number of mutations in *CHEK2* in this population. It is important to highlight that the use of different transcripts for reporting mutations makes the correlation between mutations found in different studies laborious.

This study presents a compilation of mutations in *CHEK2* with high impact in different cancer types in White, Hispanic and other populations. We also showed the necessity of performing studies in Latin American in cancer types different than breast and colorectal and a screening of other mutations in addition to the most popular mutations analyzed, such as c.1100delC.

Data availability

F1000Research: Dataset 1: A complete list of mutations, before applying filters, in *CHEK2* reported in the cBioPortal [10.5256/f1000research.9932.d142129](https://doi.org/10.5256/f1000research.9932.d142129)⁴⁸.

F1000Research: Dataset 2: A complete list of mutations, before applying filters, in *CHEK2* reported in the ICGC [10.5256/f1000research.9932.d142130](https://doi.org/10.5256/f1000research.9932.d142130)⁴⁹.

F1000Research: Dataset 3: Mutations in *CHEK2* identified in Latino American samples before applying filters (source:ExAC) [10.5256/f1000research.9932.d142131](https://doi.org/10.5256/f1000research.9932.d142131)⁵⁰.

F1000Research: Dataset 4: Variants reported in *CHEK2* that have been associated with cancer according to data in the GWAS catalog. All of these variants were found in the cBioPortal or ICGC data [10.5256/f1000research.9932.d142132](https://doi.org/10.5256/f1000research.9932.d142132)⁵¹.

F1000Research: Dataset 5: Number of individuals per cancer study and ethnicity in the TCGA cohort. Only studies in which at least one mutation in *CHEK2* was found were included [10.5256/f1000research.9932.d142133](https://doi.org/10.5256/f1000research.9932.d142133)⁵².

Author contributions

Conception and design of the work: CCL, GOS and RHAL. Data collection: RHAL and GOS. Data analysis: CCL, GOS and RHAL. Drafting of the article and critical revision: CCL, GOS, and RHAL. All authors were involved in the revision of the draft manuscript and have agreed to the final content.

Competing interests

No competing interests were disclosed.

Grant information

LCC received funding by CONADI-Universidad Cooperativa de Colombia (Grant ID1450).

References

- Brown AL, Lee CH, Schwarz JK, *et al.*: A human Cds1-related kinase that functions downstream of ATM protein in the cellular response to DNA damage. *Proc Natl Acad Sci U S A.* 1999; **96**(7): 3745–50. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zannini L, Delia D, Buscemi G: CHK2 kinase in the DNA damage response and beyond. *J Mol Cell Biol.* 2014; **6**(6): 442–57. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Jekimovs CR, Chen X, Arnold J, *et al.*: Low frequency of CHEK2 1100delC allele in Australian multiple-case breast cancer families: functional analysis in heterozygous individuals. *Br J Cancer.* 2005; **92**(4): 784–90. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lee JS, Collins KM, Brown AL, *et al.*: hCds1-mediated phosphorylation of BRCA1 regulates the DNA damage response. *Nature.* 2000; **404**(6774): 201–4. [PubMed Abstract](#) | [Publisher Full Text](#)
- Bahassi EM, Ovesen JL, Riesenberger AL, *et al.*: The checkpoint kinases Chk1 and Chk2 regulate the functional associations between hBRCA2 and Rad51 in response to DNA damage. *Oncogene.* 2008; **27**(28): 3977–85. [PubMed Abstract](#) | [Publisher Full Text](#)
- Shieh SY, Ahn J, Tamai K, *et al.*: The human homologs of checkpoint kinases Chk1 and Cds1 (Chk2) phosphorylate p53 at multiple DNA damage-inducible sites. *Genes Dev.* 2000; **14**(3): 289–300. [PubMed Abstract](#) | [Free Full Text](#)
- Stolz A, Ertch N, Kienitz A, *et al.*: The CHK2-BRCA1 tumour suppressor pathway ensures chromosomal stability in human somatic cells. *Nat Cell Biol.* 2010; **12**(5): 492–9. [PubMed Abstract](#) | [Publisher Full Text](#)
- Bell DW, Varley JM, Szlydo TE, *et al.*: Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. *Science.* 1999; **286**(5449): 2528–31. [PubMed Abstract](#) | [Publisher Full Text](#)
- Meijers-Heijboer H, van den Ouweland A, Klijn J, *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. [PubMed Abstract](#) | [Publisher Full Text](#)
- Bell DW, Kim SH, Godwin AK, *et al.*: Genetic and functional analysis of CHEK2 (CHK2) variants in multiethnic cohorts. *Int J Cancer.* 2007; **121**(12): 2661–7. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- CHEK2 Breast Cancer Case-Control Consortium: CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. *Am J Hum Genet.* 2004; **74**(6): 1175–82. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cybulski C, Wokolorczyk D, Huzarski T, *et al.*: A deletion in CHEK2 of 5,395 bp predisposes to breast cancer in Poland. *Breast Cancer Res Treat.* 2007; **102**(1): 119–22. [PubMed Abstract](#) | [Publisher Full Text](#)
- De Jong MM, van der Graaf W, Nolte IM: Increased CHEK2 1100delC genotype frequency (also) in unselected breast cancer patients. *J Clin Oncol.* 2004; **22**(suppl): 844s. [Reference Source](#)
- Ghadirian P, Robidoux A, Zhang P, *et al.*: The contribution of founder mutations to early-onset breast cancer in French-Canadian women. *Clin Genet.* 2009; **76**(5): 421–6. [PubMed Abstract](#) | [Publisher Full Text](#)
- Kleibl Z, Novotny J, Bezdicikova D, *et al.*: The CHEK2 c.1100delC germline mutation rarely contributes to breast cancer development in the Czech Republic. *Breast Cancer Res Treat.* 2005; **90**(2): 165–7. [PubMed Abstract](#) | [Publisher Full Text](#)
- Offit K, Pierce H, Kirchhoff T, *et al.*: Frequency of CHEK2*1100delC in New York breast cancer cases and controls. *BMC Med Genet.* 2003; **4**: 1. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rashid MU, Jakubowska A, Justenhoven C, *et al.*: German populations with infrequent CHEK2*1100delC and minor associations with early-onset and familial breast cancer. *Eur J Cancer.* 2005; **41**(18): 2896–903. [PubMed Abstract](#) | [Publisher Full Text](#)
- Thompson D, Seal S, Schutte M, *et al.*: A multicenter study of cancer incidence in CHEK2 1100delC mutation carriers. *Cancer Epidemiol Biomarkers Prev.* 2006; **15**(12): 2542–5. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Vahteristo P, Bartkova J, Eerola H, *et al.*: A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. *Am J Hum Genet.* 2002; **71**(2): 432–8. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Weischer M, Bojesen SE, Ellervik C, *et al.*: CHEK2*1100delC genotyping for clinical assessment of breast cancer risk: meta-analyses of 26,000 patient cases and 27,000 controls. *J Clin Oncol.* 2008; **26**(4): 542–8. [PubMed Abstract](#) | [Publisher Full Text](#)
- Weischer M, Bojesen SE, Tybjaerg-Hansen A, *et al.*: Increased risk of breast cancer associated with CHEK2*1100delC. *J Clin Oncol.* 2007; **25**(1): 57–63. [PubMed Abstract](#) | [Publisher Full Text](#)
- Zhang S, Phelan CM, Zhang P, *et al.*: Frequency of the CHEK2 1100delC mutation among women with breast cancer: an international study. *Cancer Res.* 2008; **68**(7): 2154–7. [PubMed Abstract](#) | [Publisher Full Text](#)
- Narod SA: Testing for CHEK2 in the cancer genetics clinic: ready for prime time? *Clin Genet.* 2010; **78**(1): 1–7. [PubMed Abstract](#) | [Publisher Full Text](#)
- Bak A, Janiszewska H, Junkiert-Czarnecka A, *et al.*: A risk of breast cancer in women - carriers of constitutional CHEK2 gene mutations, originating from the North - Central Poland. *Hered Cancer Clin Pract.* 2014; **12**(1): 10. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Ingvarsson S, Sigbjornsdottir BI, Huipung C, *et al.*: Mutation analysis of the CHK2 gene in breast carcinoma and other cancers. *Breast Cancer Res.* 2002; **4**(3): R4. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Siolek M, Cybulski C, Gasior-Perczak D, *et al.*: CHEK2 mutations and the risk of papillary thyroid cancer. *Int J Cancer.* 2015; **137**(3): 548–52. [PubMed Abstract](#) | [Publisher Full Text](#)
- Teodorczyk U, Cybulski C, Wokolorczyk D, *et al.*: The risk of gastric cancer in carriers of CHEK2 mutations. *Fam Cancer.* 2013; **12**(3): 473–8. [PubMed Abstract](#) | [Publisher Full Text](#)
- Zhang P, Wang J, Gao W, *et al.*: CHK2 kinase expression is down-regulated due to promoter methylation in non-small cell lung cancer. *Mol Cancer.* 2004; **3**: 14. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lek M, Karczewski KJ, Minikel EV, *et al.*: Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016; **536**(7616): 285–91. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, *et al.*: The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet.* 2013; **45**(10): 1113–20. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Cerami E, Gao J, Dogrusoz U, *et al.*: The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2012; **2**(5): 401–4. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zhang J, Baran J, Cros A, *et al.*: International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. *Database (Oxford).* 2011; **2011**: bar026. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Welter D, MacArthur J, Morales J, *et al.*: The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 2014; **42**(Database issue): D1001–6. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Shihab HA, Gough J, Cooper DN, *et al.*: Predicting the functional, molecular, and

- phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat.* 2013; **34**(1): 57–65.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Reva B, Antipin Y, Sander C: **Predicting the functional impact of protein mutations: application to cancer genomics.** *Nucleic Acids Res.* 2011; **39**(17): e118.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 36. Ng PC, Henikoff S: **Predicting deleterious amino acid substitutions.** *Genome Res.* 2001; **11**(5): 863–74.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 37. Adzhubei IA, Schmidt S, Peshkin L, *et al.*: **A method and server for predicting damaging missense mutations.** *Nat Methods.* 2010; **7**(4): 248–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 38. McLaren W, Gil L, Hunt SE, *et al.*: **The Ensembl Variant Effect Predictor.** *Genome Biol.* 2016; **17**(1): 122.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 39. R Core Team: **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria, 2013.
[Reference Source](#)
 40. Abud J, Koehler-Santos P, Ashton-Prolla P, *et al.*: **CHEK2 1100DELC germline mutation: a frequency study in hereditary breast and colon cancer Brazilian families.** *Arq Gastroenterol.* 2012; **49**(4): 273–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
 41. Carraro DM, Koike Folguedra MA, Garcia Lisboa BC, *et al.*: **Comprehensive analysis of BRCA1, BRCA2 and TP53 germline mutation and tumor characterization: a portrait of early-onset breast cancer in Brazil.** *PLoS One.* 2013; **8**(3): e57581.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 42. Felix GE, Abe-Sandes C, Machado-Lopes TM, *et al.*: **Germline mutations in BRCA1, BRCA2, CHEK2 and TP53 in patients at high-risk for HBOC: characterizing a Northeast Brazilian Population.** *Hum Genome Var.* 2014; **1**: 14012.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 43. Palmero EI, Alemar B, Schuler-Faccini L, *et al.*: **Screening for germline BRCA1, BRCA2, TP53 and CHEK2 mutations in families at-risk for hereditary breast cancer identified in a population-based study from Southern Brazil.** *Genet Mol Biol.* 2016; **39**(2): 210–22.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 44. Jablonski P, Alterman L, Pastene E, *et al.*: **Argentinean Jewish population frequencies for common mutations in BRCA1, BRCA2, and CHEK2.** *Journal of Clinical Oncology.* [Abstract]. 2014; **32**(15_suppl): 1539.
[Reference Source](#)
 45. Gonzalez-Hormazabal P, Castro VG, Blanco R, *et al.*: **Absence of CHEK2 1100delC mutation in familial breast cancer cases from a South American population.** *Breast Cancer Res Treat.* 2008; **110**(3): 543–5.
[PubMed Abstract](#) | [Publisher Full Text](#)
 46. Chaudhury A, Laukaitis C, Mauss C, *et al.*: **Abstract P3-07-05: Frequent BRCA1 and BRCA2 mutations are found in Mexican and Mexican-American women with breast cancer.** *Cancer Research.* 2013; **73**(24_suppl).
[Publisher Full Text](#)
 47. Martínez-Bouzas C, Beristain E, Guerra I, *et al.*: **CHEK2 1100delC is present in familial breast cancer cases of the Basque Country.** *Breast Cancer Res Treat.* 2007; **103**(1): 111–3.
[PubMed Abstract](#) | [Publisher Full Text](#)
 48. Guauque-Olarte S, Rivera-Herrera AL, Cifuentes-C L: **Dataset 1 in: Mutations of the CHEK2 gene in patients with cancer and their presence in the Latin American population.** *F1000Research.* 2016.
[Data Source](#)
 49. Guauque-Olarte S, Rivera-Herrera AL, Cifuentes-C L: **Dataset 2 in: Mutations of the CHEK2 gene in patients with cancer and their presence in the Latin American population.** *F1000Research.* 2016.
[Data Source](#)
 50. Guauque-Olarte S, Rivera-Herrera AL, Cifuentes-C L: **Dataset 3 in: Mutations of the CHEK2 gene in patients with cancer and their presence in the Latin American population.** *F1000Research.* 2016.
[Data Source](#)
 51. Guauque-Olarte S, Rivera-Herrera AL, Cifuentes-C L: **Dataset 4 in: Mutations of the CHEK2 gene in patients with cancer and their presence in the Latin American population.** *F1000Research.* 2016.
[Data Source](#)
 52. Guauque-Olarte S, Rivera-Herrera AL, Cifuentes-C L: **Dataset 5 in: Mutations of the CHEK2 gene in patients with cancer and their presence in the Latin American population.** *F1000Research.* 2016.
[Data Source](#)

Open Peer Review

Current Peer Review Status: ? ? ?

Version 1

Reviewer Report 24 April 2017

<https://doi.org/10.5256/f1000research.10703.r22138>

© 2017 Grzybowska E. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ewa Grzybowska

Centre of Oncology-MSK memorial Institute, Gliwice Branch, Wybrzeze Armii Krajowej 15, 44-101, Gliwice, Poland

The authors worked out the compilation of germline mutations in the CHEK2 gene in patients diagnosed with different cancer types and in different populations focusing on Latin American population.

CHEK2 mutations have been linked with Li—Fraumeni syndrome, also germline mutations are thought to confer a predisposition to sarcomas, breast cancer and brain tumors. The most frequent CHEK2 mutation c.1100 delC is the low penetrance mutation and it has low impact in breast or other cancers risk. The rest of the mutations or SNPs are much less connected with the known cancer risk. Therefore, it is difficult to evaluate the increase of the different cancers risk for the carriers of germline mutations in CHEK2 gene. It is particularly difficult to do so if there are only one or two carriers of these mutations in the population under study.

I totally agree with both reviewers especially with two issues:

1. The text is written in that way that the reader can think the authors analyze somatic mutations in CHEK2 gene in different cancer types whereas they made the search for germline mutations. In Table 4 they indicate that the blood was the tissue which was used to analyze mutations so the text should be rewritten in that way there would be no doubt that germline mutations were under study.
2. The Tables 1 and 2 also should be changed. The description of the ethnic minorities is strange. In Table 2 the data from databaseExAC do not contain the information about the disease connected with the mutation so it does not make sense to include these data if the title of the manuscript is "Mutations of the CHEK2 gene in the patients with cancer..." these data should be excluded from the analysis because they do not bring any important information about CHEK2 mutations in different cancer sites.

Is the work clearly and accurately presented and does it cite the current literature?

No

Is the study design appropriate and is the work technically sound?

No

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: cancer genetics, molecular biology of carcinogenesis, epidemiology of cancer, pharmacogenetics

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 03 April 2017

<https://doi.org/10.5256/f1000research.10703.r21485>

© 2017 Rashid M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Muhammad Usman Rashid

Shaukat Khanum Memorial Cancer Hospital and Research Centre (SKMCH & RC), Department of Basic Sciences Research, Lahore, Pakistan

The manuscript by Guauque-Olarte and colleagues is an overview of the *CHEK2* variants reported in Latin American population, searched from literature or cBioPortal, ICGC and ExAC databases. Overall the concept of the manuscript is interesting; however the data is poorly presented and the scientific writing is not up to the mark. The manuscript title also needs modification, like "An overview of *CHEK2* variants associated with cancer in Latin American population".

I have following reservations about the manuscript:

1. For missense variants or variants in 5'UTRs it is suggested to write "DNA sequence variants" instead of "mutations" throughout the manuscript, so that these can be differentiated from clear pathogenic mutations i.e. frameshift, nonsense or splice site mutations.
2. As the study objective was to compile the *CHEK2* mutations reported in Latin Americans,

Table 1 describes the *CHEK2* variants identified in other populations or even the ethnicity is unknown for majority of the variants presented in this table. The table is also not presented properly. It is suggested to omit this table or present it as a "Dataset", and just mention in the text that 78 deleterious or potentially deleterious mutations were reported in TCGA studies.

3. The authors did not state about the origin (somatic or germline) of *CHEK2* variants presented in all the tables. It would be of interest if a column is added in all tables for this information.
4. Results section: Data presented in Figure 1 and Figure 2 is not concordant as mentioned in the text. Please resolve this issue.
5. Results section: ".....after the filtering process, 38 of which were classified as with high impact" It is not clear which those 38 nucleotide variants are in Table 1? Please add a column for this information.
6. Results section: Paragraph "Two patients with three mutationsthis patient carry a frameshift and a nonsense mutation" is confusing. Is the patient with DLBC a compound heterozygous for a frameshift and nonsense *CHEK2* mutation, simultaneously?
7. Table 2: Column Genomic DNA change: The nucleotide change can't be seen in this column, there is just the nucleotide position. Please modify this column.
8. Results section: GWAS catalog. Authors should be cautious whether the SNPs rs132390-C and rs2239815-T are present in *CHEK2* gene or not?
9. Table 2: Two variants in 5'UTR are not clear, population is also not mentioned.
10. Table 2: "Effect" column; please correct that stop gain mutations are also called nonsense mutations.
11. Table 2: The data in the table is not presented properly. c.1590+2T>G and c.573+2T>G are the nucleotide changes and these are presented in column AA change. The authors should follow HGVS nomenclature, both for nucleotide change and the AA change. There should be a column for pathogenicity of missense mutations (high or medium impact) in this table.
12. Table 3: Column Consequences: I think there is no need to mention the amino acid change referring all *CHEK2* transcripts. Just follow the GenBank reference sequence for transcript variant 1 for reporting nucleotide or AA change and follow the HGVS nomenclature.
13. Discussion, paragraph 1: "....eight stop gain mutations, one frameshift mutation..." Please correct, there are four stop gain mutations (also called nonsense mutations) and five frameshift mutations.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 03 March 2017

<https://doi.org/10.5256/f1000research.10703.r20668>

© 2017 Palles C. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Claire Palles

Wellcome Trust Centre for Human Genetics, NIHR Comprehensive Biomedical Research Centre, Oxford, UK

Sandra Guauque-Olarte *et al* provide an overview of both somatic and germline mutations in CHEK2 that have been identified in Latin-American populations. The authors interrogate cBioPortal and ICGC databases to identify somatic mutations and ExAC and a review of existing literature to identify germline mutations.

My reservations about the manuscript are as follows:

Currently the authors do not make it clear throughout the manuscript whether they are describing somatic mutations or germline mutations/variants. Please add a column to table 2 to show clearly which are somatic and which are germline.

In the abstract it says: "Latin American studies have been restricted to breast and colorectal cancer and only two mutations out of four that have been interrogated in this population were identified, namely c.1100delC and c.349A>G". Table 4 which lists the mutations reported in the literature in Latin American studies does not show the c.349A>G mutation but a c.478A>G mutation and I can see no further mention of c.349A>G in the rest of the manuscript. Please resolve this.

Results:

The text description of the difference between Figure 1 and 2 in the start of the results section is unclear. As far as I can see Figure 1 shows data per cancer type in TCGA and Figure 2 shows data per study in TCGA. I don't see the need to have both figures- figure 1 is sufficient and the text should read "breast, colorectal and non small cell lung cancer had more CHEK2 mutations than other cancer type". At the start of the results section the authors describe mutations "before

filtering". Please be clearer and state before filtering steps to include only likely functional mutations.

On page 5 the sentence beginning The type of cancer with the most mutations should read "After filtering for likely functional variants the cancers with the highest numbers of mutations in CHEK2 were breast followed by uterine, non small cell (?) lung and colorectal.

Table 1 describes mutations in non Hispanic-latino samples. A98Mfs*13 and Q100*, which are found in a white Hispanic or latino sample should be removed (none of the other mutations in Latin American populations are in Table 1). The ethnicity column in table 1 also needs to be formatted properly – remove duplicated words and “_” between words.

GWAS catalogue section in results- need to insert P-values for the associations that you report. Table 4. Insert OR and P-values for associations.

Dataset 5. The %s should not have a – in front of them, it adds confusion as to what these values are.

Discussion:

The authors make reference to a CHEK2 1100delC mutation picked up in the TCGA datasets and refer to Table 1. I cant find 1100delC in table 1. I can only find it in Table 2 in ExAC. Please clarify.

The authors state that other patients with cancer types such as uterine, lung, bladder and head and neck cancer should be screened for CHEK2 mutations. Here they are trying to show that because a gene is somatically mutated in a particular cancer type that there might also be a germline mutation that increases predisposition. Some of the mutations listed in tables 1 and 2 (mutations post filtering for likely functional impact) are missense or UTR and so it would be important to show that these somatic mutations are functional. Could the authors please annotate the TCGA/ICGC mutations with information of which domain they map to.

The authors state there are 23 germline mutations which could cause cancer susceptibility. Were any of these examined for a functional affect on CHECK2 in paper by Bell et al 2007¹ or other studies of variants in CHEK2 on protein activity? I think it would be important to include this and to also state that functional assays would be helpful to determine which of these should be screened for in Latin American and other populations.

References

1. Bell DW, Kim SH, Godwin AK, Schiripo TA, et al.: Genetic and functional analysis of CHEK2 (CHK2) variants in multiethnic cohorts.*Int J Cancer*. 2007; **121** (12): 2661-7 [PubMed Abstract](#) | [Publisher Full Text](#)

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Partly

Are sufficient details of methods and analysis provided to allow replication by others?

Partly

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Partly

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research