

## A deep intronic recurrent *CHEK2* variant c.1009-118\_1009-87delinsC affects pre-mRNA splicing and contributes to hereditary breast cancer predisposition

Petra Zemankova<sup>1,2,†</sup>, Marta Cerna<sup>1,†</sup>, Klara Horackova<sup>1</sup>, Corinna Ernst<sup>3</sup>, Jana Soukupova<sup>1</sup>, Marianna Borecka<sup>1</sup>, Britta Blümcke<sup>3</sup>, Leona Cerna<sup>4</sup>, Monika Cerna<sup>5</sup>, Vaclava Curtisova<sup>6</sup>, Tatana Dolezalova<sup>1</sup>, Petra Duskova<sup>7</sup>, Lenka Dvorakova<sup>8</sup>, Lenka Foretova<sup>9</sup>, Ondrej Havranek<sup>10,11</sup>, Jan Hauke<sup>3</sup>, Eric Hahnen<sup>3</sup>, Miloslava Hodulova<sup>4</sup>, Milena Hovhannisyan<sup>1</sup>, Lucie Hruskova<sup>12</sup>, Marketa Janatova<sup>1</sup>, Maria Janikova<sup>6</sup>, Sandra Jelinkova<sup>1</sup>, Pavel Just<sup>1</sup>, Marcela Kosarova<sup>13</sup>, Monika Koudova<sup>4</sup>, Vera Krutilkova<sup>14</sup>, Eva Machackova<sup>9</sup>, Katerina Matejkova<sup>1,15</sup>, Renata Michalovska<sup>12</sup>, Adela Misove<sup>9</sup>, Petr Nehasil<sup>1,2,8</sup>, Barbora Nemcova<sup>1</sup>, Jan Novotny<sup>10</sup>, Ales Panczak<sup>10</sup>, Pavel Pesek<sup>1</sup>, Ondrej Scheinost<sup>7</sup>, Drahomira Springer<sup>1</sup>, Barbora Stastna<sup>1,16</sup>, Viktor Stranecky<sup>8</sup>, Ivan Subrt<sup>5</sup>, Spiros Tavandzis<sup>14</sup>, Eva Tureckova<sup>1</sup>, Kamila Vesela<sup>10</sup>, Zdenka Vlckova<sup>12</sup>, Michal Vocka<sup>17</sup>, Barbara Wappenschmidt<sup>3</sup>, Tomas Zima<sup>1</sup>, Zdenek Kleibl<sup>1,2</sup>, Petra Kleiblova<sup>1,10</sup>

- 1 Institute of Medical Biochemistry and Laboratory Diagnostics, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic
- 2 Institute of Pathological Physiology, First Faculty of Medicine, Charles University, Prague, Czech Republic
- 3 Center for Familial Breast and Ovarian Cancer, Center for Integrated Oncology (CIO), Medical Faculty, University of Cologne and University Hospital Cologne, Cologne, Germany
- 4 GENNET, Prague, Czech Republic
- 5 Institute of Medical Genetics, University Hospital Pilsen, Pilsen
- 6 Department of Medical Genetics, University Hospital Olomouc, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czech Republic
- 7 Hospital Ceske Budejovice, Ceske Budejovice, Czech Republic
- 8 Department of Pediatrics and Inherited Metabolic Disorders, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic
- 9 Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic
- 10 Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic
- 11 BIOCEV, First Faculty of Medicine, Charles University, Prague, Czech Republic
- 12 GHC Genetics, Prague, Czech Republic
- 13 Pronatal, Prague, Czech Republic
- 14 Department of Medical Genetics, AGEL Laboratories, AGEL Research and Training Institute, Novy Jicin, Czech Republic
- 15 Department of Genetics and Microbiology, Faculty of Science, Charles University, Prague, Czech Republic
- 16 Department of Biochemistry, Faculty of Science, Charles University, Prague, Czech Republic
- 17 Department of Oncology, First Faculty of Medicine, Charles University and General University Hospital in Prague, Prague, Czech Republic

† These authors contributed equally to this work.

**Corresponding author:** Petra Kleiblova, MD, PhD.

Laboratory of Oncogenetics, Institute of Medical Biochemistry and Laboratory Diagnostics, First Faculty of Medicine, Charles University and General University Hospital in Prague,

Address: Na Bojisti 1660/3, Prague 2, 128 00, Czech Republic. Email: [pekleje@lf1.cuni.cz](mailto:pekleje@lf1.cuni.cz)

**Key words:** deep intronic *CHEK2* variant, breast cancer, NGS, RNA analysis, genetic testing

## **Abstract**

Germline *CHEK2* pathogenic variants confer an increased risk of female breast cancer (FBC). Here we describe a recurrent germline intronic variant c.1009-118\_1009-87delinsC, which showed a splice acceptor shift in RNA analysis, introducing a premature stop codon (p.Tyr337PhefsTer37).

The variant was found in 21/10,204 (0.21%) Czech FBC patients compared to 1/3,250 (0.03%) controls ( $p=0.04$ ) and in 4/3,639 (0.11%) FBC patients from an independent German dataset. In addition, we found this variant in 5/2,966 (0.17%) Czech (but none of the 443 German) ovarian cancer patients, three of whom developed early-onset tumors.

Based on these observations, we classified this variant as likely pathogenic.

**Word count:** 100

## 1. Introduction

The *CHEK2* (checkpoint kinase 2) gene encodes a nuclear serine/threonine-protein kinase CHK2, which is one of the key mediators of cellular response to various stress stimuli [1]. Its heterozygous germline pathogenic variants (GPV) confer a moderate breast cancer (BC) risk [2]. At the same time, *CHEK2* GPV are also associated with increased risk of multiple other cancer types including colorectal, thyroid, pancreatic, kidney or hematological malignancies [3-8]. Compared to heterozygotes, carriers of homozygous or compound heterozygous *CHEK2* GPV have substantially higher cancer risk, develop tumors at younger age, but otherwise do not develop other clinical symptoms [9, 10].

Importantly, *CHEK2* is the second most frequently altered female BC predisposition gene in Western countries. Frequency of *CHEK2* GPV (truncations, splicing alterations, and large copy number variations) is approximately 1.1–1.3% in unselected female BC patients from Europe and the USA [11, 12]. Additionally, 0.5% of BC patients may be carriers of rare missense *CHEK2* GPV[13].

In this study, we have identified previously unreported frequent deep intronic *CHEK2* GPV, characterized its effect at the RNA level, and provided evidence for its contribution to increased BC risk.

## 2. Patients and methods

**Identification of c.1009-118\_1009-87delinsC *CHEK2* variant.** We have re-analyzed next generation sequencing (NGS)-based anonymized data from 10,204 female BC and 2,966 ovarian cancer (OC) Czech patients clinically tested using the CZECANCA panel (including *CHEK2*) within the Czech consortium of diagnostic laboratories ([www.czecanca.cz](http://www.czecanca.cz)) [14-17]. We have specifically searched for deep intronic germline *CHEK2* (NM\_007194.4) variants localized outside the canonical intronic splice sites. Impact of identified *CHEK2* variant on pre-mRNA processing was analyzed by CZECANCA panel-based total RNA sequencing as described previously [18, 19]. RNA was extracted from peripheral blood leukocytes with/without nonsense-mediated decay (NMD) inhibition with cycloheximide (final concentration 200 µg/mL) for four hours. For variant burden analysis (two-sided Fisher's exact test), we used data from 3,250 unselected Czech female population-matched controls (PMC) analyzed by the same NGS approach.

Variant frequency was independently assessed by analysis of data from 3,639 and 443 German BC and OC patients, respectively (described previously [20]) using GATK HaplotypeCaller variant calling. As the variant's localization was outside (but close to) the corresponding sequencing target region of the TruRisk® panel applied (Human hg19 chr22:29092810-29093050), sufficient read depth ( $\geq 30$ ) at chr22:29093063 was ensured using the samtools mpileup utility prior to variant calling.

All participants gave informed consent to germline genetic testing approved by ethics committees.

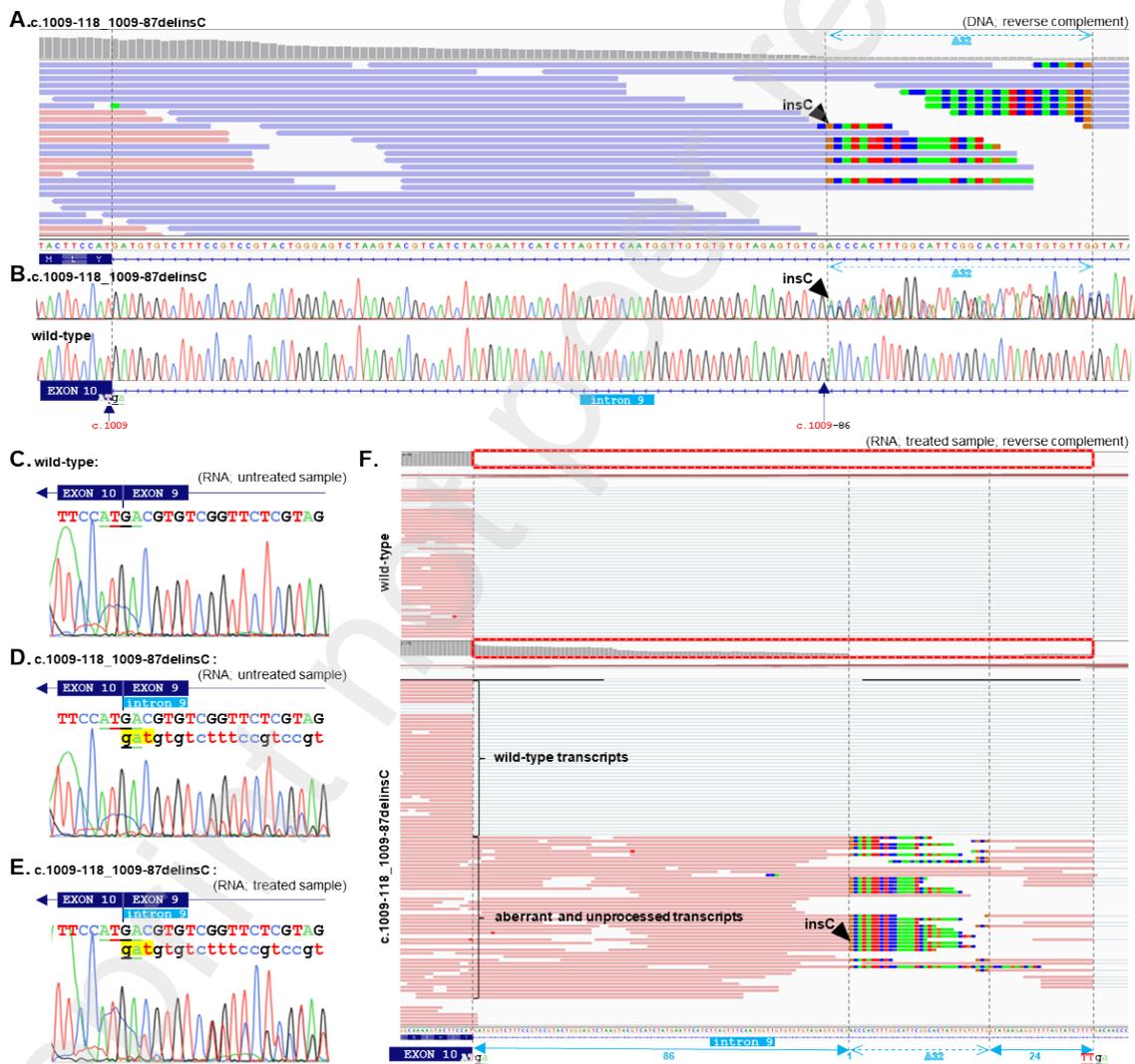
### 3. Results

By bioinformatic re-analysis of panel NGS data from Czech BC/OC patients, we identified previously unreported deep intronic *CHEK2* variant localized 87bp apart from the 3'-end of intron 9. It consisted of a 32bp deletion replaced by a single cytosine: c.1009-118\_1009-87delinsC (Figure 1A), which we confirmed by Sanger sequencing (Figure 1B). Sanger sequencing and NGS-based total RNA analysis from available variant carriers' samples showed its clear effect on *CHEK2* pre-mRNA splicing. Compared to the wild-type mRNA sequence (Figure 1C), the variant allele leads to the use of upstream alternative splice acceptor at position c.1009-142 with consequent retention of terminal part of intron 9 at the beginning of exon 10 (r.1008\_1009ins1009-142\_1009-1del1009-118\_1009-87insC). At the protein level, this aberrant *CHEK2* transcript results in premature termination of translation (p.Tyr337PhefsTer37). Importantly, proportion of this aberrant *CHEK2* transcript varied between 0.07–0.26 in RNA samples (Figure 1D), which suggests either its NMD-mediated degradation or only partial effect of investigated variant on pre-mRNA splicing. NMD inhibition in samples from variant carriers resulted in significant increase of aberrantly spliced mRNA (Figure 1E–F), therefore, confirming its partial degradation via NMD.

The c.1009-118\_1009-87delinsC variant was identified in 21/10,204 (0.21%) female BC cases (Table 1) and only in 1/3,250 (0.03%) PMC (OR=6.7; 95%CI 1.08–276.88, p=0.04). Two BC patients exhibited biallelic *CHEK2* inactivation and only one carried a co-occurring GPV in another BC-predisposition gene (Table 1). All BC c.1009-118\_1009-87delinsC variant carriers with known histology (18/21) developed ER-positive primary tumors, six developed double-primary cancers, and only two with known family cancer history (18/21) had negative or *CHEK2*-irrelevant family cancer history (Table 1). Additionally, we have identified higher frequency of c.1009-118\_1009-87delinsC variant carriers also among OC cases (5/2,966; 0.17%, Table 1) in comparison to controls, however, the difference was not statistically significant (OR=5.49; 95%CI 0.62–259.25; p=0.11).

Analysis of independent German dataset revealed four carriers of the c.1009-118\_1009-87delinsC variant among 3,639 BC patients (0.11%) and in 0/443 German OC patients. The frequency in German BC patients was significantly higher than in European (non-Finnish) from gnomAD\_v3 controls (as [22-28697075-GGGTGAAACCGTAAGCCGTGATACACAAC-G](#); 2/34,018; 0.006%; p=0.001).

**Figure 1: Characterization of the c.1009-118\_1009-87delinsC *CHEK2* variant.** (A.) NGS-based DNA sequencing visualized in the Integrative Genomics Viewer (IGV). The dashed gray lines indicate the deletion borders, and the dashed blue arrow denotes the deletion of 32bp following the nucleotide c.1009-86 at the 3'-end of intron 9 with the insertion of cytosine. (B.) DNA Sanger sequencing of the variant and wild-type samples. (C-E.) RNA (cDNA) Sanger sequencing of wild-type sample (C.), variant sample without (D.) and with (E.) NMD inhibition (cycloheximide), showing the increase of aberrant splicing variant signal peaks after NMD inhibition. (F.) RNA panel NGS from a wild-type control (top) and from a carrier of the c.1009-118\_1009-87delinsC variant after NMD inhibition (bottom). Note the difference between wild-type and variant RNA in number (coverage) of intronic retentions (red dashed-line boxes) resulting from the aberrant pre-mRNA splicing. Solid blue arrows indicate sequencing context of aberrant reads showing 86b retained from intron 9, interrupted by 32b deletion replaced by single cytosine insertion, followed by 24b from intron 9: r.1008\_1009ins1009-142\_1009-1del1009-118\_1009-87insC. We hypothesize that the reassembled primary transcript enhances the pre-existing alternative acceptor splice site TT|ga in intron 9, which precedes the canonical acceptor splice site upstream of exon 10.



**Note:** The IGV visualizes the *CHEK2* sequence in a reverse complement according to the *CHEK2* gene reverse orientation on the chromosome 22.

**Table 1. Clinical and histopathological characteristics of breast cancer (top and middle) and ovarian cancer (bottom) patients carrying germline c.1009-118\_1009-87delinsC *CHEK2* variant.**

Patient number	Age at diagnosis (years)	ER	PR	HER2	Histology	Double primary tumor (age; years)	Family cancer history
<b>CARRIERS WITH BREAST CANCER – Czech cohort (10,204 patients)</b>							
02_655a22	29	+	-	-	DCIS		CRC
02_748a22	30	+	+	+	L		CRC, LC
01_B2542 #1	33	+/+	+/+	+/+	D/L	BC (33)	BC
04_469818	34	n.a.	n.a.	n.a.	n.a.		n.a.
07_B7594 #2	39	+	+	+	D		CRC, PrC
04_1550a16	41	+	+	-	D		BC, GC
02_1101a20	44	+	-	+	D		LC
11_200451	44	+	+	+	D		PaC, PrC, RCC
07_B5877	45	+	+	+	D		neg.
04_1728421 #3	45	n.a.	n.a.	n.a.	n.a.		BC
02_1339a20	46	+	+	-	DCIS	GC	BC
01_PKM488	47	+	+	+	D		CRC, PrC
02_150a22	48	+/-	+/-	-/-	L/D	BC (58)	BC
04_1230219	49	+	+	-	D		OC, CRC, GC
01_PKM165	53	+	-	n.a.	DCIS	Hem (51)	CRC, GC
02_1080a19	58	n.a.	n.a.	n.a.	DCIS	Hem (40)	BC
04_1658119	58	+	+	+	D		BC, PrC, LC
04_972619	60	+	+	-	D	CRC (71); Hem (60)	BC, OC
04_484419	62	+	+	-	D		BC
01_PKM4433	65	+	+	+	D		n.a.
01_PKM3591	71	+	-	-	D		n.a.
<b>CARRIERS WITH BREAST CANCER – German cohort (3,639 patients)</b>							
58-10	45	+/+	+/-	+/-	D/L	BC	OC
13-24	47	n.a.	n.a.	n.a.	n.a.		BC
76-32	48	+	-	-	n.a.		BC
15-6	56	n.a.	n.a.	n.a.	n.a.		BC
<b>CARRIERS WITH OVARIAN CANCER – Czech cohort (2,966 patients)</b>							
04_39420	24	×	×	×	DYS		UBC
01_B1231	32	×	×	×	HGS		BC
08_BS670a19	34	×	×	×	n.a.	RCC (24)	OC, CRC, GC, PrC, RCC, BT, OS, Hem
02_1167a20	51	×	×	×	HGS		GC
04_355619	65	×	×	×	PAP		neg.

**Legend:** Individuals with additional pathogenic germline variant in #1 *CHEK2*: c.277del (p.Trp93GlyfsTer17); #2 *BRCA2*: c.673\_676del (p.Thr225LeufsTer4); #3 *CHEK2*: c.1009-118\_1009-87delinsC homozygote.

**BC** – breast cancer, **BT** – brain tumor, **CRC** – colorectal cancer, **D** – ductal BC, **DCIS** – ductal BC (in situ), **DYS** – dysgerminoma, **ER** – estrogen receptor, **GC** – gastric cancer, **Hem** – hematological malignancies, **HER2** – HER2 receptor, **HGS** – high-grade serous OC, **L** – lobular BC, **LC** – lung cancer, **n.a.** – not available, **neg** – negative, **OC** – ovarian cancer, **OS** – osteosarcoma, **PaC** – pancreatic cancer, **PAP** – papillary adenocarcinoma, **PR** – progesterone receptor, **PrC** – prostate cancer, **RCC** – renal cell carcinoma, **UBC** – urinary bladder cancer, + – positive, - – negative, × – not applicable.

#### 4. Discussion and conclusions

We have characterized a recurrent germline *CHEK2* deep intronic variant c.1009-118\_1009-87delinsC and showed that it leads to the formation of aberrantly spliced *CHEK2* mRNA partially subjected to

NMD, and encodes a functionally impaired prematurely terminated protein. Furthermore, we showed that this variant was overrepresented in female BC patients with frequency similar to other two Slavic founder splicing *CHEK2* GPV c.444+1G>C and c.846+4\_846+7del described in Czech population previously [21]. Importantly, germline c.1009-118\_1009-87delinsC positive BC patients had characteristics typical for *CHEK2* GPV carriers (ER-positive tumors, cancer multiplicity, and positive family history of cancer) [22]. Moreover, identified *CHEK2* variant was associated with increased risk of female BC development (OR=6.7). On the other hand, we believe that the risk level was overestimated due to lower number of PMC and may be similar as for *CHEK2* pathogenic truncations (OR=2.47–2.54) or missense variants (OR=2.83) [11-13]. There is generally conflicting evidence for OC predisposition and *CHEK2* GPV [13, 21]. Nevertheless, we recurrently observed c.1009-118\_1009-87delinsC in OC patients, including early-onset patients who may otherwise represent a specific OC subgroup with an unusually low proportion of GPV in established cancer predisposition gene [23].

In conclusion, the *CHEK2* variant c.1009-118\_1009-87delinsC results in an aberrant mRNA transcript containing premature termination codon (p.Tyr337PhefsTer37) and is partially subjected to NMD. It is significantly enriched in BC patients with a phenotype typical of known *CHEK2* GPV and can be, therefore, classified as likely pathogenic. Further studies are necessary to confirm its clinical implications and to establish its prevalence in other populations. Our study highlights the critical importance to focus on intronic regions beyond the canonical  $\pm 1/2$  splice sites within the search for RNA splicing affecting GPV.

**Word count (body text including chapter numbers and titles):** 1,013 words

**Funding:** This work has been supported by the grant projects of the Ministry of Health of the Czech Republic [NU23-03-0015, NU20-03-00283, NU20-03-00016, NU20-03-00285, DRO-VFN-64165]; Charles University projects [COOPERATIO, SVV260631, UNCE/24/MED/022]; and the Ministry of Education Youth and Sports of the Czech Republic grant [EXCELES no. LX22NPO5102, and The National Center for Medical Genomics" LM2023067]; German Cancer Aid [70114178].

**Declaration of competing interest:** All authors have no conflicts of interest to declare.

**Ethics approval:** All individuals provided written informed consent with genetic testing approved by the Ethics Committees of participating institutions and the study was performed in accordance with the Declaration of Helsinki.

**Author contributions:**

Conceptualization: PZ, ZK, PK

Data curation: BB, LC, MoC, PD, JH, LH, MariJ, MoK, VK, EM, JS, VS, ST, MV, PZ

Formal Analysis: PZ, CE, PN

Funding acquisition: ZK, PK, JH

Investigation: MB, MaC, KH, MiloH, MileH, MarkJ, SJ, PJ, AM, EM, PP, JS, ET

Resources: BB, MB, VC, LD, TD, LF, OH, EH, MiloH, MileH, MarkJ, SJ, PJ, MaK, KM, RM, AM, PN, BN, JN, AP, PP, OS, DS, BS, IS, ET, KV, MV, ZV, BW, TZ

Supervision: PK, ZK

Validation: PZ, KH

Visualization: ZK, PZ, MaC

Writing – original draft: ZK, PK, PZ, MaC

Writing – review & editing: all authors

**References**

1. Zannini, L., D. Delia, and G. Buscemi, *CHK2 kinase in the DNA damage response and beyond*. J Mol Cell Biol, 2014. **6**(6): p. 442-57.
2. Stolarova, L., et al., *CHEK2 Germline Variants in Cancer Predisposition: Stalemate Rather than Checkmate*. Cells, 2020. **9**(12): p. 2675.
3. Cybulski, C., et al., *CHEK2 is a multiorgan cancer susceptibility gene*. Am J Hum Genet, 2004. **75**(6): p. 1131-5.
4. Ma, X., B. Zhang, and W. Zheng, *Genetic variants associated with colorectal cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence*. Gut, 2014. **63**(2): p. 326-36.
5. Siolek, M., et al., *CHEK2 mutations and the risk of papillary thyroid cancer*. Int J Cancer, 2015. **137**(3): p. 548-52.
6. Wieme, G., et al., *Prevalence of Germline Pathogenic Variants in Cancer Predisposing Genes in Czech and Belgian Pancreatic Cancer Patients*. Cancers, 2021. **13**(17): p. 4430.
7. Zlowocka-Perłowska, E., S.A. Narod, and C. Cybulski, *CHEK2 Alleles Predispose to Renal Cancer in Poland*. JAMA Oncol, 2019. **5**(4): p. 576.

8. Havranek, O., et al., *Association of Germline CHEK2 Gene Variants with Risk and Prognosis of Non-Hodgkin Lymphoma*. Plos One, 2015. **10**(10): p. e0140819.
9. Sutcliffe, E.G., et al., *Differences in cancer prevalence among CHEK2 carriers identified via multi-gene panel testing*. Cancer Genet, 2020. **246-247**: p. 12-17.
10. Rainville, I., et al., *High risk of breast cancer in women with biallelic pathogenic variants in CHEK2*. Breast Cancer Res Treat, 2020. **180**(2): p. 503-509.
11. Hu, C., et al., *A Population-Based Study of Genes Previously Implicated in Breast Cancer*. N Engl J Med, 2021. **384**(5): p. 440-451.
12. Dorling, L., et al., *Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women*. N Engl J Med, 2021. **384**(5): p. 428-439.
13. Stolarova, L., et al., *ENIGMA CHEK2gether Project: A Comprehensive Study Identifies Functionally Impaired CHEK2 Germline Missense Variants Associated with Increased Breast Cancer Risk*. Clin Cancer Res, 2023. **29**(16): p. 3037-3050.
14. Soukupová, J., et al., *[CZECANCA: CZEch Cancer paNel for Clinical Application-- Design and Optimization of the Targeted Sequencing Panel for the Identification of Cancer Susceptibility in High-risk Individuals from the Czech Republic]*. Klin Onkol, 2016. **29 Suppl 1**: p. S46-54.
15. Soukupova, J., et al., *Validation of CZECANCA (CZEch Cancer paNel for Clinical Application) for targeted NGS-based analysis of hereditary cancer syndromes*. PLoS One, 2018. **13**(4): p. e0195761.
16. Lhotova, K., et al., *Multigene Panel Germline Testing of 1333 Czech Patients with Ovarian Cancer*. Cancers (Basel), 2020. **12**(4).
17. Janatova, M., et al., *Classification of germline variants identified in cancer predisposition genetic testing - consensus of the CZECANCA consortium*. Klin Onkol, 2023. **37**(6): p. 431-439.
18. Walker, L.C., et al., *Comprehensive Assessment of BARD1 Messenger Ribonucleic Acid Splicing With Implications for Variant Classification*. Front Genet, 2019. **10**: p. 1139.
19. Kleiblova, P., *Parallel DNA/RNA NGS using identical gene panel in the analysis of hereditary cancer predisposition*. Folia Biol (Praha), 2023.
20. Lepkes, L., et al., *Performance of In Silico Prediction Tools for the Detection of Germline Copy Number Variations in Cancer Predisposition Genes in 4208 Female Index Patients with Familial Breast and Ovarian Cancer*. Cancers (Basel), 2021. **13**(1).
21. Kleiblova, P., et al., *Identification of deleterious germline CHEK2 mutations and their association with breast and ovarian cancer*. Int J Cancer, 2019. **145**(7): p. 1782-1797.
22. Schwartz, C.J., et al., *Clinicopathologic and genetic analysis of invasive breast carcinomas in women with germline CHEK2 variants*. Breast Cancer Res Treat, 2024. **204**(1): p. 171-179.
23. Horackova, K., et al., *Early-Onset Ovarian Cancer <30 Years: What Do We Know about Its Genetic Predisposition?* Int J Mol Sci, 2023. **24**(23).