Universiteit
Leiden
The Netherlands

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

Boonen, R.A.C.M.

Citation

Boonen, R. A. C. M. (2023, April 4). *Functional analysis of genetic variants in PALB2 and CHEK2: linking functional impact with cancer risk*. Retrieved from https://hdl.handle.net/1887/3590202

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3590202

Note: To cite this publication please use the final published version (if applicable).

FUNCTIONAL ANALYSIS OF GENETIC VARIANTS IN *PALB2* AND *CHEK2*: LINKING FUNCTIONAL IMPACT WITH CANCER RISK

RICK BOONEN

The studies described in this thesis were performed at the Department of Human Genetics at the Leiden University Medical Centre, Leiden the Netherlands.

ISBN: 978-94-6483-006-4

The printing of this thesis was financially supported by the Leiden University Library.

Cover design: The cover and title pages have been designed by Rick A.C.M. Boonen

About the cover: With a bit of imagination, the little European robin that is carefully studied on the front cover is thought to depict a genetic variant (e.g., a variant of uncertain significance)

Thesis layout: Rick A.C.M Boonen

Printing: Ridderprint | www.ridderprint.nl

Copyright © by R.A.C.M. Boonen

All rights reserved. No parts on this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means, without permission of the author.

FUNCTIONAL ANALYSIS OF GENETIC VARIANTS IN *PALB2* AND *CHEK2*: LINKING FUNCTIONAL IMPACT WITH CANCER RISK

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof. dr. ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op dinsdag 4 april 2023

klokke 15:00 uur

door

Ricky Antonius Cornelius Martin Boonen

geboren te Boxmeer

in 1983

Promotores

Prof. dr. H. van Attikum

Prof. dr. P. Devilee

Copromotor

Dr. M.P.G. Vreeswijk

Leden promotiecommissie

Prof. dr. ir. M.K. Schmidt

Prof. dr. J.M.M. Jonkers (Nederlands Kanker Instituut)

Prof. dr. H.P.J. te Riele (Nederlands Kanker Instituut)

Prof. dr. P.M. Knipscheer (Hubrecht Instituut)

For Isabella & Odin

CONTENTS

CHAPTER 1

General Introduction

1

GENERAL INTRODUCTION

Breast cancer etiology

Breast cancer affected roughly two million women globally in 2019, resulting in an estimated death toll of 689.000 and is thereby the most common cause of cancer-related death in women (1). Although breast cancer can occur in men as well, 98.7 percent out of all breast cancer cases in 2019 were observed in female. Furthermore, it is estimated to affect roughly one in 20 women globally, and as many as one in eight women in high-income countries by the age of 85. This makes it by far the leading cancer-related cause of disease burden in women (2).

The development of breast cancer likely involves a combination of risk factors, making it an extremely heterogeneous disorder biologically. In addition to female gender and advancing age, other breast cancer risk factors include early menarche, late menopause, and first birth at 30 years of age or later (3). Most breast cancers occur in the absence of any family history of this type of cancer, meaning the underlying cause may be a combination of demographic, behavioral and environmental factors. During life, these factors may cause to somatic gene alterations that mostly occur by chance. If so, it is not possible for a person to pass on these genomic alterations to their offspring. In contrast, inherited breast cancer can occur as a result of genetic alteration in the germline. These variants can be inherited from parent to offspring and give rise to familial predisposition (3). Many of these genetic variants occur in tumor suppressor genes, such as the well-known DNA repair genes *BRCA1*, *BRCA2*, and *PALB2* or in genome caretaker genes such as *TP53*, *ATM* and *CHEK2* (4). Notably, these genes are all involved in maintaining genomic stability by acting in the DNA damage response (DDR).

DNA damage response and cancer

The inability to respond properly to DNA damage can result in a high frequency of unwanted somatic gene alterations (i.e., genomic instability), which in turn can promote the development of cancer (5). Proper regulation of the DDR is therefore crucial for cellular homeostasis and indispensable for maintaining genomic stability (6-8). During the DDR, cells can activate cell cycle checkpoints that in turn can result in cell cycle arrest, repression or activation of transcription, DNA repair, or even programmed cell death. Depending on the type and extent of DNA damage, the site of the lesion, and stage of the cell cycle, a choice is made between several DNA repair pathways to repair the DNA damage. These pathways include nucleotide excision repair (NER), base excision repair (BER), mismatch repair (MMR), single-strand break repair (SSBR), canonical non-homologous enjoining (cNHEJ), alternative non-homologous enjoining (aNHEJ), single-strand annealing (SSA) and homologous recombination (HR) (8,9). Some of these pathways are more mutagenic than others, i.e., they

have a higher probability of resulting in permanent DNA changes, and careful regulation of these pathways is therefore crucial for genomic stability.

Deregulation in the repair of DNA damage can be caused by DNA variants in genes or by aberrant activities of key proteins involved in the DDR (10). Failure to faithfully repair damaged DNA can result in a high mutational frequency within the genome of a cellular lineage (11,12). In hereditary breast cancers, it is established that pathogenic germline variants in DNA damage repair genes such as the high-risk breast cancer susceptibility genes *BRCA1*, *BRCA2* and *PALB2* (odds of developing breast cancer >5), or moderate-risk genes *ATM* and *CHEK2* (odds of developing breast cancer between 2-5), lead to a significant increase in the risk for developing breast cancer (4,13) (Fig. 1). As such, it is important that these genes are sequenced in individuals that may be at risk for developing breast cancer, so that pathogenic variants in these genes can be identified early.

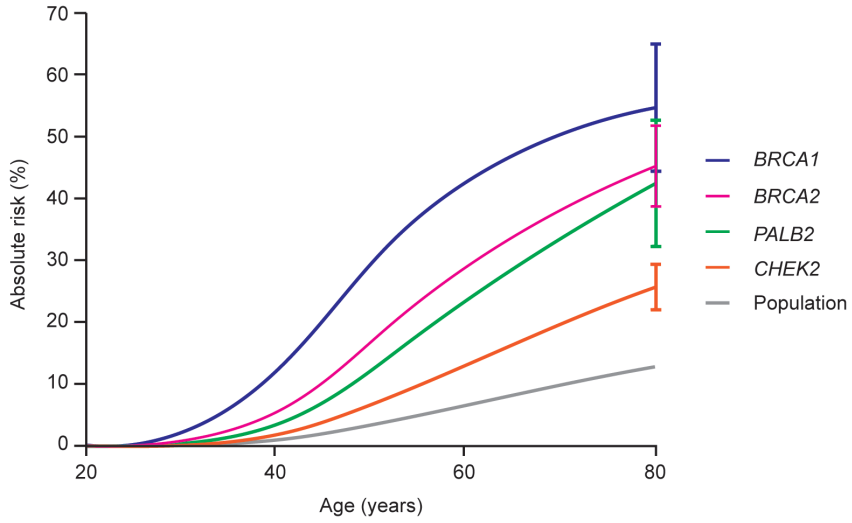


Figure 1. Estimated absolute risk of breast cancer associated with truncating variants in *BRCA1*, *BRCA2*, *PALB2* and *CHEK2* (4). The absolute risk of breast cancer is shown up to 80 years of age. The baseline estimated risk is shown in grey based on population incidences in the UK in 2016 (65). The error bars indicate 95% confidence intervals.

Homologous recombination

The three high-risk breast cancer susceptibility genes (i.e., *BRCA1*, *BRCA2* and *PALB2*) are crucial for DNA double-strand break repair by HR (Fig. 2). During HR, BRCA1 counteracts the accumulation of 53BP1, which otherwise interacts with the chromatin adjacent to the broken DNA ends to promote NHEJ (14,15). This BRCA1 activity permits DNA end-resection at the break-sites by exonucleases such as MRE11 to yield 3'-single-stranded (ss) DNA overhangs that are required for HR-mediated double-strand break repair (16). The 3'-ssDNA overhangs then become coated by replication protein A (RPA) (17), promoting the sequential accumulation of BRCA1, PALB2 and BRCA2. PALB2 is crucial in this event as it mediates the formation of the PALB2-BRCA1/2-RAD51 complex and together with BRCA2 facilitates the replacement of RPA with the RAD51 recombinase (18). The RAD51-ssDNA nucleoprotein filaments then promote the homology search using the sister chromatid and the ensuing strand exchange. As this repair pathway requires a non-damaged sister chromatid to act as a template for repair, it is mostly active during S- and G2-phase of the cell cycle and drives error-free repair of DNA double-strand breaks. As a consequence, HR is imperative for maintaining genomic stability, highlighting the importance of *BRCA1*, *BRCA2* and *PALB2* as tumor suppressor genes.

Cell cycle regulation

The moderate-risk genes *ATM* and *CHEK2* are also involved in the DNA damage response. Although their functions are linked to those of *BRCA1*, *BRCA2* and *PALB2*, they regulate the DNA damage response differently. In contrast to acting as key players in HR, their functions have predominantly been associated with the *TP53* signaling pathway. The *TP53* gene (expressing p53) represents another important tumor suppressor gene. Although somatically acquired pathogenic variants in *TP53* can be found in substantial proportions of nearly all types of cancer, germline pathogenic variants in the *TP53* gene are rare and they are associated with a significant risk for developing breast cancer (4). This is not surprising as impaired p53-mediated signaling can have a major impact on the DDR. For instance, impaired p53-mediated signaling can result in abnormal expression of numerous p53 target genes, several of which are involved in the regulation of cell cycle arrest, a process that provides cells time to repair the damaged DNA (19,20). Furthermore, defects in p53-mediated signaling may impair apoptosis. As a consequence, cells may no longer be restrained from proliferating in the presence of unrepaired DNA damage (21). It is therefore crucial for cells to have the activity of p53 carefully regulated.

Both ATM and CHK2 are important for p53-dependent signaling, as they are involved in the activation of p53 during the DDR. In fact, ATM is considered a key DNA damage signaling component in mammalian cells as it encodes a kinase that acts early in response to

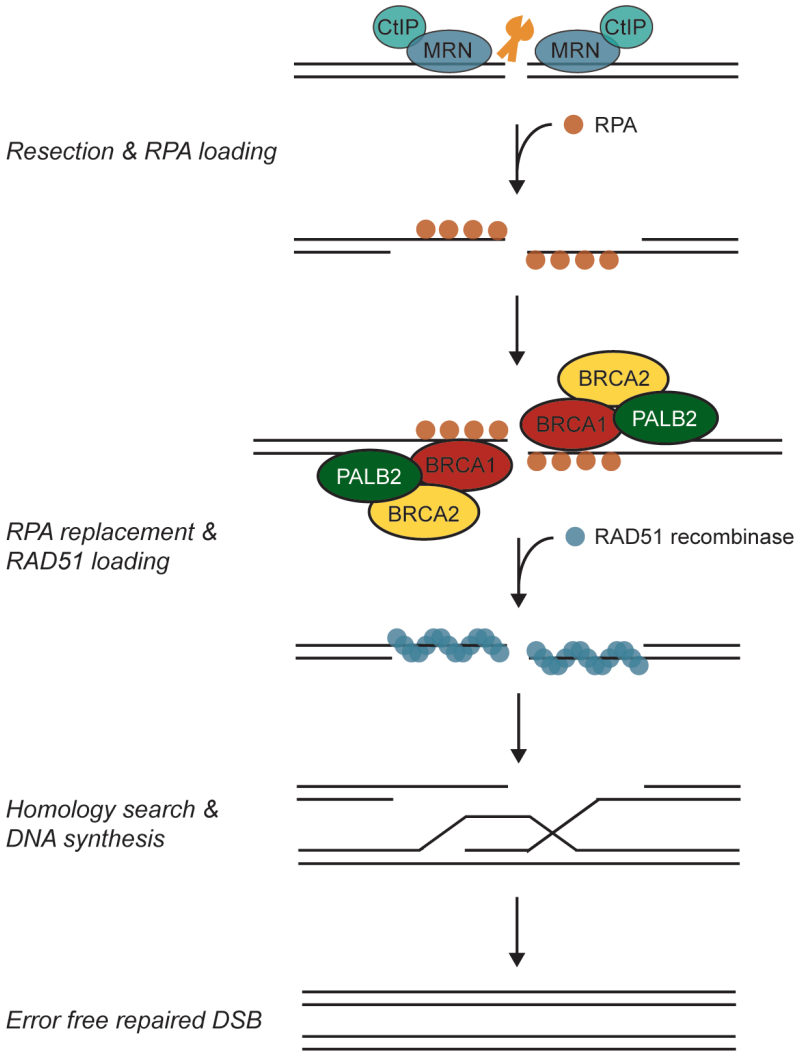


Figure 2. Schematic representation of HR-mediated repair of a DNA double-strand break. Initially, the broken DNA ends become resected by exonucleases such as MRE11 (part of the MRN complex) to yield 3'-ssDNA overhangs that are coated by RPA. Sequential recruitment of BRCA1, followed by PALB2-BRCA2 is crucial for the subsequent replacement of RPA with the RAD51 recombinase. The RAD51 nucleoprotein filaments will promote the homology search using the non-damaged sister chromatid as a template, eventually ensuing in error-free repair of the double-strand break.

DNA damage. One of the best-established downstream targets of ATM is the CHK2 kinase. CHK2, encoded by the *CHEK2* gene, functions to reduce cyclin-dependent kinase (CDK) activity by various mechanisms, including the phosphorylation and subsequent stabilization of p53. This results in an arrest in cell-cycle progression due to activation of G1-S, intra-S and possibly G2-M cell-cycle checkpoints, thereby providing time for DNA repair before cells start DNA replication and/or mitosis. These findings suggest that CHK2 and p53 act in a common pathway. Importantly, and in line with CHK2's critical role in the DDR, pathogenic variants in the *CHEK2* gene, such as the truncating c.1100delC variant, have been found to associate with a moderate risk for breast cancer (4). Consequently, it is imperative that the pathogenic potential of other type of genetic variants in *CHEK2* are well understood.

Genetic variants and clinical management

Identifying individuals who are strongly predisposed to breast cancer due to an inherited variant in a breast cancer susceptibility gene has tremendous clinical value. Such individuals may benefit from cancer prevention strategies or early detection. Several clinical features may indicate whether an individual may be at risk for breast cancer due to the presence of a genetic variant in the germline. This includes a clear positive family history (i.e., multiple (early onset) breast cancer cases), bilateral disease and distinct types of cancer (e.g., combination breast and ovarian cancer). For such cases, genetic tests can be performed that are commonly aimed at detecting variants in the *BRCA1*, *BRCA2*, *PALB2* and *CHEK2* tumor suppressor genes. Such testing can reveal the presence of different types of DNA-variants in these genes, including nonsense, frameshift, splice, missense or synonymous variants. Protein-truncating variants (PTVs, i.e., nonsense or frameshift), or variants that affect splicing, are often classified as (likely) pathogenic variants (22). These types of genetic variants are typically known to associate with high risk for breast cancer as they are expected to impair protein function. In conjunction with loss of the wildtype allele (i.e., loss of heterozygosity; LOH), which is very often seen in tumors, this means that tumor cells can no longer express a functional protein at all.

For carriers of (likely) pathogenic variants, specific recommendations for clinical management have been specified (23). For instance, measures can be taken to increase the frequency of screening for breast cancer and to consider procedures such as bilateral risk-reducing mastectomy. However, although bilateral mastectomy reduces cancer risk by at least 90% in carriers of pathogenic variants in high-risk genes such as *BRCA1* or *BRCA2* (24,25), such risk-reducing surgery is not recommended for women at moderate risk of breast cancer (e.g., due to pathogenic variants in *CHEK2*). Instead, in such carriers, annual mammography is offered on the basis that biennial screening is clinically effective in reducing advanced breast cancers and breast cancer mortality in the general population (26). Alternatively, for women

already diagnosed with breast cancer, identifying a pathogenic germline variant in *BRCA1*, *BRCA2* or *PALB2* may affect treatment options, such as surgical decisions to reduce the risk of recurrence, or the use of poly (ADP-ribose) polymerase inhibitor (PARPi) therapies. The latter option has been shown to be effective in a subset of HR-deficient tumors (27), as especially HR-deficient cells are sensitive to the inhibition of PARP (Fig. 3). It may also stimulate the testing of unaffected family members that are potentially at a similar increased risk for developing breast cancer. Taken together, it is important that carriers of (likely) pathogenic variants in the aforementioned breast cancer susceptibility genes are identified.

In contrast to PTVs or many splice variants that are clearly pathogenic, the clinical and functional impact of missense variants is often unclear. These variants are referred to as variants of uncertain significance (VUS) and for such genetic variants it is uncertain whether or not they increase the risk for developing breast cancer. This is due to the rarity of many of these missense VUS, which limits the evidence available to determine if a variant is pathogenic or benign. Accordingly, VUS cannot guide clinical decision making, complicating post-test patient counselling and clinical management (28,29). Until recently, assessment of pathogenicity of VUS relied mostly on co-segregation of the variant with cancer in families and the family history of cancer. Co-segregation is analyzed by statistical means, which usually requires multiple families to reach sufficient significance. However, the majority of VUS in breast cancer susceptibility genes occur so rarely in the general population, that they result in too few families in which the same variant can be found segregating. Hence the co-segregation as well as the associated cancer risks are difficult to assess at statistically significant levels. Furthermore, pathogenic variants in genes such as *CHEK2* are associated with a moderate risk of breast cancer. Moderate-risk alleles often cause cancer in combination with other genetic variants (such as polygenic risk alleles) and demographic, behavioral and environmental risk factors. Therefore, they can remain non-penetrant in many individuals. Accordingly, the effect of a pathogenic variant in *CHEK2* on cancer risk will often not give rise to the same sequalae seen for pathogenic variants in a high-risk gene such as *BRCA1*, *BRCA2*, and *PALB2*.

Fully realizing the clinical potential of genetic tests requires an accurate assessment of pathogenicity, even for rare genetic missense variants. To this end, additional methods for interpreting rare VUS, in both moderate- and high-risk genes, are of great value for clinical management of carriers. Knowing which VUS are damaging, or not, will help clinicians understand the test results (i.e., estimating whether a variant is pathogenic, or not) and can help to decide on the right clinical management. One powerful approach to improve the clinical classification of VUS is by using data from functional testing. Such functional evidence describes the molecular consequence of a variant on protein function and can consist of the results of either molecular or cellular experiments in vitro. When clinical data is scarce,

functional data has considerable potential to aid in variant classification, particularly VUS re-classification (30). The 2015 American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) guidelines for clinical sequence variant interpretation state that the results of a well-established functional assay can qualify as evidence to be used for clinical classification of variants (31). Aspects of the functional assay, such as calibration (by including clinically proven benign and pathogenic variants), or reproducibility of the results and the ability of the assay to reflect the tumor suppressive function of the protein, can all weigh in on the predictive power of the assay (32-34).

For *BRCA1*, *BRCA2*, and *PALB2*, assays using HR and/or resistance to DNA damaging agents have emerged as the standard for the functional characterisation of VUS in these genes (35-52). In contrast, in an effort to interpret various VUS in the moderate-risk gene *CHEK2*, several studies assessed their functional consequences, either by determining the effect on kinase activity or on cell growth (53-64). Collectively, these studies show the power of these assays in functionally characterizing many VUS, efforts that are expected to have a major impact on clinical variant interpretation.

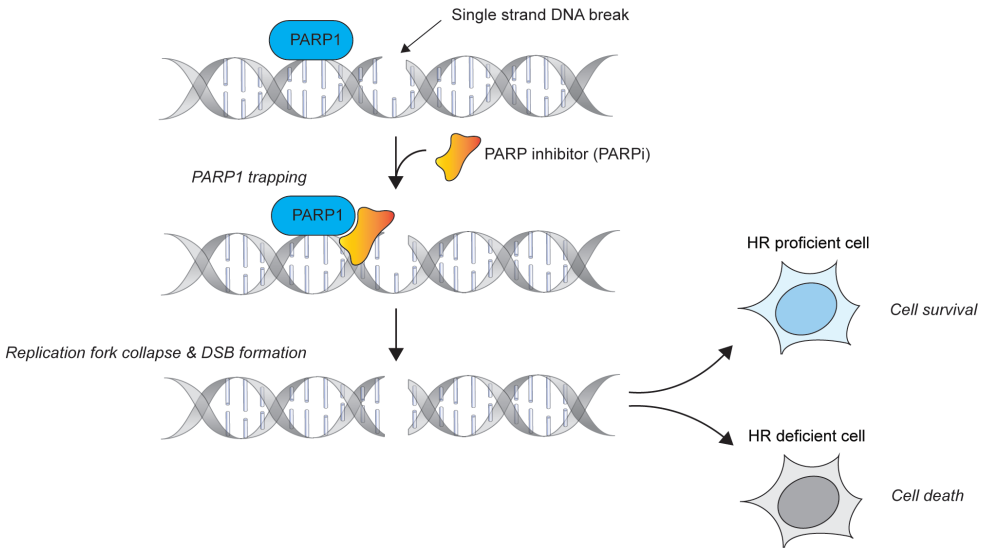


Figure 3. Mechanism of action of PARP inhibitors (PARPi). Upon the formation of single-strand breaks in the DNA, PARP1 becomes recruited and activated, resulting in the repair of these type of DNA lesions. Treatment with PARPi will result in trapping of PARP1 at the DNA lesion. This is thought to block repair of the DNA lesion and result in replication fork collapse during DNA replication, eventually ensuing in the formation of DNA double-strand breaks. This will result in lethality only in HR-deficient cells, which are unable to repair these breaks in an error free manner.

AIM AND OUTLINE OF THIS THESIS

In this thesis, I focus on the functional characterization of genetic variants in the high-risk breast cancer susceptibility gene *PALB2* and the moderate-risk gene *CHEK2.* The aim is to generate functional data for improved clinical interpretation of such variants. Quantitative assessment of the functional consequences of VUS in either gene can identify functionally damaging variants that associate with increased breast cancer risk, thereby aiding in the clinical management of patients and carriers.

In **Chapter 2** I first provide an overview of the functional analysis that have been performed by us and other research labs for variants in the *PALB2* gene. I then provide a similar overview for the *CHEK2* gene, for which different functional analysis have been used to functionally characterize numerous *CHEK2* genetic variants (**Chapter 3**). In **Chapter 4**, I present my results on the functional analysis of *PALB2* variants. Following careful validation of our newly developed functional assay, I could show that several missense VUS located in the Coiled-Coil (CC) and WD40 domains of PALB2, can result in major effects on protein function. I then present additional findings on the functional analysis of *PALB2* variants, showing that I could adapt our functional assay to allow for a high-throughput analysis of nearly all possible missense variants in the CC domain of PALB2 (**Chapter 5**). In **Chapter 6**, I discuss results on the functional analysis of *CHEK2* variants. Using a newly developed assay, I show that the degree of functional impact of variants in *CHEK2* correlates with breast cancer risk. In **Chapter 7**, I conclude my thesis with future perspectives on how the functional assays presented in this thesis can be further optimized to meet the clinical demand for functional data. Finally, I also discuss what would be required for these assays to be further implemented during clinical variant interpretation and risk assessment.

REFERENCES

1. Global Burden of Disease Cancer C, Kocarnik JM, Compton K, Dean FE, Fu W, Gaw BL*, et al.* Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life Years for 29 Cancer Groups From 2010 to 2019: A Systematic Analysis for the Global Burden of Disease Study 2019. JAMA Oncol **2021** doi 10.1001/jamaoncol.2021.6987.
2. Britt KL, Cuzick J, Phillips KA. Key steps for effective breast cancer prevention. Nat Rev Cancer **2020**;20(8):417-36 doi 10.1038/s41568-020-0266-x.
3. Society AC. Accessed on Jan 3, 2022. Breast Cancer Facts & Figures 2019-2020. <www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2019-2020.pdf>. Accessed on Jan 3, 2022.
4. Breast Cancer Association C, Dorling L, Carvalho S, Allen J, Gonzalez-Neira A, Luccarini C*, et al.* Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. N Engl J Med **2021**;384(5):428-39 doi 10.1056/NEJMoa1913948.
5. Tubbs A, Nussenzweig A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. Cell **2017**;168(4):644-56 doi 10.1016/j.cell.2017.01.002.
6. Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. Mol Cell **2010**;40(2):179-204 doi 10.1016/j.molcel.2010.09.019.
7. Jackson SP, Bartek J. The DNA-damage response in human biology and disease. Nature **2009**;461(7267):1071-8 doi 10.1038/nature08467.
8. Harper JW, Elledge SJ. The DNA damage response: ten years after. Mol Cell **2007**;28(5):739-45 doi 10.1016/j.molcel.2007.11.015.
9. Giglia-Mari G, Zotter A, Vermeulen W. DNA damage response. Cold Spring Harb Perspect Biol **2011**;3(1):a000745 doi 10.1101/cshperspect.a000745.
10. Ma J, Setton J, Lee NY, Riaz N, Powell SN. The therapeutic significance of mutational signatures from DNA repair deficiency in cancer. Nat Commun **2018**;9(1):3292 doi 10.1038/s41467-018-05228-y.
11. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X*, et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature **2016**;534(7605):47-54 doi 10.1038/nature17676.
12. Duijf PHG, Nanayakkara D, Nones K, Srihari S, Kalimutho M, Khanna KK. Mechanisms of Genomic Instability in Breast Cancer. Trends Mol Med **2019**;25(7):595-611 doi 10.1016/j.molmed.2019.04.004.
13. Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J*, et al.* Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. JAMA Oncol **2017**;3(9):1190-6 doi 10.1001/jamaoncol.2017.0424.
14. Chapman JR, Taylor MR, Boulton SJ. Playing the end game: DNA double-strand break repair pathway choice. Mol Cell **2012**;47(4):497-510 doi 10.1016/j.molcel.2012.07.029.

15. Densham RM, Garvin AJ, Stone HR, Strachan J, Baldock RA, Daza-Martin M*, et al.* Human BRCA1-BARD1 ubiquitin ligase activity counteracts chromatin barriers to DNA resection. Nat Struct Mol Biol **2016**;23(7):647-55 doi 10.1038/nsmb.3236.
16. Marini F, Rawal CC, Liberi G, Pellicioli A. Regulation of DNA Double Strand Breaks Processing: Focus on Barriers. Front Mol Biosci **2019**;6:55 doi 10.3389/fmolb.2019.00055.
17. Symington LS. Mechanism and regulation of DNA end resection in eukaryotes. Crit Rev Biochem Mol Biol **2016**;51(3):195-212 doi 10.3109/10409238.2016.1172552.
18. Jensen RB, Carreira A, Kowalczykowski SC. Purified human BRCA2 stimulates RAD51-mediated recombination. Nature **2010**;467(7316):678-83 doi 10.1038/nature09399.
19. Arizti P, Fang L, Park I, Yin Y, Solomon E, Ouchi T*, et al.* Tumor suppressor p53 is required to modulate BRCA1 expression. Mol Cell Biol **2000**;20(20):7450-9 doi 10.1128/MCB.20.20.7450-7459.2000.
20. Kannan K, Amariglio N, Rechavi G, Givol D. Profile of gene expression regulated by induced p53: connection to the TGF-beta family. FEBS Lett **2000**;470(1):77-82 doi 10.1016/s0014-5793(00)01291-6.
21. Reinhardt HC, Schumacher B. The p53 network: cellular and systemic DNA damage responses in aging and cancer. Trends Genet **2012**;28(3):128-36 doi 10.1016/j.tig.2011.12.002.
22. Lappalainen T, MacArthur DG. From variant to function in human disease genetics. Science **2021**;373(6562):1464-8 doi 10.1126/science.abi8207.
23. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS*, et al.* Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat **2008**;29(11):1282-91 doi 10.1002/humu.20880.
24. Domchek SM, Friebel TM, Singer CF, Evans DG, Lynch HT, Isaacs C*, et al.* Association of risk-reducing surgery in BRCA1 or BRCA2 mutation carriers with cancer risk and mortality. JAMA **2010**;304(9):967-75 doi 10.1001/jama.2010.1237.
25. Ludwig KK, Neuner J, Butler A, Geurts JL, Kong AL. Risk reduction and survival benefit of prophylactic surgery in BRCA mutation carriers, a systematic review. Am J Surg **2016**;212(4):660-9 doi 10.1016/j.amjsurg.2016.06.010.
26. Fracheboud J, Otto SJ, van Dijck JA, Broeders MJ, Verbeek AL, de Koning HJ*, et al.* Decreased rates of advanced breast cancer due to mammography screening in The Netherlands. Br J Cancer **2004**;91(5):861-7 doi 10.1038/sj.bjc.6602075.
27. Geenen JJJ, Linn SC, Beijnen JH, Schellens JHM. PARP Inhibitors in the Treatment of Triple-Negative Breast Cancer. Clin Pharmacokinet **2018**;57(4):427-37 doi 10.1007/s40262-017-0587-4.
28. Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G*, et al.* Variant Interpretation: Functional Assays to the Rescue. Am J Hum Genet **2017**;101(3):315-25 doi 10.1016/j.ajhg.2017.07.014.
29. Domchek SM, Bradbury A, Garber JE, Offit K, Robson ME. Multiplex genetic testing for cancer susceptibility: out on the high wire without a net? J Clin Oncol **2013**;31(10):1267-70 doi 10.1200/JCO.2012.46.9403.

30. Brnich SE, Rivera-Munoz EA, Berg JS. Quantifying the potential of functional evidence to reclassify variants of uncertain significance in the categorical and Bayesian interpretation frameworks. Hum Mutat **2018**;39(11):1531-41 doi 10.1002/humu.23609.
31. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J*, et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med **2015**;17(5):405-24 doi 10.1038/gim.2015.30.
32. Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD*, et al.* Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med **2019**;12(1):3 doi 10.1186/s13073-019-0690-2.
33. Kanavy DM, McNulty SM, Jairath MK, Brnich SE, Bizon C, Powell BC*, et al.* Comparative analysis of functional assay evidence use by ClinGen Variant Curation Expert Panels. Genome Med **2019**;11(1):77 doi 10.1186/s13073-019-0683-1.
34. Nykamp K, Anderson M, Powers M, Garcia J, Herrera B, Ho YY*, et al.* Sherloc: a comprehensive refinement of the ACMG-AMP variant classification criteria. Genet Med **2017**;19(10):1105-17 doi 10.1038/gim.2017.37.
35. Bouwman P, van der Gulden H, van der Heijden I, Drost R, Klijn CN, Prasetyanti P*, et al.* A high-throughput functional complementation assay for classification of BRCA1 missense variants. Cancer Discov **2013**;3(10):1142-55 doi 10.1158/2159-8290.CD-13-0094.
36. Chang S, Biswas K, Martin BK, Stauffer S, Sharan SK. Expression of human BRCA1 variants in mouse ES cells allows functional analysis of BRCA1 mutations. J Clin Invest **2009**;119(10):3160-71 doi 10.1172/JCI39836.
37. Ransburgh DJ, Chiba N, Ishioka C, Toland AE, Parvin JD. Identification of breast tumor mutations in BRCA1 that abolish its function in homologous DNA recombination. Cancer Res **2010**;70(3):988-95 doi 10.1158/0008-5472.CAN-09-2850.
38. Towler WI, Zhang J, Ransburgh DJ, Toland AE, Ishioka C, Chiba N*, et al.* Analysis of BRCA1 variants in double-strand break repair by homologous recombination and single-strand annealing. Hum Mutat **2013**;34(3):439-45 doi 10.1002/humu.22251.
39. Mesman RLS, Calleja F, Hendriks G, Morolli B, Misovic B, Devilee P*, et al.* The functional impact of variants of uncertain significance in BRCA2. Genet Med **2019**;21(2):293-302 doi 10.1038/s41436-018-0052-2.
40. Shimelis H, Mesman RLS, Von Nicolai C, Ehlen A, Guidugli L, Martin C*, et al.* BRCA2 Hypomorphic Missense Variants Confer Moderate Risks of Breast Cancer. Cancer Res **2017**;77(11):2789-99 doi 10.1158/0008-5472.CAN-16-2568.
41. Starita LM, Islam MM, Banerjee T, Adamovich AI, Gullingsrud J, Fields S*, et al.* A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution Variants on Protein Function. Am J Hum Genet **2018**;103(4):498-508 doi 10.1016/j.ajhg.2018.07.016.

42. Park JY, Singh TR, Nassar N, Zhang F, Freund M, Hanenberg H*, et al.* Breast cancer-associated missense mutants of the PALB2 WD40 domain, which directly binds RAD51C, RAD51 and BRCA2, disrupt DNA repair. Oncogene **2014**;33(40):4803-12 doi 10.1038/onc.2013.421.
43. Foo TK, Tischkowitz M, Simhadri S, Boshari T, Zayed N, Burke KA*, et al.* Compromised BRCA1-PALB2 interaction is associated with breast cancer risk. Oncogene **2017**;36(29):4161-70 doi 10.1038/onc.2017.46.
44. Boonen R, Rodrigue A, Stoepker C, Wiegant WW, Vroling B, Sharma M*, et al.* Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. Nat Commun **2019**;10(1):5296 doi 10.1038/s41467-019-13194-2.
45. Boonen R, Vreeswijk MPG, van Attikum H. Functional Characterization of PALB2 Variants of Uncertain Significance: Toward Cancer Risk and Therapy Response Prediction. Front Mol Biosci **2020**;7:169 doi 10.3389/fmolb.2020.00169.
46. Ng PS, Boonen RA, Wijaya E, Chong CE, Sharma M, Knaup S*, et al.* Characterisation of protein-truncating and missense variants in PALB2 in 15 768 women from Malaysia and Singapore. J Med Genet **2021** doi 10.1136/jmedgenet-2020-107471.
47. Brnich SE, Arteaga EC, Wang Y, Tan X, Berg JS. A Validated Functional Analysis of Partner and Localizer of BRCA2 Missense Variants for Use in Clinical Variant Interpretation. J Mol Diagn **2021**;23(7):847-64 doi 10.1016/j.jmoldx.2021.04.010.
48. Wiltshire T, Ducy M, Foo TK, Hu C, Lee KY, Belur Nagaraj A*, et al.* Functional characterization of 84 PALB2 variants of uncertain significance. Genet Med **2020**;22(3):622-32 doi 10.1038/s41436-019-0682-z.
49. Ducy M, Sesma-Sanz L, Guitton-Sert L, Lashgari A, Gao Y, Brahiti N*, et al.* The Tumor Suppressor PALB2: Inside Out. Trends Biochem Sci **2019**;44(3):226-40 doi 10.1016/j.tibs.2018.10.008.
50. Nepomuceno TC, Carvalho MA, Rodrigue A, Simard J, Masson JY, Monteiro ANA. PALB2 Variants: Protein Domains and Cancer Susceptibility. Trends Cancer **2021**;7(3):188-97 doi 10.1016/j.trecan.2020.10.002.
51. Rodrigue A, Margaillan G, Torres Gomes T, Coulombe Y, Montalban G, da Costa ESCS*, et al.* A global functional analysis of missense mutations reveals two major hotspots in the PALB2 tumor suppressor. Nucleic Acids Res **2019**;47(20):10662-77 doi 10.1093/nar/gkz780.
52. Southey MC, Rewse A, Nguyen-Dumont T. PALB2 Genetic Variants: Can Functional Assays Assist Translation? Trends Cancer **2020**;6(4):263-5 doi 10.1016/j.trecan.2020.01.017.
53. Bell DW, Kim SH, Godwin AK, Schiripo TA, Harris PL, Haserlat SM*, et al.* Genetic and functional analysis of CHEK2 (CHK2) variants in multiethnic cohorts. Int J Cancer **2007**;121(12):2661-7 doi 10.1002/ijc.23026.
54. Chrisanthar R, Knappskog S, Lokkevik E, Anker G, Ostenstad B, Lundgren S*, et al.* CHEK2 mutations affecting kinase activity together with mutations in TP53 indicate a functional pathway associated with resistance to epirubicin in primary breast cancer. PLoS One **2008**;3(8):e3062 doi 10.1371/journal.pone.0003062.

55. Cuella-Martin R, Hayward SB, Fan X, Chen X, Huang JW, Taglialatela A*, et al.* Functional interrogation of DNA damage response variants with base editing screens. Cell **2021**;184(4):1081-97 e19 doi 10.1016/j.cell.2021.01.041.
56. Delimitsou A, Fostira F, Kalfakakou D, Apostolou P, Konstantopoulou I, Kroupis C*, et al.* Functional characterization of CHEK2 variants in a Saccharomyces cerevisiae system. Hum Mutat **2019**;40(5):631-48 doi 10.1002/humu.23728.
57. Falck J, Mailand N, Syljuasen RG, Bartek J, Lukas J. The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. Nature **2001**;410(6830):842-7 doi 10.1038/35071124.
58. Kleiblova P, Stolarova L, Krizova K, Lhota F, Hojny J, Zemankova P*, et al.* Identification of deleterious germline CHEK2 mutations and their association with breast and ovarian cancer. Int J Cancer **2019**;145(7):1782-97 doi 10.1002/ijc.32385.
59. Lee SB, Kim SH, Bell DW, Wahrer DC, Schiripo TA, Jorczak MM*, et al.* Destabilization of CHK2 by a missense mutation associated with Li-Fraumeni Syndrome. Cancer Res **2001**;61(22):8062-7.
60. Roeb W, Higgins J, King MC. Response to DNA damage of CHEK2 missense mutations in familial breast cancer. Hum Mol Genet **2012**;21(12):2738-44 doi 10.1093/hmg/dds101.
61. Shaag A, Walsh T, Renbaum P, Kirchhoff T, Nafa K, Shiovitz S*, et al.* Functional and genomic approaches reveal an ancient CHEK2 allele associated with breast cancer in the Ashkenazi Jewish population. Hum Mol Genet **2005**;14(4):555-63 doi 10.1093/hmg/ddi052.
62. Tischkowitz MD, Yilmaz A, Chen LQ, Karyadi DM, Novak D, Kirchhoff T*, et al.* Identification and characterization of novel SNPs in CHEK2 in Ashkenazi Jewish men with prostate cancer. Cancer Lett **2008**;270(1):173-80 doi 10.1016/j.canlet.2008.05.006.
63. Wang N, Ding H, Liu C, Li X, Wei L, Yu J*, et al.* A novel recurrent CHEK2 Y390C mutation identified in high-risk Chinese breast cancer patients impairs its activity and is associated with increased breast cancer risk. Oncogene **2015**;34(40):5198-205 doi 10.1038/onc.2014.443.
64. Wu X, Webster SR, Chen J. Characterization of tumor-associated Chk2 mutations. J Biol Chem **2001**;276(4):2971-4 doi 10.1074/jbc.M009727200.
65. UK CR. 2020 Breast cancer incidence (invasive) statistics. <www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/breast-cancer/incidence-invasive>.

CHAPTER 2

Functional characterization of *PALB2* variants of uncertain significance: towards cancer risk and therapy response prediction

Rick A.C.M. Boonen, Maaike P.G. Vreeswijk, Haico van Attikum

Published in *Frontiers in Molecular Biosciences*
(PMID: 33195396)

ABSTRACT

In recent years it has become clear that pathogenic variants in *PALB2* are associated with a high risk for breast, ovarian and pancreatic cancer. However, the clinical relevance of variants of uncertain significance (VUS) in *PALB2*, which are increasingly identified through clinical genetic testing, is unclear. Here we review recent advances in the functional characterization of VUS in *PALB2*. A combination of assays has been used to assess the impact of *PALB2* VUS on its function in DNA repair by homologous recombination, cell cycle regulation and the control of cellular levels of reactive oxygen species (ROS). We discuss the outcome of this comprehensive analysis of *PALB2* VUS, which showed that VUS in PALB2's Coiled-Coil (CC) domain can impair the interaction with BRCA1, whereas VUS in its WD40 domain affect PALB2 protein stability. Accordingly, the CC and WD40 domains of PALB2 represent hotspots for variants that impair PALB2 protein function. We also provide a future perspective on the high-throughput analysis of VUS in *PALB2*, as well as the functional characterization of variants that affect *PALB2* RNA splicing. Finally, we discuss how results from these functional assays can be valuable for predicting cancer risk and responsiveness to cancer therapy, such as treatment with PARP inhibitor- or platinum-based chemotherapy.

KEYWORDS

Breast Cancer; Variant of Uncertain Significance (VUS); PALB2; DNA Repair Homologous Recombination (HR); PARP inhibitor

PALB2 is essential for DSB repair by homologous recombination

The integrity of our genome is relentlessly challenged by exogenous and endogenous insults that can induce DNA damage. To respond to such genotoxic threats, cells have evolved a number of DNA damage signalling and repair mechanisms, jointly known as the DNA damage response (DDR). The DDR is able to handle a myriad of DNA damages of which DNA double strand breaks (DSBs) are considered among the most deleterious to the cell. Human cells possess at least five pathways for DSB repair: canonical nonhomologous end joining (c-NHEJ), alternative nonhomologous end-joining (a-NHEJ), single-strand annealing (SSA), break-induced replication (BIR), and homologous recombination (HR) (1,2). c-NHEJ is the predominant DSB repair pathway in human cells and complete loss-of-function (LOF) is likely to drive cell death due to an unreasonably high DSB burden (3). In case c-NHEJ fails or is inappropriate, HR is probably the most frequently used alternative pathway for DSB repair. However, while c-NHEJ is active throughout the whole cell cycle, HR is restricted to late S/G2 phase as it relies on the presence of an undamaged sister chromatid to act as a template for error free repair (4). During HR, BRCA1 inhibits 53BP1 from interacting with the chromatin near the broken DNA ends (2,5). This permits extensive end-resection of the break by endo- and exonucleases such as MRE11, CtIP, DNA2, and EXO1, yielding 3′-single-stranded (ss) DNA overhangs that counter Ku loading and further promote DSB repair by HR (6). Following resection, the 3′-ssDNA tails become coated by the RPA heterotrimer (7). Subsequently, BRCA1, PALB2 and BRCA2 sequentially accumulate on the processed ssDNA to promote error-free repair of DSBs.

PALB2 is crucial herein as it mediates PALB2-BRCA1/2-RAD51 complex formation. That is, PALB2's N-terminal Coiled-Coil (CC) domain is required for interaction with BRCA1, whereas its C-terminal WD40 domain mediates the interaction with BRCA2 (Fig. 1) (8-12). BRCA2 possesses eight highly conserved BRC repeats and a carboxy-terminal region that have been shown to bind RAD51 (13-15). This interaction allows BRCA2 to promote HR by facilitating the replacement of RPA with the RAD51 recombinase and by stabilizing the ensuing RAD51-ssDNA filaments through blockage of ATP hydrolysis (16). Additionally, through its WD40 domain, PALB2 also interacts with the C-terminal PALB2-Interacting Domain (PID) of the RNF168 ubiquitin E3 ligase. RNF168 contains a ubiquitin-interacting motif (UIM) that allows binding of RNF168-bound PALB2 to ubiquitylated chromatin at DSBs, thereby facilitating RAD51 filament formation and HR (17). Alternatively, more recent studies suggested that RNF168 may facilitate PALB2-mediated RAD51 loading independently of BRCA1, by showing that abrogation of RNF168 activity in BRCA1-compromised cells dramatically elevated genome instability rates (18,19). Thus, it is apparent that RAD51 loading during HR, regardless of its dependency on BRCA1, RNF168, or both, requires the action of PALB2.

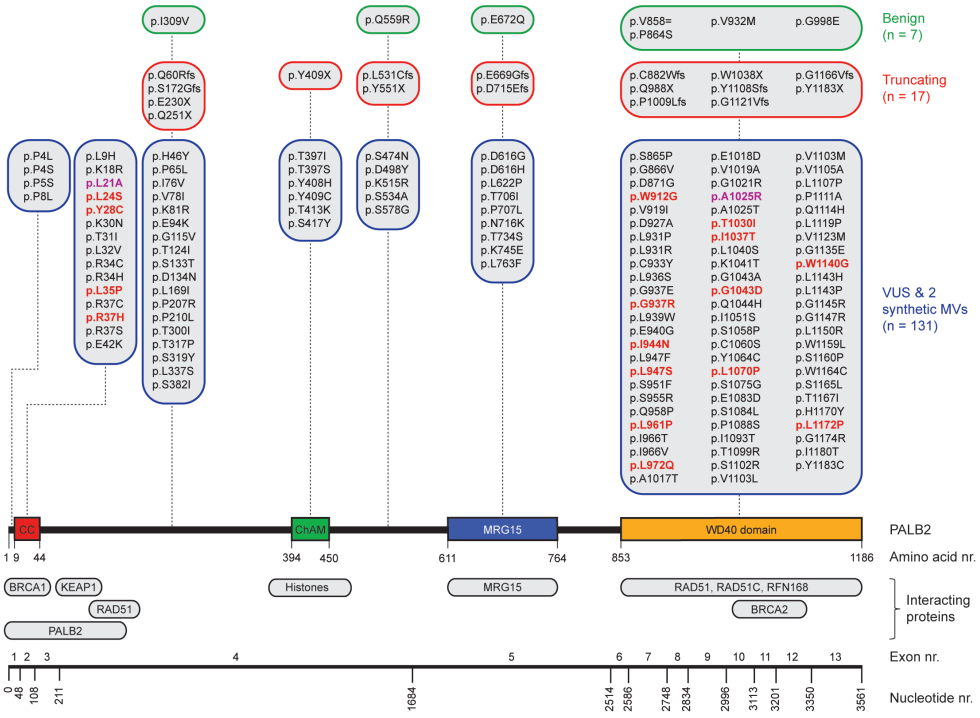


Figure 1. Schematic representation of *PALB2* variants, functional domains, interacting proteins and exons. The nucleotide numbers refer to the last nucleotide of each exon in *PALB2* cDNA (NM_024675.3). The amino acid numbers are shown to specify the evolutionarily conserved functional domains of PALB2; Coiled-coil (CC) (10-12,107), Chromatin-Association Motif (ChAM) (108), MORF-Related Gene on chromosome 15 (MRG15) binding domain (109) and WD40 domain (9,107). PALB2-interacting proteins are depicted underneath their respective PALB2 interacting domain/regions. All *PALB2* genetic variants from five functional studies (39-43) are shown and categorized per (functional) domain as benign (green framed sections), truncating (red framed sections), or VUS and synthetic missense variants (MVs) (blue framed sections) based on ClinVar. All functionally damaging *PALB2* VUS with an HR efficiency < 50% compared to wild type *PALB2* in at least one functional assay are highlighted in red. The two damaging synthetic MVs are highlighted in purple.

Genetic variants in *PALB2* and their association with cancer

Recent analysis of a metastatic pan-cancer cohort of 3504 patients, employing a strategy that relies on the presence of specific mutational footprints which are characteristic of a deficiency in HR (20,21), revealed that mutational inactivation of the *BRCA1*, *BRCA2* and *PALB2* genes, was the most common genetic cause of the observed HR signatures (22), indicative of their important role in tumor suppression. Indeed, for *BRCA1* and *BRCA2*, monoallelic LOF variants present in the germline can result in a nearly tenfold increased lifetime risk of developing breast cancer (23,24), whereas bi-allelic LOF variants cause Fanconi anemia (FA) (25,26). The

PALB2 gene, which is located on chromosome 16p12.2, comprises 13 exons and encodes a protein of 1186-amino acids (Fig. 1), was identified in 2006 as an important BRCA2-interacting protein (9,27). As it has now been established that LOF variants in *PALB2* convey a similarly high risk for breast cancer as *BRCA2* LOF variants (23,24,28), *PALB2* has become widely included in breast cancer clinical genetics practice. Consequently, a large number of people have already undergone genetic testing of *PALB2* to identify variants that may increase the risk of breast cancer susceptibility. Meanwhile, truncating *PALB2* variants have also been shown to be associated with an increased risk of familial ovarian and pancreatic cancer (29-33).

In contrast to truncating variants in *PALB2*, which are known to be deleterious to protein function, the impact of most missense variants is often unclear. Generally, assessment of pathogenicity of such variants of uncertain significance (VUS) would rely mostly on in silico analysis, co-segregation of the variant with cancer, co-occurrence with pathogenic *PALB2* variants, and family history of cancer. However, for the majority of VUS, this information is not available and hence the associated cancer risk is unknown. To extend the utility of *PALB2* genetic test results, additional methods for interpreting VUS are therefore urgently required. Accordingly, recent independent studies have developed functional assays to determine the functional impact of a large number of *PALB2* VUS (Fig. 1). Here we review this comprehensive analysis, which highlights the CC and WD40 domains of PALB2 as hotspots for variants that impair its function in HR and cell cycle checkpoint regulation. Finally, we also highlight the value of this functional analysis in predicting the associated cancer risk and therapy response for VUS in *PALB2*.

A comprehensive functional analysis of VUS in *PALB2*

Assays using HR as a read-out have emerged as the standard for the functional characterisation of VUS in *BRCA1* and *BRCA2* (34-38). More recently, VUS in *PALB2* have also been characterised in a similar manner (Fig. 1 and Fig. 2) (39-43). To identify variants that impact HR, the well described DR-GFP reporter, as well as the more recently introduced CRISPR-LMNA HR assay were used (44,45). These assays rely on HR-mediated repair of a non-functional GFP gene and HR-mediated integration of a fluorescence marker at the *LMNA* locus, respectively (Fig. 2). Furthermore, PALB2 function was assessed by exposing cells that express a *PALB2* variant to PARPi or cisplatin (Fig. 2). Catalytic inhibition of PARP1 “traps” PARP1 molecules on endogenous ssDNA breaks, resulting in replication fork collapse and DSB formation (46). Cisplatin on the other hand, induces ~90% intra-strand cross-links and ~5% inter-strand cross-links (ICLs), the latter of which are converted into DSBs and predominantly repaired through the FA pathway (47). Both PARPi- and ICL-induced DSBs are repaired by HR. Consequently, in the absence of HR (e.g. due to PALB2 LOF), PARP-trapping

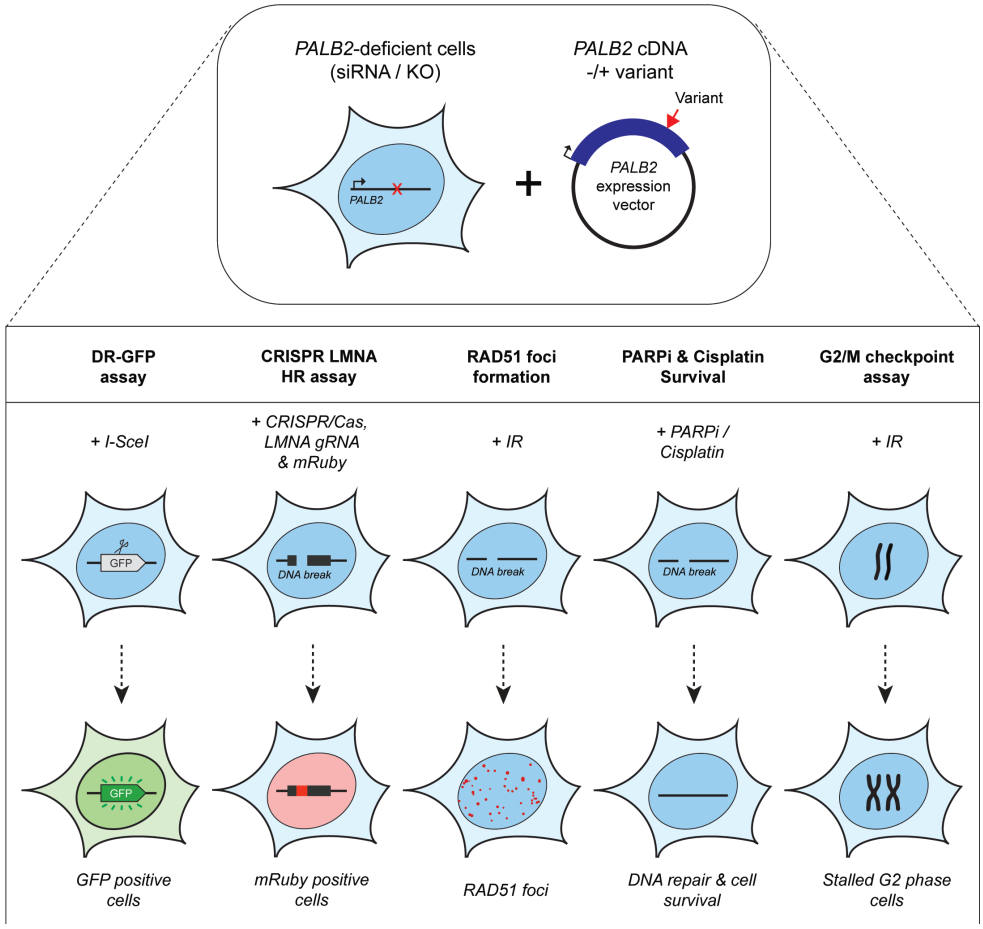


Figure 2. Overview of the functional assays used for the functional characterization of *PALB2* genetic variants. Either *Palb2* KO mouse cells or *PALB2* siRNA-depleted human cells were complemented by expressing human *PALB2* (siRNA-resistant) cDNA, without or with a variant. *PALB2* deficiency is indicated with a red cross, whereas a red arrow marks the position of a variant in the *PALB2* cDNA. Complementation was either by transient (B400 mouse cells or human cell lines) or stable expression (mES cells) (top section). *PALB2* complemented cells were subjected to multiple cell-based functional assays (bottom section). The functional assays determine in a quantitative manner: **1**) homology-directed repair of an I-*Sce*I–induced DSB in DR-GFP, which results in the restoration of a functional *GFP* gene whose expression can be monitored by fluorescence-activated cell sorting (FACS), **2**) HR-mediated integration of mRuby into the *LMNA* A/C locus (LMNA) at a break site induced by CRISPR/Cas9, **3**) the formation of IR-induced RAD51 foci, which is PALB2 dependent and provides a measurement for the HR efficiency, **4**) sensitivity to PARPi or cisplatin treatment, which leads to cell killing when HR is impaired, **5**) G2/M checkpoint maintenance after extensive DNA damage, which is dependent on PALB2-mediated HR. Deficiency in PALB2 results in progression into M-phase. Consequently, the mitotic fraction represents a measure for the functional impact of *PALB2* variants.

or ICL induction leads to persistent accumulation of DSBs. Such extensive DNA damage often results in cell cycle arrest and apoptosis, and thus, reduced proliferation and cell survival. PALB2 LOF is therefore synthetic lethal with PARPi or cisplatin treatment (48-50). Furthermore, PALB2 is also required for the repair of ionizing radiation-induced DSBs. This phenotype was used as a readout for the functional characterization of several *PALB2* variants, revealing that the expression of two variants, p.L939W and p.L1143P, impaired PALB2 functionally (41). Lastly, since PALB2 interacts with BRCA1 and BRCA2 to load RAD51 at sites of DSBs, co-immunoprecipitation, recruitment to micro irradiation-induced DNA damage, and DNA-damage-induced RAD51 foci formation were among the additional functional readouts to study the impact of PALB2 variants on HR (Fig. 2) (39-43). A complete overview of all functional assays that were performed by the three recent studies is provided in Table 1. With the above described functional assays, these studies analysed a total of 155 different *PALB2* variants (Table 2), comprising 129 VUS, 7 benign variants (as classified by ClinVar) (51), 2 synthetic missense variants with known LOF (11,52) and 17 truncating variants (Fig. 1). Sixteen VUS were identified as strongly damaging in at least one assay (i.e., >50% reduced activity compared to WT), all of which were located in the CC or WD40 domain of PALB2 (Fig. 1), highlighting the importance of these domains for PALB2's role HR. In the following sections, we review the different strategies and outcomes of these studies in more depth.

Functional characterization of VUS in *PALB2* using HR as a read-out

The largest set of *PALB2* variants, i.e., 84 patient-derived *PALB2* missense variants, was analysed by Wiltshire and colleagues (Table 2) (43). Several truncating variants (p.Q251X, p.Y551X p.D715Efs, p.Y1108Sfs, p.G1121Vfs, p.G1166Vfs and p.Y1183X) and benign missense variants as classified by ClinVar (p.I309V, p.Q559R, p.E672Q, p.P864S, p.V932M and p.G998E) were analysed to validate their functional impact. The assays were mostly performed in *Palb2*-deficient B400 mouse mammary tumour cells with a stably integrated DR-GFP reporter to measure HR. *PALB2* cDNA, with or without a variant, was transiently (over-) expressed in these cells and subsequently the effect on HR was determined. While benign variants had only a moderate or no impact on HR (<12% reduction in HR when compared to WT *PALB2*), all truncating variants strongly impacted HR (>52% reduction in HR when compared to WT *PALB2*). Moreover, four *PALB2* missense variants (p.L24S, p.L35P, p.I944N and p. L1070P) were identified that strongly disrupted HR (>65% reduction in HR compared to WT *PALB2*). To corroborate their findings, a CRISPR-*LMNA* HR assay (45) was performed in U2OS cells with endogenous *PALB2* depletion by siRNA treatment, followed by transient expression of siRNA-resistant *PALB2* cDNA with or without variant. Consistently, the same four variants disrupted HR-mediated mRuby integration into the *LMNA* locus. In this assay, the

variants exhibited a >90% reduction in HR compared to cells that were complemented with WT *PALB2* cDNA.

Table 1. Complete list of functional assays used in three independent studies.

Study	Functional assay	Nr. of variants tested (patient derived)
Boonen et al. 2019 (70 variants)	DR-GFP reporter	Complete set
	PARPi sensitivity (proliferation)	Complete set
	PALB2 expression blots	Complete set
	PARPi sensitivity (clonogenic)	8
	Cisplatin sensitivity (proliferation)	18
	RAD51 foci number after IR	5
	RAD51 foci intensity after IR	2
	CRISPR-LMNA HR	5
	G2>M checkpoint	19
	Micro-irradiation recruitment	3
	Co-immunoprecipitation	3
Rodrigue et al. 2019 (47 variants)	PARPi sensitivity proliferation assay	Complete set
	PALB2 expression blots	Complete set
	RAD51 foci number after IR	18
	RAD51 foci intensity after IR	8
	CRISPR-LMNA HR assay	18
	Micro-irradiation recruitment assay	18
	PALB2 cellular localization	18
	Mammalian two-hybrid assay (BRCA1) (1-319)	22
	Mammalian two-hybrid assay (BRCA2) (859-1186)	25
Wiltshire et al. 2019 (91 variants)	DR-GFP reporter	Complete set
	PARPi sensitivity (proliferation)	5
	PALB2 expression blots	6
	Cisplatin sensitivity (proliferation)	5
	RAD51 foci number after IR	4
	CRISPR-LMNA HR assay	4
	Micro-irradiation recruitment	4
	PALB2 cellular localization	4
	Mammalian two-hybrid assay (BRCA1) (1-319)	3
	Mammalian two-hybrid assay (BRCA2) (859-1186)	3
	Cyclohexamide chase / Stability	5
	Co-immunoprecipitation	6

Rodrigue and colleagues first tested their set of 41 *PALB2* VUS using PARPi sensitivity assays (Table 2) (42). Their assay was set up in HeLa cells in which endogenous *PALB2* was depleted by siRNA treatment. Following transient expression of siRNA-resistant YFP-*PALB2* cDNA, with or without a variant, cells were assayed for PARPi sensitivity. Although no truncating variants were assayed, several benign *PALB2* variants were included (i.e. p.P864S, p.V932M and p.G998E). As expected, expression of the benign variants rendered cells PARPi resistant, which was comparable to that observed after WT *PALB2* expression (42). The threshold for impaired PALB2 function was set based on the PARPi sensitivity observed for cells expressing p.L35P (~50% survival), which was previously reported to be damaging (40). The expression of two *PALB2* variants, p.T1030I and p.W1140G, rendered cells nearly as sensitive as those expressing p.L35P, with survival percentages of 58% and 64%, respectively, while the expression of several other variants (p.P8L, p.K18R, p.R37H, p.H46Y, p.L947F, p.L947S and p.L1119P) only resulted in a moderate, but still significant sensitivity to PARPi (~76-86% cell survival). For a more direct assessment of HR competency, the CRISPR-*LMNA* HR assay (45) was used to further characterize the effects of 18 selected *PALB2* variants on HR. Consistently, p.T1030I and p.W1140G exhibited substantially reduced HR (>65% reduction in HR when compared to WT *PALB2*), followed by p.Y28C and p.R37H (60-65% reduction in HR when compared to WT *PALB2*), whereas other variants (p.P8L, p.L947F, p.L947S and p.G1043A), showed more intermediate phenotypes (40-60% reduction in HR when compared to WT *PALB2*).

In our recent study (39), a large number of *PALB2* truncating variants was included (p.Q60Rfs, p.S172Gfs, p.E230X, p.Y409X, p.L531Cfs, p.Y551X, p.E669Gfs, p.C882Wfs, p.Q988X, p.P1009Lfs, p.W1038X and p.Y1183X), as well as several variants that were classified as benign by ClinVar (p.Q559R, p.E672Q, p.V858=, p.P864S, and p.G998E) (Table 2). Our functional analysis relied on the stable integration of *PALB2* cDNA at a safe-harbor locus in *Palb2* knockout (KO) mouse embryonic stem (mES) cells and its subsequent expression from a relatively weak promotor (39). Such a strategy avoids differences in *PALB2* expression following siRNA-mediated knockdown and reduces possible artefacts that may arise from transient overexpression of *PALB2* cDNA. While benign variants had only a moderate or no impact on HR (<20% reduction in HR when compared to WT *PALB2*), all truncating variants strongly impacted HR (>89% reduction in HR when compare to WT *PALB2*). Moreover, 48 *PALB2* VUS were analyzed, of which the expression of 15 VUS (i.e. p.L24S, p.Y28C, p.L35P, p.R37H, p.W912G, p.G937R, p.I944N, p.L947S, p.L961P, p.L972Q, p.T1030I, p.I1037T, p.G1043D, p.L1070P and p.L1172P), strongly abrogated PALB2 protein function, with HR being decreased by 55-93% in DR-GFP assays. The same variants also resulted in cellular sensitivity to PARPi (Table 2) (39).

Table 2. Complete list of human *PALB2* variants analyzed in five independent studies.

Study	Park et al. 2014	Foo et al. 2017	Boonen et al. 2019						Wiltshire et al. 2019			Rodrigue et al. 2019		
Nr. of variants	4	5	70						91			47		
Nr. of benign controls (ClinVar) + functional range from main assay	0	0	5 (80-95% in DR-GFP)						6 (88-128% in DR-GFP)			3 (102-104% in PARPi)		
Nr. of damaging controls + functional range from main assay	0	0	12 (6-11% in DR-GFP)						7 (12-48% in DR-GFP)			1 (48% for p.L35P in PARPi)		
Cell type	U2OS	B400	mES				HeLa	U2OS	B400	HeLa	U2OS	HeLa		U2OS
Protein change	**DR-GFP (%)**	**DR-GFP (%)**	**DR-GFP (%)**	**PARPi (%)**	**Cispl. (%)**	**G2>M (%)**	**RAD51 (%)**	**LMNA (%)**	**DR-GFP (%)**	**RAD51 (%)**	**LMNA (%)**	**PARPi (%)**	**RAD51 (%)**	**LMNA (%)**
p.P4L												87	NT	NT
p.P4S			98	73	NT	NT	NT	NT						
p.P5S			62	96	NT	NT	NT	NT						
p.P8L									142	NT	NT	76	68	~ 60
p.L9H									116	NT	NT			
p.K18R		~ 75	100	94	84	87	NT	NT				82	103	~ 87
p.L21A												NT	NT	NT
p.L24S*			21	55	NT	NT	NT	NT	34	~ 11	< 5			
p.Y28C*		~ 35	33	22	NT	NT	NT	NT	96	NT	NT	92	64	~ 37
p.K30N		~ 105							92	NT	NT	97	NT	NT
p.T31I			97	102	NT	NT	NT	NT						
p.L32V												87	NT	NT
p.R34C									98	NT	NT			
p.R34H									100	NT	NT			
p.L35P*		~ 5	10	10	26	261	NT	NT	16	~ 4	< 9	48	2	~ 5
p.R37C									94	NT	NT			
p.R37H*		~ 78	45	68	83	176	NT	NT	82	NT	NT	86	71	~ 38
p.R37S									90	NT	NT			
p.E42K			105	95	90	102	NT	NT						
p.H46Y												76	108	~ 82
p.Q60Rfs			9	11	NT	NT	NT	NT						
p.P65L												86	NT	NT
p.I76V												93	NT	NT
p.V78I									132	NT	NT	92	NT	NT
p.K81R												93	NT	NT
p.E94K												94	NT	NT
p.G115V												107	NT	NT
p.T124I												109	NT	NT
p.S133T									88	NT	NT			
p.D134N			91	93	98	55	NT	NT						
p.L169I									94	NT	NT	103	80	~ 77
p.S172Gfs			9	14	NT	NT	NT	NT						
p.P207R												101	NT	NT
p.P210L			85	103	NT	NT	NT	NT	130	NT	NT	108	NT	NT
p.E230X			7	18	NT	NT	NT	NT						
p.Q251X									16	NT	NT			
p.T300I									72	NT	NT			
p.I309V									116	NT	NT			
p.T317P												109	NT	NT
p.S319Y									94	NT	NT	110	79	~ 83
p.L337S			87	116	NT	NT	NT	NT	106	NT	NT			
p.S382I									146	NT	NT			
p.T397I									98	NT	NT			
p.T397S									104	NT	NT			
p.Y408H			92	109	93	80	NT	NT	98	NT	NT			
p.Y409C									80	NT	NT			

Table 2. Continued

Study	Park et al. 2014	Foo et al. 2017	Boonen et al. 2019						Wiltshire et al. 2019			Rodrigue et al. 2019		
Protein change	DR-GFP (%)	DR-GFP (%)	DR-GFP (%)	PARPi (%)	Cispl. (%)	G2>M (%)	RAD51 (%)	LMNA (%)	DR-GFP (%)	RAD51 (%)	LMNA (%)	PARPi (%)	RAD51 (%)	LMNA (%)
p.Y409X			8	17	NT	NT	NT	NT						
p.T413K									100	NT	NT			
p.S417Y			72	84	NT	NT	NT	NT	144	NT	NT			
p.S474N									102	NT	NT			
p.D498Y			94	74	NT	NT	NT	NT						
p.K515R			75	121	NT	NT	NT	NT						
p.L531Cfs			8	24	NT	NT	NT	NT						
p.S534A									94	NT	NT			
p.Y551X	NT		8	11	20	245	NT	NT	20	NT	NT			
p.Q559R			95	118	NT	NT	NT	NT	100	NT	NT			
p.S578G									104	NT	NT			
p.D616G									88	NT	NT			
p.D616H									108	NT	NT			
p.L622P			77	65	NT	NT	NT	NT						
p.E669Gfs			7	19	NT	NT	NT	NT						
p.E672Q			80	104	NT	NT	NT	NT	88	NT	NT			
p.T706I			87	78	NT	NT	NT	NT						
p.P707L			83	88	101	94	NT	NT						
p.D715Efs									12	NT	NT			
p.N716K									106	NT	NT			
p.T734S									138	NT	NT			
p.K745E									98	NT	NT			
p.L763F									92	NT	NT			
p.V858=			84	74	NT	NT	NT	NT						
p.P864S			86	87	NT	NT	NT	NT	128	NT	NT	102	90	~ 90
p.S865P			100	78	NT	NT	NT	NT						
p.G866V									112	NT	NT			
p.D871G			84	115	87	117	NT	NT	90	NT	NT			
p.C882Wfs			6	29	NT	NT	NT	NT						
p.W912G*			7	8	NT	NT	5	8						
p.V919I									96	NT	NT			
p.D927A			76	86	NT	NT	NT	NT						
p.L931P									100	NT	NT			
p.L931R			106	95	112	80	NT	NT	72	NT	NT			
p.V932M									98	NT	NT	102	84	~ 95
p.C933Y									76	NT	NT			
p.L936S									100	NT	NT			
p.G937E									88	NT	NT			
p.G937R*			17	26	NT	NT	63	25						
p.L939W	~ 85		60	91	102	218	NT	NT	96	NT	NT	97	NT	NT
p.E940G			63	81	NT	NT	NT	NT						
p.I944N*			7	15	NT	NT	NT	NT	30	~ 2	< 9			
p.L947F									74	NT	NT	77	75	~ 55
p.L947S*			30	24	NT	NT	78	47	80	NT	NT	82	69	~ 41
p.S951F									92	NT	NT			
p.S955R												103	NT	NT
p.Q958P									124	NT	NT			
p.L961P*			7	8	27	280	13	2						
p.I966T			74	79	NT	NT	NT	NT						
p.I966V												101	NT	NT
p.L972Q*			14	13	NT	NT	NT	NT						
p.Q988X			7	15	NT	NT	NT	NT						
p.G998E			95	97	80	129	NT	NT	120	NT	NT	104	94	~ 82
p.P1009Lfs			6	25	NT	NT	NT	NT						
p.A1017T												98	NT	NT

Table 2. Continued

Study	Park et al. 2014	Foo et al. 2017	Boonen et al. 2019						Wiltshire et al. 2019			Rodrigue et al. 2019		
Protein change	DR-GFP (%)	DR-GFP (%)	DR-GFP (%)	PARPi (%)	Cispl. (%)	G2>M (%)	RAD51 (%)	LMNA (%)	DR-GFP (%)	RAD51 (%)	LMNA (%)	PARPi (%)	RAD51 (%)	LMNA (%)
p.E1018D			86	84	112	73	NT	NT				95	NT	NT
p.V1019A									128	NT	NT			
p.G1021R									80	NT	NT			
p.A1025R			18	24	NT	369	NT	NT				NT	NT	NT
p.A1025T									78	NT	NT	109	NT	NT
p.T1030I*	NT		15	15	NT	NT	NT	NT	60	NT	NT	58	44	~ 24
p.I1037T*			39	52	NT	NT	NT	NT						
p.W1038X			7	12	NT	NT	NT	NT						
p.L1040S			77	100	NT	NT	NT	NT						
p.K1041T												98	NT	NT
p.G1043A									98	NT	NT	87	86	~ 47
p.G1043D*			11	10	37	276	56	22						
p.Q1044H									94	NT	NT			
p.I1051S			91	109										
p.S1058P			96	85	119	109	NT	NT						
p.C1060S									104	NT	NT			
p.Y1064C			101	87	129	100	NT	NT	88	NT	NT			
p.L1070P*			23	57	NT	NT	NT	NT	34	~ 4	< 5			
p.S1075G												99	NT	NT
p.E1083D									102	NT	NT			
p.S1084L												94	NT	NT
p.P1088S									94	NT	NT			
p.I1093T									80	NT	NT	98	96	~ 92
p.T1099R									102	NT	NT			
p.S1102R									154	NT	NT	97	NT	NT
p.V1103L									100	NT	NT			
p.V1103M									154	NT	NT			
p.V1105A												89	NT	NT
p.L1107P									92	NT	NT			
p.Y1108Sfs									12	NT	NT			
p.P1111A			103	88	NT	NT	NT	NT						
p.Q1114H									98	NT	NT	92	NT	NT
p.L1119P			94	104	NT	NT	NT	NT	98	NT	NT	84	89	~ 68
p.G1121Vfs									12	NT	NT			
p.V1123M			76	103	NT	NT	NT	NT						
p.G1135E									98	NT	NT			
p.W1140G*									98	NT	NT	64	66	~ 35
p.L1143H			70	98	NT	NT	NT	NT						
p.L1143P	~ 80								132	NT	NT	109	NT	NT
p.G1145R									104	NT	NT			
p.G1147R									96	NT	NT			
p.L1150R									150	NT	NT			
p.W1159L			91	108	NT	NT	NT	NT						
p.S1160P			92	103	NT	NT	NT	NT						
p.W1164C			81	96	NT	NT	NT	NT	100	NT	NT			
p.S1165L									98	NT	NT			
p.G1166Vfs									12	NT	NT			
p.T1167I									98	NT	NT			
p.H1170Y									94	NT	NT			
p.L1172P*			13	17	NT	NT	NT	NT						
p.G1174R			91	103	NT	NT	NT	NT						
p.I1180T			82	99	NT	NT	NT	NT	72	NT	NT	106	NT	NT
p.Y1183C			71	109	NT	NT	NT	NT						
p.Y1183X			11	10	20	345	NT	NT	48	NT	NT			

All variants are indicated at the protein level (i.e., protein change). Nucleotide annotations for each variant are available in the published manuscript, where nucleotide numbering reflects Human Genome Variation Society (HGVS) nomenclature and cDNA number +1 corresponds to the A of the ATG translation initiation codon in the reference sequence (*PALB2* NM_024675.3). The initiation codon is codon 1. Results from DR-GFP, PARPi sensitivity, cisplatin sensitivity, RAD51 foci and G2/M checkpoint assays are shown. Only data taken from bargraphs and experiments in the context of full-length PALB2 protein was used for this table. Truncating, benign (ClinVar), synthetic missense variants and strongly damaging VUS (with >50% reduced activity), are indicated in red, green, orange and with a red * in the 'protein change' column, respectively. NT stands for 'not tested'.

Effect of VUS in PALB2's CC domain on the BRCA1-interaction and HR

Formation of the PALB2-BRCA1/2-RAD51 complex is crucial for delivering RAD51 monomers to RPA-coated ssDNA overhangs and promoting strand invasion during HR (10-12,16). Variants that affect PALB2's interaction capability with BRCA1 or BRCA2 are therefore predicted to impact HR. Here we first discuss the implication of variants in PALB2's CC domain (amino acids 9 to 44) (Fig. 1). Initially it was shown by two independent studies that exchange of PALB2's CC domain residues p.L21, p.Y28 or p.L35 by an alanine (11), or p.L21 or p.L24 by a proline (12), indeed impaired HR by abolishing the interaction between PALB2 and BRCA1. Consistently, the patient-derived p.L35P missense variant in *PALB2* was more recently shown to impair the interaction with BRCA1, thereby strongly reducing HR (40). This variant was taken along by the three recent studies which all confirmed these findings (39,42,43). A similar defect in HR was observed for p.L24S, which was also attributable to an impairment in the interaction with BRCA1 (39,43). Interestingly, cycloheximide chase experiments to monitor protein stability suggested that variants that fail to interact with BRCA1 (p.L24S and p.L35P) enhanced the stability and consequently the levels of PALB2 protein (43). Consistent with this result, we and others also detected slightly higher protein levels for variants that failed to interact with BRCA1 (i.e. p.L24S, p.Y28C and p.L35P) (39,40,43). As the CC domain regulates PALB2 self-interaction in addition to the interaction with BRCA1, it is possible that an inability of PALB2 to interact with BRCA1 creates a shift towards the formation of PALB2 oligomers (53,54). Such complexes may shield PALB2 from ubiquitination-dependent degradation (55), leading to higher protein levels.

The consistency between the different studies (39,42,43), was challenged by the analysis of p.R37H, which has previously been shown to represent a variant whose expression only moderately impacts protein function (40). p.R37H was shown to reduce HR only by ~20% (40,43). Accordingly, the analysis by Foo et al. showed that p.R37H did not affect the interaction with BRCA1 (40). In contrast, Rodrigue et al. identified p.R37H as a variant whose expression led to a significant reduction in PALB2 function, both in PARPi sensitivity assays as well as the CRISPR-LMNA HR assay, with 60% reduced activity in the latter assay (42).

However, the mechanism for the reduced functionality was unclear as mammalian two-hybrid assays and laser micro-irradiation experiments suggested that this variant interacted normally with BRCA1 and was recruited to DNA damage sites, respectively. Although we reported a similar impact on HR in DR-GFP assays for this variant (55% reduction in HR when compared to WT *PALB2*) (39), we observed a partial loss of the PALB2-BRCA1 interaction in immunoprecipitation experiments, as well as the recruitment of PALB2 to sites of DNA damage induced by laser micro-irradiation (39). Thus, while all four studies consistently show the impact of p.R37H on HR, the discrepancy in the mechanistic explanation warrants further investigation of this particular variant.

Effect of VUS in PALB2's WD40 domain on protein stability and HR

In addition to the CC domain, which mediates the interaction with BRCA1, the WD40 domain of PALB2 (amino acids 853 to 1186) (Fig. 1), mediates interactions with other core HR proteins such as BRCA2 and RAD51. In our study, many damaging variants were identified in this functional domain (p.W912G, p.G337R, p.I944N, p.L947S, p.L961P, p.L972Q, p.T1030I, p.I1037T, p.G1043D, p.L1070P and p.L1172P) (39). Since all these variants exhibited strongly reduced protein expression levels, the effect on the interaction of PALB2 with other HR factors was not examined. Importantly, reverse transcription-quantitative (RT-q)PCR analysis indicated that these variants did not affect expression at the mRNA level (39), suggesting that the low abundance of PALB2 protein is likely the result of protein instability. In contrast, Rodrigue et al. performed PALB2-BRCA2 immunoprecipitation assays for damaging variants in the WD40 domain (p.L947F, p.L947S, p.T1030I, p.G1043A, p.L1119P and p.W1140G), although they similarly detected lower expression levels for these variants (42). Not surprisingly, all six variants appeared to impair the interaction with BRCA2. As these variants are scattered throughout the WD40 domain, it seems likely that they represent unstable variants rather than variants that impair specific binding sites for BRCA2. Likewise, Wiltshire and colleagues showed that the p.I944N and p.L1070P variants both decreased the interaction with BRCA2, as well as with BRCA1 (43). As the interaction motif for BRCA1 lies in PALB2's N-terminal CC domain, and not the WD40 domain in which these variants are present, these reduced interactions are more likely the result of reduced PALB2 protein stability. Although we identified several damaging variants in the WD40 domain, only the synthetic missense variant p.A1025R displayed normal expression levels, while having a major impact on HR (82% reduction in HR when compared to WT *PALB2*) (39). These results are in line with the fact that this variant impairs the PALB2-BRCA2 interaction, as shown previously by several studies (42,43,52).

In addition to the observed protein instability, it has been suggested that mis-localization of the PALB2 protein in the cytoplasm may provide an explanation for the reduced

PALB2 functionality observed for a number of variants in the WD40 domain (39,41,43). For instance, for p.I944N, Wiltshire and colleagues showed that this variant prevented nuclear localization of PALB2 and that it is retained in the cytoplasm. They observed a similar mis-localization for p.L1070P, albeit to a lesser extent. Rodrigue and colleagues additionally identified p.L947F, p.L947S, p.T1030I, p.G1043A, p.L1119P and p.W1140G as variants causing mis-localisation of PALB2. All these variants impaired PALB2 recruitment to laser-induced DSBs, an effect that was also observed for p.Y28C and p.L35P (42). However, p.Y28C and p.L35P, which both reside in the CC domain, did not negatively impact PALB2's nuclear localisation. Thus, variants in the WD40 domain that result in PALB2 instability may be signalled for degradation in the cytoplasm, providing an explanation for how such variants could impact PALB2-dependent HR.

Limitations of current assays used for the functional analysis of VUS in *PALB2*

A reasonable number of overlapping VUS was analyzed by three recent studies (39,42,43). This allows for a head-to-head comparison of the outcome of the different functional analysis, as well as the important aspects of the different experimental approaches, such as the model cell line, complementation by transient overexpression or stable expression, and the use of KO or knockdown cell lines. These differences may explain certain discrepancies, which we discuss below on the basis of several variants that were functionally characterized.

Overexpression of the *PALB2* cDNA may underestimate the functional effect of some variants. For instance, the FA-associated p.Y1183X *PALB2* variant, is located three amino acids from the end of the protein and can lead to the expression of a near full-length PALB2 protein. Stable expression of this variant impaired the HR efficiency in mES cells to a similar extent as all other truncating variants positioned throughout the gene (i.e., HR being reduced by 89-94%) (39). However, it is feasible that cDNA-based overexpression of this variant can partially rescue HR. This may have occurred in the study by Wiltshire and colleagues in which p.Y1183X reduced the HR efficiency in *Palb2*-deficient B400 mouse mammary cells by 52%, in comparison to a ~84% reduction observed for other truncating *PALB2* variants scattered throughout the gene (43). Accordingly, Rodrigue and colleagues noted that there are indeed differences in expression between variants after transient overexpression. Moreover, they showed that exogenous *PALB2* is greatly overexpressed in comparison to endogenous *PALB2* (42). Thus, we may need to take caution when variants are functionally characterized by transient overexpression, as damaging variants may still exhibit residual activity under these conditions. In fact, when we compare other overlapping variants among the three recent studies tested in DR-GFP assays (n = 26) (39,43) and PARPi sensitivity assays (n = 14) (39,42), functional defects are almost always smaller when assessed by transient overexpression compared to stable integration and expression (Fig. 3a-b). This is particularly

striking in the case of variants such as p.Y28C, p.R37H, p.L947S and p.T1030I, which may still exhibit residual activity. Consequently, this effect may lead to an underestimation of the HR defects that these variants can cause and may explain the fairly low correlation (R^2 = *~0.58*) between results from assays with transient overexpression versus stable integration and expression (Fig. 3c-d). However, this hypothesis is contradicted by the very good correlation (R^2 = *~0.91*) (Fig. 3e-f) between the effects of overlapping variants in DR-GFP and CRISPR-LMNA HR assays (n = 9) (39,42), which relied on stable and transient expression of *PALB2*, respectively. Although this result can be explained by a slightly more effective siRNA-based knockdown of endogenous *PALB2* in the U2OS cells used for the CRISPR-LMNA HR assays, it is also possible that stable versus transient expression, in a specific cellular background, impacts the outcome of the functional assays. Further research is therefore be needed to resolve these issues.

Similar to transient overexpression, *PALB2* complementation after knockdown of the gene (versus the use of KO cells), could in theory also result in an underestimation of the effects of some variants. This is because the knockdown is often incomplete, resulting in residual expression of wildtype *PALB2* in the presence of exogenously expressed *PALB2* carrying a variant. If the *PALB2* variant affects PALB2 protein function, this effect may be obscured by the presence of wildtype PALB2 protein. Also, the knockdown efficiency can differ between experiments, resulting in variability in the measured functional effects. On the other hand, with regard to the KO of genes in general, it is possible that cells can undergo adaptions in order to survive. It is possible that such adaptions can influence the functional readout that is used.

As all three recent studies employed a cDNA-based complementation approach (39,42,43), another disadvantage, specifically when analyzing truncating variants, is the absence of nonsense-mediated mRNA decay. Hypothetically, the expression of a partially functional truncated protein might mask the severe impact on protein function of such variants observed in the presence of nonsense-mediated mRNA decay, which would otherwise abrogate protein expression. A complementation method based on the use of a bacterial artificial chromosome (BAC) that contains the complete gene-of-interest would allow for inclusion of effects originating from nonsense-mediated mRNA decay. This is important, as such processes by themselves may enhance the risk for cancer and constitute an alternative mechanism for reduced protein function.

With regard to the differences in outcome between the three recent studies on *PALB2* VUS (39,42,43), one may also question whether these may originate from the use of human and mouse model cell lines. For instance, we showed that complementation of *Palb2*^{KO} mES cells with human *PALB2* cDNA resulted in a partial rescue of the HR defect (i.e., ~68% HR compared to *Trp53*^{KO} still expressing mouse *Palb2*) (39). Although it cannot be excluded that

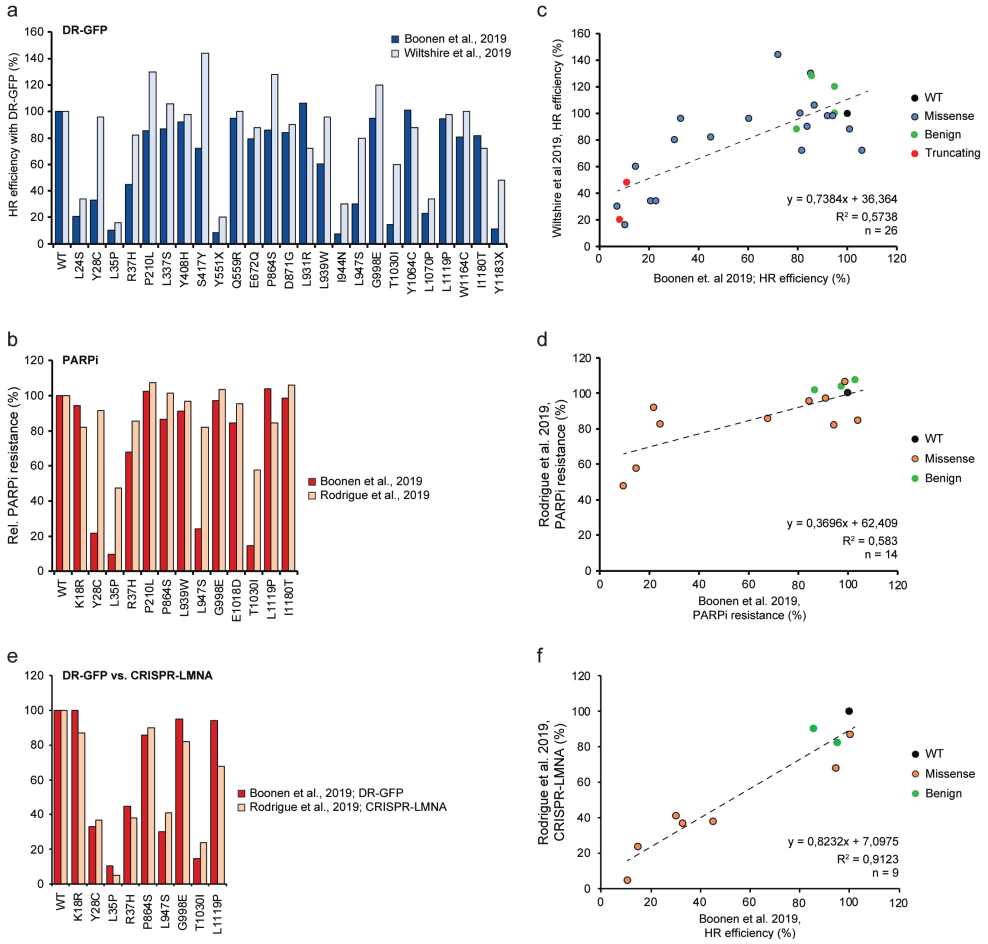


Figure 3. Comparison and correlation between DR-GFP- and PARPi-based HR assays from three different studies. a Bar graph comparing results from DR-GFP-based functional assays for 26 overlapping *PALB2* variants from studies by us and Wiltshire et al. (39,43). Mean percentages of GFP-positive cells relative to wild type *PALB2* (WT) are shown, with cells expressing WT *PALB2* being set to 100%. **b** Bar graph comparing results from PARPi-based functional assays for 14 overlapping *PALB2* variants from studies by us and Rodrigue et al. (39,42). Mean percentages of viability relative to WT *PALB2* are shown, with cells expressing WT *PALB2* being set to 100%. **c** Scatter plot showing the correlation between the results from our study and Wiltshire et al. as shown in 'a' (39,43). The color of the datapoints corresponds to the different variants/conditions: wild type (black), benign based on ClinVar (green), truncating (red), VUS (blue). **d** Scatter plot showing the correlation between the results from our study and Rodrigue et al. as shown in 'b' (39,42). The color of the datapoints corresponds to the different variants/conditions: wild type (black), benign based on ClinVar (green), VUS (orange). **e** Bar graph comparing results from DR-GFP- and CRISPR-LMNA-based HR assays for 9 overlapping *PALB2* variants from studies by us and Rodrigue et al. (39,42). Mean percentages of GFP- or mRuby-positive cells relative to WT *PALB2* are shown as in 'a'. **f** Scatter plot showing the correlation between the results from our study and Rodrigue et al. as shown in 'e' (39,42). The color of the datapoints is as shown in 'd'.

this is due to different expression levels of ectopic human *PALB2* compared to endogenous mouse *Palb2*, it is also possible that this is due to the limited homology between mouse and human *PALB2* (~59% identical and 70% similar in protein sequence). Consequently, the functional effect of some variants may be missed and this could affect the reliability of testing human variants in a mouse cell background. Nonetheless, it should be noted that so far damaging missense variants in *PALB2* have only been observed in the well conserved CC and WD40 domains, which both exhibit ~82.5% identical and ~91.5% similar protein sequence. This makes it unlikely that *PALB2* variants that have been identified as damaging in these domains in mouse cell-based assays, are not so in a human cell-based setup. Indeed, we have observed similar effects on HR for a number of VUS (p.W912G, p.G937R, p.L947S, p.L961P and G1043D) in human and mouse cell-based assays (39) .

Functional characterization of VUS in *PALB2* using checkpoint control as a read-out

Besides a critical role in promoting HR, several studies have implicated *BRCA1*, *BRCA2* and *PALB2* in DNA-damage-induced checkpoint control (56-58). Consistently, it was shown that G2/M checkpoint maintenance after IR is compromised in $Trp53^{KO}/Palb2^{KO}$ mES cells, an effect that could be rescued by expressing WT human *PALB2* (Fig. 2) (39). Interestingly, *PALB2* variants that show LOF in HR, were unable to maintain an efficient G2/M checkpoint response (p.L35P, p.L961P, p.A1025R and p.G1043D). The fact that p.L35P and p.A1025R, which are unable to interact with BRCA1 and BRCA2, respectively, were among these variants, suggests that both interactions are key to PALB2's role in regulating G2/M checkpoint control. Although checkpoint regulation could be a distinct function of *PALB2*, another possibility is that the observed defects in G2/M checkpoint maintenance could stem from defective HR. Given that a defect in HR likely leads to elevated levels of unrepaired DNA breaks, it may seem counterintuitive that G2/M checkpoint maintenance is reduced under these conditions, unless compensatory pathways take over to complete DNA repair and allow for continued progression through the cell cycle. In line with such a scenario, an inverse correlation has been observed between HR activity and a-NHEJ mediated by POLQ (59). This indicates that a-NHEJ may act as a compensatory pathway for PALB2-dependent HR. Indeed, in HR-deficient ovarian cancer cell lines POLQ was selectively upregulated, whereas restoration of HR brought back POLQ expression to normal levels (59). Based on these findings we speculate that when HR is compromised due to PALB2 LOF, activation of a-NHEJ potentially affects G2/M checkpoint maintenance in response to DNA breaks.

Control of ROS and replication stress as potential readouts for the functional analysis of VUS in *PALB2*

PALB2 has also been reported to play a role in controlling the reactive oxygen species (ROS) levels in human cells (60), which may constitute another tumor suppressive function. PALB2 suppresses ROS levels in a manner dependent on its interaction with the ubiquitin ligase KEAP1. KEAP1 functions as a cysteine-rich oxidative stress sensor, which under normal conditions, binds to and targets the antioxidant transcription factor NRF2 for degradation (60). As PALB2 bears a highly conserved ETGE-type KEAP1-binding motif (amino acids 88 to 94), that is identical to that of NRF2, PALB2 can competitively impede the inhibitory KEAP1-NRF2 interaction. Therefore, PALB2 is believed to promote NRF2 accumulation, enhance antioxidant gene expression and reduce the burden of oxidative stress. However, the truncating p.Y551X *PALB2* variant, which has been described to be associated with FA and breast cancer (61), still interacts with NRF2 as corroborated by Ma and colleagues (60), and consequently should be functional in the regulation of ROS levels. Furthermore, this truncated variant has been shown to be expressed in lymphoblasts of an individual with FA and is apparently not subjected to nonsense RNA-mediated decay (61). We therefore infer that the effect of impaired regulation of ROS levels by PALB2 may have no, or only a minor contribution to the development of FA and breast cancer, questioning the value of a more extensive analysis of the effect of VUS in *PALB2* on this process.

Besides playing a key role in HR, *BRCA1, BRCA2* and *RAD51* have also been implicated in replication fork protection and/or the recovery of stalled replication forks, which are processes that are critical for genome stability maintenance and cancer prevention, as well as cancer therapy responses (62-65). Intriguingly, it was recently shown that the interaction between BRCA1 and BARD1 promotes the protection of replication forks and that genetic variants in *BRCA1* that impair this interaction associate with cancer, even though they retain their function in HR (66). A mutational analysis of *BRCA2* revealed that a conserved C-terminal site involved in stabilizing RAD51 filaments, but not in loading RAD51 onto DNA, is essential for replication fork protection, but dispensable for HR. Consistently, the p.S3291A variant in this C-terminal region was shown to impair the protection of stalled replication forks, while leaving HR intact (65). RAD51, on the other hand, acts during DNA replication to facilitate fork reversal and protects nascent DNA strands from nuclease digestion, thereby promoting the recovery of stalled replication forks (67-69). It is plausible that PALB2 exhibits functions at the replication fork that are comparable to those of BRCA1, BRCA2 and/or RAD51. Indeed, it was previously shown that PALB2 mediates replication fork recovery after replication stress in human U2OS cells (70). Corroborating these findings, replication abnormalities, including a decreased/delayed origin firing and replication fork restart, have also been observed in blood lymphocytes heterozygous for the truncating p.L531Cfs *PALB2* variant (71). How PALB2

mechanistically facilitates these processes is still largely unclear and requires additional research. However, it is feasible that VUS in *PALB2* could have the potential to specifically impair such functions, as has been reported for *BRCA1* and *BRCA2* (65,66). Potentially, loss of PALB2's function in replication fork protection and/or recovery may associate with cancer. If so, replication fork maintenance may become another important readout for functional analysis of *PALB2* VUS.

In silico approaches predicting the functional impact of VUS are mostly unreliable

Especially with the vast accumulation of identified VUS (72,73), a variety of in silico tools, which are both publicly and commercially available, can aid in the interpretation of VUS in clinical diagnostic settings (74). However, the currently available in silico tools, such as PolyPhen-2, SIFT, MutationTaster-2, MutationAssessor, CADD and REVEL, often give rise to conflicting results and over- or underestimate the functional impact of a given variant (75,76). A systematic performance comparison between in silico prediction tools and functional assays, showed that functional assays substantially outperform every computational method examined, mostly with respect to heightened specificity (77). In this study, a panel of 26 different yeast-based complementation assays were used to measure the impact of 179 variants on 22 human disease genes. Remarkably, of the 64 non-disease-associated variants tested, 36% was predicted to be deleterious by PolyPhen, as opposed to only 13% being classified as deleterious by these functional assays (77). This high rate of false predictions is in agreement with recent data from us and Rodrigue et al., showing that in silico prediction tools all strongly overpredicted the percentage of deleterious variants in *PALB2* (39,42), Consistently, these studies observed a poor correlation between results from DR-GFP assays and predictions by CADD (R^2 = *0.08*) or REVEL (R^2 = *0.11*) (39), and between results from PARPi sensitivity assays and M-CAP (R^2 = *0.33*), VEST (R^2 = *0.07*) or REVEL (R^2 = *0.27*) (42). Due to this lack of consistency and poor performance, computational predictions are not considered strong evidence for or against pathogenicity (74). Instead, functional assays seem to represent the best strategy for overcoming the VUS challenge, as they currently constitute the strongest evidence for the functional impact of rare variants. Moreover, for genes such as *BRCA1* and *BRCA2*, for which functional assays are more established, a functional read-out such as HR can be used to improve existing computational prediction tools. In a recent study by Hart et al., the measured HR efficiency for 248 *BRCA1* and 207 *BRCA2* variants was used to recalibrate 40 in silico algorithms (78). Optimized thresholds based on such functional data significantly improved the accuracy of many of these algorithms. However, optimised algorithms for one gene may perform poorly when applied to another gene. This is perhaps not surprising as each functional domain may harbour different sensitivities to the effects of

damaging variants, explaining why different gene-specific features are important for the accuracy of in silico predictions.

Perspective on high-throughput functional analysis of *PALB2* variants

The identification of VUS has increased drastically due to the global build-up in genetic testing (79), leading to major challenges in the clinical management of carriers. To emphasize the vast number of genetic variants that are identified, 4.6 million missense variants have recently been reported in ~140000 exomes and genomes in the Genome Aggregation Database (gnomAD) (72,73) and 99% of these variants are rare with a minor allele frequency of <0.005 (80). Variant interpretation at such a scale, can currently only be addressed with computational prediction tools. However, as mentioned above, the existing tools often provide conflicting results, where functional impact is mostly overpredicted (39,77). Thus, the accelerated rate of VUS discovery makes a one-at-a-time, or even semi high-throughput, approach for functional analysis infeasible. Furthermore, as these strategies are often time-consuming, the individual in which a variant was found may not be able to take advantage of it in time.

An ambitious goal for the future is that the effect of every possible nucleotide substitution, perhaps initially only in clinically actionable genes (81,82), is functionally measured using high-throughput assays. For instance, for *PALB2* specifically, as of June 2020, 1612 distinct VUS have been reported in ClinVar. This number already makes a one-at-a-time functional analysis approach extremely challenging. High-throughput assays (i.e. multiplexed assays), aimed to address every nucleotide change in an entire gene in single experiments may provide a solution. Indeed, a saturation CRISPR/Cas9-based editing approach in haploid human HAP1 cells allowed for the assessment of more than 95% of all possible single-nucleotide variants (SNVs) in 13 exons of *BRCA1* that encode for its RING and BRCT domains (83). Importantly, this setup allowed for the functional analysis of variants in their endogenous genomic context and using cell survival as a read-out, the effect of nearly 4000 single-nucleotide variants corroborated established assessments on protein function. Furthermore, a multiplex homology-directed repair assay, which relied on stable integration of a *BRCA1* cDNA variant library, enabled the functional characterization of 1056 missense variants in the first 192 residues of *BRCA1* (38). We expect that such assays will be extended to analyzing variants in genes such as *PALB2* in the near future, ultimately leading to the development of a variant map that shows the impact of all possible *PALB2* variants on HR.

In addition to examining cell survival and HR for *PALB2* in a high-throughput setup, another more general readout might be to measure the steady-state protein abundance. Recent results from functional assays have shown that variants in PALB2's WD40 domain tend to destabilize PALB2 (39), a mechanism of protein inactivation that is in agreement with studies showing that ~75% of pathogenic variation is thought to disrupt thermodynamic stability and

alter protein levels (84-86). Therefore, high-throughput assessment of PALB2 variant protein abundance, by employing techniques such as VAMP-seq (84) or Stable-seq (87), may also prove to be highly suitable for detecting *PALB2* variants that affect protein function. Nonetheless, although such high-throughput assays provide much potential for interpreting the large number of VUS that are being identified, it should also be noted that developing variant libraries, optimizing experimental setups, and analyzing the large amount of sequencing data, can still be prohibitively time and resource intensive.

Towards the functional analysis of *PALB2* VUS in RNA splicing

It is important to note that all functional studies on VUS in *PALB2* discussed in this review (39-43), were based on expression of *PALB2* cDNAs and are therefore not suitable to assess the functional impact of *PALB2* variants that affect RNA splicing. In silico splice site prediction tools can predict the effect of variants on potential splice sites relatively well (74), but they do not provide conclusive evidence for altered splicing. One option to assess the effect of variants on splicing, is to use a minigene construct that contains a genomic segment encompassing the variant along with flanking intronic sequences (88). After transient transfection of the construct into human cells, the transcripts from the minigene can easily be analyzed and compared to transcripts derived from a wild-type construct. Although these assays can be carried out in many cell types and are fairly simple and fast, disadvantages are that variants are not measured in the context of a complete gene and that these assays do not permit downstream functional analysis. This is of course important since some splice variants can result in the expression of a transcript that may be (partially) functional. For instance, several exons in *PALB2* (exons 1, 2, 4, 6, 7, 9, 10 and 11-12 combined) can be skipped due to splice site variants and still result in an in-frame transcript (89). Such transcripts may still express an isoform of *PALB2* with an entire exon deleted, yet retain partial protein function. An example is the c.2586+1G>A (r.2515_2586del; p.T839_K862del) *PALB2* variant, which leads to an in-frame skip of exon 6. This variant appears to be a hypomorphic variant that still interacts with BRCA2 and, when overexpressed, still enables RAD51 foci formation (90). Additional research will be required to establish the functionality of other exon-skip variants in *PALB2.*

As of June 2020, 70 unique *PALB2* splice variants have been reported in ClinVar (involving canonical splice sites), the majority of which is classified as pathogenic or likely pathogenic. Generally, mRNA transcript and protein expression analysis combined with functional assays, may be needed to provide insight into the effect of variants in *PALB2* that are predicted to impact RNA splicing. Possibly, one could complement *PALB2* KO cells containing DR-GFP with a BAC containing the full length human *PALB2* gene. Such a method has previously been described for *BRCA1* and *BRCA2* (91-94) and would allow for the

introduction and functional analysis of splice variants in coding and non-coding regions, further improving their classification.

Towards estimating cancer risk associated with VUS in *PALB2*

Functional assays may aid in the classification of rare *PALB2* VUS, yet a major challenge will be to translate effects on PALB2 protein function into estimates for cancer risk. Recent studies on *BRCA2* have shown that pathogenic variants that confer high risk for breast and ovarian cancer completely abrogate BRCA2-mediated HR, whereas variants that result in a reduction of 50% in HR, i.e., hypomorphic variants, may only be associated with a moderate risk for breast cancer (Odds ratio ~2.5) (36,37). With regard to *PALB2*, truncating variants have been associated with an odds ratio of 7.46 (95% CI, 5.12-11.19) (28), whereas the frequently occurring p.L939W missense variant has been associated with an odds ratio of 1.05 (95% CI, 0.83 to 1.32) (95), which is in agreement with recent data from Wiltshire et al. and Rodrigue et al., showing that this variant does not impact the HR efficiency (~4% reduction in HR when compared to WT) (42,43). In contrast, results from us and Park et al., showed that this variant did impair HR to some degree (40% and 15% reduction in HR when compared to WT, respectively) (39,41). This may suggest that such a decrease in HR, may not considerably increase the risk for breast cancer. Future functional characterization of additional *PALB2* VUS, in combination with data from large case-control association studies, should allow for more conclusive correlations of odds ratios with HR efficiencies for *PALB2*, either for specific variants that occur frequently, or for variants as a group (i.e., damaging variants). Under the assumption that variants with similar levels of HR functionality confer the same level of cancer risk, so called burden-type of association analyses can be performed in large case-control studies, in which either genetic or clinical information of multiple variants, or joint frequencies of individual variants with similar HR levels will be pooled. Nonetheless, the fact that roles other than in HR (i.e., in replication fork stability/recovery) for all three major breast cancer susceptibility genes (*BRCA1*, *BRCA2* and *PALB2*) have been described (64-66,70), complicates the interpretation of VUS in these genes and their association with cancer risk. It should be noted, however, that only a few variants in *BRCA1* and *BRCA2,* have recently been implicated in the protection of replication forks, while having no impact on HR (65,66). To our knowledge, no such variants have yet been reported for *PALB2*. Although these *BRCA1* and *BRCA2* variants appear to associate with cancer, their exact risk needs to be further established.

The use of functional assays for predicting therapy response

Although healthy cells can often repair DNA damage by making use of their full repertoire of DNA repair mechanism, cells exhibiting deficiency in HR due to the presence of *PALB2* LOF

variants,become more reliable on alternative DNA repair mechanisms to survive and proliferate. Therefore, conventional treatment strategies (especially for HR-deficient tumours), have been developed to force DNA damage-induced cell death through synthetic lethal interactions. It is now well established that cancers that exhibit pathogenic variants in *BRCA1* or *BRCA2* respond well to treatment with PARPi (96,97), a therapeutic strategy that has emerged for *BRCA1*- and *BRCA2*-mutated breast and ovarian tumours (48,98,99). Consequently, it is of great importance to identify deleterious *PALB2* VUS that lead to HR deficiency and for which corresponding tumours may similarly respond to PARPi-based therapy.

With regard to the studies that functionally analysed VUS in *PALB2* (39-43), it is clear that within each study, the HR efficiency correlated extremely well with PARPi sensitivity, exhibiting a strong positive correlation in mES cells ($R^2 = 0.804$) (39) and human cell lines ($R^2 = 0.68$) (42). Similar results were obtained for sensitivity assays with cisplatin ($R^2 = 0.8313$) (39,43), a commonly-used chemotherapeutic for many cancers, including breast and ovarian cancer. Similar to that in many *BRCA1*- and *BRCA2*-associated tumours (100,101), many *PALB2*-associated breast cancers (i.e. 67%) show loss of the *PALB2* wild type allele via acquired pathogenic somatic variants, or via loss-of-heterozygosity (LOH) (102,103). Such *PALB2*-null cancers all exhibited HR deficiency, with some tumours even showing HR deficiency while the wild type allele was retained (102,103), suggesting that also alternative mechanisms for PALB2 LOF can be in play. With results from such studies in mind, findings from functional assays that show which VUS are damaging or functional, may prove to be valuable for predicting platinum- and/or PARPi-based therapy response in cancer patients that carry *PALB2* variants that abrogate HR.

Concluding remarks

Due to the accelerating pace by which genetic variants in *PALB2* are discovered, there is a strong need to determine which variants actually associate with disease causation. The combined effort to functionally characterize 155 *PALB2* genetic variants, for which clinical significance is unknown, represents a milestone in the reclassification of these variants. Classification of VUS to a category with a defined clinical significance is of great importance to carriers of a pathogenic variant. This will allow them to make an informed decision on how to manage their cancer risk, including increased surveillance or risk reducing surgery to reduce cancer incidence and/or offering testing of relatives at risk. Counselees carrying non-pathogenic variants may be discharged from intensive follow-up and avoid unnecessary risk-reducing surgery (104).

In this review, we have provided head-to-head comparisons of the different assays that were used for the functional characterization of variants in *PALB2*. These analyses are an

important starting point for the identification of variants that impact its major tumor suppressive function, which most likely is to be attributed to its role in HR, and whose defects correlate with significantly increased cancer risk. Although these assays were able to consistently determine effects of several variants on PALB2's function during HR, some differences in PALB2 function were also observed (Fig. 3), which may be attributed to the type of cDNA-based complementation approach being used. With regard to functional assays being used as clinical diagnostic tools, it is essential to combine results from functional assays that have been obtained by employing different experimental strategies (74,105,106). Moreover, most functional assays use HR as a read-out. However, if PALB2's role in checkpoint control, the regulation of cellular ROS levels and/or the maintenance of replication fork integrity may contribute to its tumor suppressive function as well, expanding the different read-outs of functional assays to cover these aspects of PALB2 function will be a must. Generally, these assays should also include the possibility of a combined mRNA and protein expression analysis in order to provide insight into the effect of variants in coding and non-coding regions of *PALB2* that are predicted to affect RNA splicing, further improving their classification.

Until more conclusive correlations between the level of impairment of protein function and associated cancer risk have been established, results from functional assays should be implemented with care when making a clinical assertion with regard to associated cancer risk and targeted therapies. In light of the increasing number of *PALB2* variants that will undoubtedly be identified in the future, this information will ultimately be crucial for clinical geneticists in selecting the appropriate strategy for clinical management of carriers of (rare) variants in *PALB2*.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

RACMB, MPGV and HvA conducted literature research and wrote the paper.

FUNDING

This work was financially supported by grants from the Dutch Cancer Society (5649 and 11704 to MPGV; 7473 to HvA), and received funding from the European Union's Horizon 2020 research and innovation programme under grant 634935 (BRIDGES; MPGV and HvA).

REFERENCES

1. Ciccia A, Elledge SJ. The DNA damage response: making it safe to play with knives. Mol Cell **2010**;40(2):179-204 doi 10.1016/j.molcel.2010.09.019.
2. Chapman JR, Taylor MR, Boulton SJ. Playing the end game: DNA double-strand break repair pathway choice. Mol Cell **2012**;47(4):497-510 doi 10.1016/j.molcel.2012.07.029.
3. Sishc BJ, Davis AJ. The Role of the Core Non-Homologous End Joining Factors in Carcinogenesis and Cancer. Cancers (Basel) **2017**;9(7) doi 10.3390/cancers9070081.
4. Hustedt N, Durocher D. The control of DNA repair by the cell cycle. Nat Cell Biol **2016**;19(1):1-9 doi 10.1038/ncb3452.
5. Densham RM, Garvin AJ, Stone HR, Strachan J, Baldock RA, Daza-Martin M*, et al.* Human BRCA1-BARD1 ubiquitin ligase activity counteracts chromatin barriers to DNA resection. Nat Struct Mol Biol **2016**;23(7):647-55 doi 10.1038/nsmb.3236.
6. Marini F, Rawal CC, Liberi G, Pellicioli A. Regulation of DNA Double Strand Breaks Processing: Focus on Barriers. Front Mol Biosci **2019**;6:55 doi 10.3389/fmolb.2019.00055.
7. Symington LS. Mechanism and regulation of DNA end resection in eukaryotes. Crit Rev Biochem Mol Biol **2016**;51(3):195-212 doi 10.3109/10409238.2016.1172552.
8. Prakash R, Zhang Y, Feng W, Jasin M. Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. Cold Spring Harb Perspect Biol **2015**;7(4):a016600 doi 10.1101/cshperspect.a016600.
9. Xia B, Sheng Q, Nakanishi K, Ohashi A, Wu J, Christ N*, et al.* Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. Mol Cell **2006**;22(6):719-29 doi 10.1016/j.molcel.2006.05.022.
10. Zhang F, Ma J, Wu J, Ye L, Cai H, Xia B*, et al.* PALB2 links BRCA1 and BRCA2 in the DNA-damage response. Curr Biol **2009**;19(6):524-9 doi 10.1016/j.cub.2009.02.018.
11. Sy SM, Huen MS, Chen J. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. Proc Natl Acad Sci U S A **2009**;106(17):7155-60 doi 10.1073/pnas.0811159106.
12. Zhang F, Fan Q, Ren K, Andreassen PR. PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2. Mol Cancer Res **2009**;7(7):1110-8 doi 10.1158/1541-7786.MCR-09-0123.
13. Wong AK, Pero R, Ormonde PA, Tavtigian SV, Bartel PL. RAD51 interacts with the evolutionarily conserved BRC motifs in the human breast cancer susceptibility gene brca2. J Biol Chem **1997**;272(51):31941-4 doi 10.1074/jbc.272.51.31941.
14. Bignell G, Micklem G, Stratton MR, Ashworth A, Wooster R. The BRC repeats are conserved in mammalian BRCA2 proteins. Hum Mol Genet **1997**;6(1):53-8 doi 10.1093/hmg/6.1.53.
15. Esashi F, Christ N, Gannon J, Liu Y, Hunt T, Jasin M*, et al.* CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. Nature **2005**;434(7033):598-604 doi 10.1038/nature03404.

16. Jensen RB, Carreira A, Kowalczykowski SC. Purified human BRCA2 stimulates RAD51-mediated recombination. Nature **2010**;467(7316):678-83 doi 10.1038/nature09399.
17. Luijsterburg MS, Typas D, Caron MC, Wiegant WW, van den Heuvel D, Boonen RA*, et al.* A PALB2-interacting domain in RNF168 couples homologous recombination to DNA break-induced chromatin ubiquitylation. Elife **2017**;6 doi 10.7554/eLife.20922.
18. Zong D, Adam S, Wang Y, Sasanuma H, Callen E, Murga M*, et al.* BRCA1 Haploinsufficiency Is Masked by RNF168-Mediated Chromatin Ubiquitylation. Mol Cell **2019**;73(6):1267-81 e7 doi 10.1016/j.molcel.2018.12.010.
19. Callen E, Zong D, Wu W, Wong N, Stanlie A, Ishikawa M*, et al.* 53BP1 Enforces Distinct Pre- and Post-resection Blocks on Homologous Recombination. Mol Cell **2020**;77(1):26-38 e7 doi 10.1016/j.molcel.2019.09.024.
20. Nik-Zainal S, Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X*, et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature **2016**;534(7605):47-54 doi 10.1038/nature17676.
21. Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G*, et al.* A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nat Genet **2017**;49(10):1476-86 doi 10.1038/ng.3934.
22. Nguyen L, Martens J, Van Hoeck A, Cuppen E. Pan-cancer landscape of homologous recombination deficiency. bioRxiv **2020**:2020.01.13.905026 doi 10.1101/2020.01.13.905026.
23. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL*, et al.* Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet **2003**;72(5):1117-30 doi 10.1086/375033.
24. Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkas K, Roberts J*, et al.* Breast-cancer risk in families with mutations in PALB2. N Engl J Med **2014**;371(6):497-506 doi 10.1056/NEJMoa1400382.
25. Howlett NG, Taniguchi T, Olson S, Cox B, Waisfisz Q, De Die-Smulders C*, et al.* Biallelic inactivation of BRCA2 in Fanconi anemia. Science **2002**;297(5581):606-9 doi 10.1126/science.1073834.
26. Sawyer SL, Tian L, Kahkonen M, Schwartzentruber J, Kircher M, University of Washington Centre for Mendelian G*, et al.* Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype. Cancer Discov **2015**;5(2):135-42 doi 10.1158/2159-8290.CD-14-1156.
27. Tischkowitz M, Xia B. PALB2/FANCN: recombining cancer and Fanconi anemia. Cancer Res **2010**;70(19):7353-9 doi 10.1158/0008-5472.CAN-10-1012.
28. Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J*, et al.* Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. JAMA Oncol **2017**;3(9):1190-6 doi 10.1001/jamaoncol.2017.0424.
29. Hofstatter EW, Domchek SM, Miron A, Garber J, Wang M, Componeschi K*, et al.* PALB2 mutations in familial breast and pancreatic cancer. Fam Cancer **2011**;10(2):225-31 doi 10.1007/s10689-011-9426-1.

30. Hu C, LaDuca H, Shimelis H, Polley EC, Lilyquist J, Hart SN*, et al.* Multigene Hereditary Cancer Panels Reveal High-Risk Pancreatic Cancer Susceptibility Genes. JCO Precis Oncol **2018**;2 doi 10.1200/PO.17.00291.
31. Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW*, et al.* Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science **2009**;324(5924):217 doi 10.1126/science.1171202.
32. Slater EP, Langer P, Niemczyk E, Strauch K, Butler J, Habbe N*, et al.* PALB2 mutations in European familial pancreatic cancer families. Clin Genet **2010**;78(5):490-4 doi 10.1111/j.1399-0004.2010.01425.x.
33. Yang X, Leslie G, Doroszuk A, Schneider S, Allen J, Decker B*, et al.* Cancer Risks Associated With Germline PALB2 Pathogenic Variants: An International Study of 524 Families. J Clin Oncol **2020**;38(7):674-85 doi 10.1200/JCO.19.01907.
34. Woods NT, Baskin R, Golubeva V, Jhuraney A, De-Gregoriis G, Vaclova T*, et al.* Functional assays provide a robust tool for the clinical annotation of genetic variants of uncertain significance. NPJ Genom Med **2016**;1 doi 10.1038/npjgenmed.2016.1.
35. Bouwman P, van der Gulden H, van der Heijden I, Drost R, Klijn CN, Prasetyanti P*, et al.* A high-throughput functional complementation assay for classification of BRCA1 missense variants. Cancer Discov **2013**;3(10):1142-55 doi 10.1158/2159-8290.CD-13-0094.
36. Mesman RLS, Calleja F, Hendriks G, Morolli B, Misovic B, Devilee P*, et al.* The functional impact of variants of uncertain significance in BRCA2. Genet Med **2019**;21(2):293-302 doi 10.1038/s41436-018-0052-2.
37. Shimelis H, Mesman RLS, Von Nicolai C, Ehlen A, Guidugli L, Martin C*, et al.* BRCA2 Hypomorphic Missense Variants Confer Moderate Risks of Breast Cancer. Cancer Res **2017**;77(11):2789-99 doi 10.1158/0008-5472.CAN-16-2568.
38. Starita LM, Islam MM, Banerjee T, Adamovich AI, Gullingsrud J, Fields S*, et al.* A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution Variants on Protein Function. Am J Hum Genet **2018**;103(4):498-508 doi 10.1016/j.ajhg.2018.07.016.
39. Boonen R, Rodrigue A, Stoepker C, Wiegant WW, Vroling B, Sharma M*, et al.* Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. Nat Commun **2019**;10(1):5296 doi 10.1038/s41467-019-13194-2.
40. Foo TK, Tischkowitz M, Simhadri S, Boshari T, Zayed N, Burke KA*, et al.* Compromised BRCA1-PALB2 interaction is associated with breast cancer risk. Oncogene **2017**;36(29):4161-70 doi 10.1038/onc.2017.46.
41. Park JY, Singh TR, Nassar N, Zhang F, Freund M, Hanenberg H*, et al.* Breast cancer-associated missense mutants of the PALB2 WD40 domain, which directly binds RAD51C, RAD51 and BRCA2, disrupt DNA repair. Oncogene **2014**;33(40):4803-12 doi 10.1038/onc.2013.421.

42. Rodrigue A, Margaillan G, Torres Gomes T, Coulombe Y, Montalban G, da Costa ESCS*, et al.* A global functional analysis of missense mutations reveals two major hotspots in the PALB2 tumor suppressor. Nucleic Acids Res **2019**;47(20):10662-77 doi 10.1093/nar/gkz780.
43. Wiltshire T, Ducy M, Foo TK, Hu C, Lee KY, Belur Nagaraj A*, et al.* Functional characterization of 84 PALB2 variants of uncertain significance. Genet Med **2019** doi 10.1038/s41436-019-0682-z.
44. Kass EM, Helgadottir HR, Chen CC, Barbera M, Wang R, Westermark UK*, et al.* Double-strand break repair by homologous recombination in primary mouse somatic cells requires BRCA1 but not the ATM kinase. Proc Natl Acad Sci U S A **2013**;110(14):5564-9 doi 10.1073/pnas.1216824110.
45. Ducy M, Sesma-Sanz L, Guitton-Sert L, Lashgari A, Gao Y, Brahiti N*, et al.* The Tumor Suppressor PALB2: Inside Out. Trends Biochem Sci **2019**;44(3):226-40 doi 10.1016/j.tibs.2018.10.008.
46. Murai J, Huang SY, Das BB, Renaud A, Zhang Y, Doroshow JH*, et al.* Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. Cancer Res **2012**;72(21):5588-99 doi 10.1158/0008-5472.CAN-12-2753.
47. Deans AJ, West SC. DNA interstrand crosslink repair and cancer. Nat Rev Cancer **2011**;11(7):467-80 doi 10.1038/nrc3088.
48. Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. Science **2017**;355(6330):1152-8 doi 10.1126/science.aam7344.
49. Lord CJ, Ashworth A. BRCAness revisited. Nat Rev Cancer **2016**;16(2):110-20 doi 10.1038/nrc.2015.21.
50. McCabe N, Turner NC, Lord CJ, Kluzek K, Bialkowska A, Swift S*, et al.* Deficiency in the repair of DNA damage by homologous recombination and sensitivity to poly(ADP-ribose) polymerase inhibition. Cancer Res **2006**;66(16):8109-15 doi 10.1158/0008-5472.CAN-06-0140.
51. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM*, et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res **2014**;42(Database issue):D980-5 doi 10.1093/nar/gkt1113.
52. Oliver AW, Swift S, Lord CJ, Ashworth A, Pearl LH. Structural basis for recruitment of BRCA2 by PALB2. EMBO Rep **2009**;10(9):990-6 doi 10.1038/embor.2009.126.
53. Buisson R, Masson JY. PALB2 self-interaction controls homologous recombination. Nucleic Acids Res **2012**;40(20):10312-23 doi 10.1093/nar/gks807.
54. Sy SM, Huen MS, Zhu Y, Chen J. PALB2 regulates recombinational repair through chromatin association and oligomerization. J Biol Chem **2009**;284(27):18302-10 doi 10.1074/jbc.M109.016717.
55. Orthwein A, Noordermeer SM, Wilson MD, Landry S, Enchev RI, Sherker A*, et al.* A mechanism for the suppression of homologous recombination in G1 cells. Nature **2015**;528(7582):422-6 doi 10.1038/nature16142.

56. Simhadri S, Vincelli G, Huo Y, Misenko S, Foo TK, Ahlskog J*, et al.* PALB2 connects BRCA1 and BRCA2 in the G2/M checkpoint response. Oncogene **2019**;38(10):1585-96 doi 10.1038/s41388-018-0535-2.
57. Menzel T, Nahse-Kumpf V, Kousholt AN, Klein DK, Lund-Andersen C, Lees M*, et al.* A genetic screen identifies BRCA2 and PALB2 as key regulators of G2 checkpoint maintenance. EMBO reports **2011**;12(7):705-12 doi 10.1038/embor.2011.99.
58. Cotta-Ramusino C, McDonald ER, 3rd, Hurov K, Sowa ME, Harper JW, Elledge SJ. A DNA damage response screen identifies RHINO, a 9-1-1 and TopBP1 interacting protein required for ATR signaling. Science **2011**;332(6035):1313-7 doi 10.1126/science.1203430.
59. Ceccaldi R, Liu JC, Amunugama R, Hajdu I, Primack B, Petalcorin MI*, et al.* Homologous-recombination-deficient tumours are dependent on Poltheta-mediated repair. Nature **2015**;518(7538):258-62 doi 10.1038/nature14184.
60. Ma J, Cai H, Wu T, Sobhian B, Huo Y, Alcivar A*, et al.* PALB2 interacts with KEAP1 to promote NRF2 nuclear accumulation and function. Mol Cell Biol **2012**;32(8):1506-17 doi 10.1128/MCB.06271-11.
61. Xia B, Dorsman JC, Ameziane N, de Vries Y, Rooimans MA, Sheng Q*, et al.* Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. Nat Genet **2007**;39(2):159-61 doi 10.1038/ng1942.
62. Sidorova J. A game of substrates: replication fork remodeling and its roles in genome stability and chemo-resistance. Cell Stress **2017**;1(3):115-33 doi 10.15698/cst2017.12.114.
63. Schlacher K, Wu H, Jasin M. A distinct replication fork protection pathway connects Fanconi anemia tumor suppressors to RAD51-BRCA1/2. Cancer Cell **2012**;22(1):106-16 doi 10.1016/j.ccr.2012.05.015.
64. Chaudhuri AR, Callen E, Ding X, Gogola E, Duarte AA, Lee JE*, et al.* Erratum: Replication fork stability confers chemoresistance in BRCA-deficient cells. Nature **2016**;539(7629):456 doi 10.1038/nature19826.
65. Schlacher K, Christ N, Siaud N, Egashira A, Wu H, Jasin M. Double-strand break repair-independent role for BRCA2 in blocking stalled replication fork degradation by MRE11. Cell **2011**;145(4):529-42 doi 10.1016/j.cell.2011.03.041.
66. Daza-Martin M, Starowicz K, Jamshad M, Tye S, Ronson GE, MacKay HL*, et al.* Isomerization of BRCA1-BARD1 promotes replication fork protection. Nature **2019**;571(7766):521-7 doi 10.1038/s41586-019-1363-4.
67. Zellweger R, Dalcher D, Mutreja K, Berti M, Schmid JA, Herrador R*, et al.* Rad51-mediated replication fork reversal is a global response to genotoxic treatments in human cells. J Cell Biol **2015**;208(5):563-79 doi 10.1083/jcb.201406099.
68. Godin SK, Sullivan MR, Bernstein KA. Novel insights into RAD51 activity and regulation during homologous recombination and DNA replication. Biochem Cell Biol **2016**;94(5):407-18 doi 10.1139/bcb-2016-0012.

69. Petermann E, Orta ML, Issaeva N, Schultz N, Helleday T. Hydroxyurea-stalled replication forks become progressively inactivated and require two different RAD51-mediated pathways for restart and repair. Mol Cell **2010**;37(4):492-502 doi 10.1016/j.molcel.2010.01.021.
70. Murphy AK, Fitzgerald M, Ro T, Kim JH, Rabinowitsch AI, Chowdhury D*, et al.* Phosphorylated RPA recruits PALB2 to stalled DNA replication forks to facilitate fork recovery. J Cell Biol **2014**;206(4):493-507 doi 10.1083/jcb.201404111.
71. Nikkila J, Parplys AC, Pylkas K, Bose M, Huo Y, Borgmann K*, et al.* Heterozygous mutations in PALB2 cause DNA replication and damage response defects. Nat Commun **2013**;4:2578 doi 10.1038/ncomms3578.
72. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T*, et al.* Analysis of protein-coding genetic variation in 60,706 humans. Nature **2016**;536(7616):285-91 doi 10.1038/nature19057.
73. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q*, et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv **2019**:531210 doi 10.1101/531210.
74. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J*, et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med **2015**;17(5):405-24 doi 10.1038/gim.2015.30.
75. Grimm DG, Azencott CA, Aicheler F, Gieraths U, MacArthur DG, Samocha KE*, et al.* The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity. Hum Mutat **2015**;36(5):513-23 doi 10.1002/humu.22768.
76. Miosge LA, Field MA, Sontani Y, Cho V, Johnson S, Palkova A*, et al.* Comparison of predicted and actual consequences of missense mutations. Proc Natl Acad Sci U S A **2015**;112(37):E5189-98 doi 10.1073/pnas.1511585112.
77. Sun S, Yang F, Tan G, Costanzo M, Oughtred R, Hirschman J*, et al.* An extended set of yeast-based functional assays accurately identifies human disease mutations. Genome Res **2016**;26(5):670-80 doi 10.1101/gr.192526.115.
78. Hart SN, Hoskin T, Shimelis H, Moore RM, Feng B, Thomas A*, et al.* Comprehensive annotation of BRCA1 and BRCA2 missense variants by functionally validated sequence-based computational prediction models. Genet Med **2019**;21(1):71-80 doi 10.1038/s41436-018-0018-4.
79. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C*, et al.* Pan-cancer whole-genome analyses of metastatic solid tumours. Nature **2019**;575(7781):210-6 doi 10.1038/s41586-019-1689-y.
80. Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G*, et al.* Variant Interpretation: Functional Assays to the Rescue. Am J Hum Genet **2017**;101(3):315-25 doi 10.1016/j.ajhg.2017.07.014.

81. Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL*, et al.* ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. Genet Med **2013**;15(7):565-74 doi 10.1038/gim.2013.73.
82. Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP*, et al.* Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med **2017**;19(2):249-55 doi 10.1038/gim.2016.190.
83. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M*, et al.* Accurate classification of BRCA1 variants with saturation genome editing. Nature **2018**;562(7726):217-22 doi 10.1038/s41586-018-0461-z.
84. Matreyek KA, Starita LM, Stephany JJ, Martin B, Chiasson MA, Gray VE*, et al.* Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat Genet **2018**;50(6):874-82 doi 10.1038/s41588-018-0122-z.
85. Yue P, Li Z, Moult J. Loss of protein structure stability as a major causative factor in monogenic disease. J Mol Biol **2005**;353(2):459-73 doi 10.1016/j.jmb.2005.08.020.
86. Redler RL, Das J, Diaz JR, Dokholyan NV. Protein Destabilization as a Common Factor in Diverse Inherited Disorders. J Mol Evol **2016**;82(1):11-6 doi 10.1007/s00239-015-9717-5.
87. Kim I, Miller CR, Young DL, Fields S. High-throughput analysis of in vivo protein stability. Mol Cell Proteomics **2013**;12(11):3370-8 doi 10.1074/mcp.O113.031708.
88. Gaildrat P, Killian A, Martins A, Tournier I, Frebourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. Methods Mol Biol **2010**;653:249-57 doi 10.1007/978-1-60761-759-4_15.
89. Lopez-Perolio I, Leman R, Behar R, Lattimore V, Pearson JF, Castera L*, et al.* Alternative splicing and ACMG-AMP-2015-based classification of PALB2 genetic variants: an ENIGMA report. J Med Genet **2019**;56(7):453-60 doi 10.1136/jmedgenet-2018-105834.
90. Byrd PJ, Stewart GS, Smith A, Eaton C, Taylor AJ, Guy C*, et al.* A Hypomorphic PALB2 Allele Gives Rise to an Unusual Form of FA-N Associated with Lymphoid Tumour Development. PLoS Genet **2016**;12(3):e1005945 doi 10.1371/journal.pgen.1005945.
91. Mesman RLS, Calleja F, Hendriks G, Morolli B, Misovic B, Devilee P*, et al.* The functional impact of variants of uncertain significance in BRCA2. Genetics in medicine : official journal of the American College of Medical Genetics **2018** doi 10.1038/s41436-018-0052-2.
92. Kuznetsov SG, Liu P, Sharan SK. Mouse embryonic stem cell-based functional assay to evaluate mutations in BRCA2. Nat Med **2008**;14(8):875-81 doi 10.1038/nm.1719.
93. Chang S, Biswas K, Martin BK, Stauffer S, Sharan SK. Expression of human BRCA1 variants in mouse ES cells allows functional analysis of BRCA1 mutations. J Clin Invest **2009**;119(10):3160-71 doi 10.1172/JCI39836.
94. Romy L. S. Mesman FMGRC, Miguel de la Hoya, Peter Devilee, Christi J. van Asperen MD, Harry Vrieling & Maaike P. G. Vreeswijk. Alternative mRNA splicing can attenuate the pathogenicity of presumed loss-of-function variants in BRCA2. Genetics in medicine **2020** doi https://doi.org/10.1038/s41436-020-0814-5.

95. Southey MC, Goldgar DE, Winqvist R, Pylkas K, Couch F, Tischkowitz M*, et al.* PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS. J Med Genet **2016**;53(12):800-11 doi 10.1136/jmedgenet-2016-103839.
96. Bryant HE, Schultz N, Thomas HD, Parker KM, Flower D, Lopez E*, et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature **2005**;434(7035):913-7 doi 10.1038/nature03443.
97. Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB*, et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature **2005**;434(7035):917-21 doi 10.1038/nature03445.
98. Audeh MW, Carmichael J, Penson RT, Friedlander M, Powell B, Bell-McGuinn KM*, et al.* Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial. Lancet **2010**;376(9737):245-51 doi 10.1016/S0140-6736(10)60893-8.
99. Tutt A, Robson M, Garber JE, Domchek SM, Audeh MW, Weitzel JN*, et al.* Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial. Lancet **2010**;376(9737):235-44 doi 10.1016/S0140-6736(10)60892-6.
100. Maxwell KN, Wubbenhorst B, Wenz BM, De Sloover D, Pluta J, Emery L*, et al.* BRCA locus-specific loss of heterozygosity in germline BRCA1 and BRCA2 carriers. Nat Commun **2017**;8(1):319 doi 10.1038/s41467-017-00388-9.
101. Riaz N, Blecua P, Lim RS, Shen R, Higginson DS, Weinhold N*, et al.* Pan-cancer analysis of bi-allelic alterations in homologous recombination DNA repair genes. Nat Commun **2017**;8(1):857 doi 10.1038/s41467-017-00921-w.
102. Lee JEA, Li N, Rowley SM, Cheasley D, Zethoven M, McInerny S*, et al.* Molecular analysis of PALB2-associated breast cancers. J Pathol **2018**;245(1):53-60 doi 10.1002/path.5055.
103. Li A, Geyer FC, Blecua P, Lee JY, Selenica P, Brown DN*, et al.* Homologous recombination DNA repair defects in PALB2-associated breast cancers. NPJ Breast Cancer **2019**;5:23 doi 10.1038/s41523-019-0115-9.
104. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS*, et al.* Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum Mutat **2008**;29(11):1282-91 doi 10.1002/humu.20880.
105. Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD*, et al.* Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med **2019**;12(1):3 doi 10.1186/s13073-019-0690-2.
106. Monteiro AN, Bouwman P, Kousholt AN, Eccles DM, Millot GA, Masson JY*, et al.* Variants of uncertain clinical significance in hereditary breast and ovarian cancer genes: best practices in functional analysis for clinical annotation. J Med Genet **2020** doi 10.1136/jmedgenet-2019-106368.

107. Tischkowitz M, Xia B, Sabbaghian N, Reis-Filho JS, Hamel N, Li G*, et al.* Analysis of PALB2/FANCN-associated breast cancer families. Proc Natl Acad Sci U S A **2007**;104(16):6788-93 doi 10.1073/pnas.0701724104.

108. Bleuyard JY, Buisson R, Masson JY, Esashi F. ChAM, a novel motif that mediates PALB2 intrinsic chromatin binding and facilitates DNA repair. EMBO Rep **2012**;13(2):135-41 doi 10.1038/embor.2011.243.

109. Sy SM, Huen MS, Chen J. MRG15 is a novel PALB2-interacting factor involved in homologous recombination. J Biol Chem **2009**;284(32):21127-31 doi 10.1074/jbc.C109.023937.

CHAPTER 3

3

CHEK2 variants: linking functional impact to cancer risk

Rick A.C.M. Boonen, Maaike P.G. Vreeswijk, Haico van Attikum

Published in *Trends in Cancer*
(PMID: 35643632)

ABSTRACT

Protein-truncating variants in the breast cancer susceptibility gene *CHEK2* are associated with a moderate increased risk of breast cancer. In contrast, for missense variants of uncertain significance (VUS) in *CHEK2* the associated breast cancer risk is often unclear. To facilitate their classification, functional assays that determine the impact of missense VUS on CHK2 protein function have been performed. Here we discuss these functional analyses that consistently reveal an association between impaired protein function and increased breast cancer risk. Overall, these findings suggest that damaging *CHEK2* missense VUS associate with a similar risk of breast cancer as protein-truncating variants. This indicates the urgency for expanding the functional characterization of *CHEK2* missense VUS to further understand the associated cancer risk.

KEYWORDS

Breast Cancer; *CHEK2*; Variant of Uncertain Significance (VUS); Functional Assay; Variant Classification; Cancer Risk

CHEK2 and Cancer Predisposition

The CHK2 (see Glossary) protein kinase was initially identified as the mammalian homolog of the *Saccharomyces (S.) cerevisiae* Rad53 and *Schizosaccharomyces pombe* Cds1 protein kinases (1). Its characterization revealed an important role in cell cycle control and apoptosis following exposure of cells to DNA damaging agents (1,2). This involves the phosphorylation and activation of CHK2 by ataxia-telangiectasia mutated (ATM) kinase, and the subsequent modification of downstream substrates such as p53, CDC25A, CDC25C, KAP1 and BRCA1, Collectively, this may prevent genome instability and cancer development by instructing cells to stop proliferating and repair the DNA damage, or promote apoptosis as a response to inefficient or improper repair (Fig. 1). It is perhaps not surprising that shortly after its identification, frameshift variants such as the well-known c.1100del; p.T367Mfs variant, were identified in the *CHEK2* gene and were linked to a cancer susceptibility disorder called Li-Fraumeni syndrome (LFS) (3). LFS is a rare hereditary autosomal-dominant disorder that is characterized by a wide range of malignancies that appear at an unusually early age (4). Similar to CHK2, the well-described tumor suppressor protein p53 also halts cell division in response to DNA damage and inherited mutations in the corresponding gene (TP53), account for most cases of LFS (5). Interestingly, a link between CHK2 and p53 became evident when it was shown that CHK2 phosphorylates p53 on S20, resulting in dissociation of preformed p53-Mdm2 complexes and consequently in p53 stabilization (2). These observations suggested that CHK2 is a tumor suppressor protein that acts within the p53 signaling pathway.

In recent years, several studies have confirmed CHK2's tumor suppressive function by showing that truncating variants in the *CHEK2* gene (e.g., c.1100del; p.T367Mfs) are associated with a moderate-risk for breast cancer (two- to three-fold increased risk) (6-11). For heterozygous female carriers of *CHEK2* truncating variants, this translates to a lifetime risk of ~25% to develop breast cancer before the age of 80 years (6). Furthermore, Cybulski et al. characterized *CHEK2* as a multi-organ cancer susceptibility gene (12), which was confirmed by numerous other studies (reviewed in (13)). These findings have resulted in a significant increase in genetic testing for *CHEK2,* and consequently the identification of many rare missense variants for which clinical relevance is unclear. It is now evident that besides the high-risk breast cancer susceptibility genes *BRCA1*, *BRCA2* and *PALB2*, *CHEK2*, together with *ATM*, appear to be the most commonly mutated genes in the germline of breast cancer patients (6). In fact, 1148 distinct missense VUS in *CHEK2* have currently (as of February 2022) been reported in ClinVar (14). In aggregate, many of these rare missense variants, also termed missense variants of uncertain significance (VUS), also associate with breast cancer (odds ratio [OR], 1.42; 95% CI, 1.28-1.58; p=2.5x10^{11}) (6). This association appears to be independent of their position within the gene and thus their impact on any of the functional domains of CHK2; N-terminal SQ/TQ cluster domain (residues 19-69), a fork head-associated

(FHA) domain (residues 92-205), a serine/threonine kinase domain (residues 212-501), and a nuclear localization signal (NLS) (515-522) (Fig. 2). Knowing which missense variants impact protein function, and to what extent, can help distinguish which variants associate with increased breast cancer risk. To this end, the outcomes of quantitative and well-validated functional assays for *CHEK2*, in line with ACMG guidelines (15), can help to guide clinical classification of genetic variants in this gene, thereby improving the counseling of carriers. Indeed, several recent studies described the functional characterization of *CHEK2* variants. Here we review these studies by providing an overview of the different approaches and outcomes, discussing the potential pitfalls of functional assays, and associating the functional outcomes with breast cancer risk.

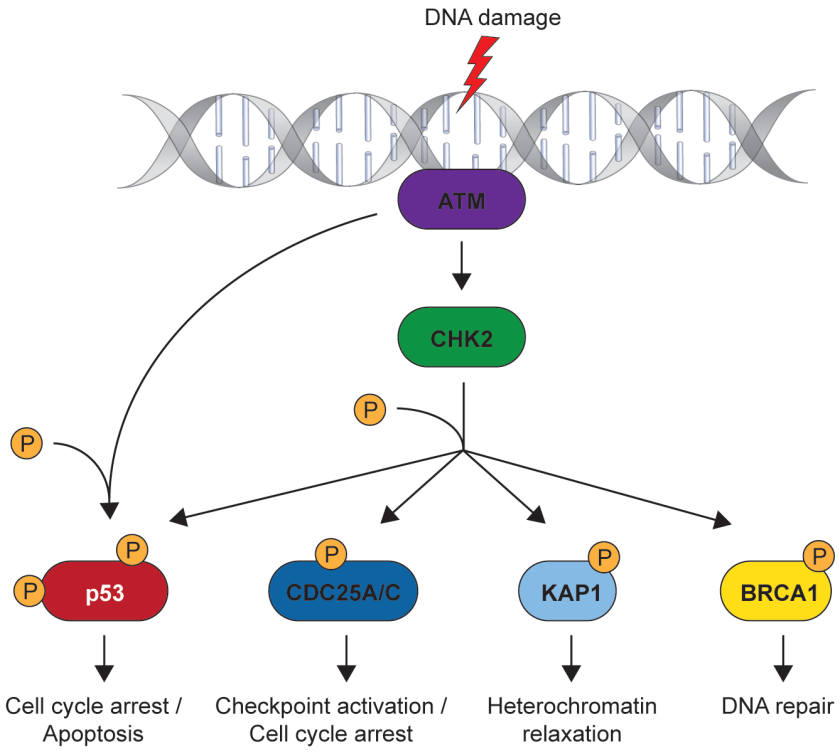


Figure 1. Schematic model displaying the regulation and function of CHK2 kinase. In response to DNA damage, ATM phosphorylates (indicated by the sphere 'P') both CHK2 and p53. ATM-dependent CHK2 phosphorylation promotes the activation of CHK2, and the subsequent CHK2-dependent phosphorylation of numerous downstream substrates such as p53, CDC25A/C, KAP1 and BRCA1. In this way, the CHK2 kinase regulates several cellular processes such as cell cycle regulation/checkpoint activation, apoptosis, heterochromatin relaxation and DNA repair.

Functional Analysis of *CHEK2* VUS

Numerous studies have set out to test the functional consequences of rare variants in the *CHEK2* gene to aid in their clinical interpretation (Table 1) (16-28). Ideally, a functional assay for a cancer predisposition gene measures a function that has been linked to the cancer phenotype. However, although it is known that CHK2 phosphorylates a wide spectrum of substrates involved in cell cycle regulation, DNA repair and apoptosis (29-34), precisely which modifications are relevant for cancer development is largely unclear. Nonetheless, CHK2's ability to phosphorylate any of these substrates may reflect its activity towards all other substrates and thus inform on its functionality in general. In the remainder of this section we discuss the different functional assays and readouts that have been used for the functional classification of missense VUS in *CHEK2* (Table 1).

Shortly after the identification of the CHK2 protein (1), the effect of the first reported missense variants that were identified in patients, were tested in functional assays (3,21,23,28). This work identified the first damaging missense variants in *CHEK2* (e.g., p.R145W), by showing a profound impact on CHK2 protein stability and/or kinase activity, as measured by *in vitro* kinase assays using CDC25A (21) or CDC25C peptides (23,28) as substrates. Three later studies similarly employed *in vitro* assays using CDC25C (18), BRCA1 (16) and KAP1 peptides (22) as substrates. These studies mostly relied on the immunoprecipitation of activated and tagged CHK2 from cells (i.e., after the induction of DNA damage) (16,18,21,23,28), or the purification of recombinant CHK2 (22). Overall, these studies resulted in the functional characterization of 39 distinct variants in the *CHEK2* gene (Fig. 2, Table 1, Supplementary table) (16,18,21-23,28).

A second system that was used for the functional analysis of *CHEK2* variants relied on the use of budding yeast *S. cerevisiae* strains that are null for *RAD53* (and *SML1* to rescue viability), which is the homolog of human *CHEK2* (1) and functional analog of *CHEK1* (35). Expressing human wild type *CHEK2* cDNA in *RAD53*-null yeast strains rescued their slow growth phenotype, likely by restoring its functions in cell cycle checkpoints (36). Accordingly, this system efficiently distinguished the damaging effect of the truncating c.1100del; p.T367Mfs variant from wild type *CHEK2*, whose expression resulted in reduced growth when compared to the wild type control (25,26). This system was later adapted by treating the cells with the DNA damaging agent Methyl methanesulfonate (MMS) (20,24), which results in cell cycle arrest due to the induction of stalled replication forks. Using this approach, two independent studies reported on the functional characterization of 132 distinct *CHEK2* variants (Fig. 2, Table 1, Supplementary table). Specifically, 35 missense VUS, which were identified in patients, two control deletion variants (p.E107_K197del and p.D265_H282del) and a catalytic-dead variant (p.D347A) that impairs kinase activity (20,24), were classified as damaging.

Figure 2. Circos plot of the CHK2 protein displaying the functional classification of 179 variants, including truncating (9), deletion (3), synonymous (7) and missense variants (160). CHK2 variants are indicated in the outer ring and are depicted clockwise, starting from the N-terminus of the CHK2 protein for which the domain structure is shown in the middle (SCD = SQ/TQ cluster domain; NLS = nuclear localization signal; FHA = forkhead-associated domain). Variants are color-coded based on type: green (synonymous variants), red (truncating variants), orange (deletion variants), and blue (missense variants). Each track, except track 1, shows the functional classification of variants from the indicated study (see also Table 1): "functional" (green sphere), "intermediate" (orange sphere), or "damaging" (red sphere). Track 1 shows the average voting score, which was calculated based on all functional classifications available for a given variant. To this end, every classification indicated in track 2-15 was given the same weight, meaning "functional" = 100%, "intermediate" = 50%, "damaging" = 0%. Using this weight, the average voting score was calculated, resulting in a classification as "functional" (green; 81 variants) ≥ 66.7%, "intermediate" (orange; 28 variants) 33.4 - 66.6%, or "damaging" (red; 70 variants) ≤ 33.3%. The data shown in this figure are also available in the Supplementary table (online manuscript only).

Table 1. List of functional studies for variants in the *CHEK2* gene.

	Study	Model system	Functional assay	Nr. of variants
n/a	Cuella-Martin et al., 2021 (19)	MCF7 and MCF10A cells	Growth after DNA damage induction using cisplatin, olaparib, doxorubicin or camptothecin	~159
2	Delimitsou et al., 2019 (20)	*RAD53*-null yeast strains	Growth after DNA damage induction using methyl methanesulfonate	122
3	Boonen et al., 2022 (17)	*Chek2* KO mES cells	Kap1 S473 phosphorylation	63
n/a	Boonen et al., 2022 (17)	*Chek2* KO mES cells	Protein stability	30
n/a	Boonen et al., 2022 (17)	*Chek2* KO mES cells	Growth after DNA damage induction using phleomycin	8
4	Kleiblova et al., 2019 (22)	*CHEK2* KO RPE1 cells	KAP1 S473 phosphorylation	28
5	Kleiblova et al., 2019 (22)	In vitro	Phosphorylation of KAP1 peptide (aa 467-478)	28
6	Kleiblova et al., 2019 (22)	In vitro	Omnia kinase assay	28
7	Roeb et al., 2012 (24)	*RAD53*-null yeast strains	Growth after DNA damage induction using methyl methanesulfonate	26
8	Bell et al., 2007 (16)	In vitro	Phosphorylation of BRCA1 peptide (aa 758-1064)	9
n/a	Bell et al., 2007 (16)	In vitro	Protein stability	
9	Lee et al., 2001 (23)	In vitro	Phosphorylation of CDC25C peptide (aa 200-256)	6
n/a	Lee et al., 2001 (23)	In vitro	Protein stability	
10	Chrisanthar et al., 2008 (18)	In vitro	Phosphorylation of CDC25C peptide	4
n/a	Chrisanthar et al., 2008 (18)	In vitro	Autophosphorylation	
11	Wu et al., 2001 (28)	In vitro	Phosphorylation of CDC25C peptide (aa 200-256)	4
n/a	Wu et al., 2001 (28)	In vitro	CHK2 T68 phosphorylation	
12	Tischkowitz et al., 2008 (26)	*RAD53*-null yeast strains	Growth	4
13	Shaag et al., 2005 (25)	*RAD53*-null yeast strains	Growth	4
14	Falck et al., 2001 (21)	In vitro	Phosphorylation of CDC25A peptide	3
15	Wang et al., 2015 (27)	Eμ-Myc p19Arf −/− B cells	Growth after DNA damage induction using cisplatin, olaparib or doxorubicin	1
n/a	Wang et al., 2015 (27)	Eμ-Myc p19Arf −/− B cells	p53 S20 and CDC25A phosphorylation	1
n/a	Wang et al., 2015 (27)	Eμ-Myc p19Arf −/− B cells	p53 protein levels	1

Tracks correspond to rings in the Circos plot (Fig. 2). Track numbers only apply to a functional readout that resulted in a functional classification by the authors (i.e., functional, intermediate and damaging). Number of variants indicates the number of unique variants that were assessed in a model system with a specific functional readout. Abbreviations: n/a, not applicable; aa, amino acid.

A third system used for functional analysis relies on the use of mammalian cell lines that were depleted of endogenous CHK2 protein, prior to complementation with human *CHEK2* cDNA carrying specific variants (17,22,27). Depletion of endogenous CHK2 was achieved through siRNA/shRNA-mediated silencing of *CHEK2* expression (i.e., knockdown) (27), or by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9-based loss of *CHEK2* expression (i.e., knockout) (17,22). *CHEK2* knockout is compatible with life, since

CHEK2 is a non-essential gene, whose absence promotes mammalian cell growth (17,19). Following loss of endogenous CHK2, the functional effects of *CHEK2* variants were measured using different readouts; i.e., CHK2 kinase activity on substrates such as CDC25A (27) or KAP1 (17,22), CHK2 protein stability (17), cell growth after DNA damage induction (17,27), or p53 protein levels (27) (Table 1). Overall, these three studies functionally characterized 81 distinct *CHEK2* variants (Fig. 2, Table 1, Supplementary table), resulting in the identification of numerous missense variants with a damaging impact (17,22,27).

Overall, the aforementioned studies resulted in the functional characterization of 179 distinct *CHEK2* variants, including 7 synonymous, 9 truncating, 3 deletion and 160 missense VUS. Importantly, an average voting score (Fig. 2, Supplementary table), revealed that 81 variants (i.e., 7 synonymous variants and 74 missense VUS) were functional, 28 variants (i.e., 1 deletion variant, 1 truncating variant and 26 missense VUS) were intermediate in function, and 70 variants (2 deletion variants, 8 truncating variants and 60 missense VUS) were damaging. Mechanistic follow-up studies further showed that some of the damaging *CHEK2* missense variants impaired autophosphorylation and thus activation of CHK2, while most of the other variants impaired function by causing protein instability (17), a mechanism also reported for pathogenic variants in other genes (37,38). Generally, most damaging missense variants were located in the FHA domain (residues 92-205) and Kinase domain (residues 212-501) of CHK2, which is perhaps not surprising as they together make up most of the protein (Fig. 2, Supplementary table). However, to gain a comprehensive view on the damaging impact of variants throughout CHK2, a more extensive functional assessment of variants located in the SCD domain (residues 19-69) and outside functional domains is needed.

Challenges in the Functional Characterization of *CHEK2* VUS

The systems that have been used thus far for the functional analysis of genetic variants each have their strengths and weaknesses, which can result in discrepancies in the outcomes and consequently the functional classification of *CHEK2* variants. Here we review these differences and highlight some future challenges.

The initial functional analysis of *CHEK2* variants relied mostly on *in vitro* kinase assays involving the expression of *CHEK2* variants in cells that still express endogenous wild type *CHEK2* (16,18,21,23,28). A limitation of such an approach is that upon activation by DNA damage, CHK2 variant proteins can form dimers with endogenous wild type CHK2 protein. This may obscure assay results as the association of CHK2 variant proteins with wild type CHK2 may impact CHK2 function. This may also apply to systems in which depletion of endogenous CHK2 relied on knockdown (27) rather than knockout, since residual wild type CHK2 protein may still be present. In contrast, purification of recombinant CHK2 variant proteins from *Escherichia coli* for use in *in vitro* kinase assays, likely influences functional

impact due to lack of posttranslational modifications that are otherwise induced in response to DNA damage in human cells (22). Moreover, *in vitro* assays are unable to detect potential defects in CHK2 protein stability or intracellular localization, and often measure CHK2 kinase activity using artificial substrates (16,18,22,23,28), which may differ from that of full-length substrates.

Most *CHEK2* variants have thus far been characterized using a yeast-based system (20,24,26). Although the overall structure of the CHK2 protein is similar in all eukaryotes, human CHK2 shows only 28% amino-acid identity with the *S. cerevisiae* Rad53 protein (39). Such differences in sequence similarity may affect functional analysis of human *CHEK2* variants in a yeast-cell context. Furthermore, yeast cells grow at 30°C rather than at 37°C, which may reduce the effect of some variants on the thermodynamic stability of CHK2. Accordingly, several unstable CHK2 variants exhibiting intermediate functional effects in mammalian cells (i.e., p.D203G, p.E239K and p.D438Y) (17), were classified as functional in a yeast-based system (20). Thus, growth temperature of a model system may therefore be an important aspect to take into account with regards to the functional characterization of human *CHEK2* variants.

Given the potential limitations of a yeast-based system, a mammalian cell-based system may be favored for the functional analysis of *CHEK2* variants. Indeed, two studies employed such a system based on stable and physiological CHK2 expression levels, rather than transient overexpression of CHK2, in *CHEK2*-deficient cells (17,22). Both studies used DNA damage-induced phosphorylation of KAP1 S473 as a functional readout for CHK2 kinase activity. Functional outcomes were generally accordant and only minor inconsistencies were observed for three (i.e., p.E64K, p.I157T and p.D438Y) out of ten variants studied. A potential limitation of this approach, however, may be that some *CHEK2* missense variants disrupt CHK2 activity against one substrate but not another. Consequently, this approach may not accurately measure the overall impact of a variant on CHK2 activity following DNA damage induction. However, correlating the results from phospho-Kap1 S473 assays to a more general functional readout (i.e., cell growth after DNA damage induction) for eight variants, showed that there is a strong and significant correlation (17). Thus, although CHK2's role in regulating cell growth after DNA damage induction likely stems from its ability to phosphorylate multiple downstream targets, these data suggest that the phosphorylation of Kap1 S473 may be a suitable readout to assess the overall function of CHK2.

When using Kap1 S473, or any other phospho-target of CHK2 as a functional readout, another aspect that also complicates the functional assessment of *CHEK2* variants is the observed kinetic defect reported for some variants, e.g., p.E64K and p.R521W (17). Examination of CHK2 kinase activity at different timepoints after IR showed that, in contrast to wild type CHK2, these two variants are unable to maintain phosphorylation of Kap1 S473 over

the course of the experiment (i.e., 6 hours compared to 2 hours after IR). This suggests that the chosen timepoint at which to asses CHK2 kinase activity after DNA damage induction, may influence functional classification. Accordingly, this may have resulted in some of the reported discrepancies for p.E64K and p.R521W (17,20,22).

In contrast to cDNA-based complementation systems, variants can also be introduced at endogenous loci using CRISPR -dependent technologies, For *BRCA1*, a CRISPR/Cas9-dependent saturation genome editing technique was used that enabled the functional characterization of nearly 4000 variants in the RING and BRCT domains of BRCA1, using cell survival as a functional readout (40). Moreover, for 86 DNA damage response genes, including *CHEK2*, a CRISPR-dependent cytosine base editing screen has been used to interrogate the functional effects of thousands of variants by examining cell growth after DNA damage induction (19). This strategy has major advantages in that it assesses the effects of variants in the context of the endogenous gene and thus at physiological expression levels. Moreover, the effects of variants located in non-coding regions can also be analyzed. Thus, potential effects on mRNA splicing from variants located in both coding and non-coding regions can be functionally assessed. Although such technological advances are anticipated to become important in the future characterizion of variants at scale, they may require optimization before they can be considered a clinical diagnostics tool. For instance, the base editor employed by Cuella-Martin and colleagues has an editing window of 6 nucleotides and often results in the introduction of multiple variants therein (19). This makes it sometimes difficult, if not impossible, to obtain and interpret results for individual variants. Moreover, the repertoire of variants that can be generated is, among others, dependent on protospacer adjacent motifs (PAM) in the DNA that is targeted by the CRISPR system, thus limiting the number of variants that can be characterized. Finally, when a general readout such as cell growth is examined, off-target effects of sgRNAs may have a major impact on the outcome of the functional assay. Nonetheless, these en masse studies will undoubtedly accelerate the path to clinical interpretation of genetic variants in a high-throughput manner.

Clinical Interpretation of *CHEK2* Variants: Functional Assays to the Rescue?

Genetic testing to identify individuals at increased risk of developing breast cancer has accelerated rapidly over the past decade and now also includes moderate-risk genes such as *CHEK2*. The clinical classification of VUS in *CHEK2*, as either pathogenic or benign, is hampered by their rare nature and the moderate breast cancer risk that is associated with pathogenic *CHEK2* variants. This precludes the use of genetic approaches, such as co-segregation analysis, that have been successfully applied in the classification of VUS in high-risk genes such as *BRCA1* and *BRCA2* (41,42). The use of validated functional assays is therefore a very attractive option to consider for improving the clinical classification of VUS in

CHEK2. Before these assays can be used for variant classification, it is essential to establish the quantitative relationship between CHK2 protein functionality and cancer risk.

To date, reliable cancer risk estimates have only been established for a few *CHEK2* variant alleles which are relatively frequent in the population (Table 2) (6,17,43,44) (6,43-45). Interestingly, the risk estimates for these variants (i.e., p.E64K, p.R117G, p.I157T, p.R180C, p.H371Y, p.T476M) show an inverse correlation with their functional impact, meaning that variants exhibiting less activity associate with higher cancer risk (Table 2). In contrast to these *CHEK2* variants, the prevalence of other missense variants is too low to determine their association with breast cancer risk empirically. Assuming that variants with a similar impact on CHK2 protein function associate with the same level of cancer risk, a burden-type association analysis based on reported protein functionality is warranted (Table 3) (17,20,22). This analysis first reveals that the *in vitro* kinase assays generally show poor correlation between functional effects and breast cancer risk, suggesting they may not adequately distinguish functional effects of *CHEK2* variants. Secondly, it shows that the yeast-based system is good at classifying damaging variants (with an OR around 2), but poor at discriminating functional variants from intermediate variants (both groups with ORs around 1.3). Finally, it confirms that the outcome of mammalian cell-based systems (17,22) show an inverse correlation between CHK2 protein function and breast cancer risk as was also reported for the unique variant alleles (Table 2, Table 3). Although the number of variants for which functional data are available is still modest, both the variant specific and the burden analysis derived ORs illustrate that there is a group of *CHEK2* missense variants that associate with a similar cancer risk as has been reported for truncating *CHEK2* variants and that those can be identified by functional analysis. Moreover, the available data thus far also show that *CHEK2* variants that do not associate with clinically relevant cancer risks up to ORs of 1.3 (e.g., p.I157T and p.R180C) do not show a functional impact (see outstanding questions).

Currently, standard guidelines for reporting *CHEK2* missense VUS are lacking, mainly due to the absence of convincing evidence of disease association. However, based on recently obtained insights (Table 3) (17,20), the existence of *CHEK2* missense variants that associate with a comparable risk of breast cancer as *CHEK2* truncating variants, including the c.1100del; p.T367Mfs variant (Table 2), is highly likely. It is therefore crucial that functional assays are used to discriminate between missense variants that affect protein function and are associated with breast cancer risk from those that do not. In this way, functional analysis will provide an essential contribution to reliable variant classification and improved clinical management for carriers and their families.

In addition to functional assays, computational tools may be useful in the clinical interpretation of *CHEK2* missense variants (at scale). For instance, the *in silico* prediction tool Helix (46,47) has been shown to perform well in predicting functionality of *CHEK2* missense

variants (17). These *in silico* predictions should, however, be handled with caution as they have been shown to overestimate the number of damaging variants (38,48,49). Therefore, computational tools might specifically aid in predicting functionality of missense variants that require further analysis of their impact, either because functional outcomes were inconsistent across different studies, or because functional analysis have yet to be performed.

Table 2. Breast cancer risk associated with genetic variants in *CHEK2*.

Nucleotide change	Amino acid change	Average voting score (Fig.2)	Odds ratio	95% CI	*p*-value	Reference
c .190G>A	p.E64K	Intermediate	1,78	1,14 - 2,77	0,0112	Dorling et al., 2021 (6), Boonen et al., 2022 (17)
c.349A>G	p.R117G	Damaging	2,22	1,34 - 3,68	0,002	Dorling et al., 2021 (6), Boonen et al., 2022 (17)
			2,26	1.29 - 3.95	0,003	Southey et al., 2016 (44)
c.470T>C	p.I157T	Functional	1,37 (iCOGS)	1,21 - 1,55	<0,0001	Michailidou et al., 2017 (43)
			1.26 (OncoArray)	1,11 - 1,42	0,0002	
			0.96 (GWAS)	0.72 - 1.28	0,77	
c.538C>T	p.R180C	Functional	1,33	1,05 - 1,67	0,016	Southey et al., 2016 (44)
c.1100delC	p.T367Mfs	Damaging	2,66	2,27 - 3,11	<0,0001	Dorling et al., 2021 (6)
c.1111C>T	p.H371Y	Functional	1,01	0,64 - 1,59	0,9618	Dorling et al., 2021 (6), Boonen et al., 2022 (17)
c.1427C>T	p.T476M	Damaging	1,60	1,10 - 2,35	0,0145	Dorling et al., 2021 (6)

Abbreviations: GWAS, genome-wide association study; iCOGS, International Collaborative Oncological Gene–Environment Study.

Table 3. Burden-type cancer risk association analysis for human *CHEK2* variants.

Study	Variant group based on function	Nr. cases	Nr. controls	OR	95% CI	*p*-value
Boonen et al.,. 2021 (17); *Chek2* KO mES cells	Functional variants	117	108	1,13	0,87-1,46	0,378
	Intermediate variants	110	70	1,63	1,21-2,20	0,0014
	Intermediate variants (excl. p.E64K)	57	39	1,52	1,01-2,28	0,0448
	Damaging variants	118	55	2,23	1,62-3,07	<0,0001
	Damaging variants (excl. p.R117G)	71	33	2,23	1,48-3,38	<0,0001
Delimitsou et al., 2019 (20); *RAD53*-null yeast strains	Functional variants	397	304	1,36	1,17 - 1,58	0,0001
	Functional variants (excl. p.E64K):	344	273	1,31	1,12 - 1,53	0,0009
	Intermediate variants	138	109	1,31	1,02 - 1,69	0,0329
	Intermediate variants (excl. p.T476M)	70	65	1,12	0,80 - 1,57	0,5165
	Damaging variants	116	58	2,08	1,52 - 2,85	<0,0001
	Damaging variants (excl. p.R117G)	69	36	1,99	1,33 - 2,98	0,0008
Kleiblova et al., 2019 (22); *CHEK2* KO RPE1 cells	Functional variants	173	133	1,35	1,08 - 1,69	0,0092
	Functional variants (excl. p.T476M)	105	89	1,23	0,92 - 1,63	0,1592
	Intermediate variants	31	20	1,61	0,92 - 2,82	0,0971
	Damaging variants	91	54	1,75	1,25 - 2,45	0,0011
	Damaging variants (excl. p.E64K)	38	23	1,72	1,02 - 2,88	0,0411
Kleiblova et al., 2019 (22); pKap1 in vitro	Functional variants	153	107	1,48	1,16 - 1,90	0,0017
	Functional variants (excl. p.E64K):	100	76	1,37	1,01 - 1,84	0,0404
	Intermediate variants	38	34	1,16	0,73 - 1,84	0,5282
	Damaging variants	104	66	1,64	1,20 - 2,23	0,0018
	Damaging variants (excl. p.T476M)	36	22	1,7	1,00 - 2,89	0,0501
Kleiblova et al., 2019 (22); in vitro Omnia assay	Functional variants	131	90	1,51	1,16 - 1,98	0,0017
	Functional variants (excl. p.E64K):	78	59	1,37	0,98 - 1,93	0,0404
	Intermediate variants (only p.R406H)	14	12	1,21	0,56 - 2,62	0,6258
	Damaging variants	150	105	1,48	1,16 - 1,90	0,002
	Damaging variants (excl. p.T476M)	82	61	1,4	1,00 - 1,94	0,0487

Abbreviations: KO, knockout; mES cells, mouse embryonic stem cells.

Concluding Remarks and Future Perspectives

Due to the accelerating pace by which germline *CHEK2* variants are discovered, there is a strong need to determine which variants are associated with increased cancer risk. To this end, functional assays have been developed and used to characterize a substantial set of *CHEK2* missense variants, resulting in the identification of rare *CHEK2* variants that exhibit damaging effects on protein function (Fig. 2). These analyses have allowed for a burden-type association analysis, allowing us to correlate the level of functional impact of rare *CHEK2* missense variants to breast cancer risk (Table 3) (17). Importantly, extension of the current cDNA-based methods to genome editing-based methods will provide insight into the effect of

coding and non-coding variants on RNA splicing and downstream functional consequences, further improving the clinical classification of variants in *CHEK2*. Future assays aimed to address the functional effect of every possible nucleotide change in *CHEK2* in a high-throughput manner, such as those performed for *BRCA1* (50), should ultimately result in publicly available resources displaying the quantitative functional output from validated and calibrated functional assays for all *CHEK2* variants. Finally, a ClinGen variant curation expert panel (VCEP) will establish *CHEK2*-specific specifications of the ACMG-based clinical variant interpretation guidelines and provide recommendations for the implementation of results from functional analysis in the classification of missense variants in *CHEK2*. Ultimately, the addition of functional data from validated assays will improve their clinical interpretation and aid in the counseling of carriers and their families.

HIGHLIGHTS

Functional assays have been developed that can determine the impact of missense variants of uncertain significance (VUS) on CHK2 protein function.

Functional analyses of *CHEK2* missense VUS reveal an association between impaired protein function and increased breast cancer risk.

Damaging *CHEK2* missense VUS may associate with a similar risk of breast cancer as protein-truncating variants.

A comprehensive functional characterization of *CHEK2* missense VUS is needed to determine the associated cancer risk.

Functional analysis of missense VUS in *CHEK2* will improve the clinical management of carriers and their family members.

OUTSTANDING QUESTIONS

What is an ideal system for functional analysis of genetic variants in CHEK2? The ideal system may study the functional impact of variants in human cells and in the context of the endogenous gene. With the availability of multiplex assays and CRISPR/Cas9-mediated base-editing, the introduction of all possible variants in *CHEK2* is within reach but is certainly not a standard approach yet. Whether non-cancerous or (breast) cancer cells should be used is debatable, as differences in cell type, tissue and genetic context may affect the functional impact of *CHEK2* variants. Finally, how loss of CHK2's function relates to cancer development is presently unclear. Consequently, a functional readout that captures CHK2 defects that are causally linked to cancer remains to be established.

Can functional analysis keep up with the overwhelming number of CHEK2 variants that have been, and are being, identified by genetic tests? Using a one-by-one approach for functional analysis of *CHEK2* variants is too time-consuming to address the vast number of identified variants (1148 distinct missense VUS in *CHEK2* have currently (as of February 2022 been reported in ClinVar). High-throughput approaches, such as those performed for *BRCA1* and *PTEN* (37,50), may provide answers to this challenge. As protein instability causes most *CHEK2* missense VUS to be damaging (45), using an experimental strategy such as variant abundance by massively parallel sequencing (VAMP-seq) (37) may provide a good means to identify unstable, and thus damaging *CHEK2* missense variants en masse. Alternatively, the combining FACS-based phospho-Kap1 S473 measurements with VAMP-seq (45), may be a means to identify damaging variants that rather impact CHK2's kinase function.

What about functional analysis of CHEK2 splice variants? Generating variants at the endogenous *CHEK2* locus, high-throughput or not, may allow studying their impact on RNA splicing, CHK2 expression and CHK2 functionality. Alternatively, *Chek2*^{KO} mES cells could be complemented with a bacterial artificial chromosome (BAC) containing the human *CHEK2* gene, as has also been performed for *BRCA2* (51). In such a scenario, it is imperative that RNA analysis is performed to show that splicing of human *CHEK2* RNA in mES cells is comparable to that in human cells.

Can CHEK2 functional assays be used for breast cancer risk prediction? A major challenge is to establish the quantitative relationship between CHK2 protein functionality and breast cancer risk. Association analysis (Table 2, Table 3) (45) showed that the degree of CHK2 dysfunction correlates with increased breast cancer risk and that functional analysis can identify missense variants associated with cancer risks similar to those associated with *CHEK2* truncating

variants (OR ≥ 2). However, the exact risk calculations differ slightly per study. This may be related to the fact that these variants are rare, requiring burden-type analyses to estimate cancer risk for groups of variants. Therefore, data from larger or additional case-control association studies than those currently available (e.g., from the Breast Cancer Association Consortium (6)), as well as functional analysis of additional *CHEK2* variants will be pivotal to better understand the extent to which functional defects in CHK2 associate with cancer risk. These analyses might even enable the development of a 'continuous risk' model whereby a variant-specific risk (also with OR < 2) is calculated and can serve as a risk prediction factor on the basis of its impact on functionality.

How to establish the functional threshold for pathogenicity? Guidelines published by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) suggest the use of 'well established' functional studies that provide strong support for or against pathogenicity of a variant (52). However, since the number of pathogenic *CHEK2* missense variants is insufficient, a threshold for pathogenicity cannot be set on basis of such variants. Under the assumption that missense variants with similar levels of functionality associate with the same level of cancer risk, a burden-type association analysis can be performed using large case-control studies (6,53). This analysis will reveal if a group of missense variants (defined by similar levels of functionality) is associated with a risk similar to that of pathogenic variants (i.e., OR>2). While this threshold may be used to identify pathogenic missense variants, its reliability has to be confirmed in the future with (missense) variants that will be classified as pathogenic independent of functional analysis.

Can functional assays guide therapy choice for patients with CHK2-deficient tumors? Currently it is unclear precisely how CHK2 loss of function leads to increased cancer risk and if this deficiency leads to a targetable vulnerability in cancer cells. Consequently, the potential of *CHEK2* functional assays in guiding therapy choice or predicting therapy response for patients with CHK2 related cancer, remains to be elucidated.

GLOSSARY

Checkpoint kinase 2 (*CHEK2*): a tumor suppressor gene that encodes the serine/threonine kinase CHK2, which is involved in DNA repair, cell cycle arrest or apoptosis.

Saccharomyces (S.) cerevisiae: A unicellular eukaryotic organism that constitutes a valuable model for fundamental research.

Rad53: a serine/threonine kinase from *S. cerevisiae* required for DNA damage and replication checkpoints, promoting cell cycle arrest and DNA repair

c.1100delC/p.T367Mfs: HGVS descriptions of a genetic variant at the nucleotide and protein level. 'c' refers to cDNA sequence, while 'p' refers to protein sequence. The numbers reflect nucleotide or codon positions of the wild-type reference sequence. 'del' Refers to deletion of the nucleotide 'C' (cytosine). 'T' Refers to the original wild type amino acid Threonine. 'M' denotes the change of a Threonine to a Methionine at amino acid position 367 in this example. 'fs' Indicates that the nucleotide change results in a frameshift in codon usage at amino acid position 367.

TP53: A tumor suppressor gene that encodes the transcription factor Tumor Protein P53 (p53), which is involved in cell cycle regulation and apoptosis.

Odds ratio (OR): A measure of association between a variable (e.g., a genetic variant) and an outcome (e.g. breast cancer). An OR indicates the odds that breast cancer will occur when carrying a specific variant, compared to the odds of breast cancer occurring in the absence of that specific variant.

Variant of Uncertain Significance (VUS): Genetic variant that cannot be used for clinical decision making or cancer risk assessment due to insufficient clinical and/or functional data needed to assess pathogenicity.

BReast CAncer 1/2 (*BRCA1/2*): The two most commonly affected high-risk breast cancer susceptibility genes, which are involved in the repair of DNA double-strand breaks and the protection of (stalled) DNA replication forks.

Partner and localizer of BRCA2 (*PALB2*): A high-risk gene breast cancer susceptibility gene, which is involved in the repair of DNA double-strand breaks by linking the actions of BRCA1 and BRCA2 therein.

Ataxia telangiectasia-mutated (*ATM*): A moderate-risk breast cancer susceptibility gene that encodes *the* serine/threonine kinase ATM, which is recruited and activated by DNA double-strand breaks to regulate cell cycle progression and DNA repair. Autosomal recessive mutations in *ATM* lead to Ataxia telangiectasia, which is a rare disorder characterized by for instance neurodegeneration, immunodeficiency, radiosensitivity and cancer.

ClinVar: A freely accessible public archive that aggregates information about genomic variation and its relationship to human health (https://www.ncbi.nlm.nih.gov/clinvar/).

ACMG guidelines: Recommendations of the American College of Medical Genetics and Genomics (ACMG) for the clinical interpretation of sequence variants.

Clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9: Molecular biological tool used for genomic editing with the Cas9 nuclease.

Functional assay: Molecular and cellular experiments that can produce data describing the functional impact of a variant on a gene product.

Cell Division Cycle 25 (CDC25)A/C: Two crucial cell cycle regulators and homologs that act as a phosphatase by removing the inhibitory phosphorylation of cyclin-dependent kinases (CDKs), thereby positively regulating the activity of CDKs in promoting cell cycle progression.

Methyl-methanesulfonate (MMS): An alkylating agent that induces replication fork stalling by modifying both guanine (to 7-methylguanine) and adenine (to 3-methlyladenine) bases in the DNA.

Knockdown: Experimental condition that reduces the expression of one or more genes in a cell or organism.

Knockout: Experimental condition by which the genomic DNA of a cell or organism is perturbed to permanently prevent the expression of one or more genes in a cell or organism.

Forkhead-associated (FHA) domain: A protein modular domain that binds phospho-peptides.

Kinase domain: A structurally conserved protein domain harboring the catalytic activity of protein kinases.

SQ/TQ cluster domain (SCD): A protein domain that is defined by the presence of multiple SQ/TQ motifs within a variable stretch of amino acids. SCDs are recognized targets for kinases involved in the DDR.

Escherichia (E.) coli: A gram-negative, rod-shaped bacterium of the genus *Escherichia* that is commonly found in the lower intestine of warm-blooded organisms. It constitutes an important species in the fields of biotechnology and microbiology, where it can serve as the host organism for work with recombinant DNA.

DNA damage response (DDR): An extensive surveillance network that maintains genome integrity and stability, and is thus critical for cellular homeostasis and disease prevention.

ACKNOWLEDGEMENTS

This work is supported by grants from the Dutch Cancer Society (KWF/Pink Ribbon 11704 and KWF-CRAFT 12754) to M.P.G.V and H.v.A..

REFERENCES

1. Matsuoka S, Huang M, Elledge SJ. Linkage of ATM to cell cycle regulation by the Chk2 protein kinase. Science **1998**;282(5395):1893-7 doi 10.1126/science.282.5395.1893.
2. Chehab NH, Malikzay A, Appel M, Halazonetis TD. Chk2/hCds1 functions as a DNA damage checkpoint in G(1) by stabilizing p53. Genes Dev **2000**;14(3):278-88.
3. Bell DW, Varley JM, Szydlo TE, Kang DH, Wahrer DC, Shannon KE*, et al.* Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. Science **1999**;286(5449):2528-31 doi 10.1126/science.286.5449.2528.
4. Schneider K, Zelley K, Nichols KE, Garber J. Li-Fraumeni Syndrome. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Gripp KW*, et al.*, editors. GeneReviews((R)). Seattle (WA)1993.
5. McBride KA, Ballinger ML, Killick E, Kirk J, Tattersall MH, Eeles RA*, et al.* Li-Fraumeni syndrome: cancer risk assessment and clinical management. Nat Rev Clin Oncol **2014**;11(5):260-71 doi 10.1038/nrclinonc.2014.41.
6. Breast Cancer Association C, Dorling L, Carvalho S, Allen J, Gonzalez-Neira A, Luccarini C*, et al.* Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. N Engl J Med **2021**;384(5):428-39 doi 10.1056/NEJMoa1913948.
7. Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J*, et al.* Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. JAMA Oncol **2017**;3(9):1190-6 doi 10.1001/jamaoncol.2017.0424.
8. Decker B, Allen J, Luccarini C, Pooley KA, Shah M, Bolla MK*, et al.* Rare, protein-truncating variants in ATM, CHEK2 and PALB2, but not XRCC2, are associated with increased breast cancer risks. J Med Genet **2017**;54(11):732-41 doi 10.1136/jmedgenet-2017-104588.
9. Hauke J, Horvath J, Gross E, Gehrig A, Honisch E, Hackmann K*, et al.* Gene panel testing of 5589 BRCA1/2-negative index patients with breast cancer in a routine diagnostic setting: results of the German Consortium for Hereditary Breast and Ovarian Cancer. Cancer Med **2018**;7(4):1349-58 doi 10.1002/cam4.1376.
10. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R*, et al.* Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet **2002**;31(1):55-9 doi 10.1038/ng879.
11. Weischer M, Bojesen SE, Tybjaerg-Hansen A, Axelsson CK, Nordestgaard BG. Increased risk of breast cancer associated with CHEK2*1100delC. J Clin Oncol **2007**;25(1):57-63 doi 10.1200/JCO.2005.05.5160.
12. Cybulski C, Gorski B, Huzarski T, Masojc B, Mierzejewski M, Debniak T*, et al.* CHEK2 is a multiorgan cancer susceptibility gene. Am J Hum Genet **2004**;75(6):1131-5 doi 10.1086/426403.
13. Stolarova L, Kleiblova P, Janatova M, Soukupova J, Zemankova P, Macurek L*, et al.* CHEK2 Germline Variants in Cancer Predisposition: Stalemate Rather than Checkmate. Cells **2020**;9(12) doi 10.3390/cells9122675.

14. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM*, et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res **2014**;42(Database issue):D980-5 doi 10.1093/nar/gkt1113.
15. Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD*, et al.* Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med **2019**;12(1):3 doi 10.1186/s13073-019-0690-2.
16. Bell DW, Kim SH, Godwin AK, Schiripo TA, Harris PL, Haserlat SM*, et al.* Genetic and functional analysis of CHEK2 (CHK2) variants in multiethnic cohorts. Int J Cancer **2007**;121(12):2661-7 doi 10.1002/ijc.23026.
17. Boonen R, Wiegant WW, Celosse N, Vroling B, Heijl S, Kote-Jarai Z*, et al.* Functional Analysis Identifies Damaging CHEK2 Missense Variants Associated with Increased Cancer Risk. Cancer Res **2022**;82(4):615-31 doi 10.1158/0008-5472.CAN-21-1845.
18. Chrisanthar R, Knappskog S, Lokkevik E, Anker G, Ostenstad B, Lundgren S*, et al.* CHEK2 mutations affecting kinase activity together with mutations in TP53 indicate a functional pathway associated with resistance to epirubicin in primary breast cancer. PLoS One **2008**;3(8):e3062 doi 10.1371/journal.pone.0003062.
19. Cuella-Martin R, Hayward SB, Fan X, Chen X, Huang JW, Taglialatela A*, et al.* Functional interrogation of DNA damage response variants with base editing screens. Cell **2021**;184(4):1081-97 e19 doi 10.1016/j.cell.2021.01.041.
20. Delimitsou A, Fostira F, Kalfakakou D, Apostolou P, Konstantopoulou I, Kroupis C*, et al.* Functional characterization of CHEK2 variants in a Saccharomyces cerevisiae system. Hum Mutat **2019**;40(5):631-48 doi 10.1002/humu.23728.
21. Falck J, Mailand N, Syljuasen RG, Bartek J, Lukas J. The ATM-Chk2-Cdc25A checkpoint pathway guards against radioresistant DNA synthesis. Nature **2001**;410(6830):842-7 doi 10.1038/35071124.
22. Kleiblova P, Stolarova L, Krizova K, Lhota F, Hojny J, Zemankova P*, et al.* Identification of deleterious germline CHEK2 mutations and their association with breast and ovarian cancer. Int J Cancer **2019**;145(7):1782-97 doi 10.1002/ijc.32385.
23. Lee SB, Kim SH, Bell DW, Wahrer DC, Schiripo TA, Jorczak MM*, et al.* Destabilization of CHK2 by a missense mutation associated with Li-Fraumeni Syndrome. Cancer Res **2001**;61(22):8062-7.
24. Roeb W, Higgins J, King MC. Response to DNA damage of CHEK2 missense mutations in familial breast cancer. Hum Mol Genet **2012**;21(12):2738-44 doi 10.1093/hmg/dds101.
25. Shaag A, Walsh T, Renbaum P, Kirchhoff T, Nafa K, Shiovitz S*, et al.* Functional and genomic approaches reveal an ancient CHEK2 allele associated with breast cancer in the Ashkenazi Jewish population. Hum Mol Genet **2005**;14(4):555-63 doi 10.1093/hmg/ddi052.
26. Tischkowitz MD, Yilmaz A, Chen LQ, Karyadi DM, Novak D, Kirchhoff T*, et al.* Identification and characterization of novel SNPs in CHEK2 in Ashkenazi Jewish men with prostate cancer. Cancer Lett **2008**;270(1):173-80 doi 10.1016/j.canlet.2008.05.006.

27. Wang N, Ding H, Liu C, Li X, Wei L, Yu J*, et al.* A novel recurrent CHEK2 Y390C mutation identified in high-risk Chinese breast cancer patients impairs its activity and is associated with increased breast cancer risk. Oncogene **2015**;34(40):5198-205 doi 10.1038/onc.2014.443.
28. Wu X, Webster SR, Chen J. Characterization of tumor-associated Chk2 mutations. J Biol Chem **2001**;276(4):2971-4 doi 10.1074/jbc.M009727200.
29. Ahn J, Urist M, Prives C. The Chk2 protein kinase. DNA Repair (Amst) **2004**;3(8-9):1039-47 doi 10.1016/j.dnarep.2004.03.033.
30. Bartek J, Lukas J. Chk1 and Chk2 kinases in checkpoint control and cancer. Cancer Cell **2003**;3(5):421-9 doi 10.1016/s1535-6108(03)00110-7.
31. Kastan MB, Bartek J. Cell-cycle checkpoints and cancer. Nature **2004**;432(7015):316-23 doi 10.1038/nature03097.
32. Li J, Williams BL, Haire LF, Goldberg M, Wilker E, Durocher D*, et al.* Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2. Mol Cell **2002**;9(5):1045-54 doi 10.1016/s1097-2765(02)00527-0.
33. Zhang J, Willers H, Feng Z, Ghosh JC, Kim S, Weaver DT*, et al.* Chk2 phosphorylation of BRCA1 regulates DNA double-strand break repair. Mol Cell Biol **2004**;24(2):708-18 doi 10.1128/MCB.24.2.708-718.2004.
34. Hu C, Zhang S, Gao X, Gao X, Xu X, Lv Y*, et al.* Roles of Kruppel-associated Box (KRAB)-associated Co-repressor KAP1 Ser-473 Phosphorylation in DNA Damage Response. J Biol Chem **2012**;287(23):18937-52 doi 10.1074/jbc.M111.313262.
35. Lanz MC, Dibitetto D, Smolka MB. DNA damage kinase signaling: checkpoint and repair at 30 years. EMBO J **2019**;38(18):e101801 doi 10.15252/embj.2019101801.
36. Zhao X, Chabes A, Domkin V, Thelander L, Rothstein R. The ribonucleotide reductase inhibitor Sml1 is a new target of the Mec1/Rad53 kinase cascade during growth and in response to DNA damage. EMBO J **2001**;20(13):3544-53 doi 10.1093/emboj/20.13.3544.
37. Matreyek KA, Starita LM, Stephany JJ, Martin B, Chiasson MA, Gray VE*, et al.* Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat Genet **2018**;50(6):874-82 doi 10.1038/s41588-018-0122-z.
38. Boonen R, Rodrigue A, Stoepker C, Wiegant WW, Vroling B, Sharma M*, et al.* Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. Nat Commun **2019**;10(1):5296 doi 10.1038/s41467-019-13194-2.
39. Bartek J, Falck J, Lukas J. CHK2 kinase--a busy messenger. Nat Rev Mol Cell Biol **2001**;2(12):877-86 doi 10.1038/35103059.
40. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M*, et al.* Accurate classification of BRCA1 variants with saturation genome editing. Nature **2018**;562(7726):217-22 doi 10.1038/s41586-018-0461-z.
41. Parsons MT, Tudini E, Li H, Hahnen E, Wappenschmidt B, Feliubadalo L*, et al.* Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: An ENIGMA resource to support clinical variant classification. Hum Mutat **2019**;40(9):1557-78 doi 10.1002/humu.23818.

42. Caputo SM, Golmard L, Leone M, Damiola F, Guillaud-Bataille M, Revillion F*, et al.* Classification of 101 BRCA1 and BRCA2 variants of uncertain significance by cosegregation study: A powerful approach. Am J Hum Genet **2021**;108(10):1907-23 doi 10.1016/j.ajhg.2021.09.003.
43. Michailidou K, Lindstrom S, Dennis J, Beesley J, Hui S, Kar S*, et al.* Association analysis identifies 65 new breast cancer risk loci. Nature **2017**;551(7678):92-4 doi 10.1038/nature24284.
44. Southey MC, Goldgar DE, Winqvist R, Pylkas K, Couch F, Tischkowitz M*, et al.* PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS. J Med Genet **2016**;53(12):800-11 doi 10.1136/jmedgenet-2016-103839.
45. Boonen RA, Wiegant WW, Celosse N, Vroling B, Heijl S, Kote-Jarai Z*, et al.* Functional analysis identifies damaging CHEK2 missense variants associated with increased cancer risk. Cancer Res **2021** doi 10.1158/0008-5472.CAN-21-1845.
46. Heijl S, Vroling B, Bergh Tvd, Joosten H. Mind the gap: preventing circularity in missense variant prediction (https://doi.org/10.1101/2020.05.06.080424). **2020**.
47. Vroling B, Heijl S. White paper: The Helix Pathogenicity Prediction Platform (https://arxiv.org/abs/2104.01033). **2021**.
48. Miosge LA, Field MA, Sontani Y, Cho V, Johnson S, Palkova A*, et al.* Comparison of predicted and actual consequences of missense mutations. Proc Natl Acad Sci U S A **2015**;112(37):E5189-98 doi 10.1073/pnas.1511585112.
49. Rodrigue A, Margaillan G, Torres Gomes T, Coulombe Y, Montalban G, da Costa ESCS*, et al.* A global functional analysis of missense mutations reveals two major hotspots in the PALB2 tumor suppressor. Nucleic Acids Res **2019**;47(20):10662-77 doi 10.1093/nar/gkz780.
50. Starita LM, Islam MM, Banerjee T, Adamovich AI, Gullingsrud J, Fields S*, et al.* A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution Variants on Protein Function. Am J Hum Genet **2018**;103(4):498-508 doi 10.1016/j.ajhg.2018.07.016.
51. Mesman RLS, Calleja F, de la Hoya M, Devilee P, van Asperen CJ, Vrieling H*, et al.* Alternative mRNA splicing can attenuate the pathogenicity of presumed loss-of-function variants in BRCA2. Genet Med **2020**;22(8):1355-65 doi 10.1038/s41436-020-0814-5.
52. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J*, et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med **2015**;17(5):405-24 doi 10.1038/gim.2015.30.
53. Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J*, et al.* A Population-Based Study of Genes Previously Implicated in Breast Cancer. N Engl J Med **2021**;384(5):440-51 doi 10.1056/NEJMoa2005936.

CHAPTER 4

4

Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene *PALB2*

Rick A.C.M. Boonen, Amélie Rodrigue, Chantal Stoepker, Wouter W. Wiegant, Bas Vroling, Milan Sharma, Magdalena B. Rother, Nandi Celosse, Maaike P.G. Vreeswijk, Fergus Couch, Jacques Simard, Peter Devilee, Jean-Yves Masson and Haico van Attikum

Published in *Nature Communications*
(PMID: 31757951)

ABSTRACT

Heterozygous carriers of germ-line loss-of-function variants in the DNA repair gene *PALB2* are at a highly increased lifetime risk for developing breast cancer. While truncating variants in *PALB2* are known to increase cancer risk, the interpretation of missense variants of uncertain significance (VUS) is in its infancy. Here we describe the development of a relatively fast and easy cDNA-based system for the semi high-throughput functional analysis of 48 VUS in human *PALB2*. By assessing the ability of *PALB2* VUS to rescue the DNA repair and checkpoint defects in *Palb2* knockout mouse embryonic stem (mES) cells, we identify various VUS in *PALB2* that impair its function. Three VUS in the coiled-coil domain of PALB2 abrogate the interaction with BRCA1, whereas several VUS in the WD40 domain dramatically reduce protein stability. Thus, our functional assays identify damaging VUS in *PALB2* that may increase cancer risk.

KEYWORDS

Breast Cancer; Variant of Uncertain Significance (VUS); PALB2; DNA Repair Homologous Recombination (HR); PARP inhibitor.

INTRODUCTION

Germline loss-of-function (LOF) variants in the breast cancer susceptibility genes *BRCA1* and *BRCA2* are known to result in an approximately tenfold increased lifetime risk of developing breast cancer (1). Similar to these genes, mono-allelic LOF variants in the gene encoding partner and localizer of *BRCA2* (*PALB2*) also increase the risk of breast cancer (2), whereas bi-allelic LOF variants cause Fanconi anemia (FA) (3). It is now well established that women who carry pathogenic variants in *PALB2* are at a similar risk for breast cancer as those who carry pathogenic variants in *BRCA2* (1,4). Therefore, *PALB2* takes a valid place on breast cancer predisposition gene panel tests and is becoming widely included in breast cancer clinical genetics practice. This has already led to the identification of numerous variants in *PALB2*, which may associate with breast cancer (as of September 2019, 1301 *PALB2* VUS have already been reported in ClinVar). However, current risk estimates for *PALB2* variants have so far only been based on truncating variants that are predicted to fully inactivate the protein (5). For most missense variants the impact on protein function is unclear and therefore the associated cancer risk is unknown. Assessment of pathogenicity of such variants of uncertain significance (VUS), therefore relies mostly on co-segregation with disease, co-occurrence with known pathogenic variants, and family history of cancer. To extend the utility of *PALB2* genetic test results, additional methods for interpreting VUS are urgently required.

A key facet of interpreting VUS in *PALB2* is understanding their impact on PALB2 protein function. PALB2 exists as oligomers that can form a complex with BRCA1 and BRCA2 and the recombinase RAD51 (6,7). This involves PALB2's N-terminal coiled-coiled domain for interaction with BRCA1 (7) and its C-terminal WD40 domain for interaction with BRCA2 (8). The PALB2-BRCA1/2-RAD51 complex plays an essential role in homologous recombination (HR), which is a critical pathway for the repair of highly-deleterious DNA double-strand breaks (DSBs). Following their detection, the ends of a DSB are resected to generate stretches of 3' single-stranded DNA (ssDNA), which are bound by the ssDNA-binding protein RPA. PALB2 becomes recruited to these resected DSB ends in a manner dependent on BRCA1 to facilitate the assembly of BRCA2 and RAD51 onto broken DNA ends. RAD51 in turn catalyzes strand invasion and DNA transfer, usually from a sister chromatid available in S/G2 phase (6,7,9), ultimately leading to error-free repair of DSBs.

Germline nonsense and frameshift variants in *BRCA1*, *BRCA2* and *PALB2* give rise to a characteristic genome instability signature that is associated with HR deficiency (10). Targeting this HR deficiency has proven to be effective in PARP inhibitor (PARPi)-based cancer treatment, during which the ensuing DSBs can be repaired by HR in healthy cells, but not in HR-deficient cancer cells (11,12). While PARP inhibitor-based therapy holds great promise for the treatment of HR-deficient cancers, a major obstacle is that clinical testing of

these tumors often reveals numerous VUS in *BRCA1*, *BRCA2* and *PALB2*, for which the effect on HR and the response to PARP inhibitor-based therapy is often unclear.

For *BRCA1* and *BRCA2*, functional assays that mostly use HR as a read-out have been established to assess the effect of VUS on protein function (13-17). These assays have successfully determined the functional consequences and potential therapy response of a variety of VUS. However, with regard to *PALB2*, the functional analysis of variants is still in its infancy even though there is a clear clinical demand. Here, we fill this gap by describing the development of a robust functional assay for the analysis of VUS in *PALB2*. The assay allows a semi high-throughput analysis of VUS in human *PALB2* cDNA in *Palb2* knockout mouse embryonic stem (mES) cells using HR, PARPi sensitivity and G2/M checkpoint maintenance as read-outs. We identify at least 14 *PALB2* VUS that strongly abrogate PALB2 function. Moreover, *PALB2* VUS located in the WD40 domain have a high tendency to impair PALB2 protein function by affecting its stability, whereas *PALB2* variants located in the coiled-coil domain tend to impair its interaction with BRCA1. Thus, we report on the development of a relatively rapid and easy functional assay that can determine the functional consequences of VUS in *PALB2*, thereby facilitating cancer risk assessment and predicting therapy response.

RESULTS

A cell-based functional assay for *PALB2* variants

For the analysis of *PALB2* variants we envisioned a cell-based assay that allows for reliable semi high-throughput testing of variants in human *PALB2*. This cell-based approach should combine efficient integration and equal expression of human *PALB2* cDNA carrying these variants in a cellular background devoid of endogenous *Palb2* and with the ability to assess their effect on HR. To this end, we introduced the well-established DR-GFP reporter into IB10 mES cells, which are highly proficient in HR (Fig.1a, Supplementary Fig. 1a-c) (18). The HR efficiency was nearly identical in all 3 correctly targeted clones (~10%) (Supplementary Fig. 1d) and clone 5 was selected for further experiments.

Next, we introduced the recombination-mediated cassette exchange (RMCE) system into cells from clone 5 (13). One component of this system, which consists of an acceptor cassette with F3 and Frt sites (Fig. 1a, Supplementary Fig. 2), was correctly integrated at the *Rosa26* locus in 1 out of 6 targeted clones (Supplementary Fig. 2a,b). The other component is an exchange cassette that carries a promoterless neomycin selection marker and an *EF1α* promotor fused to human *PALB2* cDNA flanked by F3 and Frt sites. This exchange cassette can be used for FlpO-mediated, site-specific integration of human *PALB2* cDNA at the RMCE acceptor cassette (Fig. 1a) (19). This would allow for stable expression of human *PALB2,*

which we envisioned in a cellular background devoid of endogenous *Palb2*.

Since knockout (KO) of *PALB2* is embryonic lethal (20-22), it has been notoriously difficult to generate *PALB2*^{KO} cells. However, since p53 deficiency could partially rescue *in utero* development of *Palb2*^{KO} mice, we decided to generate *Palb2*^{KO} mES cells in a p53-deficient background. In addition to facilitating the KO of *Palb2*, deficiency in both *p53 and Palb2* may also mimic tumor settings, as somatic *TP53* mutations are common in breast cancer associated with *BRCA1/2 (23,24)* and *PALB2 (25)*. We first employed CRISPR/Cas9-based genome editing to knockout mouse *Trp53* in cells harboring DR-GFP and the RMCE acceptor cassette (Fig. 1a,b, Supplementary Fig. 3a,b). Subsequent analysis of 4 *Trp53*^{KO} clones revealed that HR remained unaffected in these cells (Fig. 1b), allowing functional analysis in this genomic background using HR as a read-out. *Trp53*^{KO} clone-3 had the highest percentage of cells (~50%) with a normal chromosome number (*i.e*. 40 chromosomes) (Fig. 1c) and was therefore selected for further experiments.

Finally, we applied CRISPR/Cas9-mediated genome-editing to knockout mouse *Palb2* (Fig. 1d, Supplementary Fig. 3c,d). As expected, the efficiency of HR in the DR-GFP reporter assay was strongly reduced (by ~95%) in *Trp53*^{KO}/*Palb2*^{KO} cells when compared to that in *Trp53*^{KO} cells alone (Fig. 1d). To test whether human wild-type *PALB2* can complement this defect, we stably expressed wild-type human *PALB2* cDNA using RMCE (Fig. 1a). Importantly, due to site-specific integration, the promoterless neomycin gene will be driven by the endogenous *Rosa26* promoter, which enhances targeting efficiency and allows for selection of integrants on medium containing neomycin. Indeed, we observed *PALB2* expression in all individual neomycin resistant clones that were tested for *PALB2* expression (Supplementary Fig. 4a). However, since some differences in *PALB2* expression were observed between single clones, we pooled the neomycin-resistant clones (~500 clones) prior to examining the HR efficiency (Supplementary Fig. 4b), ruling out any effects on HR caused by differences in *PALB2* expression. We found that HR was efficiently rescued (by ~68%) following expression of human *PALB2* in the *Trp53*^{KO}/*Palb2*^{KO} cells compared to the *Trp53*^{KO} cells (Fig. 1d). Thus, we have developed a highly efficient cDNA-based complementation system for the functional analysis of variants in human *PALB2*.

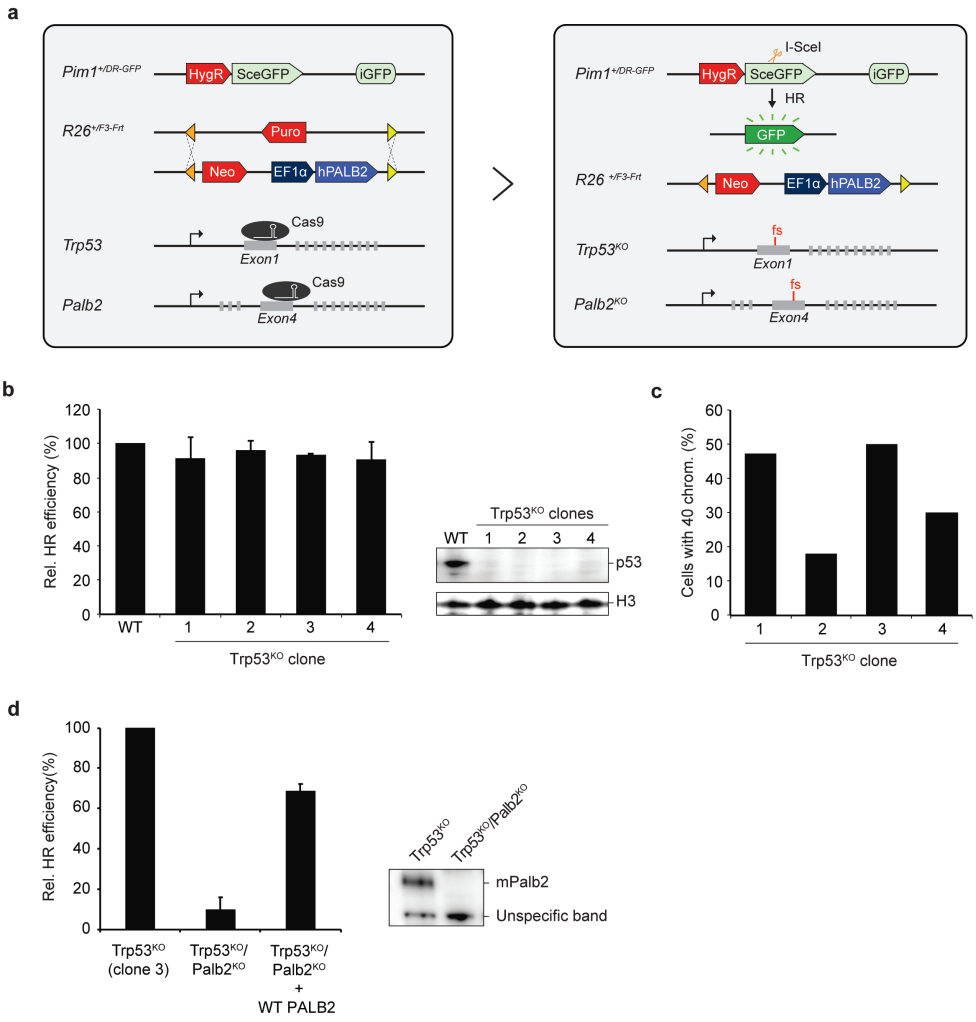


Figure 1. Development of a cDNA-based complementation system for the functional analysis of human *PALB2.* **a** Schematic of the cDNA-based complementation system for functional analysis of human *PALB2.* The DR-GFP reporter for HR and Recombination-Mediated Cassette Exchange system (RMCE) for site-specific integration and expression of a human *PALB2* cDNA were incorporated at the mouse *Pim1 and Rosa26* loci, respectively. Endogenous mouse *Trp53* was targeted with CRISPR/Cas with a gRNA for exon 1, whereas endogenous *Palb2* was targeted with a gRNA against exon 4 (left). Transient expression of the I-SceI endonuclease in *Trp53*^{KO}/*Palb2*^{KO} cells expressing human *PALB2* cDNA (with or without a variant) allows for assessment of the HR efficiency using the DR-GFP reporter (right). **b** DR-GFP assay in *Trp53*^{KO} mES cell clones co-transfected with I-SceI and mCherry expression vectors and GFP expression was monitored by FACS. Data represent mean percentages (± SEM) of GFP-positive cells among the mCherry-positive cells relative to that for the wild type (WT), which was set to 100%, from 2 independent experiments (left). Western blot analysis of *Trp53* expression in *Trp53*^{KO} 4 mES cell clones. Histone 3 (H3) was a loading control (right). **c** Karyotyping of *Trp53*^{KO} mES clones from **b**. The bar graph shows the percentages of cells with 40 chromosomes (n = 50 cells per condition). **d** DR-GFP assay in *Trp53*^{KO} and *Trp53*^{KO}/*Palb2*^{KO} mES cells expressing wild-type *PALB2* or

not. Cells were co-transfected with I-SceI and mCherry expression vectors and GFP expression was monitored by FACS. Data represent mean percentages (± SEM) of GFP-positive cells among the mCherry-positive cells relative to that for *Trp53*^{KO} cells, which was set to 100%, from 4 independent experiments (left). Western blot analysis of Palb2 expression in *Trp53*^{KO} and *Trp53*^{KO}/*Palb2*^{KO} (clone 3) mES cells (right). An unspecific band was a loading control (right). Source data are provided as a Source Data file.

Validation of a cell-based functional assay for *PALB2* variants

To evaluate our system, we selected 12 truncating *PALB2* variants (Fig. 2a, *red*) that are known to be deleterious and associate with cancer and/or Fanconi anemia (3,4,26-28). In addition, we selected 8 missense variants from the dbSNP database (Fig. 2a, *green*), which we expect to be benign/neutral because of their frequency in the general population (between 0.1-15% based on the 1000 Genomes Project). Site-directed mutagenesis was used to introduce these variants, as well as a synonymous variant (c.2574T>C, p.V858=), into the RMCE vector that carries human *PALB2* cDNA. Sequence-verified constructs were introduced by RMCE into the *Trp53*^{KO}/*Palb2*^{KO} mES cells, which were then subjected to DR-GFP assays. As expected, HR was dramatically reduced in cells carrying the empty vector (Ev) when compared to cells expressing human *PALB2* cDNA (*i.e.* reduction in HR of ~90-95%) (Fig. 2b). Similarly, cells expressing human *PALB2* with a truncating variant displayed strong defects in HR. In contrast, cells that expressed either the benign/neutral variants or the synonymous variant showed HR levels comparable to that of cells expressing wild-type *PALB2* (Fig. 2b).

To corroborate these findings, we also examined whether cells expressing benign/neutral or truncating *PALB2* variants display sensitivity to PARPi. As expected, we found that *Trp53*^{KO}/*Palb2*^{KO} cells complemented with the Ev were hypersensitive to PARPi when compared to those expressing wild-type human *PALB2* cDNA (Fig. 3a, Supplementary Figs. 5, 6 and 7). Moreover, the expression of truncating *PALB2* variants led to a dramatically increased sensitivity to PARPi (at least by ~70%), while that of the benign/neutral variants did not (Fig. 3a, Supplementary Fig. 5). Thus, by measuring HR efficiencies using DR-GFP and PARPi sensitivity, our cell-based system reproducibly classifies benign/neutral and pathogenic/truncating variants based on their effect on PALB2 function in HR.

Functional analysis of *PALB2* VUS

In contrast to truncating variants in *PALB2*, the contribution of missense variants with respect to cancer risk is largely unclear. We therefore analyzed the effect of 48 *PALB2* VUS and one synthetic missense variant (p.A1025R) (Fig. 2a, *blue*) (29). Many of these VUS have been identified during a multigene panel analysis for a large case-control association study performed by the BRIDGES consortium. In addition, several VUS were gathered from ClinVar (p.I944N, p.L24S and p.L1070P) and literature (p.K18R, p.Y28C, p.L35P, p.R37H) (30,31).

Interestingly, we observed strong HR defects in DR-GFP assays for p.L35P-, p.W912G-, p.I944N-, p.L961P-, p.G1043D-PALB2, exhibiting a ~90-95% reduction in HR, comparable to the truncating *PALB2* variants and the empty vector conditions (Fig. 2b). In addition, we also observed strong effects on HR for several other VUS (p.L24S, p.Y28C, p.G937R, p.L947S, p.L972Q, p.T1030I, p.I1037T, p.L1070P, p.L1172P), as well as the synthetic missense variant p.A1025R in PALB2, reducing HR by ~60-90% when compared to wild-type PALB2 (Fig. 2b). A FACS-based cell cycle analysis for 33 selected *PALB2* variants showed no effect on cell cycle distribution (Supplementary Fig. 8), excluding the possibility that effects on HR were due to differences in cell-cycle progression.

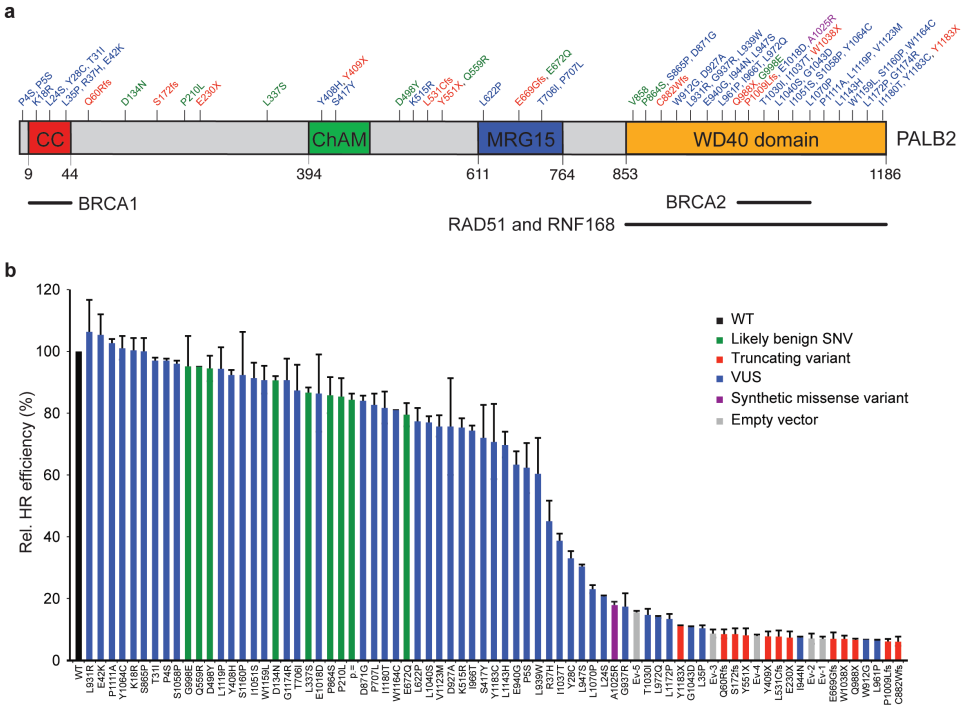


Figure 2. Human *PALB2* variants and their effect on HR. **a** Schematic representation of the PALB2 protein with variant positions indicated and categorized as either neutral (green), truncating (red), VUS (blue) and synthetic missense variant (purple). The amino acid numbers are shown to specify the evolutionarily conserved functional domains of PALB2. PALB2 regions involved in the interactions with BRCA1, BRCA2, RNF168 and RAD51 are indicated. **b** DR-GFP assay in *Trp53*^{KO}/*Palb2*^{KO} mES cells expressing human *PALB2* variants (or an empty vector control, Ev). Cells were co-transfected with I-SceI and mCherry expression vectors and GFP expression was monitored by FACS. Data represent mean percentages (± SEM) of GFP-positive cells among the mCherry-positive cells relative to wild type (WT), which was set to 100%, from 2 independent experiments, except for p.L939W and p.G998E for which data from 3 independent experiments are presented. Variants/conditions are categorized by color

as either wild type (WT, black), likely benign SNV (green), truncating variant (red), VUS (blue), synthetic missense variant (purple) or empty vector (Ev, grey). Ev1-5 refer to Ev controls from 5 different replicates. Source data are provided as a Source Data file.

Next, we examined the effect of the 48 selected VUS and p.A1025R on PARPi sensitivity using a cellular proliferation assay. We observed that 11 VUS (p.Y28C, p.L35P, p.W912G, p.G937R, p.I944N, p.L947S, p.L961P, p.L972Q, p.T1030I, p.G1043D and p.L1172P), as well as p.A1025R, displayed sensitivity to PARPi treatment comparable to that observed for *PALB2* truncating variants (Fig. 3a, Supplementary Fig. 6,7). Importantly, when comparing the HR efficiency measured by DR-GFP and PARPi sensitivity assays, a strong positive correlation was observed for all variants tested (R^2*=0.804*) (Fig. 3b). These results indicate that our complementary cell-based assays can determine the functional consequences of VUS in human *PALB2*. Most notably, taking the data from both assays into account, we identified at least 5 VUS (p.L35P, p.W912G, p.L961P, p.I944N and p.G1043D) that affect PALB2 function to a similar extent as the truncating variants. The effect of these VUS on PARPi sensitivity was further evaluated using a clonogenic survival assay. This revealed that 4 *PALB2* VUS (p.W912G, p.L961P, p.I944N and p.G1043D) also render cells hypersensitive to prolonged treatment with lower concentrations of PARPi (Fig. 3c). Consequently, such VUS may confer an increased cancer risk and serve as a target for PARPi-based therapy.

While PARPi treatment holds great promise for the treatment of HR-deficient tumors, an alternative strategy may be to treat with interstrand crosslink (ICL)-inducing chemotherapeutic drugs, since ICLs require HR for their repair (32). We therefore analyzed several *PALB2* variants in their response to the ICL-inducing agent cisplatin. As expected, two truncating variants p.Y551X and p.Y1183X displayed strong sensitivity to cisplatin comparable to the empty vector condition (Fig. 3d). Consistent with the effects observed in the HR and PARPi assays, three PALB2 VUS (p.L35P, p.L961P and p.G1043D) were also sensitive to cisplatin. When comparing the HR efficiency measured by DR-GFP to cisplatin sensitivity, a strong correlation (R^2*=0.8313)* was observed (Fig. 3e). Thus, VUS in *PALB2* that impair HR may serve as targets for both PARPi- and ICL-based chemotherapy.

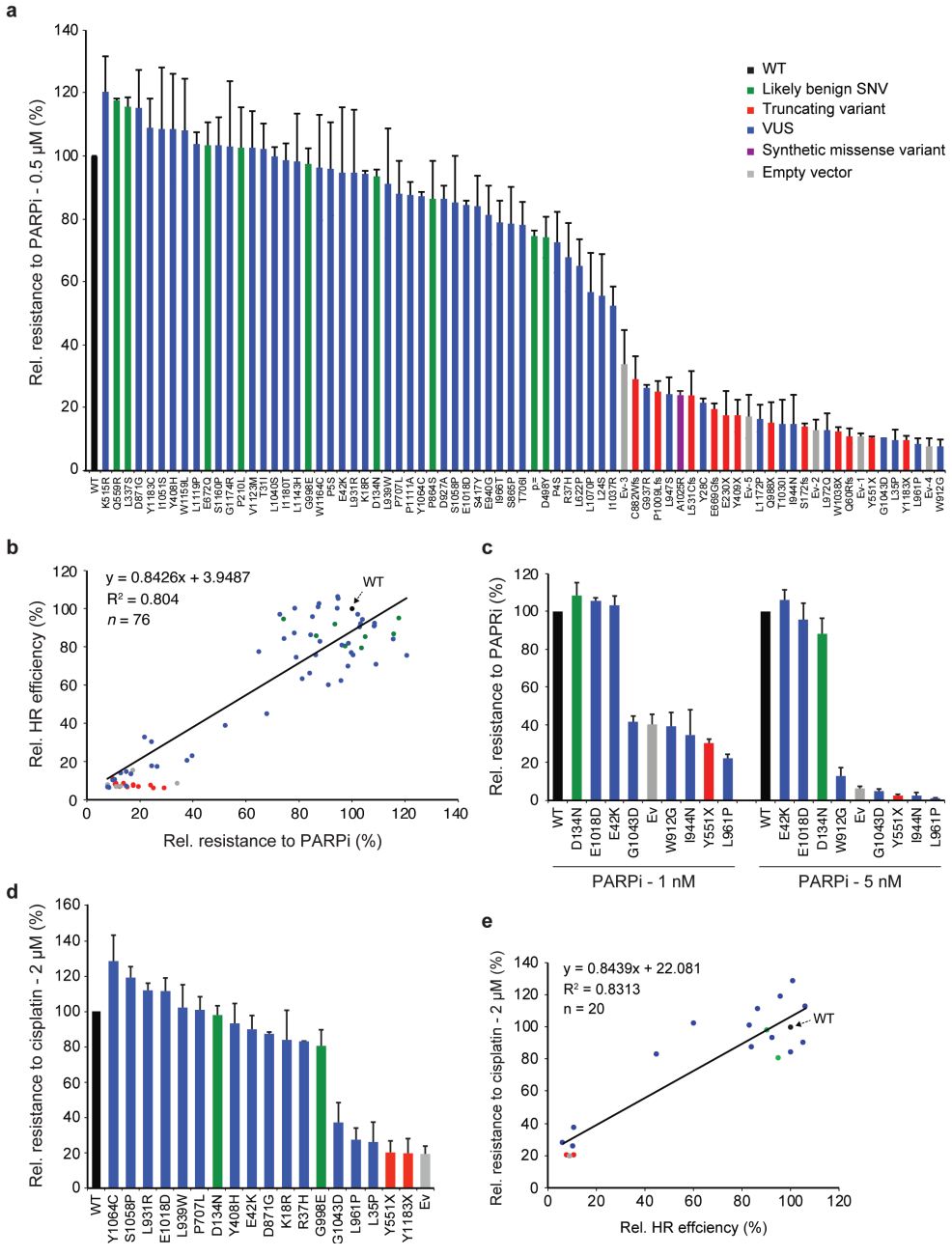


Figure 3. Functional analysis of *PALB2* VUS using PARP inhibitor and cisplatin sensitivity assays. **a** Proliferation-based PARP inhibitor (PARPi) sensitivity assay using *Trp53*^{KO}/*Palb2*^{KO} mES cells expressing human *PALB2* variants (or an empty vector control, Ev). Cells were exposed to 0.5 µM PARPi for two days. Cell viability was measured 1 day later using FACS. Data represent the mean percentage of viability relative to wild type (± SEM), which was set to 100%, from 2 independent

experiments, except for p.P4S, p.P210L, p.L939W and p.V1123M, for which data from three independent experiments is presented, and p.L24S, p.L1070P for which data from four independent experiments is presented. Variants/conditions are categorized by color. **b** Scatter plot showing the correlation between HR efficiencies and PARPi sensitivity measured in Fig. 2a and **b**, respectively. Variants/conditions are categorized by color as in **a**. The trendline indicates the positive correlation between the outcome of DR-GFP and PARPi sensitivity assays. **c** Clonogenic PARP inhibitor survival assay using *Trp53*^{KO}*/Palb2*^{KO} mES cells expressing human *PALB2* variants (or an empty vector control, Ev). Cells were exposed to the indicated concentrations of PARPi for 7-9 days after which surviving colonies were counted. Data represent the mean percentage of survival (± SEM) relative to cells expressing wild-type *PALB2*, which were set to 100%, from 3 independent experiments upon in case of treatment with 1 nM PARPi, and 4 experiments in case of treatment with 5 nM PARPi. Variants/conditions are categorized by color as in **a**. **d** As in **a**, except that cells were exposed to 2 μM cisplatin. Data represent the mean percentage of viability relative to wild type (± SEM), which was set to 100%, from 2 independent experiments. **e** Scatter plot showing the correlation between HR efficiencies and cisplatin sensitivity measured in Fig. 2b and **d**. The trendline indicates the positive correlation between the outcome of DR-GFP and cisplatin sensitivity assays. Variants/conditions are categorized by color as in **a**. Source data are provided as a Source Data file.

Correlation of functional analysis and *in silico* prediction

We next compared the outcome of our functional assays with the predictions of several *in silico* algorithms for all missense variants. For the prediction tools that give categorical results for missense variants, including PolyPhen (33), SIFT (34), and AlignGVGD (35), we observed little to no correlation with the outcome of DR-GFP and PARPi sensitivity assays (Supplementary Data 1). For instance, if we assume an HR efficiency of 40% or lower as damaging in the DR-GFP assay, then 24.1% of the missense variants (likely benign and VUS) are classified as damaging in our functional assay. However, we observed a gross overrepresentation of damaging variants when using PolyPhen (86.2%), SIFT (77.6%) and AlignGVGD (36.2%, counting C55 and C65). With respect to the latter, extreme caution should be taken as AlignGVGD classified at least two variants, which we found to be similarly damaging as truncating variants, as likely benign (p.W912G (C0) and p.I944N (C15); Supplementary Data 1). For *in silico* prediction tools that assign a continuous prediction score, such as (CADD (36) and REVEL (37)), we similarly observed a poor correlation with the outcome of DR-GFP and PARPi sensitivity assays (Supplementary Fig. 9). For instance, based on cut-offs of 0.0-0.5 for benign variants and 0.5-1.0 for damaging variants, REVEL would only categorize three of the PALB2 VUS (p.D871G, p.W912G and p.L931R) as damaging. However, both p.D871G and p.L931R appear to be fully functional in our assays. Thus, while REVEL severely underestimates the effects of VUS on protein function, it may also lead to false-positive predictions. Based on these observations, we conclude that predictive algorithms, as opposed to our functional analysis, are poor in predicting the effect of VUS on PALB2 protein function.

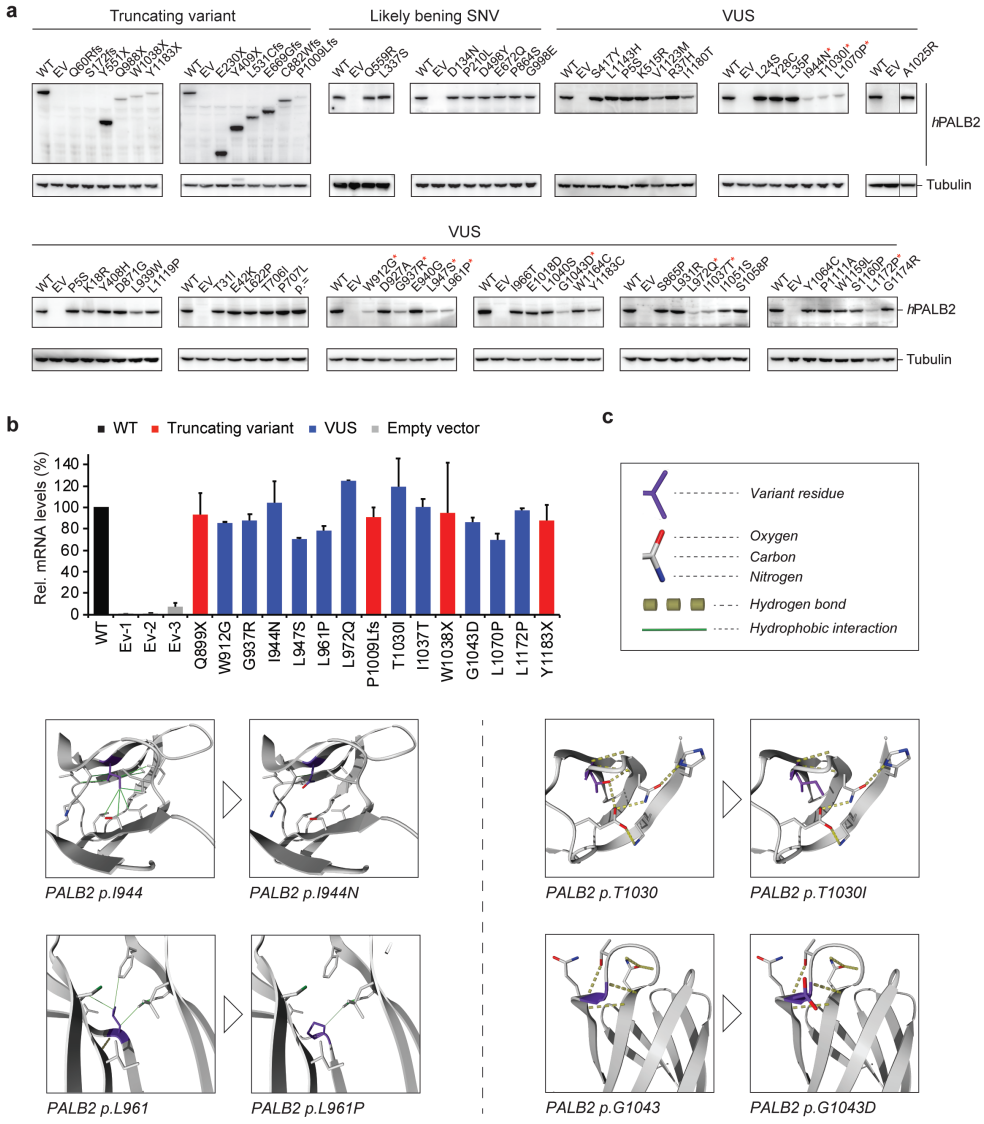


Figure 4. Effect of *PALB2* variants on protein expression and/or stability. **a** Western blot analysis of the expression of human *PALB2* variants in *Trp53*^{KO}/*Palb2*^{KO} mES cells using an antibody directed against the N-terminus of PALB2. Wild-type (WT) human PALB2 and empty vector (Ev) served as controls on each blot. Tubulin was used as a loading control. Marked PALB2 variants (red *) showed low levels of protein expression. **b** RT-qPCR analysis of human *PALB2* variants from A with low expression levels (red *). Primers specific for human *PALB2* cDNA and the *Pim1* control locus were used. Data represent the mean percentage (± SEM) of *PALB2* mRNA relative to wild type, which was set to 100%, from 2 independent RNA isolation experiments. Variants/conditions are categorized by color as either wild type (WT, black), truncating variant (red), VUS (blue) or empty vector (Ev, grey). Ev-1, -2, -3 refer to Ev controls from 3 different replicates. **c** Partial structures of the PALB2 WD40 domain showing the effect of 4 PALB2 variants exhibiting low protein expression as shown in **a**. Partial structures without and with

variant are shown side by side for each variant, indicating loss of stabilizing interactions (but not any possible conformational changes). Source data are provided as a Source Data file.

VUS in the PALB2 WD40 domain affect protein stability

Having identified *PALB2* variants that affect HR, we sought to address their mechanism of action. To this end, we first examined their effect on PALB2 expression by western blot analysis. For all benign variants, PALB2 expression was comparable to that of wild-type PALB2 (Fig. 4a). Similarly, most truncating and missense variants were unaffected in their expression levels, although the truncating variants resulted in the expression of the expected smaller proteins. However, for some truncating variants (p.Q899X, p.P1009Lfs, p.W1038X and p.Y1183X) and VUS located in the C-terminal WD40 domain (p.W912G, p.G937R, p.I944N, p.L947S, p.L961P, p.L972Q, p.T1030I, p.I1037T, p.G1043D, p.L1070P, p.L1172P), low levels of expression were observed (Fig. 4a, *red asterisk*). Reverse transcription-quantitative (RT-q)PCR analysis indicated that these variants did not affect expression at the mRNA level (Fig. 4b). This suggests that the low abundance of PALB2 protein is likely the result of protein misfolding and/or instability.

Crystal structure studies of the PALB2 C-terminal WD40 domain suggested that loss of the last 3 amino acids of PALB2 caused by the FA-associated p.Y1183X variant disrupts the hydrogen bonding in the seventh blade of the WD40 domain (3,29). Consistently, we also observed strongly reduced expression of PALB2 carrying this variant (p.Y1183X) (Fig. 4a). Thus, p.Y1183X may lead to in an incompletely folded PALB2 protein that is likely to be degraded rapidly. As such, it is not surprising that other truncating variants in the WD40 domain result in expression of a truncated protein that is unstable and degraded quickly. However, truncating PALB2 variants that lack the entire WD40 domain (p.E230X, p.Y409X, p.L531Cfs, p.Y551X, p.E669Gfs) appeared to express well (Figs. 2a and 4a). Nevertheless, they have likely lost all of their ability to interact with BRCA2 and RAD51, thereby impairing HR completely. Consistently, we observed almost no difference in the extent to which the different truncated forms of PALB2 affect HR.

Our results suggest that the WD40 domain of PALB2 is extremely sensitive to variants that affect protein folding and/or stability. Using the crystal structure of the WD40 domain (2W18) (29), *in silico* modeling of all PALB2 VUS that display low expression levels indeed showed that all these amino acid substitutions are extremely unfavorable for correct folding of this domain. Starting with p.I944N, we see that this isoleucine is a well-conserved hydrophobic residue that is located in an antiparallel β-sheet and whose side-chain is part of a tightly packed hydrophobic environment (Fig. 4c). Replacement of this isoleucine with an asparagine will lead to the loss of stabilizing hydrophobic interactions due to the energetically unfavorable presence

of a hydrophilic residue in a very hydrophobic environment. These opposed effects may destabilize the local environment and/or lead to folding problems. Comparable effects are predicted for p.L947S, p.L972Q and p.I1037T (Supplementary Fig. 10). L961 is another example of a residue that is located in a β-sheet and is involved in several hydrophobic interactions (Fig. 4c). When it changes into a proline (p.L961P), all of these local interactions are lost. Furthermore, proline is unfavored, because it results in the loss of a backbone hydrogen bond, thereby destabilizing the β-sheet. Comparable effects are predicted for p.W912G, p.L1070P and p.L1172P (Supplementary Fig. 10). However, for p.W912G the change into a very small glycine is also thought to result in excess flexibility at a position where this is not desired.

The side-chain of the hydrophilic residue p.T1030 is involved in an extensive network of hydrogen bonds and electrostatic interactions that extends across all 4 strands of the β-sheet (Fig. 4c). This variant will impair the formation of hydrogen bonds as isoleucine is not capable of these bonds through its sidechain. Consistent with our findings (Fig. 4a), an earlier study also reported protein instability for p.T1030I (31). Finally, p.G937 and p.G1043 are examples of glycine residues that provide structural flexibility at the beginning of a loop structure (Fig. 4c, Supplementary Fig. 10). Changing these residues into a larger and charged arginine (p.G937R) or aspartate (p.G1043D), will lead to deformation of the loop structure and probable loss of surrounding hydrogen bonds in the case of p.G1043D. Altogether, this *in silico* modeling may provide explanations for how these PALB2 VUS affect protein stability/expression levels. Nonetheless, some VUS for which similar destabilizing effects are predicted (p.D871G, p.L931R, p.E1018D and p.W1164C) are fully functional in our HR-based assays, underpinning the importance of functional analysis of VUS.

VUS in the PALB2 CC-domain disrupt the interaction with BRCA1

In addition to the damaging VUS in PALB2's WD40 domain, we also found 4 PALB2 VUS (p.L24S, p.Y28C, p.L35P, p.R37H) exhibiting strong effects on HR and PARPi sensitivity (Figs. 2a and 3a, Supplementary Fig. 6). These variants were all located in PALB2's N-terminal coiled-coil domain, which is required for interaction with BRCA1 (6,9). Indeed, the previously reported p.Y28C and p.L35P variants affected HR by impairing the interaction with BRCA1 (30). However, exactly how p.L24S and p.R37H impact HR is unclear, also because p.R37H has previously been reported not to affect the PALB2-BRCA1 interaction (30). To examine this further, we transiently expressed YFP-tagged PALB2 carrying p.L24S, p.L35P or p.R37H in U2OS cells and performed pull-downs using GFP Trap beads. p.L24S, similar to p.L35P, failed to co-precipitate any endogenous BRCA1, whereas p.R37H partially affected the co-precipitation of BRCA1 (Fig. 5a). Additionally, we examined whether these VUS have

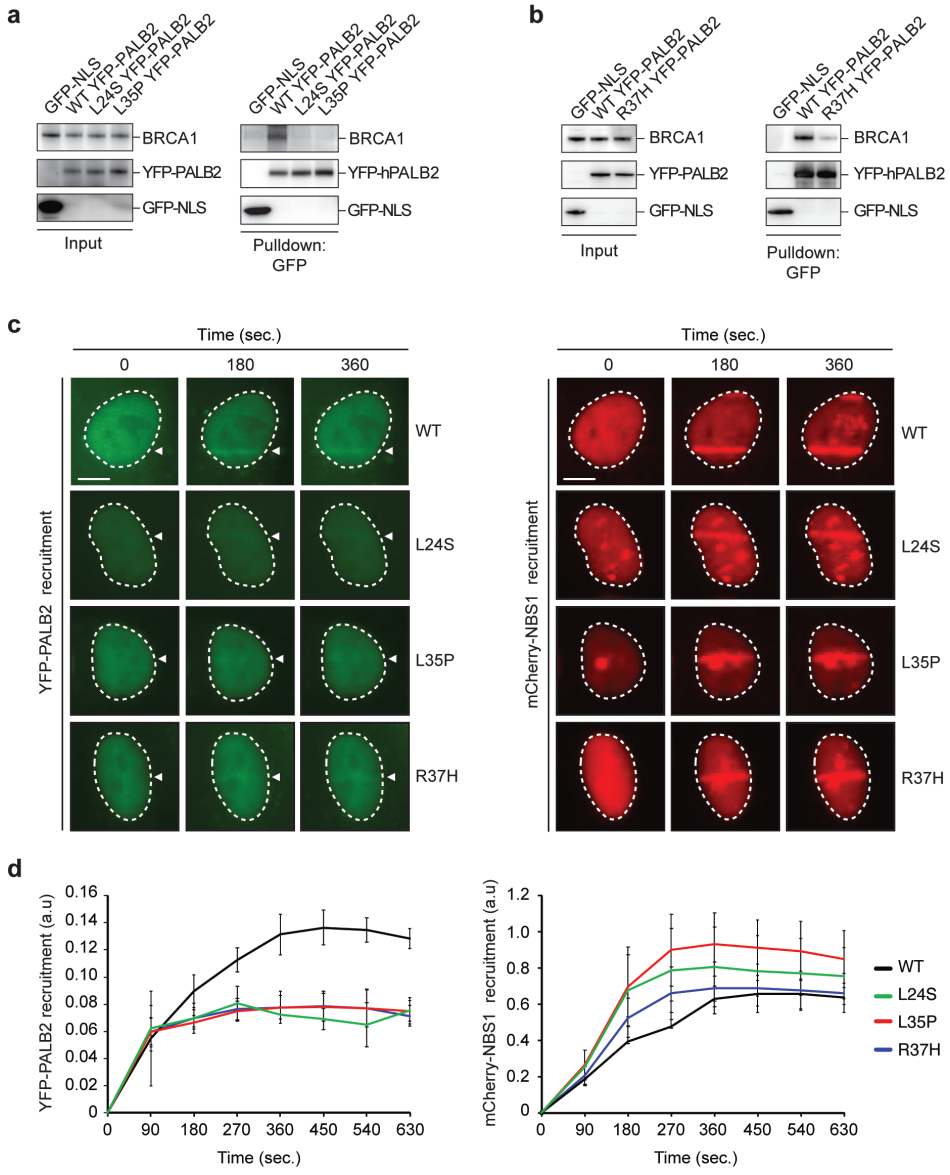


Figure 5. Effect of PALB2 VUS on the BRCA1 interaction and recruitment to DNA damage sites. **a** YPF/GFP pulldowns of the indicated proteins following transient expression in U2OS cells. GFP-NLS and YFP-PALB2-L35P served as negative controls. Western blot analysis was performed using antibodies against GFP and BRCA1. **b** As in **a**, except for p.R37H. **c** Live cell imaging of the recruitment of the indicated YFP-PALB2 proteins to DNA damage tracks generated by laser micro-irradiation in U2OS cells. mCherry-Nbs1, which was co-expressed with the indicated YPF-PALB2 proteins, served as a DNA damage marker. Representative images are shown. White triangles indicate irradiated regions. Scale bars: 5 µm. **d** Quantification of the recruitment of the indicated YPP-PALB2 proteins and mCherry-Nbs1 to DNA damage tracks in cells from. Data represent the mean values (± SEM) from 3 independent experiments. **c**. Source data are provided as a Source Data file.

an impact on the BRCA1-dependent localization of PALB2 to sites of DNA damage. To this end, YFP-tagged PALB2 carrying p.L24S, p.L35P or p.R37H were transiently expressed in U2OS cells and examined for their localization at DNA damage-containing tracks generated by laser micro-irradiation. We found that all three VUS impaired the recruitment of PALB2 to sites of DNA damage (Fig. 5c,d). The effect of these VUS on PALB2's interaction with BRCA1 and localization at sites of DNA damage are highly consistent with the observed HR defect (Figs. 2b and 3a, Supplementary Fig. 6). Taken together, we identified p.L24S and R37H as VUS that impair PALB2's function in HR by abrogating its interaction with BRCA1, and consequently its BRCA1-dependent recruitment to DNA damage sites.

***PALB2* VUS affect G2/M-phase progression after DNA damage**

While PALB2 is essential for HR, two independent genetic screens identified *PALB2* as a critical regulator of the DNA damage-induced G2/M checkpoint response (38,39). Another study demonstrated that *PALB2* plays a role in maintaining a proper G2/M checkpoint response in human cancer cells exposed to ionizing radiation (IR) (40). We therefore addressed if VUS in *PALB2* would affect the DNA damage-induced checkpoint by measuring the mitotic fraction of *Trp53*^{KO} and *Trp53*^{KO}/*Palb2*^{KO} mES cells following exposure to IR. One hour after exposure to 3 or 10 Gy of IR, both *Trp53*^{KO} *and Trp53*^{KO}/*Palb2*^{KO} mES cells showed an almost complete loss of mitotic cells, indicating efficient activation of the G2/M checkpoint in both cell types (Fig. 6a). While at 6 hours after 3 Gy of IR the mitotic fraction of both *Trp53*^{KO} and *Trp53*^{KO}/*Palb2*^{KO} mES cells dramatically increased, we only observed this increase in *Trp53*^{KO}/*Palb2*^{KO} mES after exposure to 10 Gy (Fig. 6a). Thus, *PALB2* is also required for the maintenance of the IR-induced G2/M checkpoint in mES cells.

This prompted us to assess the effect of 19 different *PALB2* variants on G2/M checkpoint maintenance. We expressed these variants, which were selected based on their differential impact on HR (Fig. 2a), in *Trp53*^{KO}/*Palb2*^{KO} mES cells and determined the mitotic fraction 6 hours after exposure to 10 Gy of IR. Importantly, expression of wild-type human *PALB2* rescued the G2/M checkpoint maintenance defect observed in *Trp53*^{KO}/*Palb2*^{KO} mES cells, whereas expressing the empty vector or either of two truncating variants (p.Y551X and p.Y1183X) resulted in a checkpoint defect (Fig. 6b). Two benign variants (p.D134N and p.G998E) and 9 different VUS (p.K18R, p.E42K, p.Y408H, p.P707L, p.D871G, p.L931R, p.1018D, p.Y1046C and p.S1058P) that did not impair HR, also did not impact the maintenance of the IR-induced G2/M checkpoint. In contrast, strong defects in G2/M checkpoint maintenance were observed for 3 VUS (p.L35P, p.L961P and p.G1043D) and the synthetic missense variant p.A1025R that also abrogated HR (Fig. 2b), whereas p.R37H and p.L939W exhibited a moderate effect (Fig. 6b), consistent with their mild impact on HR (Fig.

2b). Accordingly, we found a strong correlation between the impact of *PALB2* variants on HR and G2/M checkpoint maintenance (R^2=*0.8577)* (Fig. 6c). Interestingly, p.L35P and p.A1025R have been shown to abrogate the interaction of PALB2 with BRCA1 (Fig. 5a) (30) and BRCA2 (29), respectively. This indicates that both the interaction with BRCA1 and BRCA2 is crucial for PALB2's function in controlling G2/M-phase progression following DNA damage, which is in accordance with observations in human cancer cells (40).

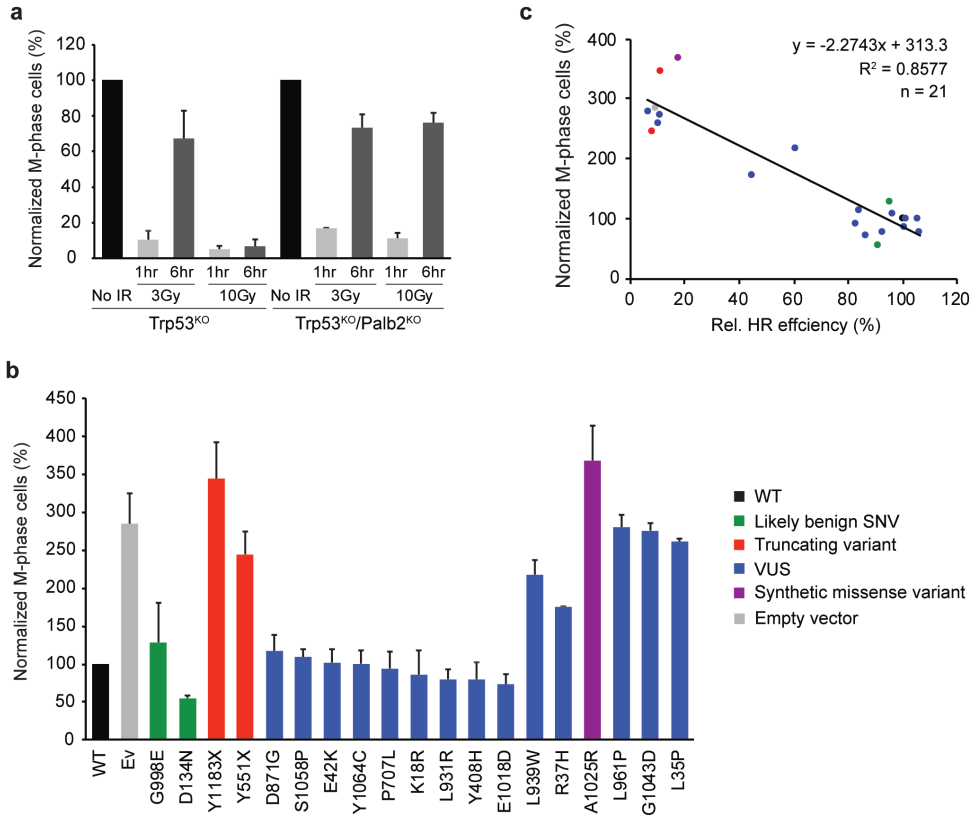


Figure 6. Effect of *PALB2* variants on the DNA damage-induced G2/M checkpoint. a *Trp53*^{KO}/*Palb2*^{KO} mES cells were irradiated with 3 or 10 Gy of IR and collected at the indicated time points after radiation exposure to assess the mitotic index by phospho-histone H3 (Ser10) staining and flowcytometry analysis. Data represent the mean percentage of mitotic cells (± SEM) relative to the unirradiated cells, which was set to 100%, from 2 independent experiments. **b** *Trp53*^{KO}/*Palb2*^{KO} mES cells expressing the indicated *PALB2* variants were irradiated with 10 Gy of IR and collected 6 hours after radiation exposure to assess the mitotic index by phospho-histone H3 (Ser10) staining and flowcytometry analysis. For each variant, the mean percentage of mitotic cells (± SEM) from 2 independent experiments is shown relative to unirradiated cells, except for p.L939W and p.G998E for which data from three independent experiments is presented. **c** Scatter plot showing the correlation between the HR efficiencies and the mitotic index after IR as measured in Fig. 2b and **b**, respectively.

Variants/conditions are categorized by color as in **b**. The trendline indicates the negative correlation between the HR efficiency and mitotic index after IR, revealing a strong positive correlation between the impact of *PALB2* variants on HR and G2/M checkpoint maintenance. Source data are provided as a Source Data file.

Functional analysis of *PALB2* VUS in human cell-based assays

To validate results from our mES cell-based assays, we selected 5 LOF VUS located in the WD40 domain of PALB2 (p.W912G, p.G937R, p.L947S, p.L961P and p.G1043D) and tested their effect on HR in human cell-based assays. To this end, we first employed the CRISPR-LMNA HR assay, which monitors the integration of mRuby, into the Lamin A/C locus (LMNA) by CRISPR/Cas9-mediated HR (Supplementary Fig. 11a,b) (41). Following siRNA-mediated knockdown of *PALB2* in U2OS cells, plasmids encoding the mRuby2-LMNA donor, Cas9 and a *LMNA* gRNA, and siRNA-resistant YFP-PALB2 with or without VUS, were co-transfected into these cells (Supplementary Fig. 11c). Four PALB2 VUS (p.W912G, p.G937R, p.L961P and p.G1043D) showed a dramatic impact on the HR-mediated integration of mRuby (Fig. 7a). One VUS (p.L947S), had a moderate effect, although this is likely explained by the slightly higher transient expression of this variant (Supplementary Fig. 11c). We then assessed whether these VUS would affect PARPi sensitivity. To this end, siRNA-resistant YFP-PALB2 constructs carrying these VUS were expressed in PALB2-depleted HeLa cells (Supplementary Fig. 11d). Four PALB2 VUS (p.W912G, p.G937R, p.L961P and p.G1043D), showed a dramatic increase in PAPRi sensitivity, while 1 VUS (p.L947S) had a more moderate effect, consistent with findings from the CRISPR-LMNA HR assay (Fig. 7b). Altogether, these results corroborate our findings from the DR-GFP and PARPi sensitivity assays in mES cells (Figs. 2a and 3a, Supplementary Fig. 6).

Finally, PALB2 drives HR by promoting the accumulation of RAD51 at DSB sites. To further assess the impact of the 5 selected VUS on PALB2, we examined whether they affected the accumulation of RAD51 at IR-induced DSBs by measuring the formation RAD51 foci. HeLa cells were treated with siRNAs against endogenous *PALB2* and complemented by transient expression of siRNA-resistant YFP-PALB2, with or without VUS. Following exposure to IR, the average number of RAD51 foci was scored in cyclin-A- and YFP-PALB2-expressing S-phase cells (Fig. 7c,d). While 3 VUS (p.W912G, p.L961P and p.G1043D) had a dramatic impact on the percentage of cells showing RAD51 foci, 2 VUS (p.G937R and p.L947S) displayed a more minor effect. However, for these 2 VUS, we found that the intensity of RAD51 foci was dramatically reduced (Fig. 7e). As all 5 variants displayed problems in protein stability in mES cells, we believe that the defects observed in RAD51 foci formation and/or intensity mostly stem from impaired RAD51 recruitment due to reduced PALB2 protein levels. Overall, our findings in human cell-based assays solidify those obtained in the mES cell-based assays,

indicating that our system in mES cells is robust and suited for semi-high throughput functional analysis of VUS in human *PALB2*.

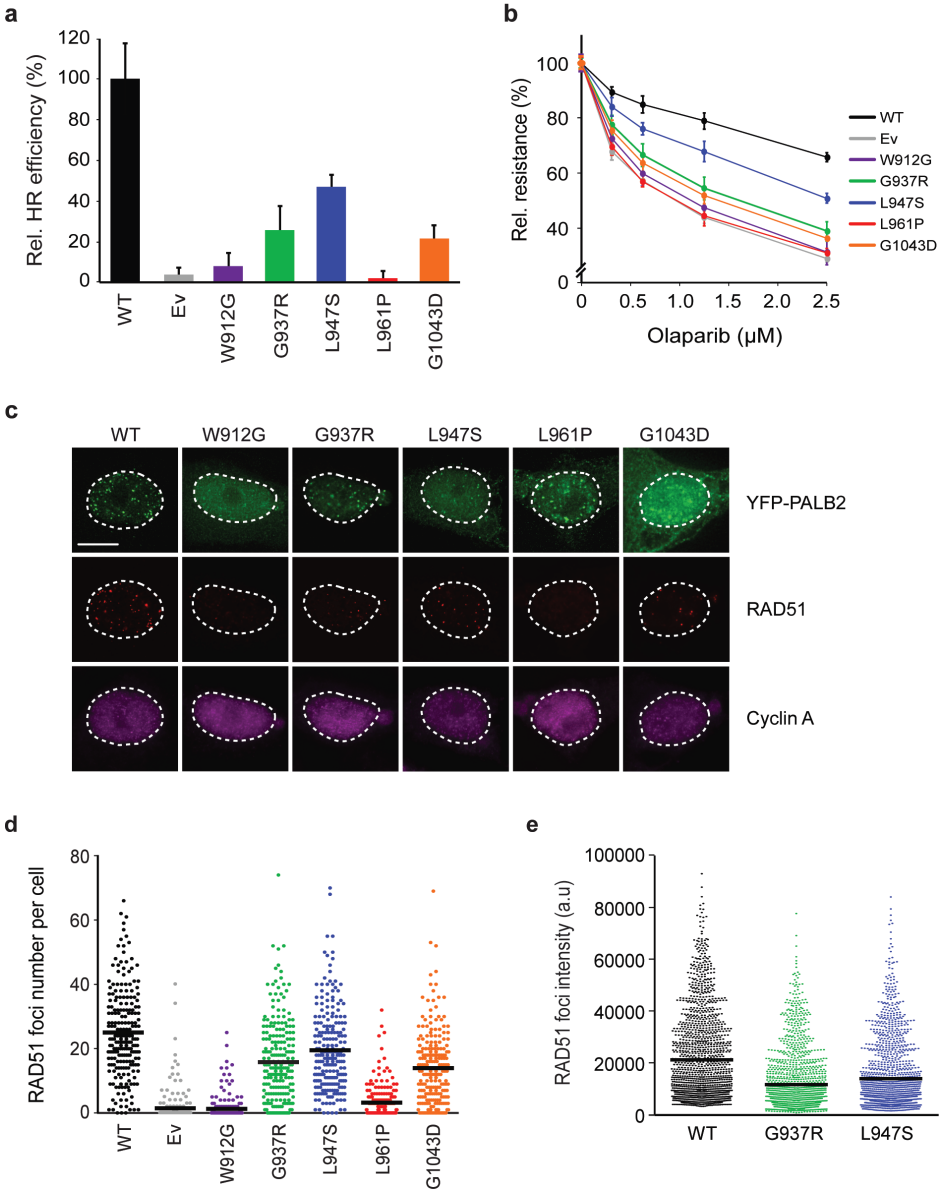


Figure 7. Functional analysis of damaging *PALB2* variants in human cells. a CRISPR-LMNA HDR assay in siRNA-treated U2OS *PALB2* knockdown cells expressing siRNA-resistant human *PALB2* cDNA with the indicated variants (or an empty vector control, Ev). Data represent the mean percentage (± SD) of mRuby2-positive cells among the YFP-positive cells from 3 independent experiments (n >300 YFP-positive cells per condition) relative to wild type (WT), which was set to 100%. **b** PARP inhibitor

(PARPi) sensitivity assay using siRNA-treated HeLa *PALB2* knockdown cells expressing siRNA-resistant human *PALB2* cDNA with the indicated variants (or an empty vector control, Ev). Survival curves were determined after 72 hours of PARPi treatment. Data represent the mean percentage of viability relative to untreated cells (± SD), which was set to 100%, of 3 independent experiments, each performed in triplicate. **c** Representative images of RAD51 foci 4 hours after 2 Gy of ionizing radiation in siRNA-treated HeLa *PALB2* knockdown cells expressing siRNA-resistant human *PALB2* cDNA with the indicated variants (or an empty vector control, Ev). Scale bar: 5 µm. **d** Quantification of the results from **c**. Scatter dot plot shows the number of RAD51 foci in cyclin A-positive cells expressing the indicated variant, with the horizontal lines designating the mean values (± SD) of 3 independent experiments (n>200 cells per condition). **e** Quantification of the results from **c**. Scatter dot plot shows the intensity of RAD51 foci in cyclin A-positive cells expressing the indicated variant, with the horizontal lines designating the mean values (± SD) of 3 independent experiments (n>500 cells per condition). Source data are provided as a Source Data file.

DISCUSION

To address the impact of *PALB2* VUS on protein function, we developed a mES cell-based system that allows a rapid and robust functional classification of genetic variants in human *PALB2*. Out of the 49 PALB2 missense variants tested in this study (Supplementary Data 1), we identified 15 variants (p.L24S, p.Y28C, p.L35P, p.W912G, p.G937R, p.I944N, p.L947S, p.L961P, p.L972Q, p.A1025R, p.T1030I, p.I1037T, p.G1043D, p.L1070P, p.L1172P) as damaging, reducing HR by >60%. For three variants that have been described previously (p.Y28C, p.L35P and p.T1030I), our results are highly consistent with published data, showing that these variants which confer increased risk for breast cancer, strongly impact HR (30,31). Furthermore, we observed a strong positive correlation between the DR-GFP and PARPi or cisplatin sensitivity assays, suggesting that carriers of the identified damaging VUS may benefit from PARPi- or cisplatin-based treatment. Lastly, our data from the human cell-based assays further verify the results from the mES-based cell assays, indicating that our system in mES cells is well-suited for the rapid, semi-high throughput functional analysis of VUS in human PALB2.

In addition to p.Y28C and p.L35P, which have both been reported to impair the interaction with BRCA1 (30), p.R37H also resides in the N-terminal coiled-coil domain and impairs the HR activity by more than 55% in our DR-GFP assay (Fig. 2b). In contrast to an earlier report showing that p.R37H did not affect the interaction with BRCA1 (30), we found that this variant impaired the PALB2-BRCA1 interaction and the BRCA1-dependent recruitment of PALB2 to sites of DNA damage, which is highly consistent with its moderate impact on HR. Our results on the identified p.L24S variant, are in line with a previous study in which the CC6 PALB2 variant, for which the amino acids LKK at position 24-26 are changed to AAA, impairs the interaction with BRCA1 and consequently abrogates HR (42). Thus, our HR and protein-protein association studies for both p.L24S and p.L35P further underline the importance of the BRCA1-PALB2 interaction for efficient HR and likely tumor suppression.

The C-terminal WD40 domain of PALB2 is an important regulatory platform that mediates interactions with several important HR pathway components, such as BRCA2 and RAD51. Crystal structure studies of the WD40 domain showed that it forms a seven-bladed β-propeller-like structure of which correct folding is crucial for PALB2 function (29). As such, it is likely that variants in this region are prone to interfere with the structure and/or biochemical properties of this domain. For example, although it has been reported that p.W1038X exposes a nuclear export signal leading to cytoplasmic localization (43), we see in our assays that the expression levels of this variant are dramatically reduced compared to wild-type PALB2 (Fig. 4a), probably due to instability/misfolding and rapid degradation in the cytoplasm. Indeed, we see similar effects for three other truncating variants (p.Q899X, p.P1009Lfs, p.Y1183X), which includes p.Y1183X that lacks only the last 3 amino acids. Consistent with the WD40 domain being prone to 'destabilizing' variants, we identified 11 damaging VUS in the WD40 domain that exhibited strongly reduced PALB2 protein levels, and consequently strongly reduced HR (~60-95%). Importantly, 5 of these 11 VUS are *bona fide* null variants that abrogate the HR activity to the same extent as the *PALB2* truncating variants. These results indicate that that the WD40 domain is a 'hotspot' for deleterious LOF variants that affect protein stability. Consistently, a recent study on PTEN, showed that 64% of the pathogenic missense variants reduce its expression level (44). This suggests that protein instability due to LOF variants in tumor suppressor genes, including *PALB2,* constitutes a mechanism of pathogenicity.

Several studies have implicated *BRCA1*, *BRCA2* and *PALB2* in DNA-damage-induced checkpoint control (38-40). Accordingly, we found that G2/M checkpoint maintenance after IR is compromised in $Trp53^{KO}/Palb2^{KO}$ mES cells, an effect that could be rescued by expressing wild-type human *PALB2*. Interestingly, *PALB2* variants that show LOF in HR, were unable to maintain an efficient G2/M checkpoint response. Both p.L35P and p.A1025R, which are unable to interact with BRCA1 and BRCA2, respectively, were among these variants, suggesting that these interactions are key to PALB2's checkpoint function. Moreover, we infer that the observed defects in G2/M checkpoint maintenance could stem from defective HR. In line with such a scenario, an inverse correlation has been observed between HR activity and POLQ-mediated DSB repair (45). This indicates that POLQ-mediated DSB repair may act as a compensatory pathway for PALB2-dependent HR that potentially affects G2/M checkpoint maintenance in response to DNA breaks.

Although our functional assays may aid in the classification of rare *PALB2* VUS, a major challenge will be to translate effects on PALB2 protein function into estimates for cancer risk. Whereas the truncating *PALB2* variants have been associated with an odds ratio of 7.46 (5), the p.L939W variant has been associated with an odds ratio of 1.05 (46). This would suggest that a decrease of 40% in HR in our DR-GFP assay, as shown for the p.L939W variant (Fig. 2b), would barely increase the risk for breast cancer. It will therefore be interesting to see

whether the extent to which variants affect HR is proportional to increased cancer risk and at which level of HR deficiency, cancer risk significantly increases. Finally, it will be important to examine whether *PALB2* VUS, either in coding or non-coding sequences, affect *PALB2* splicing. For all missense variants presented in this study *in silico* splice site prediction analysis was performed using five different algorithms (Splice Site Finder-like, MaxEntScan, GeneSplicer, NNSplice, Human Splicing Finder) in Alamut (http://www.interactive-biosoftware.com/). For all VUS an effect on RNA splicing was unlikely, with the exception of c.53A>G (p.K18R) for which NNSplice predicted the introduction of a new weak acceptor recognition site in exon 2. Complementation of our *Trp53*^{KO}/*Palb2*^{KO} cells with a bacterial artificial chromosome (BAC) containing the full length human *PALB2* gene, as has been previously described for *BRCA1* and *BRCA2* (14,47,48), would also enable us to address the functional effect of splice variants in *PALB2*. Ultimately, the results from functional assays for VUS can be incorporated into multifactorial risk models to allow for better clinical classification in the future. Indeed, multiple pieces of evidence, in addition to functional assay results, will be required to enable clinical classification of VUS.

MATERIALS AND METHODS

Cell lines and culture conditions

129/Ola E14 IB10 mES cells (49) were cultured on gelatin-coated dishes in 50% BRL/50% complete medium (13) with 0.1 mM beta-mercaptoethanol (Merck) and 10^3 Units/ml ESGRO LIF (Millipore). STR genotyped U2OS and HeLa human cells (ATCC) were maintained, respectively, in McCoy's 5A (Wisent) and DMEM (ThermoFischer) supplemented with 10% Fetal Bovine Serum (FBS) and 1% penicillin and streptomycin.

Generation of *Trp53*^{KO}/*Palb2*^{KO} mES cells with DR-GFP and RMCE

Trp53^{KO}/*Palb2*^{KO} mES cells carrying the DR-GFP reporter and RMCE system were generated as follows. 75 µg of the plasmids carrying *Pim1*:DR-GFP (p59X DRGFP) (50) or the *Rosa26:* RMCE acceptor cassette (pTT5-Puro) (TaconicArtemis GmbH) were linearized with *Xho*I and *Pvu*I respectively. *Pim1*:DR-GFP was transfected into mES cells (49) using Lipofectamine 2000 (Invitrogen). Integration of DR-GFP at *Pim1* was verified using PCR and Southern blot analysis. Similarly, the RMCE acceptor cassette was integrated at *Rosa26* in cells carrying DR-GFP. Integration of the RMCE acceptor cassette at *Rosa26* was verified using PCR and Southern blot analysis. *Trp53*^{KO} cells were generated by transfection of 1 µg of pSpCas9(BB)-2A-GFP (*pX458*) (51), which encodes a gRNA that targets exon 1 (5'-CGAGCTCCCTCTGAGCCAGG-3'), into mES cells carrying DR-GFP and the RMCE acceptor

cassette. GFP-positive cells were FACS-sorted and seeded. Individual clones were examined by TIDE and western blot analysis for loss of p53 expression. Similarly, the *Palb2*^{KO} was generated in *Trp53*^{KO} mES cells carrying DR-GFP and RMCE acceptor cassette using a gRNA that targets exon 4 (5'-GGGGACAACAAAGACGCCGT-3'), and verified by TIDE and western blot analysis for loss of Palb2 expression.

Cloning and site-directed mutagenesis of human *PALB2* cDNA

pBudCE4.1 (ThermoFisher, V53220), which contains an EF1α promotor, was modified by cloning two different oligonucleotides with *Pac*I restriction sites into the *Nhe*I (5'-CTAGGACTTAATTAAGTCGATCGCCGG-3') and *Bgl*II restriction sites (5'-GATCTCTTAATTAAGACTG-3'), respectively. Human Flag-tagged *PALB2* cDNA was obtained from pcDNA3-Flag-PALB2 and subcloned into pBudCE4.1-*Pac*I using the *Acc*65I and *Xho*I restriction sites. An Ef1α-*PALB2*-containing fragment from pBudCE4.1-*Pac*I-PALB2 was then cloned into the RMCE vector pRNA 251-MCS RMCE) (TaconicArtemis GmbH) using the *Pac*I restriction sites in both vectors. *PALB2* variants were introduced by site-directed mutagenesis using the Quick-Change Lightning protocol (Agilent Technologies). Constructs were verified by Sanger sequencing and used for mES cell-based assays. For human cell-based assays, siRNA-resistant pEYFP-PALB2 construct was generated by site-directed mutagenesis using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) as per the manufacturer's protocol with the following primers: forward primer - 5'-GATCTTATTGTTCTACCAGGAAAATC-3' and reverse primer - 5'-TTCCTCTAAGTCCTCCATTTCTG-3'. PALB2 variants were introduced in the siRNA-resistant pEYFP-C1-PALB2 plasmid by site-directed mutagenesis using the same kit. All primers used for site-directed mutagenesis are listed in Supplementary Data 1.

Karyotyping

mES cells (50% confluency) were incubated with 0.05 µg/ml colcemid (Gibco) for 2.5 hours. After trypsinization, 2.5 ml of 0.4% Na-citrate, 0.4% KCL (1:1) was added in a dropwise manner. Cells were centrifuged at 120 g after which the supernatant was aspirated and 2.5 ml fixative consisting of methanol and acidic acid (4:1) was added while slowly vortexing. This step was repeated twice. Using ultrathin pipet tips, a small number of cells was dropped onto a cleaned microscopy slide (VWR, 631-1551) and left to air-dry. DAPI was used for visualizing the chromosomes, which were counted using a Zeiss microscope Imager M2 (63x) and ZEN 2012 microscopy software.

Western blot analysis

Expression of endogenous mouse PALB2 and human PALB2 in mES was monitored by

protein extraction and western blot. Briefly, samples were generated by taking up ~1.5x10^6 cells in 75 µl Laemmli buffer and boiling them at 95°C for 5 minutes. Samples were incubated with 1.5 µl benzonase (Merck Millipore 70746-3, 25 U/µl) for 10 minutes at room temperature and then loaded for gel electrophoresis and immunoblotting. Primary antibodies used were a rabbit polyclonal antibody against the N-terminus of human PALB2 (1:1000, kindly provided by Cell Signaling Technology prior to commercialization), a homemade rabbit antibody against the N-terminus of mouse PALB2 (42) (NB3 anti-mPalb2, 1:2000, kind gift form Bing Xia) and a mouse monoclonal antibody against alpha tubulin (1:10000, Sigma, T6199 clone DM1A). Peroxidase-AffiniPure Goat Anti-Rabbit secondary antibody (Jackson laboratories) and SuperSignal West Femto Maximum Sensitivity Substrate (ThermoFisher) were used for development of blots on the Amersham Imager 600 (GE Healthcare Life Sciences).
Western blotting was performed by separating U2OS and HeLa protein extracts on 12% SDS-PAGE gels at 100V and transferred to nitrocellulose membrane during 1.5 hour at 100V. Membranes were blocked for 1 hour in 5% milk in Tris-buffered saline (TBS)-Tween. Primary antibodies applied were mouse monoclonal anti-GFP (1:1000, Roche, #11814460001**),** anti-alpha tubulin (1:200000, Abcam, #ab7291) and a home-made rabbit polyclonal antibody against human PALB2 (1:5000). Horseradish peroxidase-conjugated anti-mouse IgG (1:10000, Jackson ImmunoResearch) were used as secondary antibodies.

RT-qPCR analysis

RNA was isolated from mES cells on 6-well plates using Trizol (ThermoFisher, 15596026) as per the manufacturer's protocol. For each condition, 3 µg RNA was treated with RQ1 RNAse-free DNAse (Promega, M6101) and cDNA was synthesized from 0.2 µg DNAse-treated RNA using hexamer primers (ThermoFisher, N8080127) and AMV Reverse Transcriptase (ThermoFisher, 12328019) as per the manufacturer's protocols. RT-qPCRs were carried out using GoTaq qPCR Master mix (Promega, A6002), a CFX384 Real-Time System (Bio-Rad) and the following qPCR primers directed at the human *PALB2* cDNA or the mouse control gene *Pim1*; PALB2-Fw - 5'-GATTACAAGGATGACGACGATAAGATGGAC-3', PALB2-Rv - 5'-CCTTTTCAAGAATGCTAATTTCTCCTTTAACTTTTCC-3'. Pim1-exon4-Fw - 5'-GCGGCGAAATCAAACTCATCGAC-3' and Mouse Pim1-exon5-Rv - 5'-GTAGCGATGGTAGCGAATCCACTCTGG-3'.

HR Reporter Assays

2x10^6 *Trp53*^{KO}*/Palb2*^{KO} mES cells carrying the DR-GFP reporter and RMCE system were subjected to RMCE by co-transfecting 1 µg FlpO expression vector (pCAGGs-FlpO-IRES-puro) (19) with 1 µg RMCE exchange vector. Neomycin-resistant cells from ~500 resistant clones were pooled and expanded for DR-GFP reporter assays. 1 µg of a plasmid for co-

expression of I-SceI and mCherry (pCMV-Red-Isce, kind gift from Jos Jonkers) was transfected in $1x10^6$ cells using Lipofectamine 2000 (ThermoFisher) (13). A co-transfection of 1 µg pCAGGs (53) with 0.05 µg of an mCherry expression vector was included as control. Two days after transfection, mCherry/GFP double-positive cells were scored using a Novocyte Flow Cytometer (ACEA Biosciences, Inc.).

For the CRISPR-LMNA HR assay (43), U2OS cells were seeded in 6-well plates at $2x10^5$ cells per well. Knockdown of *PALB2* was performed 6 hours later with 50 nM siRNA against PALB2 (5'-CUUAGAAGAGGACCUUAUU-3'; Dharmacon) using Lipofectamine RNAiMAX (Invitrogen). Twenty-four hours post-transfection, $1.5x10^6$ cells were pelleted for each condition and resuspended in 100 µl complete nucleofector solution (SE Cell Line 4D-Nucleofector™ X Kit, Lonza) to which 1µg of pCR2.1-mRuby2LMNAdonor, 1 µg pX330-LMNAgRNA, 1 µg peYFP-C1 or the indicated siRNA-resistant YFP-PALB2 construct, and 150 pmol siRNA was added. Once transferred to a 100 µl Lonza certified cuvette, cells were transfected using the 4D-Nucleofector X-unit, program CM-104 and transferred to a 10 cm dish. After 48 hours, cells were trypsinized and plated onto glass coverslips. Cells were fixed with 4% paraformaldehyde and analyzed for mRuby2 and YFP expression on a Leica CTR 6000 inverted microscope using a 63x/1.40 oil immersion objective 72 hours post-nucleofection.

PARPi and cisplatin sensitivity assays

For proliferation-based PARPi and cisplatin sensitivity assays, mES cells were seeded in triplicate at 10.000 cells per well of a 96-well plate. The next day, cells were treated with PARP inhibitor Olaparib (Selleckchem, S1060) or cisplatin (Accord Healthcare, 15683354) for two days, after which the medium was refreshed and cells were cultured for one more day. Viable cells were subsequently counted using the Novocyte Flow Cytometer (ACEA Biosciences, Inc.).

For clonogenic PARPi survival assays, mES cells were seeded on p60 plates at the following densities: 250 cells without PARPi, 400 cells for functional variants with 1 or 5 nM PARPi, and 3000 cells for damaging variants with 1 or 5 nM PARPi. Cells were treated for 7-9 days allowing the visible formation of surviving colonies which were counted following methylene blue staining (2.5 gr/L in 5% ethanol). HeLa cells were seeded at 240000 cells per well of a 6-well plate before being transfected 6 hours later with 50 nM control or *PALB2* siRNA using Lipofectamine RNAiMAX (Invitrogen). The next day, cells were complemented with 0.8 µg of EYFP-PALB2 plasmid DNA using Lipofectamine 2000 (Invitrogen) for 24 hours and then seeded in triplicates into a Corning 3603 black-sided clear bottom 96-well microplate at a density of 3000 cells per well. After 3 days of treatment with Olaparib (Selleckchem, S1060), nuclei were stained with Hoechst 33342 (Invitrogen) at 10 µg/ml in media for 45 minutes at

37°C. Images of entire wells were acquired at 4x with a Cytation 5 Cell Imaging Multi-Mode Reader followed by quantification of Hoechst-stained nuclei with the Gen5 Data Analysis Software v3.03 (BioTek Instruments).

Cell cycle analysis and G2/M checkpoint assays

For cell cycle profile analysis cells were fixed in 70% ethanol. After 15 minutes incubation on ice, cells were pelleted and resuspended in 500 µl PBS containing 50 µg/ml propidium iodide (PI) (ThermoFisher, P1304MP), 0.1 mg/ml RNase A and 0.05% Triton X-100, followed by 40 minutes incubation at 37°C. Cells were then washed with PBS and analyzed using the Novocyte Flow Cytometer (ACEA Biosciences, Inc.).

For G2/M checkpoint assays, $1x10^6$ mES cells were seeded on p60 dishes one day before exposure to 3 or 10 Gy of IR. One or 6 hours later, cells were fixed as described for cell cycle profile analysis and incubated overnight at -20°C. Fixed cells were then permeabilized for 15 minutes on ice using 0.25% Triton X-100 in PBS, after which mitotic cells were stained in 100 µl PBS with 1 µl anti-phospho-H3 Ser10 (1µg/µl, Sigma-Aldrich, 06-570) for 3 hours at room temperature. Alexa-488 goat α-rabbit (1:100 in 100µl PBS; ThermoFisher, 11034) was used as a secondary antibody. Cells were analyzed using the Novocyte Flow Cytometer (ACEA Biosciences, Inc.).

Pulldown assays

20 µg YFP-PALB2 plasmid DNA (previously described (54)) was transfected into ~$10x10^6$ U2OS cells on a 15 cm dish using Lipofectamine 2000. The next day cells were trypsinized, washed with cold PBS, and transferred to LoBind Eppendorf tubes. Cells were then lysed in 1 ml EBC buffer (50 mM Tris pH 7.3, 150 mM NaCl, 0.5% NP-40, 2.5 mM $MgCl_2$), containing 1 tablet protease inhibitor (Roche) per 10 ml buffer. 500 Units benzonase was then added to each condition and cells were incubated for 60 minutes at 4 °C on a rotating wheel. The lysate was subsequently centrifuged for 10 minutes at 18400 g at 4°C. The supernatant was then added to 25 µl of pre-washed GFP-trap beads (ChromoTek) in LoBind Eppendorf tubes and incubated for 1.5 hours at 4 °C on a rotating wheel. The beads were washed 5-6 times with EBC buffer with spinning steps of 1 minute at 3380 g at 4°C. Beads were eventually resuspended in 25 µl Laemmli buffer after which about half of each sample was analyzed by western blot analysis using a homemade rabbit antibody against human BRCA1 (55) (1:1000, kind gift form Dan Durocher).

Laser micro-irradiation and PALB2 recruitment

U2OS cells were grown on 18-mm coverslips and sensitized with 10 μM 5′ -bromo-2-deoxyuridine (BrdU) for 24 h before micro-irradiation. Cells were co-transfected with 1 μg pYFP-PALB2, with or without a variant, and 0.5 μg mCherry-NBS1 expression vector using lipofectamine 2000 (Invitrogen). For micro-irradiation, cells were placed in a live-cell imaging chamber set to 37 °C in CO2-independent Leibovitz's L15 medium supplemented with 10% FCS and penicillin–streptomycin (Invitrogen). Live cell imaging and micro-irradiation experiments were carried out with a Zeiss Axio Observer microscope driven by ZEN software using a 63x/1.4 oil immersion objective coupled to a 355 nm pulsed DPSS UV-laser (Rapp OptoElectronic). To monitor the recruitment of YFP-PALB2 to laser-induced DNA damage sites, cells were imaged before and after laser irradiation at 90 seconds time intervals over a period of 10.5 minutes. The fluorescence intensity of YFP-PALB2 and mCherry-NBS1 at DNA damage sites relative to that in an unirradiated region of the nucleus was quantified and plotted over time. Kinetic curves were obtained by averaging the relative fluorescence intensity of cells displaying positive recruitment (n>30 cells per condition).

RAD51 foci analysis

HeLa cells were seeded on glass coverslips in 6-well plates at 225.000 cells per well. Knockdown of *PALB2* was performed 18 hours later with 50 nM PALB2 siRNA using Lipofectamine RNAiMAX (Invitrogen). After 5 hours, cells were subjected to a double thymidine block. Briefly, cells were treated with 2 mM thymidine for 18 hours and released into fresh medium for 9 hours. During the release time, 0.8 μg YFP-PALB2 plasmid DNA (with or without variant) was transfected into the cells using Lipofectamine 2000. Cells were then treated with 2 mM thymidine for 17 hours and protected from light from this point on. After 2 hours of release from the second block, cells were irradiated with 2 Gy and processed for immunofluorescence 4 hours post-irradiation. Unless otherwise stated, all immunofluorescence dilutions were prepared in PBS and incubations performed at room temperature with intervening washes in PBS. Cell fixation was carried out by incubation with 4% paraformaldehyde for 10 minutes followed by 100% ice-cold methanol for 5 minutes at -20 °C. This was succeeded by permeabilization in 0.2% Triton X-100 for 5 minutes and a quenching step using 0.1% sodium borohydride for 5 minutes. After blocking for 1 hour in a solution containing 10% goat serum and 1% BSA, cells were incubated for 1 hour with primary antibodies anti-RAD51 (1 :7000, B-bridge International, #70-001) and anti-cyclin A (1 :400, BD Biosciences, # 611268) diluted in 1% BSA. Secondary antibodies Alexa Fluor 568 goat anti-rabbit (Invitrogen, #A-11011) and Alexa Fluor 647 goat anti-mouse (Invitrogen, #A-21235) were diluted 1 :1000 in 1% BSA and applied for 1 hour. Nuclei were stained for 10 minutes with 1 μg/mL DAPI prior to mounting onto slides with 90% glycerol containing 1 mg/ml paraphenylenediamine anti-fade reagent. Z-

stack images were acquired at 63X magnification on a Leica CTR 6000 microscope, then deconvolved and analyzed for RAD51 foci. The number and intensity of RAD51 foci in cyclin A-positive cells expressing the indicated YFP-PALB2 constructs were scored using automatic spot counting in Volocity software v6.0.1 (Perkin-Elmer Improvision).

DATA AVAILABILITY

All data generated or analysed during this study are included in this published article (and its supplementary information files).

AUTHOR CONTRIBUTIONS

R.B. cloned *PALB2* cDNA in the RMCE exchange construct, generated *PALB2* variants using site-directed mutagenesis, generated *Trp53*^{KO}/*Palb2*^{KO} mES cells harboring the DR-GFP reporter and RMCE acceptor cassette, and performed DR-GFP, PARPi, Cisplatin, G2/M checkpoint, and pull-down assays, as well as Southern and western blot analysis, and PCR and DNA sequencing analysis in mES cells. Amélie Rodrigue generated YFP-PALB2 variants using site-directed mutagenesis and performed CRISPR-LMNA HR, PARPi sensitivity and RAD51 foci assays, as well as western blot analysis in human cells. C.S. assisted with the PARPi assays in mES cells. W.W. performed G2/M checkpoint assays. B.V. performed *in silico* modeling of *PALB2* variants. M.S. generated PALB2 p.A1025R using site-directed mutagenesis and performed DR-GFP and PARPi proliferation assays for this variant. M.R. studied YFP-PALB2 recruitment to laser-induced DNA damage. N.C. performed cell cycle profile analysis in mES cells. M.V. helped with gathering *PALB2* variants from databases and literature and assisted in the *in silico* analysis. F.C. provided p.L24S, p.I944N and p.L1070P variants. J.S. and J.-Y.M. supervised experiments performed in human cells. M.V., P.D. and H.v.A. conceived the project. H.v.A. supervised the project. R.B and H.v.A. wrote the paper.

COMPETING INTERESTS

The authors declare no competing interests.

ACKNOWLEDGEMENTS

The authors would like to thank Jos Jonkers and Peter Bouwman for providing the pTT5-Puro (RMCE acceptor cassette), pRNA-251-MCS (RMCE exchange vector) and pCMV-Red-I-Sce constructs, as well as Dan Durocher and Bing Xia for sharing BRCA1 and PALB2 antibodies,

respectively. We are grateful to Maria Jasin and Francis Stewart for sharing the DR-GFP reporter and FlpO constructs, respectively. We thank Patrick van Vliet and Richard Lemmers for help with the Southern blot analysis, and Cell Signaling and Dan Durocher for providing antibodies directed against PALB2 and BRCA1, respectively. Finally, we would like to thank Jamie Allen and the BRIDGES consortium for providing *PALB2* variants. This work was financially supported by the Government of Canada through Genome Canada and the Canadian Institutes of Health Research, the Ministère de l'Économie, de la Science et de l'Innovation du Québec through Genome Québec and the Quebec Breast Cancer Foundation (J.S. and J.-Y.M.), as well as by grants from the Ministère de l'Économie, de la Science et de l'Innovation du Québec through the PSR-SIIRI-949 program (J.S. and J.-Y.M), the CIHR (Foundation grant to J.-Y.M), European Union (BRIDGES grant to P.D., M.V. and H.v.A.) and the Dutch Cancer Society (P.D. and H.v.A.).

REFERENCES

1. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL*, et al.* Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. American journal of human genetics **2003**;72(5):1117-30 doi 10.1086/375033.
2. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A*, et al.* PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nature genetics **2007**;39(2):165-7 doi 10.1038/ng1959.
3. Reid S, Schindler D, Hanenberg H, Barker K, Hanks S, Kalb R*, et al.* Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. Nature genetics **2007**;39(2):162-4 doi 10.1038/ng1947.
4. Antoniou AC, Casadei S, Heikkinen T, Barrowdale D, Pylkas K, Roberts J*, et al.* Breast-cancer risk in families with mutations in PALB2. The New England journal of medicine **2014**;371(6):497-506 doi 10.1056/NEJMoa1400382.
5. Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J*, et al.* Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. JAMA Oncol **2017**;3(9):1190-6 doi 10.1001/jamaoncol.2017.0424.
6. Zhang F, Fan Q, Ren K, Andreassen PR. PALB2 functionally connects the breast cancer susceptibility proteins BRCA1 and BRCA2. Molecular cancer research : MCR **2009**;7(7):1110-8 doi 10.1158/1541-7786.MCR-09-0123.
7. Zhang F, Ma J, Wu J, Ye L, Cai H, Xia B*, et al.* PALB2 links BRCA1 and BRCA2 in the DNA-damage response. Curr Biol **2009**;19(6):524-9 doi 10.1016/j.cub.2009.02.018.
8. Xia B, Sheng Q, Nakanishi K, Ohashi A, Wu J, Christ N*, et al.* Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. Molecular cell **2006**;22(6):719-29 doi 10.1016/j.molcel.2006.05.022.
9. Sy SM, Huen MS, Chen J. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. Proceedings of the National Academy of Sciences of the United States of America **2009**;106(17):7155-60 doi 10.1073/pnas.0811159106.
10. Polak P, Kim J, Braunstein LZ, Karlic R, Haradhavala NJ, Tiao G*, et al.* A mutational signature reveals alterations underlying deficient homologous recombination repair in breast cancer. Nature genetics **2017**;49(10):1476-86 doi 10.1038/ng.3934.
11. Farmer H, McCabe N, Lord CJ, Tutt AN, Johnson DA, Richardson TB*, et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. Nature **2005**;434(7035):917-21 doi 10.1038/nature03445.
12. Buisson R, Dion-Cote AM, Coulombe Y, Launay H, Cai H, Stasiak AZ*, et al.* Cooperation of breast cancer proteins PALB2 and piccolo BRCA2 in stimulating homologous recombination. Nature structural & molecular biology **2010**;17(10):1247-54 doi 10.1038/nsmb.1915.

13. Bouwman P, van der Gulden H, van der Heijden I, Drost R, Klijn CN, Prasetyanti P*, et al.* A high-throughput functional complementation assay for classification of BRCA1 missense variants. Cancer discovery **2013**;3(10):1142-55 doi 10.1158/2159-8290.CD-13-0094.
14. Mesman RLS, Calleja F, Hendriks G, Morolli B, Misovic B, Devilee P*, et al.* The functional impact of variants of uncertain significance in BRCA2. Genetics in medicine : official journal of the American College of Medical Genetics **2018** doi 10.1038/s41436-018-0052-2.
15. Shimelis H, Mesman RLS, Von Nicolai C, Ehlen A, Guidugli L, Martin C*, et al.* BRCA2 Hypomorphic Missense Variants Confer Moderate Risks of Breast Cancer. Cancer research **2017**;77(11):2789-99 doi 10.1158/0008-5472.CAN-16-2568.
16. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M*, et al.* Accurate classification of BRCA1 variants with saturation genome editing. Nature **2018**;562(7726):217-22 doi 10.1038/s41586-018-0461-z.
17. Starita LM, Islam MM, Banerjee T, Adamovich AI, Gullingsrud J, Fields S*, et al.* A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution Variants on Protein Function. American journal of human genetics **2018**;103(4):498-508 doi 10.1016/j.ajhg.2018.07.016.
18. Kass EM, Helgadottir HR, Chen CC, Barbera M, Wang R, Westermark UK*, et al.* Double-strand break repair by homologous recombination in primary mouse somatic cells requires BRCA1 but not the ATM kinase. Proceedings of the National Academy of Sciences of the United States of America **2013**;110(14):5564-9 doi 10.1073/pnas.1216824110.
19. Kranz A, Fu J, Duerschke K, Weidlich S, Naumann R, Stewart AF*, et al.* An improved Flp deleter mouse in C57Bl/6 based on Flpo recombinase. Genesis **2010**;48(8):512-20 doi 10.1002/dvg.20641.
20. Rantakari P, Nikkila J, Jokela H, Ola R, Pylkas K, Lagerbohm H*, et al.* Inactivation of Palb2 gene leads to mesoderm differentiation defect and early embryonic lethality in mice. Hum Mol Genet **2010**;19(15):3021-9 doi 10.1093/hmg/ddq207.
21. Bouwman P, Drost R, Klijn C, Pieterse M, van der Gulden H, Song JY*, et al.* Loss of p53 partially rescues embryonic development of Palb2 knockout mice but does not foster haploinsufficiency of Palb2 in tumour suppression. J Pathol **2011**;224(1):10-21 doi 10.1002/path.2861.
22. Blomen VA, Majek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, Staring J*, et al.* Gene essentiality and synthetic lethality in haploid human cells. Science **2015**;350(6264):1092-6 doi 10.1126/science.aac7557.
23. Greenblatt MS, Chappuis PO, Bond JP, Hamel N, Foulkes WD. TP53 mutations in breast cancer associated with BRCA1 or BRCA2 germ-line mutations: distinctive spectrum and structural distribution. Cancer research **2001**;61(10):4092-7.
24. Rhei E, Bogomolniy F, Federici MG, Maresco DL, Offit K, Robson ME*, et al.* Molecular genetic characterization of BRCA1- and BRCA2-linked hereditary ovarian cancers. Cancer research **1998**;58(15):3193-6.

25. Li A, Geyer FC, Blecua P, Lee JY, Selenica P, Brown DN*, et al.* Homologous recombination DNA repair defects in PALB2-associated breast cancers. NPJ Breast Cancer **2019**;5:23 doi 10.1038/s41523-019-0115-9.
26. Xia B, Dorsman JC, Ameziane N, de Vries Y, Rooimans MA, Sheng Q*, et al.* Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. Nature genetics **2007**;39(2):159-61 doi 10.1038/ng1942.
27. Susswein LR, Marshall ML, Nusbaum R, Vogel Postula KJ, Weissman SM, Yackowski L*, et al.* Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. Genetics in medicine : official journal of the American College of Medical Genetics **2016**;18(8):823-32 doi 10.1038/gim.2015.166.
28. Kraus C, Hoyer J, Vasileiou G, Wunderle M, Lux MP, Fasching PA*, et al.* Gene panel sequencing in familial breast/ovarian cancer patients identifies multiple novel mutations also in genes others than BRCA1/2. International journal of cancer **2017**;140(1):95-102 doi 10.1002/ijc.30428.
29. Oliver AW, Swift S, Lord CJ, Ashworth A, Pearl LH. Structural basis for recruitment of BRCA2 by PALB2. EMBO reports **2009**;10(9):990-6 doi 10.1038/embor.2009.126.
30. Foo TK, Tischkowitz M, Simhadri S, Boshari T, Zayed N, Burke KA*, et al.* Compromised BRCA1-PALB2 interaction is associated with breast cancer risk. Oncogene **2017**;36(29):4161-70 doi 10.1038/onc.2017.46.
31. Park JY, Singh TR, Nassar N, Zhang F, Freund M, Hanenberg H*, et al.* Breast cancer-associated missense mutants of the PALB2 WD40 domain, which directly binds RAD51C, RAD51 and BRCA2, disrupt DNA repair. Oncogene **2014**;33(40):4803-12 doi 10.1038/onc.2013.421.
32. Michl J, Zimmer J, Tarsounas M. Interplay between Fanconi anemia and homologous recombination pathways in genome integrity. The EMBO journal **2016**;35(9):909-23 doi 10.15252/embj.201693860.
33. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P*, et al.* A method and server for predicting damaging missense mutations. Nat Methods **2010**;7(4):248-9 doi 10.1038/nmeth0410-248.
34. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: predicting effects of amino acid substitutions on proteins. Nucleic Acids Res **2012**;40(Web Server issue):W452-7 doi 10.1093/nar/gks539.
35. Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB*, et al.* Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. J Med Genet **2006**;43(4):295-305 doi 10.1136/jmg.2005.033878.
36. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res **2019**;47(D1):D886-D94 doi 10.1093/nar/gky1016.

37. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S*, et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. American journal of human genetics **2016**;99(4):877-85 doi 10.1016/j.ajhg.2016.08.016.
38. Cotta-Ramusino C, McDonald ER, 3rd, Hurov K, Sowa ME, Harper JW, Elledge SJ. A DNA damage response screen identifies RHINO, a 9-1-1 and TopBP1 interacting protein required for ATR signaling. Science **2011**;332(6035):1313-7 doi 10.1126/science.1203430.
39. Menzel T, Nahse-Kumpf V, Kousholt AN, Klein DK, Lund-Andersen C, Lees M*, et al.* A genetic screen identifies BRCA2 and PALB2 as key regulators of G2 checkpoint maintenance. EMBO reports **2011**;12(7):705-12 doi 10.1038/embor.2011.99.
40. Simhadri S, Vincelli G, Huo Y, Misenko S, Foo TK, Ahlskog J*, et al.* PALB2 connects BRCA1 and BRCA2 in the G2/M checkpoint response. Oncogene **2019**;38(10):1585-96 doi 10.1038/s41388-018-0535-2.
41. Ducy M, Sesma-Sanz L, Guitton-Sert L, Lashgari A, Gao Y, Brahiti N*, et al.* The Tumor Suppressor PALB2: Inside Out. Trends Biochem Sci **2019** doi 10.1016/j.tibs.2018.10.008.
42. Simhadri S, Peterson S, Patel DS, Huo Y, Cai H, Bowman-Colin C*, et al.* Male fertility defect associated with disrupted BRCA1-PALB2 interaction in mice. The Journal of biological chemistry **2014**;289(35):24617-29 doi 10.1074/jbc.M114.566141.
43. Pauty J, Couturier AM, Rodrigue A, Caron MC, Coulombe Y, Dellaire G*, et al.* Cancer-causing mutations in the tumor suppressor PALB2 reveal a novel cancer mechanism using a hidden nuclear export signal in the WD40 repeat motif. Nucleic Acids Res **2017**;45(5):2644-57 doi 10.1093/nar/gkx011.
44. Matreyek KA, Starita LM, Stephany JJ, Martin B, Chiasson MA, Gray VE*, et al.* Multiplex assessment of protein variant abundance by massively parallel sequencing. Nature genetics **2018**;50(6):874-82 doi 10.1038/s41588-018-0122-z.
45. Ceccaldi R, Liu JC, Amunugama R, Hajdu I, Primack B, Petalcorin MI*, et al.* Homologous-recombination-deficient tumours are dependent on Poltheta-mediated repair. Nature **2015**;518(7538):258-62 doi 10.1038/nature14184.
46. Southey MC, Goldgar DE, Winqvist R, Pylkas K, Couch F, Tischkowitz M*, et al.* PALB2, CHEK2 and ATM rare variants and cancer risk: data from COGS. J Med Genet **2016**;53(12):800-11 doi 10.1136/jmedgenet-2016-103839.
47. Kuznetsov SG, Liu P, Sharan SK. Mouse embryonic stem cell-based functional assay to evaluate mutations in BRCA2. Nat Med **2008**;14(8):875-81 doi 10.1038/nm.1719.
48. Chang S, Biswas K, Martin BK, Stauffer S, Sharan SK. Expression of human BRCA1 variants in mouse ES cells allows functional analysis of BRCA1 mutations. J Clin Invest **2009**;119(10):3160-71 doi 10.1172/JCI39836.
49. Robanus-Maandag E, Dekker M, van der Valk M, Carrozza ML, Jeanny JC, Dannenberg JH*, et al.* p107 is a suppressor of retinoblastoma development in pRb-deficient mice. Genes & development **1998**;12(11):1599-609.
50. Moynahan ME, Pierce AJ, Jasin M. BRCA2 is required for homology-directed repair of chromosomal breaks. Molecular cell **2001**;7(2):263-72.

51. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nature protocols **2013**;8(11):2281-308 doi 10.1038/nprot.2013.143.
52. Cruz C, Castroviejo-Bermejo M, Gutierrez-Enriquez S, Llop-Guevara A, Ibrahim YH, Gris-Oliver A*, et al.* RAD51 foci as a functional biomarker of homologous recombination repair and PARP inhibitor resistance in germline BRCA-mutated breast cancer. Ann Oncol **2018**;29(5):1203-10 doi 10.1093/annonc/mdy099.
53. Niwa H, Yamamura K, Miyazaki J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene **1991**;108(2):193-9.
54. Bleuyard JY, Buisson R, Masson JY, Esashi F. ChAM, a novel motif that mediates PALB2 intrinsic chromatin binding and facilitates DNA repair. EMBO reports **2012**;13(2):135-41 doi 10.1038/embor.2011.243.
55. Noordermeer SM, Adam S, Setiaputra D, Barazas M, Pettitt SJ, Ling AK*, et al.* The shieldin complex mediates 53BP1-dependent DNA repair. Nature **2018**;560(7716):117-21 doi 10.1038/s41586-018-0340-7.

SUPPLEMENTARY INFORMATION

Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene *PALB2*

Rick A.C.M. Boonen, Amélie Rodrigue, Chantal Stoepker, Wouter W. Wiegant, Bas Vroling, Milan Sharma, Magdalena B. Rother, Nandi Celosse, Maaike P.G. Vreeswijk, Fergus Couch, Jacques Simard, Peter Devilee, Jean-Yves Masson and Haico van Attikum.

The supplementary information contains:

- Supplementary Data 1
- Supplementary Figure 1-11

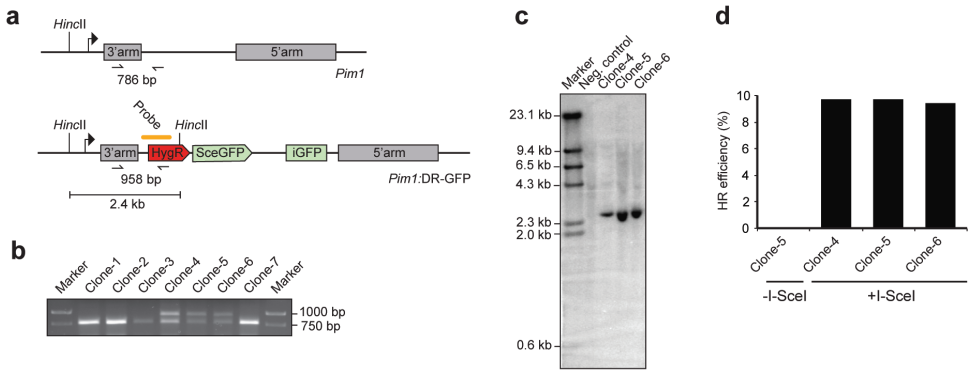
Supplementary Data 1. Complete list of human *PALB2* variants analyzed in this study.

Protein change	Mutation type	Prior classification	Align GVGD	CADD (phred)	PolyPhen	SIFT	REVEL	HR (%)	PARPi (%)	Cispl. (%)	Norm. M-phase (%)
p.P4S	Missense	VUS	C0	16,53	0,09	0,35	0,04	97,61	72,70	X	X
p.P5S	Missense	VUS	C0	16,21	0,02	0,49	0,03	62,31	95,74	X	X
p.K18R	Missense	VUS	C0	24,30	1,00	0,03	0,18	100,19	94,47	84,05	86,76
p.L24S*	Missense	VUS	C65	23,80	1,00	0,01	0,17	20,67	55,40	X	X
p.Y28C*	Missense	VUS	C65	26,70	1,00	0,01	0,22	32,92	21,70	X	X
p.T31I	Missense	VUS	C65	26,50	1,00	0,01	0,21	97,16	102,22	X	X
p.L35P*	Missense	VUS	C65	31,00	1,00	0,10	0,35	10,40	9,68	26,03	260,91
p.R37H	Missense	VUS	C25	24,20	0,97	0,01	0,16	44,90	67,82	83,16	175,61
p.E42K	Missense	VUS	C15	34,00	1,00	0,02	0,14	105,41	94,84	89,72	101,81
p.Q60Rfs	Frameshift	Pathogenic	X	X	X	X	X	8,55	11,06	X	X
p.D134N	Missense	Likely benign	C0	10,56	0,02	0,47	0,04	90,64	93,44	97,73	54,70
p.S172fs	Frameshift	Pathogenic	X	X	X	X	X	8,55	13,87	X	X
p.P210L	Missense	Likely benign	C0	9134,00	0,02	0,66	0,10	85,37	102,70	X	X
p.E230X	Nonsense	Pathogenic	X	X	X	X	X	7,42	17,50	X	X
p.L337S	Missense	Likely benign	C0	11,35	0,29	0,23	0,04	86,67	115,71	X	X
p.Y408H	Missense	VUS	C65	29,40	1,00	N/A	-	92,32	108,51	93,33	79,77
p.Y409X	Nonsense	Pathogenic	X	X	X	X	X	7,82	17,43	X	X
p.S417Y	Missense	VUS	C15	26,10	1,00	0,00	0,33	72,20	83,96	X	X
p.D498Y	Missense	Likely benign	C0	21,00	0,90	0,07	0,09	94,49	74,18	X	X
p.K515R	Missense	VUS	C0	15,96	0,20	0,15	0,01	75,45	120,54	X	X
p.L531Cfs	Frameshift	Pathogenic	X	X	X	X	X	7,75	23,96	X	X
p.Y551X	Nonsense	Pathogenic	X	X	X	X	X	8,12	10,68	20,08	244,69
p.Q559R	Missense	Likely benign	C0	0,08	0,00	0,75	0,02	95,02	117,58	X	X
p.L622P	Missense	VUS	C65	28,90	1,00	0,01	0,34	77,45	64,92	X	X
p.E669Gfs	Frameshift	Pathogenic	X	X	X	X	X	7,03	19,46	X	X
p.E672Q	Missense	Likely benign	C0	10,78	0,23	0,28	0,03	79,52	103,53	X	X
p.T706I	Missense	VUS	C15	24,40	1,00	0,01	0,25	87,35	78,20	X	X
p.P707L	Missense	VUS	C65	27,70	1,00	0,00	0,33	82,83	87,96	100,97	93,92
p.V858=	Synonymous	Likely benign	X	X	X	X	X	84,43	74,36	X	X
p.P864S	Missense	Likely benign	C0	19,49	0,58	0,38	0,06	85,80	86,54	X	X
p.S865P	Missense	VUS	C0	28,20	1,00	0,03	0,19	100,10	78,30	X	X
p.D871G	Missense	VUS	C35	27,60	1,00	0,02	0,52	84,07	115,45	87,40	116,75
p.C882Wfs	Frameshift	Pathogenic	X	X	X	X	X	6,10	29,04	X	X
p.W912G*	Missense	VUS	C0	25,60	1,00	0,00	0,56	6,66	7,73	X	X
p.D927A	Missense	VUS	C0	32,00	0,96	0,02	0,25	75,71	86,26	X	X
p.L931R	Missense	VUS	C65	27,30	1,00	0,00	0,57	106,25	94,55	112,08	79,77
p.G937R*	Missense	VUS	C65	28,40	1,00	0,00	0,48	17,35	26,39	X	X
p.L939W	Missense	VUS	C55	29,40	1,00	0,00	0,36	60,28	91,12	102,20	218,25

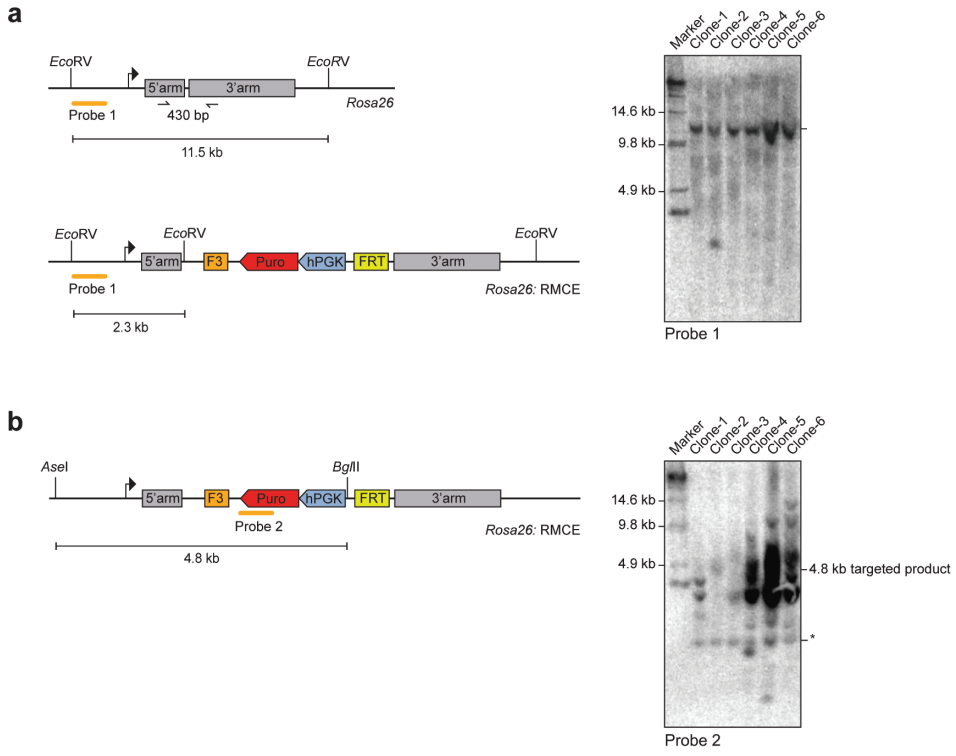
Supplementary Data 1. Continued

Protein change	Mutation type	Prior classification	Align GVGD	CADD (phred)	PolyPhen	SIFT	REVEL	HR (%)	PARPi (%)	Cispl. (%)	Norm. M-phase (%)
p.E940G	Missense	VUS	C65	29,60	1,00	0,00	0,43	63,40	81,17	X	X
p.I944N*	Missense	VUS	C15	26,70	1,00	0,00	0,45	7,27	14,78	X	X
p.L947S*	Missense	VUS	C65	24,90	1,00	0,00	0,38	30,27	24,31	X	X
p.L961P*	Missense	VUS	C25	25,50	1,00	0,02	0,27	6,53	8,41	27,29	280,28
p.I966T	Missense	VUS	C0	26,40	1,00	0,22	0,21	74,49	78,91	X	X
p.L972Q*	Missense	VUS	C35	28,20	1,00	0,00	0,23	14,02	12,77	X	X
p.Q988X	Nonsense	Pathogenic	X	X	X	X	X	6,70	15,23	X	X
p.G998E	Missense	Likely benign	C65	27,70	1,00	0,01	0,29	95,16	97,37	80,44	129,12
p.P1009Lfs	Frameshift	Pathogenic	X	X	X	X	X	6,16	25,09	X	X
p.E1018D	Missense	VUS	C0	23,40	1,00	0,05	0,12	86,41	84,41	111,51	73,37
p.A1025R	Missense	VUS	C65	23,10	1,00	0,05	0,13	17,62	24,27	X	368,90
p.T1030I*	Missense	VUS	C65	31,00	1,00	0,00	0,43	14,68	14,80	X	X
p.I1037T*	Missense	VUS	C25	26,10	1,00	0,00	0,41	38,86	52,23	X	X
p.W1038X	Nonsense	Pathogenic	X	X	X	X	X	6,98	12,35	X	X
p.L1040S	Missense	VUS	C0	31,00	1,00	0,03	0,31	76,97	99,69	X	X
p.G1043D*	Missense	VUS	C65	29,80	1,00	0,00	0,30	10,59	10,42	37,28	275,71
p.I1051S	Missense	VUS	C35	27,30	1,00	0,00	0,22	91,24	108,56	X	X
p.S1058P	Missense	VUS	C0	28,10	0,99	0,01	0,19	95,88	85,13	119,03	108,97
p.Y1064C	Missense	VUS	C65	31,00	1,00	0,00	0,44	101,04	87,27	128,59	100,42
p.L1070P*	Missense	VUS	C65	26,40	1,00	0,00	0,47	23,09	56,67	X	X
p.P1111A	Missense	VUS	C25	28,10	1,00	0,02	0,40	102,61	87,51	X	X
p.L1119P	Missense	VUS	C65	-	1,00	0,00	0,47	94,33	103,83	X	X
p.V1123M	Missense	VUS	C0	26,70	1,00	0,00	0,23	75,85	102,57	X	X
p.L1143H	Missense	VUS	C0	26,70	1,00	0,14	0,22	69,83	98,43	X	X
p.W1159L	Missense	VUS	C0	28,20	1,00	0,01	0,37	90,82	108,27	X	X
p.S1160P	Missense	VUS	C0	26,90	1,00	0,01	0,26	92,22	103,21	X	X
p.W1164C	Missense	VUS	C65	33,00	1,00	0,00	0,43	80,95	96,22	X	X
p.L1172P*	Missense	VUS	C65	28,90	1,00	0,00	0,48	13,46	16,60	X	X
p.G1174R	Missense	VUS	C65	31,00	1,00	0,00	0,43	90,57	102,88	X	X
p.I1180T	Missense	VUS	C25	24,20	1,00	0,01	0,34	81,82	98,64	X	X
p.Y1183C	Missense	VUS	C55	27,90	1,00	0,00	0,41	70,86	109,07	X	X
p.Y1183X	Nonsense	Pathogenic	X	X		X	X	11,12	9,67	19,53	344,61

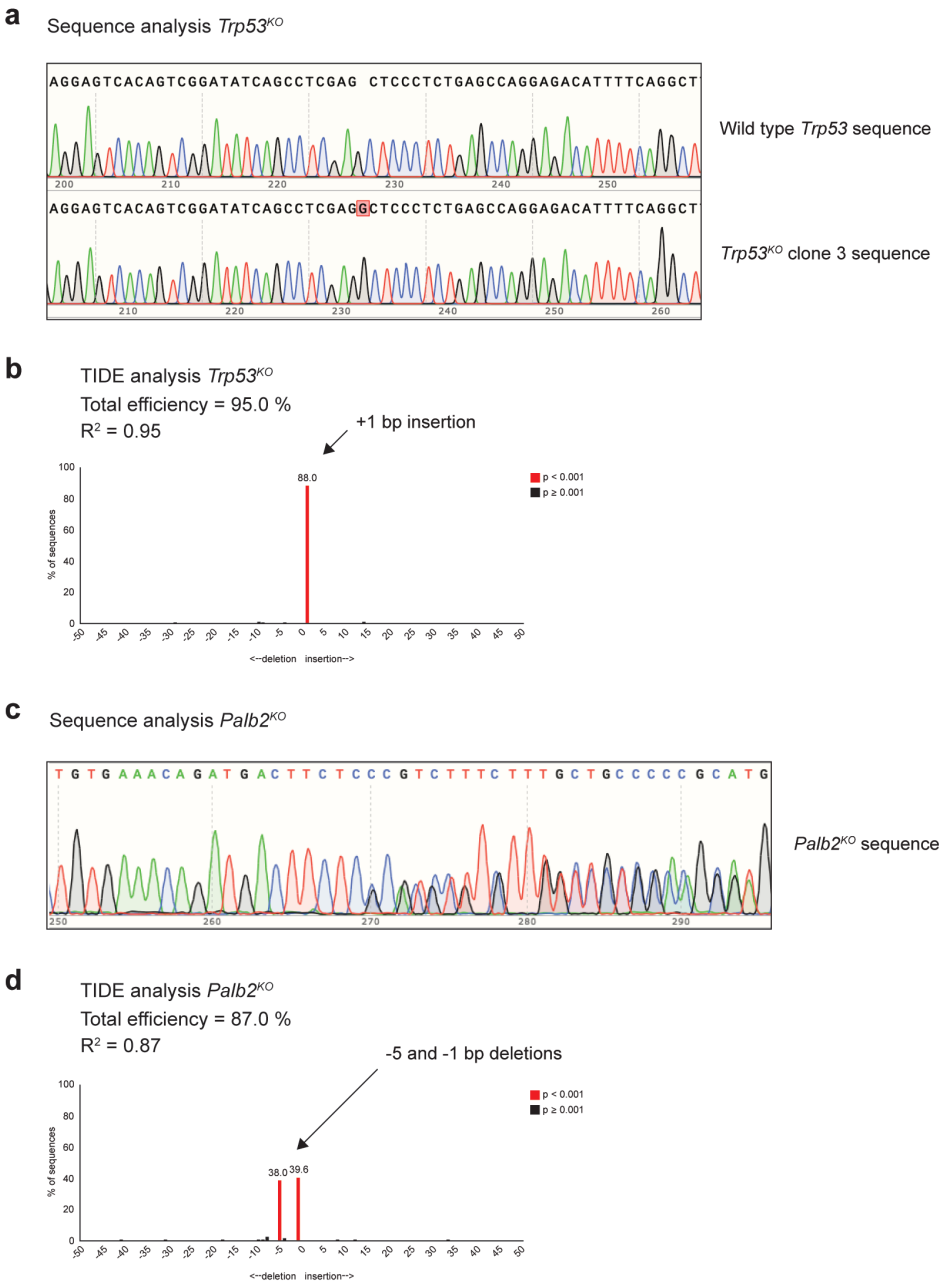
All variants are indicated at the protein level (i.e., protein change). Nucleotide annotations for each variant are available in the published manuscript, where nucleotide numbering reflects Human Genome Variation Society (HGVS) nomenclature and cDNA number +1 corresponds to the A of the ATG translation initiation codon in the reference sequence (*PALB2* NM_024675.3). The initiation codon is codon 1. In silico predictions, results from DR-GFP, PARPi sensitivity, cisplatin sensitivity and G2/M checkpoint assays in mES cells are included. Strongly damaging variants from the functional assays (HR >60% reduced) are indicated in the 'protein change' column (red *).



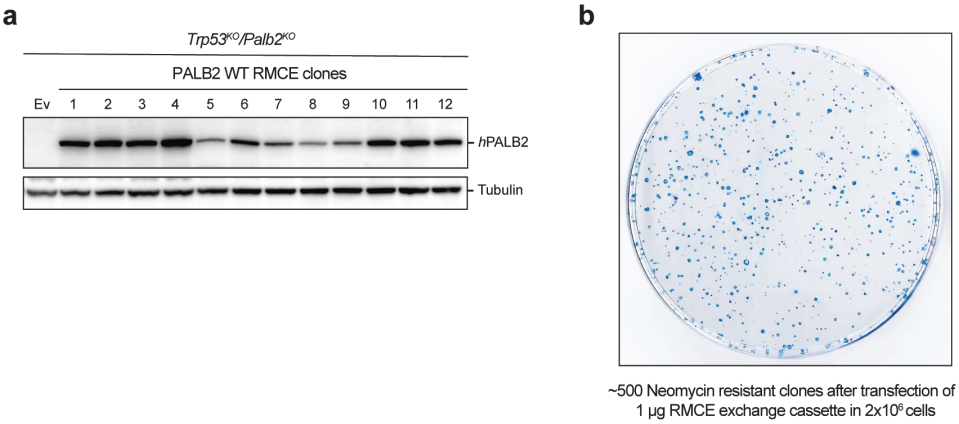
Supplementary Figure 1. Stable integration of the DR-GFP reporter at the *Pim1* locus in mES cells. **a** Schematic showing the *Pim1* locus (upper) and *Pim1* locus with an integrated DR-GFP reporter (*Pim1:*DR-GFP; lower) in mES cells. Integration is directed by the 3' and 5' homology arms. Correct integration of the reporter results in expression of a hygromycin resistance marker under control of the endogenous *Pim1* promoter (not shown). Correct integration was examined by PCR and Southern blot analysis using the indicated primers, as well as probe and restriction enzymes, respectively. **b** PCR analysis of genomic DNA from hygromycin-resistant mES cell clones obtained after targeting the *Pim1* locus with a DR-GFP cassette using primers indicated in A. Clone 4-6 show correct integration of DR-GFP at a *Pim1* allele (as evidence by the appearance of a 958 bp band). **c** Southern blot analysis of *Hinc*II-digested genomic DNA from mES cell clones 4-6 from B using the probe shown in A. Single copy genomic integration at a *Pim1* allele is observed in all three clones (as evidence by the appearance of a 2.4 kb band). **d** DR-GFP assay DR-GFP assay in clone 4-6 from **b** and **c**. Cells were co-transfected with I-SceI and mCherry expression vectors, or mCherry expression vector only, and GFP expression was monitored by FACS. Data represent the absolute percentage of GFP-positive cells among the mCherry-positive-cells. Source data are provided as a Source Data file.



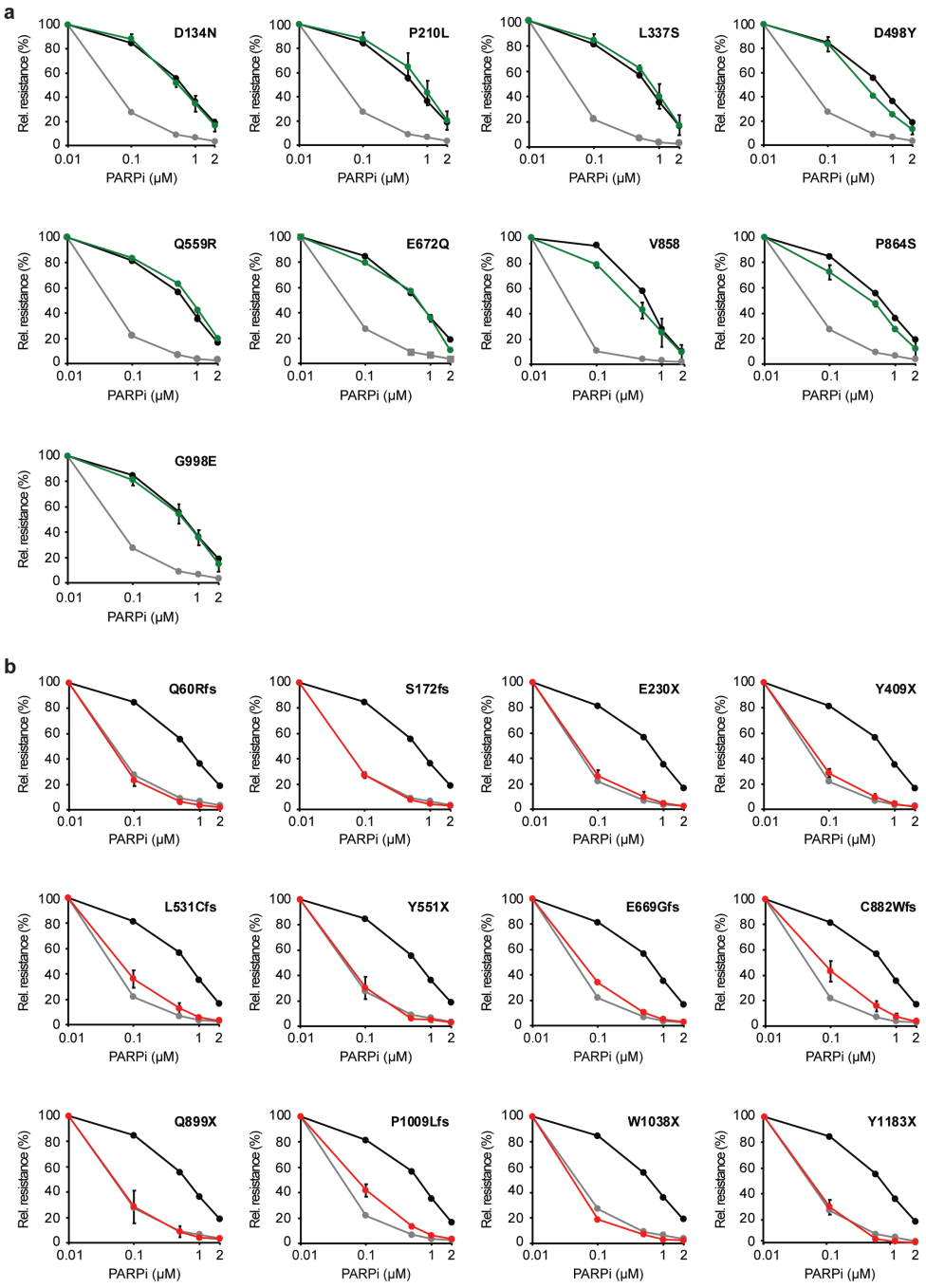
Supplementary Figure 2. Stable integration of the RMCE acceptor cassette at the *Rosa26* locus in mES cells carrying DR-GFP. **a** Schematic showing the *Rosa26* locus (upper left) and *Rosa26* locus with an integrated RMCE acceptor cassette (*Rosa26:*RMCE; lower left) in mES cells. Integration is directed by the 3' and 5' homology arms. Correct integration of the RMCE acceptor cassette results in expression of a puromycin resistance marker under control of the PGK1 promoter. Correct integration was examined by Southern blot analysis of *Eco*RV-digested genomic DNA from mES cell clones 1-6 using the indicated probe (right). Single copy genomic integration at a *Rosa26* allele is observed in clone 2 (as evidence by the appearance of a 2.3 kb band). **b** Schematic as in **a**, except that a different probe and different restriction sites for Southern blot analysis are shown (left). Correct integration was examined by Southern blot analysis of *Ase*I- and *Bgl*II-digested genomic DNA from mES cell clones 1-6 using the indicated probe (right). Single copy genomic integration at a *Rosa26* allele is observed in clone 2 (as evidence by the appearance of a 4.8 kb band).



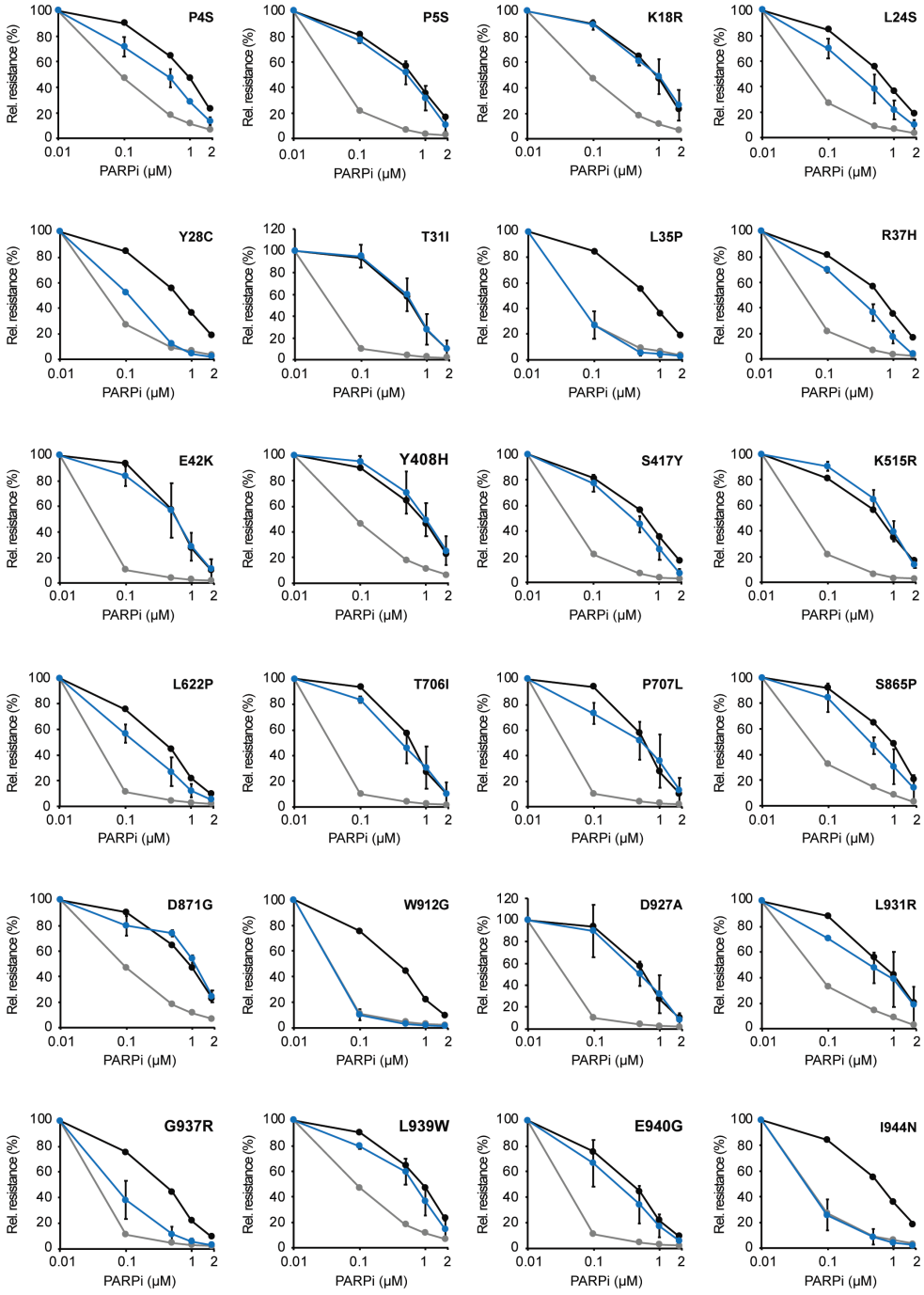
Supplementary Figure 3. Validation of *Trp53*^{KO}/*Palb2*^{KO} mES cells. **A**, Sequence alignment of a fragment of exon 1 of the *Trp53* gene showing a +1 bp (guanine) insertion in *Trp53*^{KO} clone 3. **b** TIDE analysis confirming the +1 bp insertion in exon 1 of the *Trp53* gene in *Trp53*^{KO} clone 3. **c** Sequence alignment of a fragment of exon 4 of the *Palb2* gene showing -5 bp and -1 bp deletions in the *Palb2*^{KO} clone. **d** TIDE analysis confirming -1 and -5 bp deletions in exon 4 of the *Palb2* gene in the *Palb2*^{KO} clone.



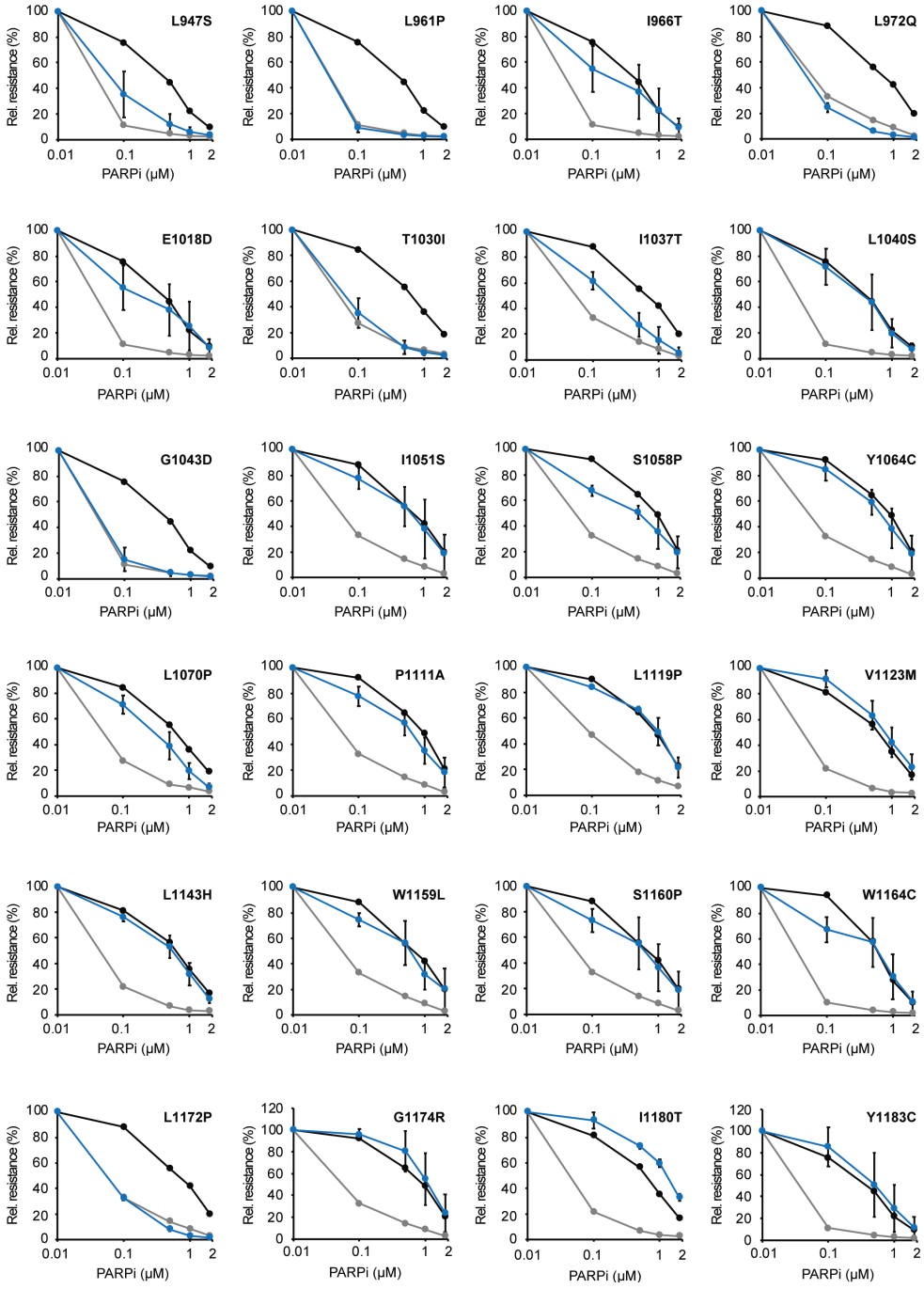
Supplementary Figure 4. RMCE efficiency in mES cells. **a** Western blot analysis of the expression of wild-type human *PALB2* in 12 individual *Trp53*^{KO}/*Palb2*^{KO} mES cell clones using an antibody directed against the N-terminus of PALB2. An empty vector (Ev) served as negative control. Tubulin was used as a loading control. **b** Representative image of a culture dish with methylene-stained neomycin resistant clones after transfection of *Trp53*^{KO}/*Palb2*^{KO} mES cells using RMCE exchange cassette. Source data are provided as a Source Data file.


a
D134N
P210L
L337S
D498Y
Q559R
E672Q
V858
P864S
G998E
b
Q60Rfs
S172fs
E230X
Y409X
L531Cfs
Y551X
E669Gfs
C882Wfs
Q899X
P1009Lfs
W1038X
Y1183X
Rel. resistance (%)
PARPi (μM)
0.01
0.1
1
2
0
20
40
60
80
100

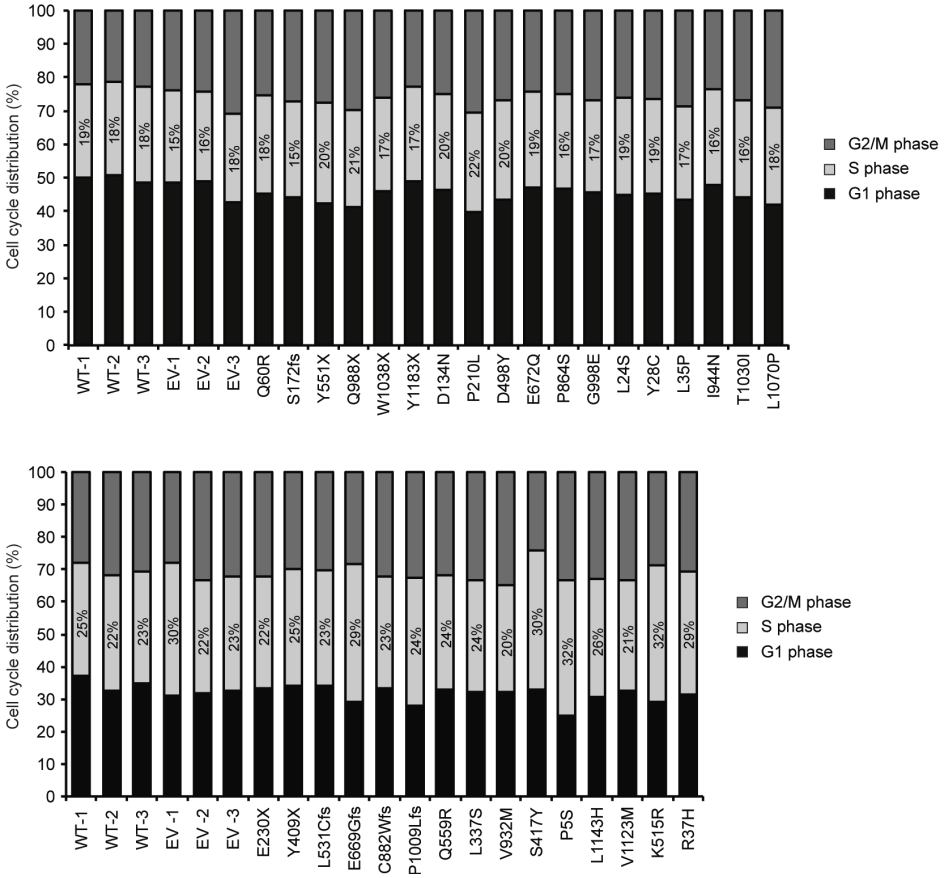
Supplementary Figure 5. Functional analysis of benign and truncating variants in human *PALB2* by PARPi sensitivity assays. **a** PARP inhibitor (PARPi) sensitivity assay using *Trp53*^{KO}/*Palb2*^{KO} mES cells expressing human *PALB2* variants (or an empty vector control, Ev). Cells were exposed to the indicated concentrations of PARPi for two days. Cell viability was measured 1 day later using FACS. Data represent the mean percentage of viability/resistance relative to untreated cells (± SEM) from 2 independent experiments, except for p.P210L for which data from three independent experiments is presented. Variants/conditions are categorized by color as either wild type (WT, black), likely benign SNV (green), or empty vector (Ev, grey). Data from the 0.5 μM PARPi concentration are shown in Fig. 3a. **b** as in **a**, except for truncating variants (red). Source data are provided as a Source Data file.


P4S
P5S
K18R
L24S
Y28C
T31I
L35P
R37H
E42K
Y408H
S417Y
K515R
L622P
T706I
P707L
S865P
D871G
W912G
D927A
L931R
G937R
L939W
E940G
I944N
Rel. resistance (%)
PARPi (μM)
0.01
0.1
1
2

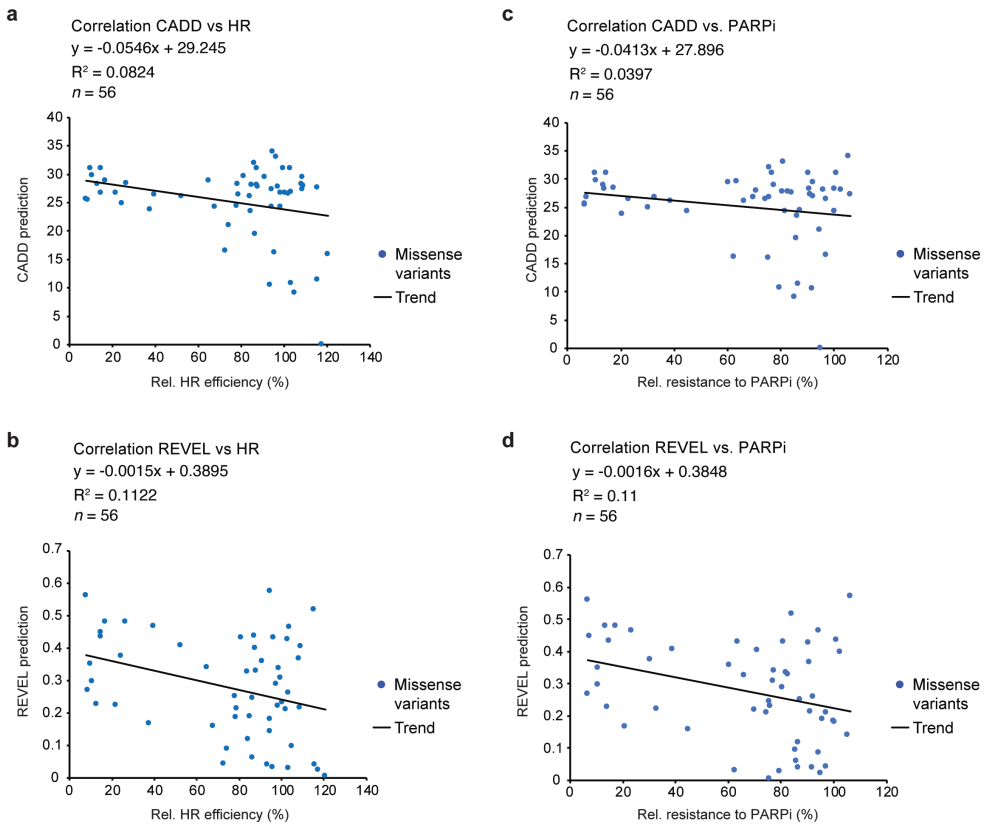
Supplementary Figure 6. Functional analysis of selected VUS in human *PALB2* by PARPi sensitivity assays. PARP inhibitor (PARPi) sensitivity assay using *Trp53*^{KO}/*Palb2*^{KO} mES cells expressing human *PALB2* variants (or an empty vector control, Ev). Cells were exposed to the indicated concentrations of PARPi for two days. Cell viability was measured 1 day later using FACS. Data represent the mean percentage of viability/resistance relative to untreated cells (± SEM) from 2 independent experiments, except for p.P4S and p.L939W, for which data from three independent experiments is presented, and p.L24S for which data from four independent experiments is presented. Variants/conditions are categorized by color as either wild type (WT, black), VUS (blue), or empty vector (Ev, grey). Data from the 0.5 μM PARPi concentration are shown in Fig. 3a. Source data are provided as a Source Data file.


L947S
L961P
I966T
L972Q
E1018D
T1030I
I1037T
L1040S
G1043D
I1051S
S1058P
Y1064C
L1070P
P1111A
L1119P
V1123M
L1143H
W1159L
S1160P
W1164C
L1172P
G1174R
I1180T
Y1183C
Rel. resistance (%)
PARPi (μM)

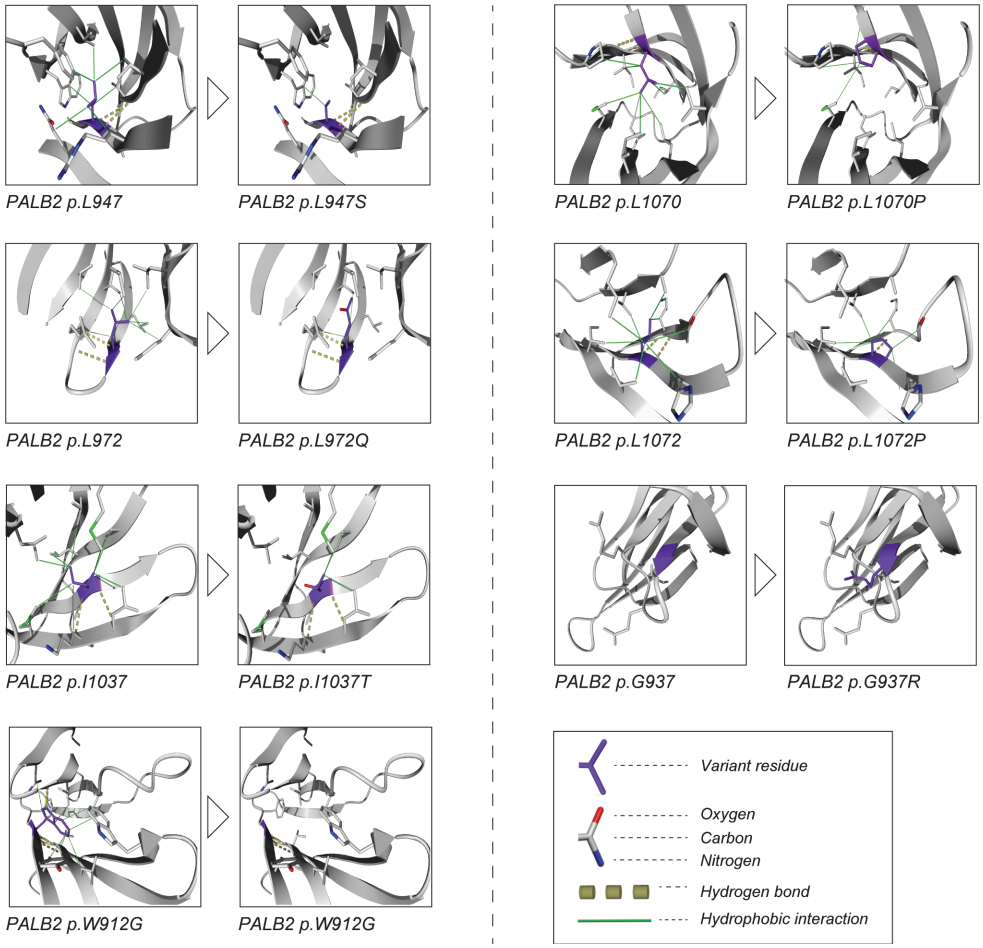
Supplementary Figure 7. Functional analysis of selected VUS in human *PALB2* by PARPi sensitivity assays. PARP inhibitor (PARPi) sensitivity assay using *Trp53*^{KO}/*Palb2*^{KO} mES cells expressing human *PALB2* variants (or an empty vector control, Ev). Cells were exposed to the indicated concentrations of PARPi for two days. Cell viability was measured 1 day later using FACS. Data represent the mean percentage of viability/resistance relative to untreated cells (± SEM) from 2 independent experiments, except for p.V1123M for which data from three independent experiments is presented, and p.L1070P for which data from four independent experiments is presented. Variants/conditions are categorized by color as either wild type (WT, black), VUS (blue), or empty vector (Ev, grey). Data from the 0.5 μM PARPi concentration are shown in Fig. 3a. Source data are provided as a Source Data file.



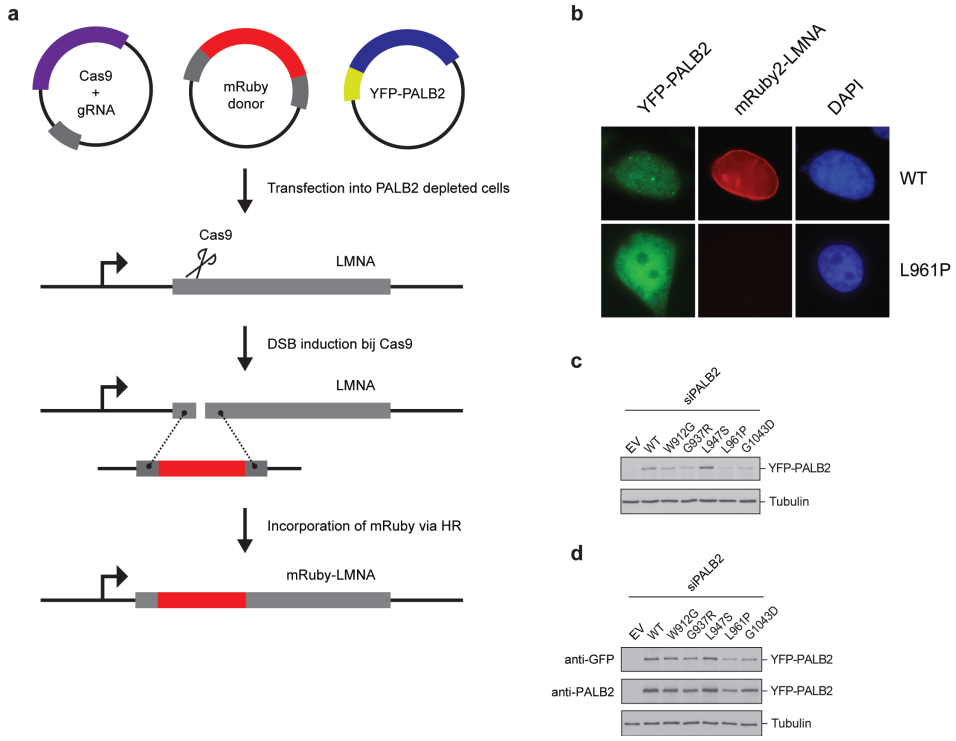
Supplementary Figure 8. Cell cycle profiles of *Trp53*^{KO}/*Palb2*^{KO} mES cells expressing human *PALB2* variants. Cell cycle profiles are from cells in Fig. 2b. Cells were treated with propidium staining (PI) and analyzed by FACS. Data represent the mean percentage of cell cycle phase distributions from 2 independent measurements. Source data are provided as a Source Data file.



Supplementary Figure 9. Correlation between *in silico* predictions and the outcome of functional assays for missense variants in human *PALB2*. **a** Scatter plot showing correlation between the *in silico* prediction from CADD and results from the DR-GFP assay in Fig. 2b. **b** Scatter plot showing correlation between the *in silico* prediction from REVEL and results from the DR-GFP assay in Fig. 2b. **c** Scatter plot showing correlation between the *in silico* prediction from CADD and results from the PARPi sensitivity assay in Fig. 3a. **d** Scatter plot showing correlation between the *in silico* prediction from REVEL and results from the PARPi sensitivity assay in Fig. 3a.



Supplementary Figure 10. Effect of *PALB2* variants on protein stability. Partial structures of the PALB2 WD40 domain showing the effect of 7 PALB2 variants exhibiting low protein expression as shown in Fig. 4a. Partial structures without and with variant are shown side by side for each variant, indicating loss of stabilizing interactions (but not any possible conformational changes).



Supplementary Figure 11. Functional analysis of damaging *PALB2* variants in human cells. **a** Schematic of the CRISPR-LMNA HDR assay in human cells. Homology-directed repair of the Cas9-induced DSB will result in the in-frame integration of mRuby in the first exon of *LMNA*, leading to expression of red fluorescent mRuby-LMNA. The number of mRuby-positive cells is a measure of the HR efficiency. **b** Representative fluorescence microscopy images of mRuby2-LMNA expression after successful homology directed repair (HDR) in a PALB2-depleted U2OS cell complemented with YFP-PALB2-WT (upper), and a cell negative for mRuby2-LMNA expression after complementation with the damaging YFP-PALB2-L961P variant (lower). **c** Western blot analysis of the expression of human PALB2 variants in si*PALB2*-treated U2OS cells 24 hours after complementation with the indicated siRNA-resistant YFP-*PALB2* variant cDNA constructs. Tubulin was used as a loading control. **d** Western blot analysis of the expression of human PALB2 variants in si*PALB2*-treated HeLa cells 24 hours after complementation with the indicated siRNA-resistant *YFP-PALB2* variant cDNA constructs. Tubulin was used as a loading control. Source data are provided as a Source Data file.

CHAPTER 5

5

Functional interpretation of *PALB2* missense variants and their association with breast cancer risk

Rick A.C.M. Boonen, Sabine C. Knaup, Roberta Menafra, Dina Ruano, Magdalena B. Rother, Emile J. de Meijer, Pei Sze Ng, Soo Hwang Teo, Noel F. de Miranda, Maaike P.G. Vreeswijk, Susan L. Kloet and Haico van Attikum

Figure 1 in this chapter is published in *Journal of Medical Genetics* (PMID: 33811135)

ABSTRACT

Genetic testing for sequence alterations in genes that associate with cancer, frequently reveals missense variants of uncertain significance (VUS) for which the effects on protein function and associated cancer risk and are unclear. To extend the utility of genetic tests for the high-risk breast cancer gene *PALB2*, functional assays can be performed to determine the effects of variants in this gene. Here we employ both semi high-throughput and high-throughput approaches for the functional analysis of genetic variants in *PALB2*. Our semi high-throughput approach identified four novel damaging missense variants in the WD40 domain of PALB2, and furthermore showed that the ChAM and MRG15 domains are dispersible for PALB2's function in homologous recombination (HR). Our high-throughput assay allowed us to functionally interrogate 603 variants in the Coiled-Coil (CC) domain of *PALB2*, which may provide evidence for the re-classification of over 60 *PALB2* CC missense VUS reported in ClinVar. Correlation of functional data from the semi high-throughput approach with breast cancer risk, shows for the first time that reduced homologous recombination (HR) as a result of patient-derived missense variants in *PALB2,* correlates with increased breast cancer risk. We therefore predict that the results presented here will eventually be useful for the clinical interpretation of many *PALB2* missense variants, and that this approach can be extended to overcome the challenge of managing VUS carriers.

INTRODUCTION

Genetic testing for genes that have been associated with hereditary breast cancer has led to the identification of a plethora of genetic variants for which the impact on protein function is often not clear. Many of these variants, of which most constitute rare missense variants, are reported as variants of uncertain significance (VUS). As accurate quantification of cancer risk for such rare variants is generally not possible, even after extensive worldwide sharing of clinical data, they can result in a lot of distress for clinical geneticists and carriers, and even result in unnecessary surgeries [1; 2]. Therefore, to complement genetic test results, additional methods for interpreting the molecular effects of VUS are urgently required.

For the high-risk breast cancer susceptibility gene *PALB2*, currently 2202 VUS have been reported in ClinVar (as of August 2022), of which 1985 VUS constitute (rare) missense variants. One way to interpret such a large number of variants, is to perform computational predictions for impaired protein function. Although it is feasible to perform computational predictions *en masse*, many of these computational algorithms exhibit a high rate of false predictions, as has also been shown for missense variants in *PALB2* [3-6]. Another way to interpret *PALB2* VUS, is to perform functional analysis. As DNA double-stranded break repair by homologous recombination (HR) is a key tumour suppressive function of PALB2 [7-9], one commonly used assay to measure the functional effects of *PALB2* variants, is to measure their impact on HR efficiency. In an effort to address the functional consequences of genetic variants in *PALB2*, three recent studies have functionally characterized 155 unique missense variants in total, with most assays examining DNA repair by HR [3-6; 10; 11]. Although these assays have successfully identified several damaging missense variants in *PALB2*, these 'one-at-a-time' or semi high-throughput approaches, are often time and resource intensive. In addition, functional assays are generally performed after a variant is encountered in an individual, with results probably becoming public years later. For individuals carrying a damaging *PALB2* variant, functional results may then no longer be beneficial, at least with regards to therapeutic options. Lastly, whether an identified damaging missense variant in *PALB2* will actually associate with increased (breast) cancer risk is also unclear, since many of the identified damaging variants are present in only a few carriers, making it extremely difficult to associate these rare variants with breast cancer risk.

Here we aim to address these issues by linking the functional impact of *PALB2* missense variants to breast cancer risk using a burden type association analysis. For this, we initially employed our reported semi high-throughput approach [3], to functionally characterise 18 *PALB2* VUS identified in an Asian cohort [12], as well as 58 *PALB2* VUS identified within 44 BCAC studies combined [13]. In order to address the large numbers of functionally uncharacterized *PALB2* missense VUS, we further developed this approach to allow for functional analysis of *PALB2* variants *en masse*. To this end, we employed a cDNA variant

library for the Coiled-Coil (CC) domain of PALB2, in which several damaging missense variants have previously been identified [3; 5; 6], and used sensitivity to PARP inhibition (PARPi) as a functional readout. This allowed for the identification of numerous damaging missense variants in this domain of PALB2. Based on the case-control association study performed by Dorling et al. [14], and functional results from the semi high-throughput approach, we then show that functional impact of *PALB2* missense variants can be linked to increased breast cancer risk. Notably, fully damaging missense variants in *PALB2* appear to associate with a similar high risk for breast cancer as *PALB2* truncating variants.

RESULTS

Functional characterisation of rare *PALB2* missense variants identified in South East Asian populations

In a population-based study of 7,840 breast cancer cases and 7,928 healthy Chinese, Malay and Indian women from Malaysia and Singapore, 18 rare *PALB2* missense VUS were identified [12]. We evaluated the functional impact of these missense variants in our previously published mouse embryonic stem (mES) cell-based functional assay [3]. These results, which are presented in Figure 1, have been previously published [12]. Briefly, mES cells in which *Palb2* was deleted using CRISPR-Cas9 technology were complemented with human *PALB2* cDNA, with or without *PALB2* variant, through stable integration at the *Rosa26* locus [3]. By using the well-established DR-GFP reporter [15], which was integrated at the *Pim1* locus, HR was measured to evaluate the functional impact of all 18 missense variants in *PALB2* [3]. Two other variants (p.A38G and p.A38V) were included for comparison purposes. Of the 20 missense variants (Supplementary Table) in total, two variants (p.R37C and p.R37H) exhibited moderate HR activity (50-60%) (Fig. 1a). An impaired PALB2-BRCA1 interaction likely explains this defect, as well as the reduced recruitment of p.R37H to sites of DNA damage induced by laser micro-irradiation [3]. Interestingly, two other *PALB2* missense variants (p.L1027R and p.G1043V) exhibited >80% reduction in HR (Fig. 1a), indicating that they are similarly damaging as truncating *PALB2* variants [3].

As HR defects have been associated with sensitivity to PARPi [16], we next evaluated the effect of five *PALB2* missense variants that exhibited the largest defect in HR in DR-GFP assays, using a cellular proliferation assay. We found that the three variants exhibiting a mild to moderate impact on HR (i.e., p.R37C, p.R37H and p.A38V) (Fig. 1a), did not have a major impact on PARPi sensitivity. In contrast, p.L1027R and p.G1043V displayed strong sensitivity to PARPi (Fig. 1b), which is consistent with the HR efficiency measured with the DR-GFP reporter (Fig. 1a). As a consequence of the functional impact observed for both p.L1027R and

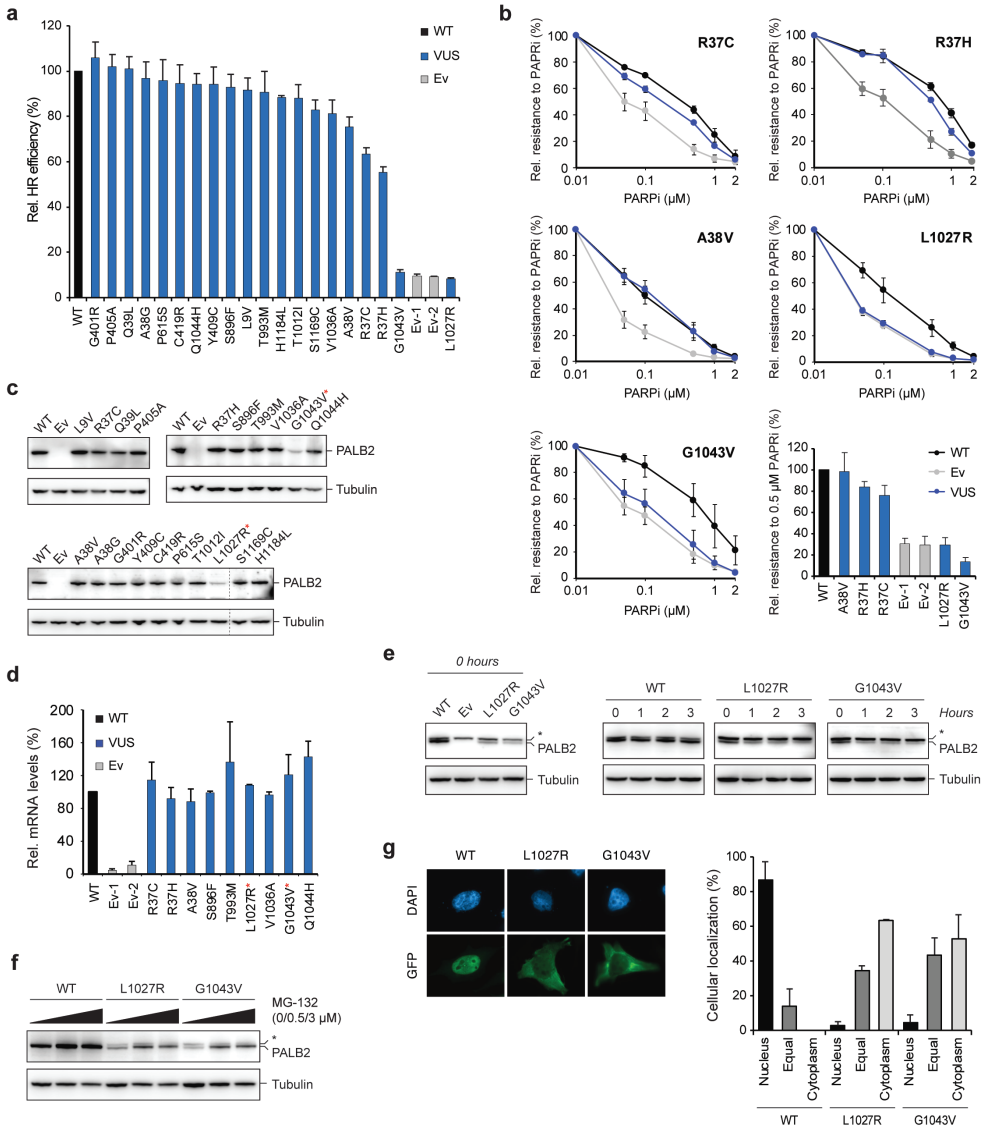


Figure 1. Functional analysis of *PALB2* missense variants from an Asian cohort. **a** HR assay (DR-GFP) in *Trp53*^{KO}/*PALB2*^{KO} mES cells expressing the indicated *PALB2* variants (or an empty vector, Ev). Normalized values are plotted with the wild type (WT) condition set to 100% (absolute HR efficiencies for cells expressing wild type *PALB2* were in the range ~7-10% [3]). **b** Proliferation-based PARP inhibitor (PARPi) sensitivity assay using mES cells expressing the indicated *PALB2* variants (or an empty vector, Ev). The bar graph showed the relative viability/resistance to 0.5 μM PARPi treatment, for all 5 variants. **c** Western blot analysis for the expression of all *PALB2* variants analysed. **d** RT-qPCR analysis of selected *PALB2* variants. Primers specific for human *PALB2* cDNA and the mouse Pim1 control locus were used. Tubulin is a loading control. **e** Western blot analysis of PALB2 protein abundance for the indicated variants in the absence of cycloheximide (CHX) and after the indicated time of incubation in

the presence of 100 µg/ml CHX. Tubulin is a loading control. Asterisk indicates an nonspecific band. **f** Western blot analysis of PALB2 protein abundance for the indicated variants after 24-hour incubation with the indicated concentrations of MG-132. Tubulin is a loading control. Asterisk indicates an nonspecific band. **g** Immunofluorescence analysis and quantification of the nucleocytoplasmic distribution of EGFP-PALB2, with or without the indicated variants, following transient expression in HeLa cells. Data represent the mean percentages (±SEM) from at least 3 independent experiments. For all bar plots in (a), (b) and (d), data represent the mean percentages (±SEM) of parameter under investigation with value relative to wild-type, which was set at 100% (i.e., GFP positive cells (a), viability/resistance (b) and mRNA (d) from at least 2 independent experiments). Variants/conditions are categorized by colour as either wild-type (black), VUS (blue) or Ev (grey). Ev1-2 refer to Ev controls from 2 different replicates. Variants with low expression levels are indicated in red *.

p.G1043V, they may associate with increased risk of breast cancer and serve as targets for PARPi-based therapy.

To complement the DR-GFP and PARPi sensitivity assays, we examined protein expression levels for all 20 *PALB2* missense variants. Consistent with the functional impact observed for p.L1027R and p.G1043V, both variants showed strongly reduced expression levels in comparison to wild type *PALB2* (Fig. 1c), suggesting that these two variants negatively affect *PALB2* protein levels. mRNA analysis subsequently showed that the transcript levels of several variants, including p.L1027R and p.G1043V, were similar to that of the wild type complemented condition, suggesting that the weak expression of p.L1027R and p.G1043V is likely due to protein instability (Fig. 1d). To examine this further, we performed cycloheximide chase experiments to halt protein synthesis and assess *PALB2* protein levels over time. While wild type *PALB2* protein levels remained stable over a 3 hour time span after cycloheximide treatment, both p.L1027R and p.G1043V showed marked reductions in protein levels compared to the 0 hour timepoint (Fig. 1e). These data provide evidence that p.L1027R and p.G1043V impair *PALB2* protein function through protein instability. Treatment with the proteasome inhibitor MG-132 further showed that *PALB2*, with or without the p.L1027R or p.G1043V variant, is subjected to proteasome-dependent degradation (Fig. 1f). Most likely as a result of protein instability and subsequent proteasomal degradation in the cytoplasm, both the p.L1027R and p.G1043V variants mis-localised in the cytoplasm (Fig. 1g). These data are concordant with previous localisation data for *PALB2* variants in the WD40 domain, such as p.I944N and p.T1030I, which have also been reported to be unstable and mis-localise in the cytoplasm [3; 5; 6], thereby impacting HR. However, given that several proteins involved in HR, including BRCA2 and RNF168, interact with PALB2's WD40 domain [7; 17; 18], we cannot exclude the possibility that these variants also impact HR by affecting the interaction between PALB2 and these proteins. Nonetheless, the defects for p.L1027R and p.G1043V in HR and PARPi sensitivity are similar to those observed for the empty vector conditions and compare to those previously reported for pathogenic *PALB2* truncating variants [3]. Accordingly, these variants

may associate with a high risk for breast cancer similar to that observed for *PALB2* truncating variants.

Functional characterisation of *PALB2* missense variants identified in 44 studies of the Breast Cancer Association Consortium (BCAC)

In order to estimate the risks of breast cancer associated with rare germline missense variants in genes such as *PALB2*, germline DNA samples from 60.466 women with breast cancer and 53.461 controls participating in 44 BCAC studies (14 family-based and 30 population-based studies), were sequenced [13]. These efforts led to the identification of 567 distinct *PALB2* missense variants of which most are considered VUS with unknown effects on PALB2 protein function. Out of these 567 missense variants, we selected 58 *PALB2* missense VUS (Supplementary Table) for semi high-throughput functional analysis. Selection was based on one or more of the following criteria; (i) position throughout the *PALB2* protein sequence (ii) frequencies of these variants in cases and controls in the 44 BCAC studies [13] and (iii) computational predictions from Helix (i.e., mostly variants that were predicted to be damaging) [19-21]. Four additional VUS (p.R239del, p.M416V, p.S771G, p.R976S) and one truncating variant (p.S201fs), were gathered from ClinVar. Interestingly, two damaging missense VUS (p.W912S and p.L1026P) were identified with HR efficiencies comparable to truncating variants (i.e., <12% HR) [3]. In addition, 7 missense VUS (p.L24W, p.R34L, p.L897R, p.G937E, p.R976G, p.R976S and p.Y1183D) exhibited intermediate functionality (i.e., 12-75% HR). All other *PALB2* VUS exhibited HR efficiencies comparable to cells expressing wild type *PALB2* (Fig. 2a), or previously studied likely benign missense variants [3].

Next, we examined the effect of 25 selected *PALB2* variants on PARPi sensitivity. Their selection was based on the observation that these variants exhibited variable degrees of functional impact in the DR-GFP assay (Fig. 2a). We observed that three VUS (p.L897R, p.W912S, and p.L1026P), displayed sensitivity to PARPi treatment comparable to that observed for empty vector conditions and *PALB2* truncating variants [3], while four VUS (p.L24W, p.G937E, p.R976G and p.Y1183D) displayed intermediate sensitivity (i.e., 35-75% resistance to PARPi) (Fig. 2b). Consequently, we observed a strong positive correlation (R^2 = 0.76, p = <0.0001) between DR-GFP and PARPi sensitivity assays for these selected *PALB2* VUS (Fig. 2c).

Consistent with results observed for *PALB2* variants such as p.L1027R and p.G1043V (Fig. 1a-d), western blot analysis for selected *PALB2* variants from this set showed low expression levels for all variants residing in the WD40 domain that were functionally damaging or intermediate (p.L897R, p.W912S p.G937E, p.R976G, p.L1026P and p.Y1183D) (Fig. 2d). The expression levels for functionally intermediate p.L24W was comparable to the expression level of wild type *PALB2,* and we therefore hypothesize that the functional impact of this variant

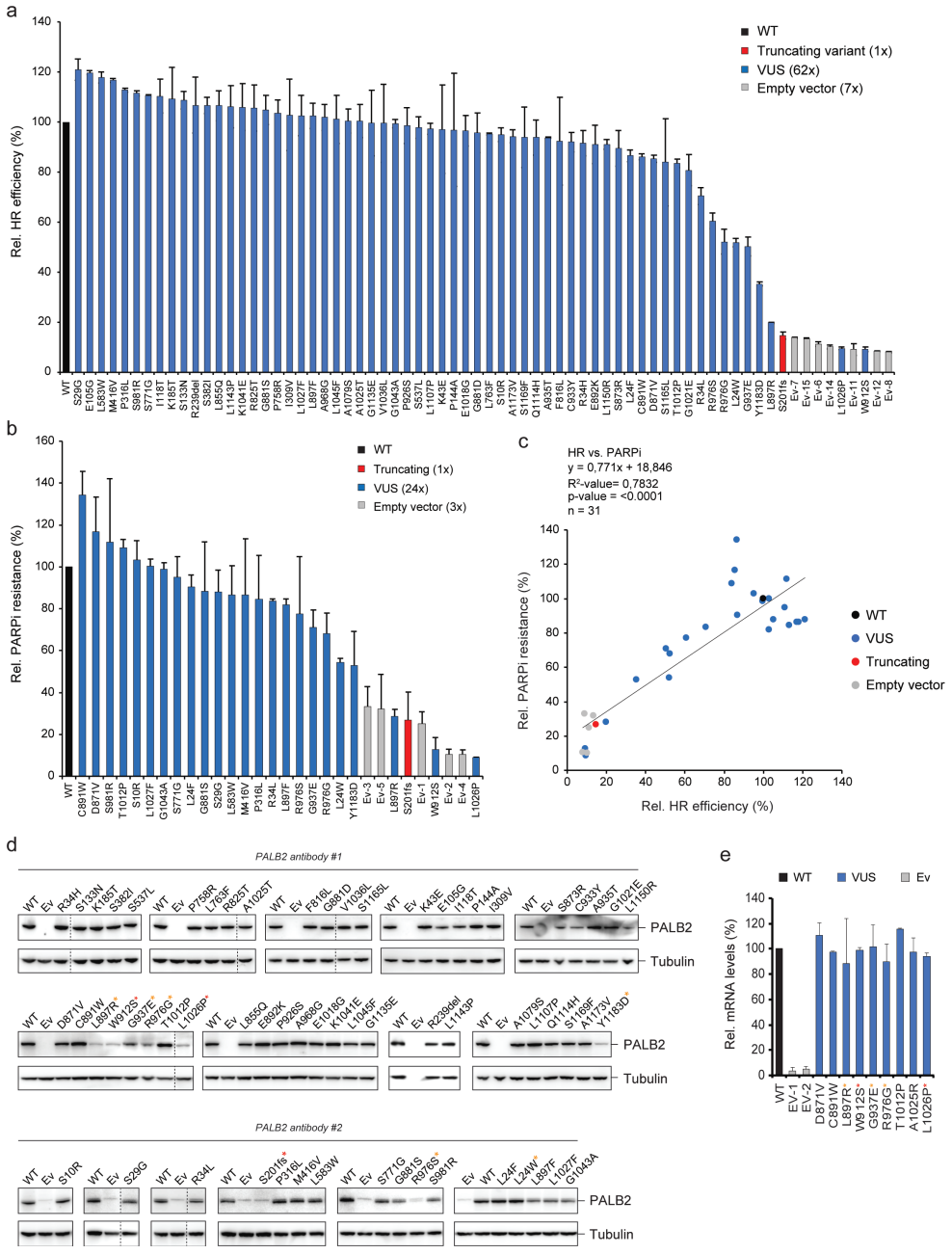


Figure 2. Functional analysis of *PALB2* missense variants identified in 44 BCAC studies. **a** HR assay as in (Fig. 1a). **b** Proliferation-based PARP inhibitor (PARPi) sensitivity assay as in (Fig. 1b). **c** Scatter plot showing the correlation between HR efficiencies and PARPi sensitivity for variants measured in both (a) and (b), respectively. Variants/conditions are categorized by colour as indicated. **d** Western blot analysis for the expression of all *PALB2* variants analysed in (a). Tubulin is a loading control. **e** RT-

qPCR analysis of selected *PALB2* variants as in (Fig. 1d). The asterisk indicates variant functionality observed in Figure 1a; none (functional), orange (intermediate), red (damaging).

may be due to somewhat reduced interaction with BRCA1, as previously shown for p.L24S [3]. For nine selected *PALB2* VUS, including the five VUS that displayed reduced protein levels (Fig. 2d, red asterisk), we subsequently quantified mRNA transcript levels, which for all variants compared well to those of wild type *PALB2* (Fig. 2e). Again, this suggests that for the five variants that are located in the WD40 domain and display low abundance of PALB2 protein levels (Fig. 2d, red asterisk), the variants result in protein instability, as we have confirmed for p.L1027R and p.G1043V using cycloheximide assays (Fig. 1e). Overall these data suggest that the WD40 domain of PALB2 is exceptionally sensitive to variants that affect protein stability and consequently HR.

A multiplex assay for measuring the functional effect of *PALB2* missense variants in the CC domain

Currently 1985 *PALB2* missense VUS have been reported in ClinVar (as of August 2022). As a one-by-one approach for functionally characterizing such a large number of *PALB2* missense variants is not feasible, high-throughput assays, such as those performed for *BRCA1* [22; 23], are strongly desired [24]. Here we developed a high-throughput strategy for the analysis of missense variants in *PALB2* (Fig. 3a). To this end, we obtained a variant library for the CC domain of PALB2 (amino acid 9-43), containing 667 variants out of the 700 variant possible nonsense and missense variants that can be introduced in this domain (Fig. 3b). We introduced this variant library in our *Palb2*^{KO} mES cells by RMCE and pooled the neomycin resistant clones each expressing a single *PALB2* variant. On the pool of cells, we performed PARPi sensitivity assays in triplicate and included non-treated cells as control conditions (Fig. 3a). The region of the *PALB2* cDNA coding for the CC domain was then amplified and sequenced. For each variant, depletion scores and standard errors were calculated by the computational Enrich2 software tool, which are based on the ratio of variant frequencies before and after PARPi treatment and the consistency between replicate measurements. The scores calculated by Enrich2 include a normalization to wild type *PALB2,* which was set to '0', followed by a normalization to the average score of the nonsense variants, which was set to '-1'.

A characteristic of high-throughput assays to functionally measure variant effects is that they are inherently noisy and that the variance in scores is particularly high for variants with low read counts. For each integration experiment, we therefore excluded such variants from the analysis by using a threshold based on the three PARPi replicates and the standard

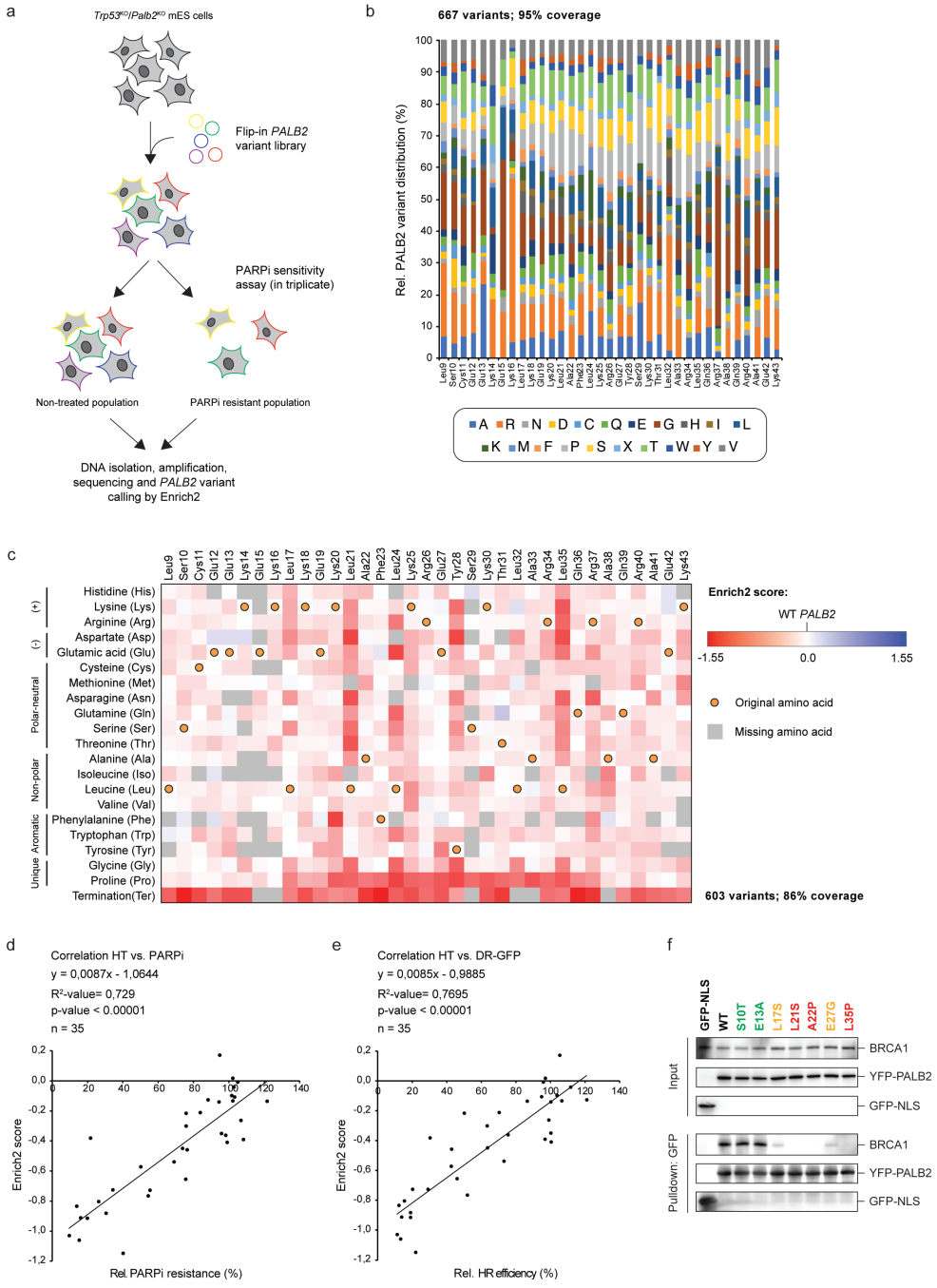


Figure 3. High-throughput analysis of *PALB2* variants in the CC domain. **a** Schematic flow of the high-throughput functional analysis employed in this study. **b** Bar graph showing the variant diversity distribution of the CC-variant library containing 667 distinct *PALB2* variants. **c** Amino acid function map

of the CC domain of *PALB2*. Amino acid characteristics are indicated at the left of the plot. Dark red squares represent variants that were depleted in PARPi treated conditions versus untreated conditions. Blue squares represent variants that were (potentially) enriched. Grey squares represent variants for which data is not available. Orange dots represent the original wild type amino acids. **d** The correlation between single PARPi sensitivity assays for previously characterized CC-variants and scores from the high-throughput (HT) assay in (c). The correlation and significance is indicted at the top of the plot. **e** The correlation between DR-GFP assays for previously characterized CC-variants and scores from the high-throughput assay in (c). The correlation and significance is indicted at the top of the plot. **f** YPF/GFP pulldowns of the indicated PALB2 variant proteins following transient expression in U2OS cells. *PALB2* variants are indicated in three colours reflecting their functional outcome in the high-throughput analysis in (c); green is functional, orange is intermediate, red is damaging. GFP-NLS and YFP-PALB2-L35P served as negative controls. Western blot analysis was performed using antibodies against GFP and BRCA1.

error (SE) calculated by Enrich2 (i.e., variants with an SE >0.5 were excluded). For each integration, the number of variants passing this SE-based filter varied. However, for 603 variants we were able to obtain scores from all six library integration experiments, which translates to a variant coverage of 86%. As expected, synonymous *PALB2* variants as a group were barely depleted, if at all, after treatment with PARPi (i.e., Enrich2 score of -0.11; SE ±0.07). In contrast, all recovered nonsense variants (n=29) displayed strong depletion after treatment with PARPi (i.e., Enrich2 scores <-0.58) (Fig. 3c). Among the *PALB2* missense variants, 67 exhibited scores that were within the range of the 29 nonsense *PALB2* variants; i.e., scores below that of p.Y28X, which was the least depleted variant of the 29 nonsense variants. This suggests that these *PALB2* missense variants may be just as damaging as the nonsense variants. Consistently, this list includes p.L35P which is listed as likely pathogenic in ClinVar. For further validation, we correlated the high-throughput Enrich2 scores to the relative PARPi resistance levels measured in semi high-throughput assays (Fig. 1a, Fig. 2a) and in two previous studies (Supplementary Table; n=35) [3; 12]. This showed that there is a good and significant correlation between the outcomes of the high-throughput and semi high-throughput approaches (R^2=0.73, p<0.0001) (Fig. 3d). Consistently, we observed a similarly good correlation between the PARPi sensitivity-based high-throughput outcomes and those obtained with the semi high-throughput DR-GFP reporter-based approach (R^2=0.77, p<0.0001) (Fig. 3e). Lastly, we show that variants that impact PALB2 protein function, do so by affecting the interaction with BRCA1, as shown in co-immunoprecipitation experiments (Fig. 3g). While two functional variants had no effect on this interaction, two intermediate variants (p.L17S and p.E27G) had a moderate effect on the interaction. Moreover, two damaging variants (p.L21S and p.A22P) completely impaired the interaction to the same extent as p.L35P, the latter of which was included as a negative control [3; 25]. Altogether, these data validate our high-throughput assay and its value in functionally characterizing *PALB2* missense variants in the CC domain.

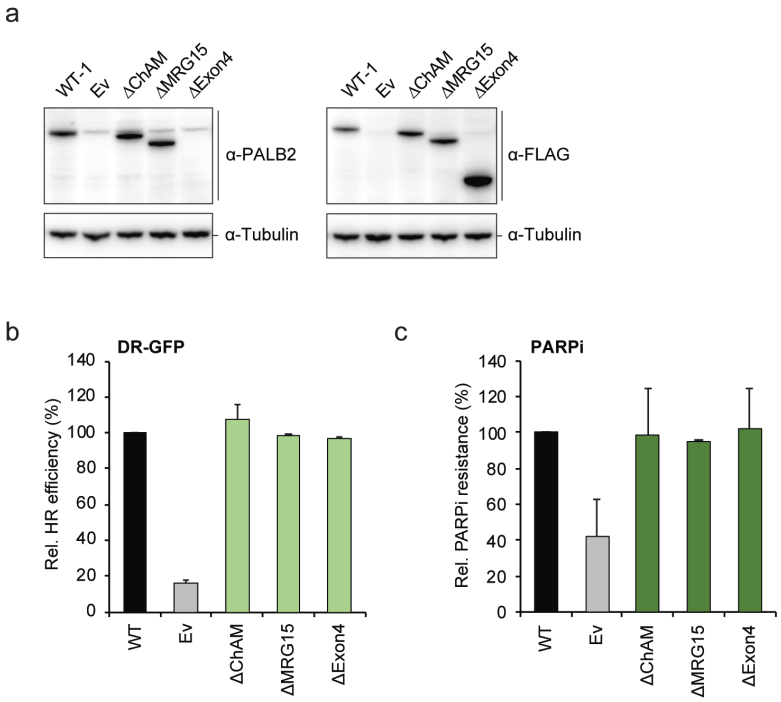


Figure 4. Functional analysis of *PALB2* deletion variants. **a** Western blot analysis for the expression of three *PALB2* domain deletion variants as indicated. Tubulin is a loading control. **b** HR assay as in (Fig. 1a). **c** Proliferation-based PARP inhibitor (PARPi) sensitivity assay as in (Fig. 1b).

The ChAM and MRG15 functional domains are dispensable for HR

No missense variants outside of the CC and WD40 domains of PALB2 have thus far been identified as damaging [3; 5; 6]. To assess the requirement of the ChAM and MRG15 domain of PALB2 for HR, or of less conserved regions that are part of PALB2's large exon 4, we generated three PALB2 deletions constructs, ΔChAM, ΔMRG15 and ΔExon4 (Supplementary Table), and assessed HR using the DR-GFP reporter. All three *PALB2* deletion variants exhibited HR efficiencies comparable to that in cells expressing wildtype *PALB2* (Fig. 4a, b). Consistently, the expression of these deletions constructs also did not confer PARPi sensitivity (Fig. 4c). These data suggest that these regions are dispensable for PALB2's function in HR, and decreases the likelihood that damaging missense variants in these regions will be identified. However, we cannot rule out that variants in these regions impact protein functionality by affecting mRNA splicing.

Association between functional defects in PALB2 and breast cancer risk

Having determined the functional impact of VUS in *PALB2*, we next investigated whether the observed impact correlates with increased cancer risk. For this, we considered all 60.466 breast cancer cases and 53.461 controls of the case-control association study performed by Dorling et al. [14]. Out of all *PALB2* missense VUS functionally characterized here (Fig. 1a and 2a) or in two previous studies [3; 12], case-control carrier frequencies were reported for 89 VUS [13]. In order to allow for correlation of PALB2 functional defects with breast cancer risk, we next combined the case-control frequencies for several groups of *PALB2* VUS, where grouping was based on the measured HR efficiency. *PALB2* variant groups exhibiting 12-50% HR, or a higher efficiency in HR, all associated with an OR close to 1, suggesting there is no increased risk (Table 1). Interestingly, *PALB2* VUS that can be considered completely damaging (Fig. 5, HR <12% 'pathogenic' threshold [3]), based on similar HR efficiencies as measured for *PALB2* truncating variants, associated with an OR comparable to that what has been reported for *PALB2* truncating variants (OR 6.19; 95% CI, 0.76-50.31; p = 0.0882) [13; 26]. Including *PALB2* VUS with an HR efficiency up to ~20% in this group, strongly reduced the associated risk (OR 3.54; 95% CI, 0.75-16.66; p = 0.1101) (Table 1). Although none of these *PALB2* variant groups associated with significantly increased breast cancer risk (Table 1), this burden-type association analysis suggests that decreased HR efficiency correlates with increased breast cancer risk. It should, however, be noted that 19.3% of the 60.466 breast cancer cases and 5.2% of the 53.461 controls from the BRIDGES case-control association study stem from family-based studies in which patients were oversampled [13]. This may have resulted in a bias in the calculated cancer risk. Nonetheless, based on these data we estimate that only variants exhibiting <20% HR will associate with a moderate to high risk for breast cancer.

Table 1. Burden-type cancer risk association analysis for human *PALB2* variants.

Variant group based on HR range (%)	Nr. of distinct variants	Nr. cases	Nr. controls	Odds ratio	95% CI	*p*-value
6% - 121%	64	176	157	0.99	0.80-1.23	0.9355
75% - 121%	46	90	100	0.80	0.60-1.06	0.1161
50% - 75%	9	72	50	1.27	0.89-1.83	0.1898
12% - 75%	13	79	56	1.25	0.89-1.76	0.2062
12% - 50%	4	7	6	1.03	0.35-3.07	0.9555
6% - 50%	9	14	7	1.77	0.71-4.38	0.2182
6% - 20%	7	8	2	3.54	0.75-16.66	0.1101
6% - 12%	5	7	1	6.19	0.76-50.31	0.0882

Variants are grouped based on their efficiency in HR, as measured with the DR-GFP reporter. Variants and data previously reported in Boonen et al., 2019 (ref 3) has been included in this analysis. The case-control frequencies reflect those from all 44 BCAC studies (60466 cases and 53461 controls); i.e., 30

population-based studies and 14 familiy-based studies reported in Dorling et al., (ref 13). 27 *PALB2* missense VUS that were selected for functional analysis on the basis of their reported case-control frequencies were excluded from this analysis.

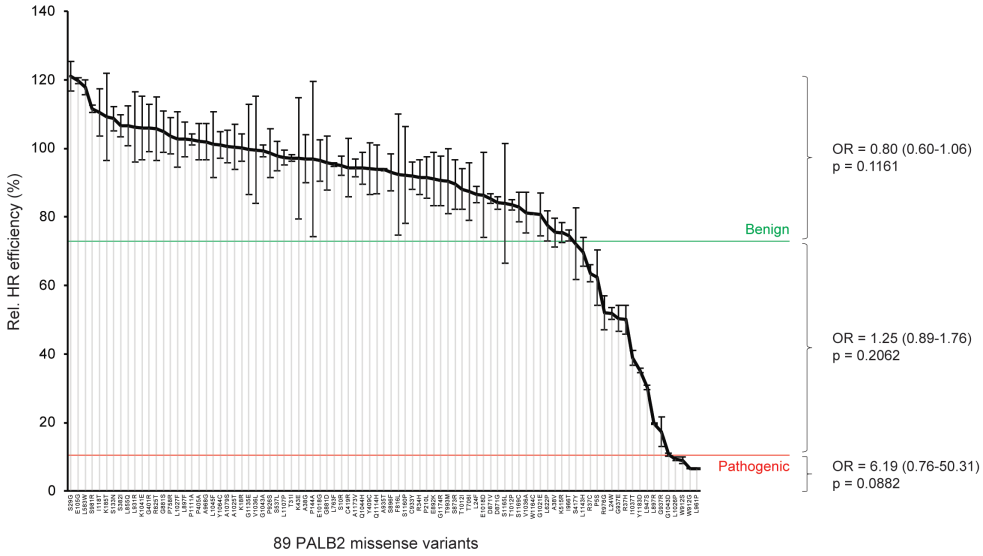


Figure 5. Association of PALB2 HR efficiency with breast cancer risk. Bar graph showing results from HR assays (DR-GFP) in *Trp53*^{KO}/*PALB2*^{KO} mES cells complemented with 89 distinct human *PALB2* variants. Previously published results [3], as well as those from (Fig. 1a) and (Fig. 2a) are shown. OR estimates are based on the case-control association study from the Breast Cancer Association Consortium [13] and are shown for three *PALB2* variant groups, based on HR efficiency, as indicated.

DISCUSSION

The three recent studies that functionally characterized a total of 155 unique *PALB2* variants [3-6; 10; 11], represent a milestone for the clinical management of individuals carrying *PALB2* genetic VUS. However, many more VUS in *PALB2* remain functionally uncharacterized and an actual correlation between functional impact of *PALB2* missense variants and cancer risk is still lacking. To build on these previous studies and address this issue, we present here additional data from different approaches aimed at interpreting (rare) *PALB2* genetic missense variants.

Using a semi high-throughput approach, we systematically assessed the HR activities of 82 *PALB2* missense variants and one truncating variant by performing HR-based assays (Fig. 1a, Fig. 2a). The four damaging missense variants identified with this approach (p.W912S, p.L1026P, p.L1027R, p.G1043V) all locate to the C-terminal WD40 domain of

PALB2. This is consistent with previous studies in which damaging missense variants in *PALB2* have only been identified in the CC and WD40 domain [3-6; 25]. Although these studies are in strong support of PALB2 protein instability as a consequence of these variants [3], here we provide more conclusive evidence for such a mechanism of action using cycloheximide, proteasome inhibitor and cellular localization assays for PALB2 p.L1027R and p.G1043V. Most likely as a result of protein instability (Fig. 2e), these variants are subjected to proteasomal degradation (Fig. 2f) and mis-localize to the cytoplasm (Fig. 1g). This prevents PALB2's transport into the nucleus and consequently hampers HR-mediated DNA repair. These data are highly consistent with previous reports on *PALB2* WD40 variant 'p.I944N', for which instability and mis-localization was also shown [6].

Our high-throughput functional analysis allowed us to measure the functional effects of 603 variants (574 missense and 29 nonsense) in the CC domain of *PALB2*, by assessing sensitivity to PARPi treatment. These results showed a strong correlation with those from DR-GFP assays (Fig. 3e), thereby validating this approach. Furthermore, most damaging missense variants concerned amino acid residues (i.e., p.L21, p.L24, p.Y28 and p.L35) for which damaging variants have already been reported [3-6; 25; 27; 28]. Altogether, these results allowed for the functional characterization of 62 out of 65 *PALB2* CC missense VUS that are listed in ClinVar. For instance, 6 VUS (p.L21S, p.A22P, p.L24S, p.Y28N, p.L32P, p.A33P; Enrich2 scores <0.58) appeared to be just as damaging as nonsense variants in the CC domain (Fig. 3c), whereas 13 VUS (p.E19D, p.K20I, p.E27G, p.Y28C, p.K30E, p.R37C, p.R37G, p.R37S, p.R37L, p.A38V, p.R40I, p.K43E, p.K43N; depletion scores between -0.30 and -0.58) showed intermediate functionality. Importantly, our high-throughput results may contribute to the clinical re-classification of these VUS in ClinVar. However, with regard to functional analysis being used as clinical diagnostic tools, especially those involving 'relatively noisy' high-throughput assays, it is important to consider using results from several distinct functional assays. Ideally, these assays have been performed in different research labs, have used different experimental strategies, and include the possibility of mRNA transcript analysis in order to provide insight into the effect of variants on RNA splicing. In that regard, it is important to note that for all variants analyzed here, possible effects on splicing were not examined.

Although we have established assays allowing the functional characterization of *PALB2* VUS, a major challenge is still to translate functional effects into estimates for cancer risk. The burden-type association analysis presented in this study suggests that damaging missense variants as a group (exhibiting 6-12% HR [3]) (Fig. 5, Table 1) may be associated with an increased risk for breast cancer (OR 6.19; 95% CI, 0.76-50.31; p=0.065) that is comparable to that reported for truncating *PALB2* variants[13; 26]. However, due to the low number of case control frequencies associated with this variant group, the increased risk was not significant

and thus the exact risk remains to be established. In order to address this further, data from larger case control association studies (compared to Dorling et al., [13]) are required, or data from large case control association studies need to be combined. Additionally, the burden-type association analysis would improve with the identification of more intermediate and damaging *PALB2* missense variants. Extending the high-throughput strategy that we applied to the CC domain of PALB2 (Fig. 3c) to other regions, may result in the identification of more damaging missense variants for which the associated cancer risk could be established. This is exemplified by the observation that only 19 out of the 567 missense variants identified in the study from Dorling et al. [13], located within the CC domain of PALB2 and yielded functional data through our high-throughput analysis (Fig. 3c). Disappointingly, this resulted in the identification of only one additional uncharacterized VUS with a large functional impact (i.e., p.L24W). Therefore, it is imperative that high-throughput assays are performed for the WD40 domain of PALB2, which is ten times larger than the CC domain and which was previously shown to be a "hotspot" for damaging variants [3; 5; 6]. In this domain, Dorling et al. identified 176 missense variants [13]. Moreover, 580 out of the 1985 *PALB2* missense VUS listed in ClinVar (as of August 2022), locate to the WD40 region. High-throughput analysis may be a feasible way to functionally characterize such a large number of variants, provide evidence for re-classification and pave the way for cancer risk association analysis.

To facilitate clinical classification of genetic variants, the American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) have proposed variant interpretation guidelines that incorporate different types of evidence (including functional assessment) at various levels of strength. These guidelines also provide rules for combining the different types of evidence to result in a final classification (benign, likely benign, uncertain significance, likely pathogenic, pathogenic), each with defined clinical significance [29; 30]. So when clinical evidence such as phenotypic data, population frequency and segregation analysis in scarce or insufficient, functional data can be extremely valuable for clinical classification of genetic variants. Accordingly, high-throughput results may contribute to the clinical re-classification of many reported VUS, as well as variants that will undoubtedly be identified in the future. Furthermore, the association of functional impact of missense VUS with cancer risk, will ultimately be crucial for clinical interpretation of these rare missense variants. Classification of VUS to a category with a defined clinical significance is of great importance to carriers of these variants. This will help them to make an informed decision on how to manage their cancer risk. The work presented here for the *PALB2* gene may aid in making such informed decisions.

MATERIALS AND METHODS

Cell culture and generation of *Trp53*^{KO}/*Palb2*^{KO} mES cells with DR-GFP and RMCE

Trp53^{KO}/*Palb2*^{KO} mES cells carrying the DR-GFP reporter and RMCE system at the *Pim1* and *Rosa26* locus, respectively, were generated previously [3] and cultured as previously described[19].

Cloning and site-directed mutagenesis of human *PALB2* cDNA variants

The RMCE vector (pRNA-251-MCS-RMCE) (TaconicArtemis GmbH) containing *PALB2* cDNA driven by an Ef1α promotor was generated previously [3]. *PALB2* variants were introduced by site-directed mutagenesis using the Quick-Change Lightning protocol (Agilent Technologies). Constructs were verified by Sanger sequencing and used for downstream mES cell-based assays.

HR reporter assays

HR assays using 2x10^6 *Trp53*^{KO}/*Palb2*^{KO} mES cells carrying the DR-GFP reporter and RMCE system were performed as previously described [3]. Briefly, cells that were complemented with human *PALB2* cDNA with or without a variant (or an empty vector), were treated with neomycin to select for cells with integrated *PALB2* variant cDNA. Two days after transfection of an *I-SceI* and mCherry co-expression vector [31], GFP expression was measured using fluorescence-activated cell sorting (FACS).

Western blot analysis

Expression of all *PALB2* variants was examined by Western blot analysis as previously described [3]. Two different primary rabbit polyclonal antibodies directed against the N-terminus of human PALB2 (1:1000, kindly provided by Cell Signalling Technology prior to commercialization) were used. Wild type human *PALB2* and empty vector (Ev) were used as controls on the blot, while tubulin (Sigma, T6199 clone DM1A) was used as loading control.

For protein stability and degradation assays, cells were treated with 100 μg/ml cycloheximide (Sigma, C7698-1G) for up to 3 hours, or 0.5 or 3 μM MG-132 (Selleckchem, S2619) for 24 hours, after which western blot samples were collected and analysed.

Cellular localization assay

Quantification of EGFP-PALB2 subcellular localization was based on transient expression in HeLa cells that were fixed using 4% formaldehyde and permeabilized using Triton X-100. Cells were immunostained with anti-GFP and DAPI prior to immunofluorescence analysis and

quantification (based on ~25 cells per condition per replicate). Assays were conducted in duplicate and average values and SEM were calculated to generate the respective plots.

RT-qPCR analysis

RT-qPCR was performed for a selected panel of *PALB2* variants as previously described [3]. Briefly, RNA was isolated using Trizol (ThermoFisher, 15596026), and DNAse (Promega, M6101). Subsequently, reverse transcriptase (ThermoFisher, 12328019) reactions were performed. GoTaq qPCR Master mix (Promega, A6002) and the following qPCR primers directed at the human *PALB2* cDNA or the mouse control gene *Pim1* were used; human PALB2-N-term-Flag-Fw—5'-GATTACAAGGATGACGACGATAAGATGGAC-3', human PALB2-exon2-Rv—5'-CCTTTTCAAGAATGCTAATTTCTCCTTTAACTTTTCC-3', mouse Pim1-exon4-Fw—5'-GCGGCGAAATCAAACTCATCGAC-3', and mouse Pim1-exon5-Rv—5'-GTAGCGATGGTAGCGAATCCACTCTGG-3'.

Pulldown assays

Pulldown assays were performed as previously described [3]. Briefly, 20 µg pYFP-PALB2 plasmid [32] was transfected into ~10 x 10^6 U2OS cells on a 15 cm dish using Lipofectamine 2000. The next day cells were trypsinized, and lysed in 1 ml EBC buffer (50mM Tris pH 7.3, 150mM NaCl, 0.5% NP-40, 2.5mM MgCl2) containing 1 tablet protease inhibitor (Roche) per 10 ml buffer. Lysates were incubated with benzonase and centrifuged. The supernatant was then added to 25 µl of pre-washed GFP-trap beads (ChromoTek) and incubated for 1.5 hours at 4 °C on a rotating wheel. The beads were washed 5–6 times with EBC buffer and eventually resuspended in 25 µl Laemmli buffer after which about half of each sample was analysed by western blot analysis using an antibody against human BRCA1 (1:1000).

***PALB2* CC-variant library integration**

The *PALB2* CC-variant library concerning amino acid residues 9-43, was integrated in 100x10^6 *Trp53*^{KO}/*Palb2*^{KO} mES cells. Cells were divided in fractions of 10x10^6 cells for which each fraction was subjected to co-transfection of 1 µg FlpO expression vector (pCAGGs-FlpO-IRES-puro) [33] with 1 µg RMCE exchange vector (i.e., CC-variant library) as previously described [3]. Transfected cells were divided over twenty 10 cm tissue-culture plates and treated one day later with 50 mg/ml neomycin/G418 sulfate (ThermoFisher, 10131035) for 6-7 days. Resistant colonies expressing *PALB2* variant cDNAs were pooled (estimation of 50-100x10^3 colonies per CC-library integration), mixed well and plated over three 10 cm tissue-culture plates containing neomycin. Two plates were trypsined and stored at -80 degrees as backup and one plate was used for three replicate PARPi sensitivity assays.

PARPi sensitivity assays

Functional analysis of single *PALB2* variants using semi high-throughput proliferation-based PARPi Olaparib (Selleckchem, S1060) sensitivity assays was performed for selected *PALB2* missense variants as previously described [3]. Briefly, cells were exposed to various concentrations of PARPi for two days. Thereafter, cells were incubated for one more day in drug free media, after which viability was measured using FACS (using only forward scatter and sideways scatter).

PARPi sensitivity assays after *PALB2* CC-variant library integration were performed using $0.57x10^6$ cells seeded on a 6 cm tissue-culture plates. One day after seeding, cells were treated with PARPi Olaparib (Selleckchem, S1060) for two days, after which the medium was refreshed with drug-free medium and cells were cultured for one more day. A non-treated plate was taken along as a control at the start of seeding. DNA was eventually isolated from the surviving cells and subjected to next-generation sequencing.

***PALB2* CC-variant library amplification and sequencing**

The CC-region of the integrated human *PALB2* cDNA was amplified from 100ng genomic DNA. Reactions contained 2* Kapa HiFi MasterMix polymerase (KR0370), a forward primer located in front of the CC-region (5'-GATGTGTATAAGAGACAGCGAGCTCGGATCCACTAGTAACG-3'), and a reverse primer located in behind of the CC-region (5'-CGTGTGCTCTTCCGATCTCTGAGTGTTTTAGCTGCGGTGAG-3'). PCR was performed under the following conditions; 98 °C for 1 minute; 18 cycles of 98 °C for 20 seconds, 65 °C for 30 seconds, and 72 °C for 30 seconds; and 72 °C for 2 minutes. The reactions produced a 283 base amplicon specifically from integrated human *PALB2* cDNA. After clean up with Ampure XP beads (Beckman Coulter) the PCR product was checked on a Agilent BioAnalyzer 2100 HS chip. A second PCR with Illumina index primers was performed under the following conditions; 98 °C for 1 minute; 10 cycles of 98 °C for 20 seconds, 60 °C for 30 seconds, and 72 °C for 30 seconds; and 72 °C for 2 minutes. The resulting PCR products were equimolar pooled. All samples were sequenced on an Illumina MiSeq.

Variant scoring and analysis

FASTQ files for each sample were used as input for the software package Enrich2 [34]. Enrich2 was used to translate and count both the unique nucleotide and unique amino acid variants. Reads containing insertions, deletions or multiple amino acid substitutions were removed from the analysis. Amino acid variants producing unreliable/noisy results over the three PARPi-treatment replicates were filtered out based on the standard error (SE) calculated by Enrich2; i.e., variants with an SE >0,5 were excluded. The counts for each protein variant were

translated into an abundance score by Enrich2. These scores are based on the ratio of the frequency of each variant in the PARPi-treated population over its frequency in the non-treated population, and include a normalization to the wild type *PALB2* abundance, which was set to '0'. Six independent CC-library integration experiments were performed. Only variants that passed the SE-based filtering and were scored in all six replicate library integration experiments were retained in the analysis. This included 29 nonsense variants for which an average abundance score was calculated for each integration assay. All variant scores for each integration experiment were then normalized by setting the average score of the 29 nonsense variants to '-1' by using the following formula:

$$'\text{Norm. Enrich2 score}' = 2\,\frac{'Enrich2\ score' - 'Mean\ nonsense\ score\ (as\ neg.value)'}{'Mean\ nonsense\ score\ (as\ pos.value)' - 'Mean\ nonsense\ score\ (as\ neg.value)'} - 1$$

A final abundance score per variant was calculated by taking the mean of the normalized abundance scores across the six replicate library integration experiments. A standard error for each abundance score was calculated by dividing the standard deviation of the normalized values for each variant by the square root of the number of replicate library integration experiments (i.e., six). Final abundance scores were plotted in a heatmap using the matrix visualization and analysis software MORPHEUS [35].

REFERENCES

1. Murray, M.L., Cerrato, F., Bennett, R.L., and Jarvik, G.P. (2011). Follow-up of carriers of BRCA1 and BRCA2 variants of unknown significance: variant reclassification and surgical decisions. Genet Med 13, 998-1005.
2. Welsh, J.L., Hoskin, T.L., Day, C.N., Thomas, A.S., Cogswell, J.A., Couch, F.J., and Boughey, J.C. (2017). Clinical Decision-Making in Patients with Variant of Uncertain Significance in BRCA1 or BRCA2 Genes. Ann Surg Oncol 24, 3067-3072.
3. Boonen, R., Rodrigue, A., Stoepker, C., Wiegant, W.W., Vroling, B., Sharma, M., Rother, M.B., Celosse, N., Vreeswijk, M.P.G., Couch, F., et al. (2019). Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. Nat Commun 10, 5296.
4. Boonen, R., Vreeswijk, M.P.G., and van Attikum, H. (2020). Functional Characterization of PALB2 Variants of Uncertain Significance: Toward Cancer Risk and Therapy Response Prediction. Front Mol Biosci 7, 169.
5. Rodrigue, A., Margaillan, G., Torres Gomes, T., Coulombe, Y., Montalban, G., da Costa, E.S.C.S., Milano, L., Ducy, M., De-Gregoriis, G., Dellaire, G., et al. (2019). A global functional analysis of missense mutations reveals two major hotspots in the PALB2 tumor suppressor. Nucleic Acids Res 47, 10662-10677.
6. Wiltshire, T., Ducy, M., Foo, T.K., Hu, C., Lee, K.Y., Belur Nagaraj, A., Rodrigue, A., Gomes, T.T., Simard, J., Monteiro, A.N.A., et al. (2020). Functional characterization of 84 PALB2 variants of uncertain significance. Genet Med 22, 622-632.
7. Xia, B., Sheng, Q., Nakanishi, K., Ohashi, A., Wu, J., Christ, N., Liu, X., Jasin, M., Couch, F.J., and Livingston, D.M. (2006). Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. Mol Cell 22, 719-729.
8. Nepomuceno, T.C., De Gregoriis, G., de Oliveira, F.M.B., Suarez-Kurtz, G., Monteiro, A.N., and Carvalho, M.A. (2017). The Role of PALB2 in the DNA Damage Response and Cancer Predisposition. Int J Mol Sci 18.
9. Zhang, F., Ma, J., Wu, J., Ye, L., Cai, H., Xia, B., and Yu, X. (2009). PALB2 links BRCA1 and BRCA2 in the DNA-damage response. Curr Biol 19, 524-529.
10. Nepomuceno, T.C., Carvalho, M.A., Rodrigue, A., Simard, J., Masson, J.Y., and Monteiro, A.N.A. (2021). PALB2 Variants: Protein Domains and Cancer Susceptibility. Trends Cancer 7, 188-197.
11. Southey, M.C., Rewse, A., and Nguyen-Dumont, T. (2020). PALB2 Genetic Variants: Can Functional Assays Assist Translation? Trends Cancer 6, 263-265.
12. Ng, P.S., Boonen, R.A., Wijaya, E., Chong, C.E., Sharma, M., Knaup, S., Mariapun, S., Ho, W.K., Lim, J., Yoon, S.Y., et al. (2021). Characterisation of protein-truncating and missense variants in PALB2 in 15 768 women from Malaysia and Singapore. J Med Genet.
13. Breast Cancer Association, C., Dorling, L., Carvalho, S., Allen, J., Gonzalez-Neira, A., Luccarini, C., Wahlstrom, C., Pooley, K.A., Parsons, M.T., Fortuno, C., et al. (2021). Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. N Engl J Med 384, 428-439.

14. Breast Cancer Association, C., Dorling, L., Carvalho, S., Allen, J., Gonzalez-Neira, A., Luccarini, C., Wahlstrom, C., Pooley, K.A., Parsons, M.T., Fortuno, C., et al. (2021). Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. The New England journal of medicine.
15. Kass, E.M., Helgadottir, H.R., Chen, C.C., Barbera, M., Wang, R., Westermark, U.K., Ludwig, T., Moynahan, M.E., and Jasin, M. (2013). Double-strand break repair by homologous recombination in primary mouse somatic cells requires BRCA1 but not the ATM kinase. Proc Natl Acad Sci U S A 110, 5564-5569.
16. Li, A., Geyer, F.C., Blecua, P., Lee, J.Y., Selenica, P., Brown, D.N., Pareja, F., Lee, S.S.K., Kumar, R., Rivera, B., et al. (2019). Homologous recombination DNA repair defects in PALB2-associated breast cancers. NPJ Breast Cancer 5, 23.
17. Ducy, M., Sesma-Sanz, L., Guitton-Sert, L., Lashgari, A., Gao, Y., Brahiti, N., Rodrigue, A., Margaillan, G., Caron, M.C., Cote, J., et al. (2019). The Tumor Suppressor PALB2: Inside Out. Trends Biochem Sci 44, 226-240.
18. Luijsterburg, M.S., Typas, D., Caron, M.C., Wiegant, W.W., van den Heuvel, D., Boonen, R.A., Couturier, A.M., Mullenders, L.H., Masson, J.Y., and van Attikum, H. (2017). A PALB2-interacting domain in RNF168 couples homologous recombination to DNA break-induced chromatin ubiquitylation. Elife 6.
19. Boonen, R., Wiegant, W.W., Celosse, N., Vroling, B., Heijl, S., Kote-Jarai, Z., Mijuskovic, M., Cristea, S., Solleveld-Westerink, N., van Wezel, T., et al. (2022). Functional Analysis Identifies Damaging CHEK2 Missense Variants Associated with Increased Cancer Risk. Cancer Res 82, 615-631.
20. Heijl, S., Vroling, B., Bergh, T.v.d., and Joosten, H. (2020). Mind the gap: preventing circularity in missense variant prediction (https://doi.org/10.1101/2020.05.06.080424).
21. Vroling, B., and Heijl, S. (2021). White paper: The Helix Pathogenicity Prediction Platform (https://arxiv.org/abs/2104.01033).
22. Starita, L.M., Islam, M.M., Banerjee, T., Adamovich, A.I., Gullingsrud, J., Fields, S., Shendure, J., and Parvin, J.D. (2018). A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution Variants on Protein Function. Am J Hum Genet 103, 498-508.
23. Findlay, G.M., Daza, R.M., Martin, B., Zhang, M.D., Leith, A.P., Gasperini, M., Janizek, J.D., Huang, X., Starita, L.M., and Shendure, J. (2018). Accurate classification of BRCA1 variants with saturation genome editing. Nature 562, 217-222.
24. Starita, L.M., Ahituv, N., Dunham, M.J., Kitzman, J.O., Roth, F.P., Seelig, G., Shendure, J., and Fowler, D.M. (2017). Variant Interpretation: Functional Assays to the Rescue. Am J Hum Genet 101, 315-325.
25. Foo, T.K., Tischkowitz, M., Simhadri, S., Boshari, T., Zayed, N., Burke, K.A., Berman, S.H., Blecua, P., Riaz, N., Huo, Y., et al. (2017). Compromised BRCA1-PALB2 interaction is associated with breast cancer risk. Oncogene 36, 4161-4170.

26. Couch, F.J., Shimelis, H., Hu, C., Hart, S.N., Polley, E.C., Na, J., Hallberg, E., Moore, R., Thomas, A., Lilyquist, J., et al. (2017). Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. JAMA Oncol 3, 1190-1196.
27. Sy, S.M., Huen, M.S., and Chen, J. (2009). PALB2 is an integral component of the BRCA complex required for homologous recombination repair. Proc Natl Acad Sci U S A 106, 7155-7160.
28. Oliver, A.W., Swift, S., Lord, C.J., Ashworth, A., and Pearl, L.H. (2009). Structural basis for recruitment of BRCA2 by PALB2. EMBO Rep 10, 990-996.
29. Brnich, S.E., Abou Tayoun, A.N., Couch, F.J., Cutting, G.R., Greenblatt, M.S., Heinen, C.D., Kanavy, D.M., Luo, X., McNulty, S.M., Starita, L.M., et al. (2019). Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med 12, 3.
30. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (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, 405-424.
31. Bouwman, P., van der Gulden, H., van der Heijden, I., Drost, R., Klijn, C.N., Prasetyanti, P., Pieterse, M., Wientjens, E., Seibler, J., Hogervorst, F.B., et al. (2013). A high-throughput functional complementation assay for classification of BRCA1 missense variants. Cancer Discov 3, 1142-1155.
32. Bleuyard, J.Y., Buisson, R., Masson, J.Y., and Esashi, F. (2012). ChAM, a novel motif that mediates PALB2 intrinsic chromatin binding and facilitates DNA repair. EMBO Rep 13, 135-141.
33. Kranz, A., Fu, J., Duerschke, K., Weidlich, S., Naumann, R., Stewart, A.F., and Anastassiadis, K. (2010). An improved Flp deleter mouse in C57Bl/6 based on Flpo recombinase. Genesis 48, 512-520.
34. Rubin, A.F., Gelman, H., Lucas, N., Bajjalieh, S.M., Papenfuss, A.T., Speed, T.P., and Fowler, D.M. (2017). A statistical framework for analyzing deep mutational scanning data. Genome Biol 18, 150.
35. Morpheus, https://software.broadinstitute.org/morpheus.

SUPPLEMENTARY INFORMATION

Functional interpretation of *PALB2* missense variants and their association with breast cancer risk

Rick A.C.M. Boonen, Sabine C. Knaup, Roberta Menafra, Dina Ruano, Pei Sze Ng, Soo Hwang Teo, Noel F. de Miranda, Maaike P.G. Vreeswijk, Susan L. Kloet and Haico van Attikum

The supplementary information contains:

- Supplementary Table

Supplementary Table. Complete list of human *PALB2* variants analyzed in this study.

cDNA annotation	Variant (aa)	Variant type	Average HR	SEM (HR)	Average PARPi	SEM (PARPi)	Nr. cases	Nr. controls
Variants Figure 1a,b (CC variants also used in Fig 3d,e)								
c.25C>G	L9V	Missense	91,68	5,24	n/a	n/a	n/a	n/a
c.109C>T	R37C	Missense	63,51	2,49	75,73	9,84	3	3
c.110G>A	R37H	Missense	55,27	2,37	83,85	4,88	5	2
c.113C>T	A38V	Missense	75,46	4,22	98,28	17,69	0	1
c.113C>G	A38G	Missense	96,95	7,04	n/a	n/a	4	5
c.117A>T	Q39L	Missense	100,98	5,39	n/a	n/a	n/a	n/a
c.1201G>C	G401R	Missense	105,92	6,87	n/a	n/a	1	3
c.1213C>G	P405A	Missense	101,95	5,34	n/a	n/a	5	4
c.1226A>G	Y409C	Missense	94,08	7,61	n/a	n/a	0	1
c.1255T>C	C419R	Missense	94,36	8,57	n/a	n/a	2	0
c.1843C>T	P615S	Missense	95,85	9,28	n/a	n/a	1	0
c.2687C>T	S896F	Missense	93,00	5,42	n/a	n/a	3	0
c.2978C>T	T993M	Missense	90,50	9,51	n/a	n/a	5	1
c.3035C>T	T1012I	Missense	88,09	5,94	n/a	n/a	5	16
c.3080T>G	L1027R	Missense	8,15	0,50	29,61	7,09	n/a	n/a
c.3107T>C	V1036A	Missense	81,24	5,94	n/a	n/a	1	0
c.3128G>T	G1043V	Missense	11,06	1,15	13,92	3,58	n/a	n/a
c.3132A>T	Q1044H	Missense	94,19	4,74	n/a	n/a	0	1
c.3506C>G	S1169C	Missense	82,90	4,36	n/a	n/a	0	1
c.3549_3552delCCACinsTTTG	H1184L	Missense	88,21	0,95	n/a	n/a	n/a	n/a
x	Ev-1	Empty vector	9,39	0,92	30,59	5,47	n/a	n/a
x	Ev-2	Empty vector	9,27	0,00	29,62	8,00	n/a	n/a
Variants Figure 2a-c (CC variants also used in Fig 3d-e)								
c.30C>G	S10R	Missense	95,02	2,77	103,22	9,41	0	1
c.71T>G	L24W	Missense	51,87	1,72	54,36	1,97	1	0
c.72G>C	L24F	Missense	86,56	2,38	90,57	5,55	1	1
c.85A>G	S29G	Missense	121,03	4,25	88,16	10,39	2	4
c.101G>A	R34H	Missense	91,60	4,99	n/a	n/a	4	1
c.101G>T	R34L	Missense	70,45	3,40	83,81	0,96	n/a	n/a
c.127A>G	K43E	Missense	97,15	17,68	n/a	n/a	3	0
c.314A>G	E105G	Missense	119,79	0,83	n/a	n/a	2	0
c.353T>C	I118T	Missense	110,42	6,86	n/a	n/a	5	1
C.398G>A	S133N	Missense	108,72	3,55	n/a	n/a	2	0
c.430C>G	P144A	Missense	96,89	22,56	n/a	n/a	2	0
c.554A>C	K185T	Missense	109,15	12,65	n/a	n/a	3	0
c.601dup	S201fs	Truncating	14,64	1,48	26,78	13,53	n/a	n/a
c.715_717delAGA	R239del	Missense	106,79	11,20	n/a	n/a	n/a	n/a

Supplementary Table. Continued

cDNA annotation	Variant (aa)	Variant type	Average HR	SEM (HR)	Average PARPi	SEM (PARPi)	Nr. cases	Nr. controls
c.925A>G	I309V	Missense	102,90	14,32	n/a	n/a	n/a	n/a
c.947C>T	P316L	Missense	112,82	0,72	84,72	20,72	n/a	n/a
C.1145G>T	S382I	Missense	106,61	3,27	n/a	n/a	5	2
c.1246A>G	M416V	Missense	116,76	0,84	86,60	26,90	n/a	n/a
c.1610C>T	S537L	Missense	97,75	4,29	n/a	n/a	6	0
c.1748T>G	L583W	Missense	117,85	2,11	86,65	13,90	2	3
c.2273C>G	P758R	Missense	103,67	5,19	n/a	n/a	6	2
c.2289G>C	L763F	Missense	95,28	0,46	n/a	n/a	22	13
c.2311A>G	S771G	Missense	110,59	0,51	95,20	9,67	n/a	n/a
c.2448C>G	F816L	Missense	92,40	17,62	n/a	n/a	2	0
c.2474G>C	R825T	Missense	105,75	9,25	n/a	n/a	39	22
c.2564T>A	L855Q	Missense	106,58	5,90	n/a	n/a	0	1
c.2612A>T	D871V	Missense	85,29	1,38	116,75	16,65	1	0
c.2619T>G	S873R	Missense	89,48	7,23	n/a	n/a	3	1
c.2641G>A	G881S	Missense	104,88	5,96	88,27	23,66	4	3
c.2642G>A	G881D	Missense	95,73	7,83	n/a	n/a	2	0
c.2673C>G	C891W	Missense	86,10	1,19	134,47	11,04	n/a	n/a
c.2674G>A	E892K	Missense	91,09	7,70	n/a	n/a	11	9
c.2689C>T	L897F	Missense	102,60	5,10	82,09	2,56	1	0
c.2690T>G	L897R	Missense	19,69	0,18	28,66	3,30	1	0
c.2735G>C	W912S	Missense	9,03	0,90	12,86	5,60	1	0
c.2776C>T	P926S	Missense	98,63	7,11	n/a	n/a	2	0
c.2798G>A	C933Y	Missense	92,01	4,00	n/a	n/a	0	1
c.2803G>A	A935T	Missense	93,79	0,34	n/a	n/a	1	0
c.2810G>A	G937E	Missense	50,34	3,78	71,14	8,47	0	1
c.2903C>G	A968G	Missense	101,90	5,29	n/a	n/a	2	2
c.2926A>G	R976G	Missense	52,13	4,97	68,13	9,77	1	0
c.2928G>T	R976S	Missense	60,49	3,06	77,64	27,37	n/a	n/a
c.2941A>C	S981R	Missense	111,57	1,06	111,70	30,43	1	0
c.3034A>C	T1012P	Missense	83,53	1,57	109,29	3,74	0	1
c.3053A>G	E1018G	Missense	96,52	6,01	n/a	n/a	0	1
c.3062G>A	G1021E	Missense	80,74	6,32	n/a	n/a	1	0
c.3073G>A	A1025T	Missense	100,43	6,57	n/a	n/a	3	0
c.3077T>C	L1026P	Missense	9,43	0,50	9,04	0,07	0	1
c.3079C>T	L1027F	Missense	102,63	8,08	100,32	3,31	1	1
c.3107T>C	V1036L	Missense	99,57	15,66	n/a	n/a	4	1
c.3121A>G	K1041E	Missense	105,95	9,34	n/a	n/a	1	0
c.3128G>C	G1043A	Missense	99,34	1,73	98,83	3,05	0	2
c.3133C>T	L1045F	Missense	101,12	9,60	n/a	n/a	1	0

Supplementary Table. Continued

cDNA annotation	Variant (aa)	Variant type	Average HR	SEM (HR)	Average PARPi	SEM (PARPi)	Nr. cases	Nr. controls
c.3235G>T	A1079S	Missense	100,53	4,79	n/a	n/a	6	1
c.3320T>C	L1107P	Missense	97,39	2,16	n/a	n/a	10	3
c.3342G>C	Q1114H	Missense	93,81	7,10	n/a	n/a	3	0
c.3404G>A	G1135E	Missense	99,62	13,18	n/a	n/a	1	0
c.3428T>C	L1143P	Missense	106,09	8,47	n/a	n/a	n/a	n/a
c.3449T>G	L1150R	Missense	91,01	1,98	n/a	n/a	n/a	n/a
c.3494C>T	S1165L	Missense	83,93	17,39	n/a	n/a	3	0
c.3506C>T	S1169F	Missense	94,02	12,07	n/a	n/a	n/a	n/a
c.3518C>T	A1173V	Missense	94,28	2,65	n/a	n/a	4	0
c.3547T>G	Y1183D	Missense	35,25	0,70	52,99	16,15	3	0
x	Ev-1	Empty vector	11,20	0,95	25,09	5,80	n/a	n/a
x	Ev-2	Empty vector	13,72	0,31	n/a	n/a	n/a	n/a
x	Ev-3	Empty vector	7,98	0,17	10,54	2,41	n/a	n/a
x	Ev-4	Empty vector	9,10	2,34	33,26	9,58	n/a	n/a
x	Ev-5	Empty vector	8,36	0,19	n/a	n/a	n/a	n/a
x	Ev-6	Empty vector	10,13	0,76	10,44	2,21	n/a	n/a
x	Ev-7	Empty vector	13,4	0,3	32,16	16,59	n/a	n/a
Variants Figure 3d,e								
c.29G>C	S10T	Missense	103,91	3,98	101,54	0,19	n/a	n/a
c.33T>G	C11W	Missense	100,39	6,03	98,97	9,40	n/a	n/a
c.38A>C	E13A	Missense	106,63	2,38	102,28	9,42	n/a	n/a
c.50T>C	L17S	Missense	45,92	3,30	75,63	7,84	n/a	n/a
c.56A>T	E19V	Missense	96,01	2,76	104,93	15,18	n/a	n/a
c.59A>T	K20I	Missense	73,13	3,20	69,05	18,03	n/a	n/a
c.62T>C	L21S	Missense	19,14	0,67	30,52	4,21	n/a	n/a
c.64G>C	A22P	Missense	15,61	0,08	26,25	0,04	n/a	n/a
c.65C>A	A22E	Missense	86,69	0,90	121,53	23,96	n/a	n/a
c.73A>G	K25E	Missense	42,90	2,06	76,31	6,38	n/a	n/a
c.77G>A	R26K	Missense	111,83	0,27	102,78	10,41	n/a	n/a
c.80A>G	E27G	Missense	63,60	1,91	73,77	11,19	n/a	n/a
c.82T>G	Y28D	Missense	22,02	0,25	40,05	15,26	n/a	n/a
c.85A>T	S29C	Missense	100,28	5,77	95,82	14,26	n/a	n/a
c.86G>C	S29T	Missense	99,11	0,10	106,84	3,28	n/a	n/a
c.88A>G	K30E	Missense	96,96	2,63	108,28	4,54	n/a	n/a
c.91A>C	T31P	Missense	14,02	0,20	16,02	0,50	n/a	n/a
c.95T>C	L32P	Missense	12,30	0,28	13,91	1,08	n/a	n/a
c.97G>C	A33P	Missense	13,43	0,68	15,33	2,15	n/a	n/a
c.101G>C	R34P	Missense	19,10	3,24	19,82	0,56	n/a	n/a
c.104T>A	L35H	Missense	42,60	7,76	50,08	8,75	n/a	n/a

Supplementary Table. Continued

cDNA annotation	Variant (aa)	Variant type	Average HR	SEM (HR)	Average PARPi	SEM (PARPi)	Nr. cases	Nr. controls
c.107A>C	**Q36P**	Missense	29,23	0,46	34,24	10,46	n/a	n/a
c.128A>T	**K43M**	Missense	114,60	9,07	n/a	n/a	n/a	n/a
x	**Ev-1**	Empty vector	10,13	0,76	10,44	2,21	n/a	n/a
x	**Ev-2**	Empty vector	13,4	0,3	32,16	16,59	n/a	n/a
Variants Figure 4b-c								
ChAM deletion 4x FLAG	**ΔChAM**	domain deletion	98,85	0,38	88,67	16,03	n/a	n/a
MRG15 deletion 4x FLAG	**ΔMRG15**	domain deletion	96,5	1,29	87,82	6,98	n/a	n/a
Exon 4 deletion 4x FLAG	**ΔEx4**	exon deletion	107,81	7,95	91,93	13,08	n/a	n/a
Ev-13	**Ev**	Empty vector	15,99	1,69	36,98	15,90	n/a	n/a

Nucleotide numbering reflects Human Genome Variation Society (HGVS) nomenclature where cDNA numbering +1 corresponds to the A of the ATG translation initiation codon in the reference sequence (*PALB2* NM_024675.3). The initiation codon is codon 1. For each variant, results from DR-GFP assays, PARPi sensitivity assays, and population-based case-control frequencies are shown. The population-based case-control frequencies are based on a study from the BRIDGES consortium in collaboration with the BCAC 13. x, not applicable; n/a, not available.

CHAPTER 6

6

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

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

Published in *Cancer Research*
(PMID: 34903604)

ABSTRACT

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

STATEMENT OF SIGNIFICANCE

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

KEYWORDS

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

INTRODUCTION

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

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

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

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

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

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

RESULTS

A cell-based functional assay for *CHEK2* variants

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

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

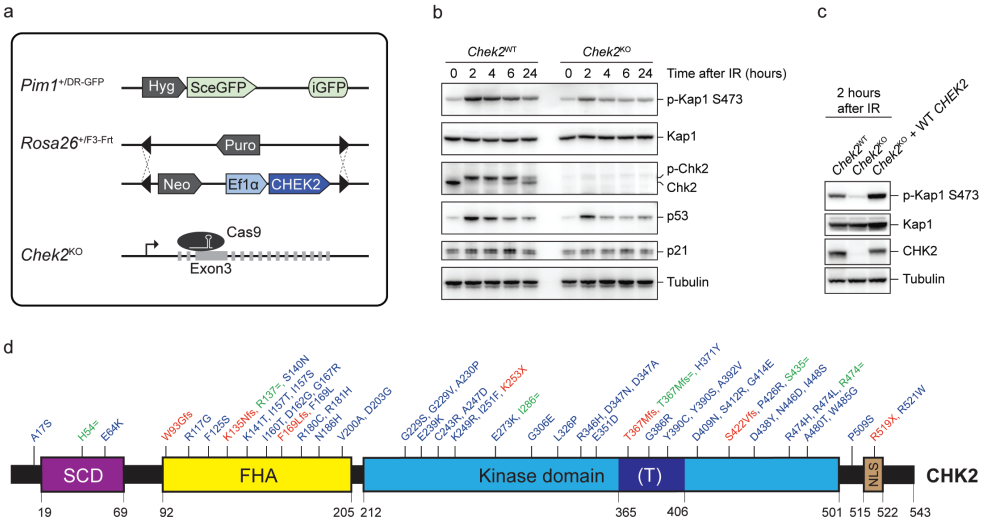


Figure 1. Generation of a cDNA-based complementation system for the functional analysis of human *CHEK2* variants. **a** Schematic representation of the mES cell- and cDNA-based complementation system for functional analysis. The DR-GFP reporter and Recombination-Mediated Cassette Exchange system (RMCE)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

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

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

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

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

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

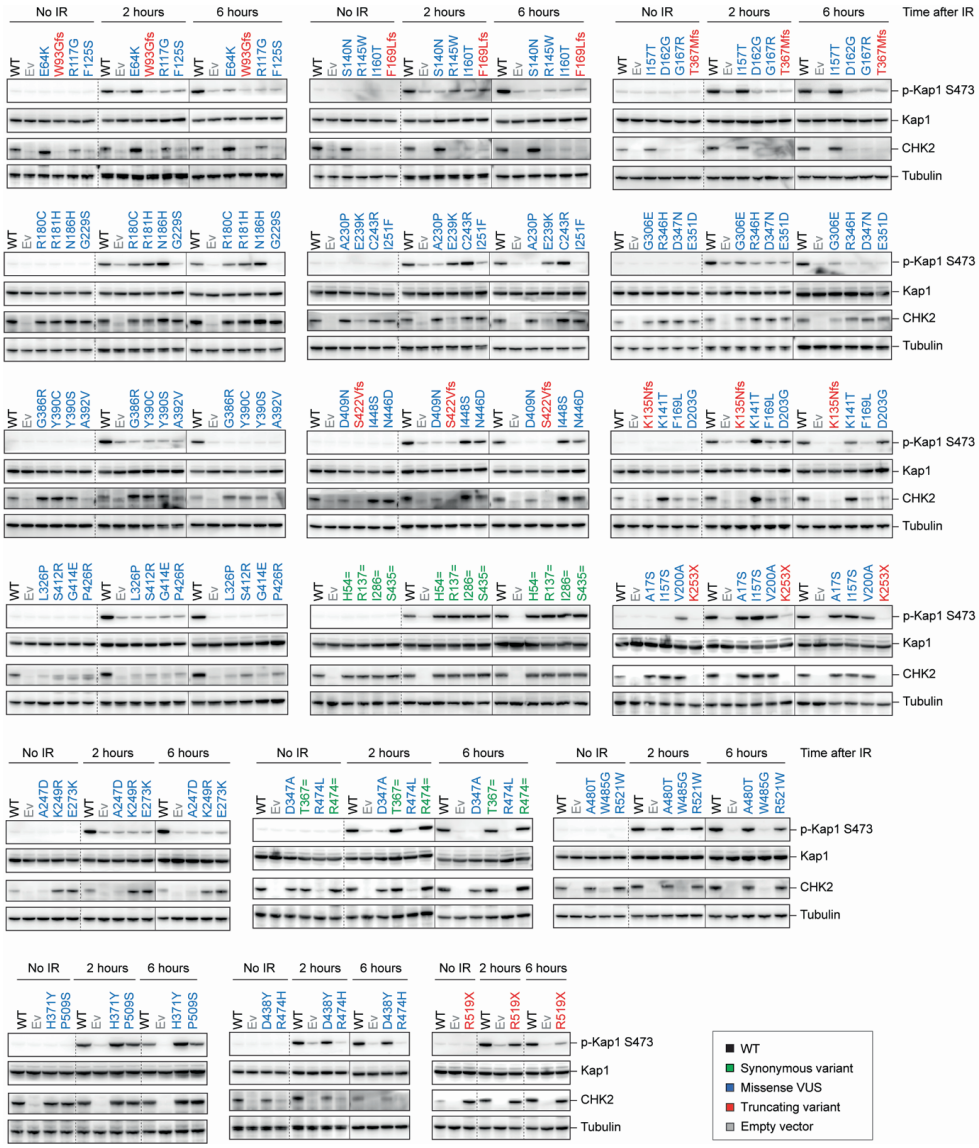


Figure 2. Human *CHEK2* variants and their effect on CHK2 expression and kinase activity toward Kap1 p.S473. Western blot analysis of the indicated proteins from *Chek2*^{KO} mES cells expressing wild type (WT, black) human untagged CHK2, empty vector (Ev, grey), or the indicated untagged CHK2 variants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

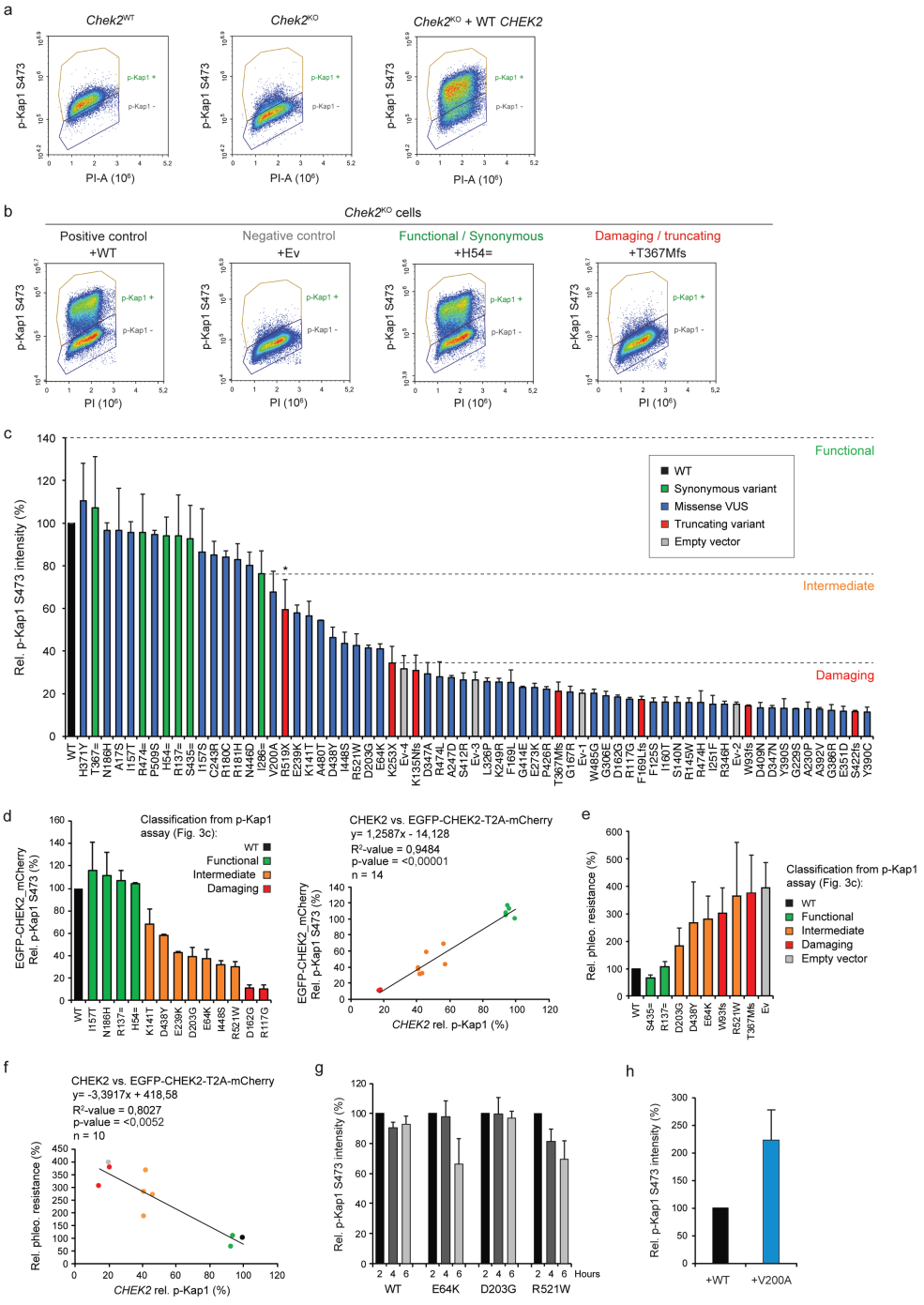

a
Chek2WT
Chek2KO
Chek2KO + WT CHEK2
p-Kap1 S473
PI-A (10^6)
p-Kap1 +
p-Kap1 -
b
Chek2KO cells
Positive control
+WT
Negative control
+Ev
Functional / Synonymous
+H54=
Damaging / truncating
+T367Mfs
PI (10^6)
c
Rel. p-Kap1 S473 intensity (%)
Functional
Intermediate
Damaging
WT
Synonymous variant
Missense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

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

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

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

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

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

Functional analysis of *CHEK2* missense VUS

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

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

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

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

Several variants alter the kinetics of CHK2

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

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

Correlation between computational predictions and functionality of variants

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

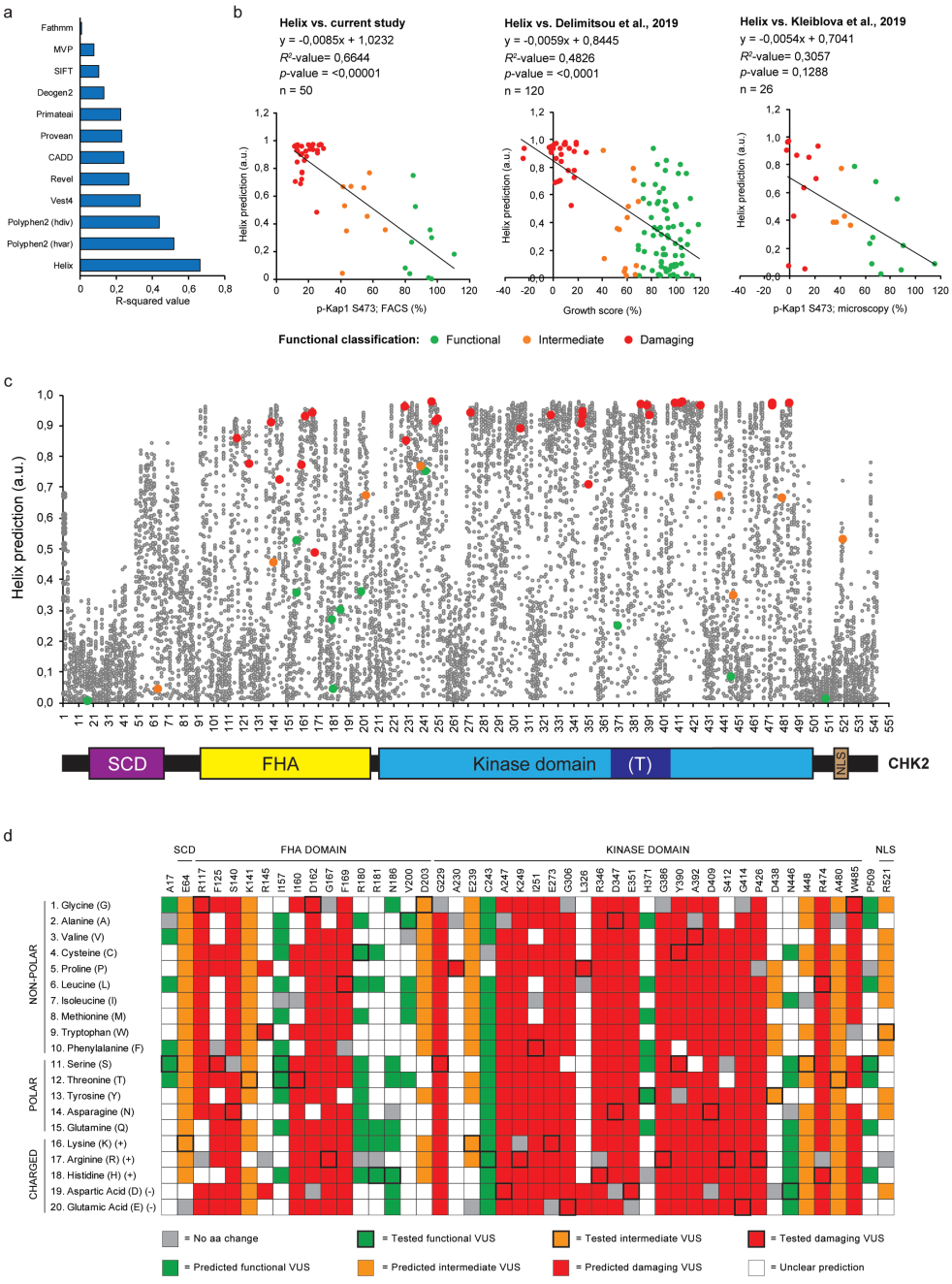


Figure 4. Correlation between computational predictions and functionality of *CHEK2* variants. **a** Bar plot showing the R^2-correlation values between the FACS-based analysis of Kap1 p.S473 phosphorylation as shown in Fig. 3c and computational predictions from twelve different prediction algorithms.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

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

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

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

CHEK2 VUS affect protein function through distinct mechanisms

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

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

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

Association of CHK2 functional defects with breast cancer risk

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

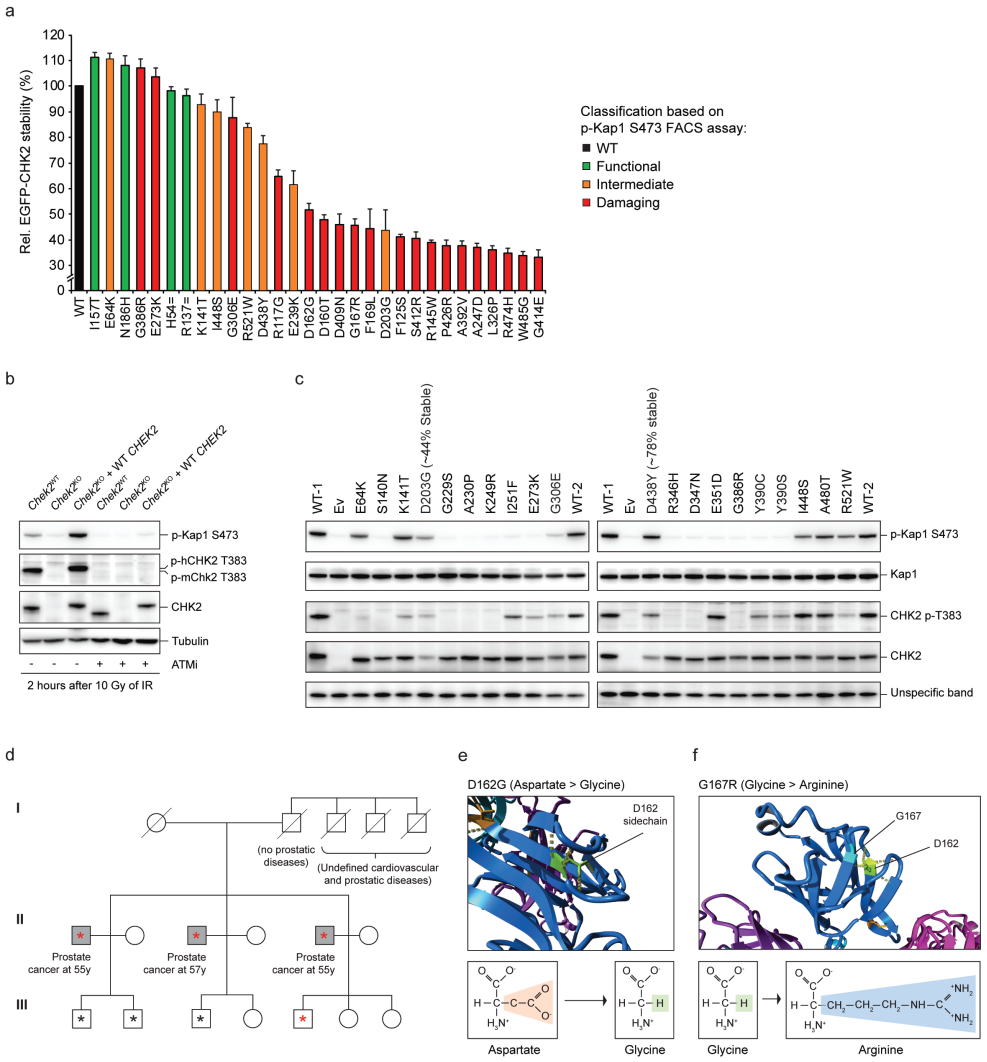


Figure 5. Analysis of pathogenic mechanisms of *CHEK2* VUS and the association of two VUS with prostate cancer. **a** Quantification of FACS measurements of the average EGFP intensity in *Chek2*^{KO} mES cells complemented with EGFP-CHEK2-T2A-mCherry, with or without the indicated *CHEK2* variants. EGFP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

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

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

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

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

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

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

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

Table 1. Continued

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

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

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

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

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

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

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

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

VAF refers to variant allele frequency.

DISCUSSION

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

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

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

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

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

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

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

MATERIALS AND METHODS

Cell lines and cell culture conditions

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

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

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

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

Selection of human *CHEK2* variants

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

Cloning and generation of human *CHEK2* variants

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

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

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

Western blot analysis

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

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

HR Reporter Assays

1-2x10^6 *Chek2*^{KO} mES cells carrying the DR-GFP reporter and RMCE system were subjected to HR assays by transfecting 1 µg of plasmid that co-expresses I-*Sce*I and mCherry (pCMV-Red-Isce, kind gift from Jos Jonkers) using Lipofectamine 2000 (ThermoFisher) (43). A co-transfection of 1 µg pCAGGs (44) with 0,05 µg of an mCherry expression vector was included as control. Two days after transfection, mCherry/GFP double-positive cells were scored using a Novocyte Flow Cytometer (ACEA Biosciences, Inc.).

Phleomycin sensitivity assays

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

RT-qPCR analysis

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

Flow cytometry (FACS) analysis

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

Exome sequencing in prostate cancer family members

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

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

LOH assessment

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

Ethics declaration

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

ACKNOWLEDGEMENTS

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

AUTHOR CONTRIBUTIONS

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

REFERENCES

1. Ahn J, Urist M, Prives C. The Chk2 protein kinase. DNA Repair (Amst) **2004**;3(8-9):1039-47 doi 10.1016/j.dnarep.2004.03.033.
2. Bartek J, Lukas J. Chk1 and Chk2 kinases in checkpoint control and cancer. Cancer cell **2003**;3(5):421-9 doi 10.1016/s1535-6108(03)00110-7.
3. Hirao A, Cheung A, Duncan G, Girard PM, Elia AJ, Wakeham A*, et al.* Chk2 is a tumor suppressor that regulates apoptosis in both an ataxia telangiectasia mutated (ATM)-dependent and an ATM-independent manner. Mol Cell Biol **2002**;22(18):6521-32 doi 10.1128/mcb.22.18.6521-6532.2002.
4. Meijers-Heijboer H, van den Ouweland A, Klijn J, Wasielewski M, de Snoo A, Oldenburg R*, et al.* Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet **2002**;31(1):55-9 doi 10.1038/ng879.
5. Vahteristo P, Bartkova J, Eerola H, Syrjakoski K, Ojala S, Kilpivaara O*, et al.* A CHEK2 genetic variant contributing to a substantial fraction of familial breast cancer. American journal of human genetics **2002**;71(2):432-8 doi 10.1086/341943.
6. Couch FJ, Shimelis H, Hu C, Hart SN, Polley EC, Na J*, et al.* Associations Between Cancer Predisposition Testing Panel Genes and Breast Cancer. JAMA Oncol **2017**;3(9):1190-6 doi 10.1001/jamaoncol.2017.0424.
7. Breast Cancer Association C, Dorling L, Carvalho S, Allen J, Gonzalez-Neira A, Luccarini C*, et al.* Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. The New England journal of medicine **2021** doi 10.1056/NEJMoa1913948.
8. Cybulski C, Gorski B, Huzarski T, Masojc B, Mierzejewski M, Debniak T*, et al.* CHEK2 is a multiorgan cancer susceptibility gene. American journal of human genetics **2004**;75(6):1131-5 doi 10.1086/426403.
9. Cybulski C, Huzarski T, Gorski B, Masojc B, Mierzejewski M, Debniak T*, et al.* A novel founder CHEK2 mutation is associated with increased prostate cancer risk. Cancer research **2004**;64(8):2677-9 doi 10.1158/0008-5472.can-04-0341.
10. Dong X, Wang L, Taniguchi K, Wang X, Cunningham JM, McDonnell SK*, et al.* Mutations in CHEK2 associated with prostate cancer risk. American journal of human genetics **2003**;72(2):270-80 doi 10.1086/346094.
11. Le Calvez-Kelm F, Lesueur F, Damiola F, Vallee M, Voegele C, Babikyan D*, et al.* Rare, evolutionarily unlikely missense substitutions in CHEK2 contribute to breast cancer susceptibility: results from a breast cancer family registry case-control mutation-screening study. Breast Cancer Res **2011**;13(1):R6 doi 10.1186/bcr2810.
12. Dufault MR, Betz B, Wappenschmidt B, Hofmann W, Bandick K, Golla A*, et al.* Limited relevance of the CHEK2 gene in hereditary breast cancer. International journal of cancer **2004**;110(3):320-5 doi 10.1002/ijc.20073.

13. Ingvarsson S, Sigbjornsdottir BI, Huiping C, Hafsteinsdottir SH, Ragnarsson G, Barkardottir RB*, et al.* Mutation analysis of the CHK2 gene in breast carcinoma and other cancers. Breast Cancer Res **2002**;4(3):R4 doi 10.1186/bcr435.
14. Schutte M, Seal S, Barfoot R, Meijers-Heijboer H, Wasielewski M, Evans DG*, et al.* Variants in CHEK2 other than 1100delC do not make a major contribution to breast cancer susceptibility. American journal of human genetics **2003**;72(4):1023-8 doi 10.1086/373965.
15. Sodha N, Bullock S, Taylor R, Mitchell G, Guertl-Lackner B, Williams RD*, et al.* CHEK2 variants in susceptibility to breast cancer and evidence of retention of the wild type allele in tumours. British journal of cancer **2002**;87(12):1445-8 doi 10.1038/sj.bjc.6600637.
16. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM*, et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic acids research **2014**;42(Database issue):D980-5 doi 10.1093/nar/gkt1113.
17. Stolarova L, Kleiblova P, Janatova M, Soukupova J, Zemankova P, Macurek L*, et al.* CHEK2 Germline Variants in Cancer Predisposition: Stalemate Rather than Checkmate. Cells **2020**;9(12) doi 10.3390/cells9122675.
18. Guo X, Ward MD, Tiedebohl JB, Oden YM, Nyalwidhe JO, Semmes OJ. Interdependent phosphorylation within the kinase domain T-loop Regulates CHK2 activity. J Biol Chem **2010**;285(43):33348-57 doi 10.1074/jbc.M110.149609.
19. Schwarz JK, Lovly CM, Piwnica-Worms H. Regulation of the Chk2 protein kinase by oligomerization-mediated cis- and trans-phosphorylation. Mol Cancer Res **2003**;1(8):598-609.
20. Ahn JY, Li X, Davis HL, Canman CE. Phosphorylation of threonine 68 promotes oligomerization and autophosphorylation of the Chk2 protein kinase via the forkhead-associated domain. J Biol Chem **2002**;277(22):19389-95 doi 10.1074/jbc.M200822200.
21. Kastan MB, Bartek J. Cell-cycle checkpoints and cancer. Nature **2004**;432(7015):316-23 doi 10.1038/nature03097.
22. Li J, Williams BL, Haire LF, Goldberg M, Wilker E, Durocher D*, et al.* Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2. Mol Cell **2002**;9(5):1045-54 doi 10.1016/s1097-2765(02)00527-0.
23. Zhang J, Willers H, Feng Z, Ghosh JC, Kim S, Weaver DT*, et al.* Chk2 phosphorylation of BRCA1 regulates DNA double-strand break repair. Mol Cell Biol **2004**;24(2):708-18 doi 10.1128/mcb.24.2.708-718.2004.
24. Hu C, Zhang S, Gao X, Gao X, Xu X, Lv Y*, et al.* Roles of Kruppel-associated Box (KRAB)-associated Co-repressor KAP1 Ser-473 Phosphorylation in DNA Damage Response. J Biol Chem **2012**;287(23):18937-52 doi 10.1074/jbc.M111.313262.
25. Cann KL, Dellaire G. Heterochromatin and the DNA damage response: the need to relax. Biochem Cell Biol **2011**;89(1):45-60 doi 10.1139/O10-113.
26. Czerwinska P, Mazurek S, Wiznerowicz M. The complexity of TRIM28 contribution to cancer. J Biomed Sci **2017**;24(1):63 doi 10.1186/s12929-017-0374-4.
27. Bolderson E, Savage KI, Mahen R, Pisupati V, Graham ME, Richard DJ*, et al.* Kruppel-associated Box (KRAB)-associated co-repressor (KAP-1) Ser-473 phosphorylation regulates

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

28. Lemaitre C, Soutoglou E. Double strand break (DSB) repair in heterochromatin and heterochromatin proteins in DSB repair. DNA Repair (Amst) **2014**;19:163-8 doi 10.1016/j.dnarep.2014.03.015.
29. Bell DW, Kim SH, Godwin AK, Schiripo TA, Harris PL, Haserlat SM*, et al.* Genetic and functional analysis of CHEK2 (CHK2) variants in multiethnic cohorts. International journal of cancer **2007**;121(12):2661-7 doi 10.1002/ijc.23026.
30. Lee SB, Kim SH, Bell DW, Wahrer DC, Schiripo TA, Jorczak MM*, et al.* Destabilization of CHK2 by a missense mutation associated with Li-Fraumeni Syndrome. Cancer research **2001**;61(22):8062-7.
31. Roeb W, Higgins J, King MC. Response to DNA damage of CHEK2 missense mutations in familial breast cancer. Hum Mol Genet **2012**;21(12):2738-44 doi 10.1093/hmg/dds101.
32. Tischkowitz MD, Yilmaz A, Chen LQ, Karyadi DM, Novak D, Kirchhoff T*, et al.* Identification and characterization of novel SNPs in CHEK2 in Ashkenazi Jewish men with prostate cancer. Cancer Lett **2008**;270(1):173-80 doi 10.1016/j.canlet.2008.05.006.
33. Wang N, Ding H, Liu C, Li X, Wei L, Yu J*, et al.* A novel recurrent CHEK2 Y390C mutation identified in high-risk Chinese breast cancer patients impairs its activity and is associated with increased breast cancer risk. Oncogene **2015**;34(40):5198-205 doi 10.1038/onc.2014.443.
34. Delimitsou A, Fostira F, Kalfakakou D, Apostolou P, Konstantopoulou I, Kroupis C*, et al.* Functional characterization of CHEK2 variants in a Saccharomyces cerevisiae system. Hum Mutat **2019**;40(5):631-48 doi 10.1002/humu.23728.
35. Kleiblova P, Stolarova L, Krizova K, Lhota F, Hojny J, Zemankova P*, et al.* Identification of deleterious germline CHEK2 mutations and their association with breast and ovarian cancer. International journal of cancer **2019**;145(7):1782-97 doi 10.1002/ijc.32385.
36. Cuella-Martin R, Hayward SB, Fan X, Chen X, Huang JW, Taglialatela A*, et al.* Functional interrogation of DNA damage response variants with base editing screens. Cell **2021**;184(4):1081-97 e19 doi 10.1016/j.cell.2021.01.041.
37. Robanus-Maandag E, Dekker M, van der Valk M, Carrozza ML, Jeanny JC, Dannenberg JH*, et al.* p107 is a suppressor of retinoblastoma development in pRb-deficient mice. Genes & development **1998**;12(11):1599-609.
38. Boonen R, Rodrigue A, Stoepker C, Wiegant WW, Vroling B, Sharma M*, et al.* Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. Nature communications **2019**;10(1):5296 doi 10.1038/s41467-019-13194-2.
39. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc **2013**;8(11):2281-308 doi 10.1038/nprot.2013.143.
40. Susswein LR, Marshall ML, Nusbaum R, Vogel Postula KJ, Weissman SM, Yackowski L*, et al.* Pathogenic and likely pathogenic variant prevalence among the first 10,000 patients referred for next-generation cancer panel testing. Genet Med **2016**;18(8):823-32 doi 10.1038/gim.2015.166.

41. Matsuoka S, Ballif BA, Smogorzewska A, McDonald ER, 3rd, Hurov KE, Luo J*, et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. Science **2007**;316(5828):1160-6 doi 10.1126/science.1140321.
42. Kranz A, Fu J, Duerschke K, Weidlich S, Naumann R, Stewart AF*, et al.* An improved Flp deleter mouse in C57Bl/6 based on Flpo recombinase. Genesis **2010**;48(8):512-20 doi 10.1002/dvg.20641.
43. Bouwman P, van der Gulden H, van der Heijden I, Drost R, Klijn CN, Prasetyanti P*, et al.* A high-throughput functional complementation assay for classification of BRCA1 missense variants. Cancer discovery **2013**;3(10):1142-55 doi 10.1158/2159-8290.CD-13-0094.
44. Niwa H, Yamamura K, Miyazaki J. Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene **1991**;108(2):193-9.
45. de Jonge MM, Auguste A, van Wijk LM, Schouten PC, Meijers M, Ter Haar NT*, et al.* Frequent Homologous Recombination Deficiency in High-grade Endometrial Carcinomas. Clinical cancer research : an official journal of the American Association for Cancer Research **2019**;25(3):1087-97 doi 10.1158/1078-0432.CCR-18-1443.
46. Parameswaran B, Chiang HC, Lu Y, Coates J, Deng CX, Baer R*, et al.* Damage-induced BRCA1 phosphorylation by Chk2 contributes to the timing of end resection. Cell Cycle **2015**;14(3):437-48 doi 10.4161/15384101.2014.972901.
47. Takai H, Naka K, Okada Y, Watanabe M, Harada N, Saito S*, et al.* Chk2-deficient mice exhibit radioresistance and defective p53-mediated transcription. EMBO J **2002**;21(19):5195-205 doi 10.1093/emboj/cdf506.
48. Hirao A, Kong YY, Matsuoka S, Wakeham A, Ruland J, Yoshida H*, et al.* DNA damage-induced activation of p53 by the checkpoint kinase Chk2. Science **2000**;287(5459):1824-7 doi 10.1126/science.287.5459.1824.
49. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T*, et al.* Analysis of protein-coding genetic variation in 60,706 humans. Nature **2016**;536(7616):285-91 doi 10.1038/nature19057.
50. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q*, et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv **2019**:531210 doi 10.1101/531210.
51. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J*, et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med **2015**;17(5):405-24 doi 10.1038/gim.2015.30.
52. Vroling B, Heijl S. White paper: The Helix Pathogenicity Prediction Platform (https://arxiv.org/abs/2104.01033). **2021**.
53. Heijl S, Vroling B, Bergh Tvd, Joosten H. Mind the gap: preventing circularity in missense variant prediction (https://doi.org/10.1101/2020.05.06.080424). **2020**.

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

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

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

57. Ahn JY, Schwarz JK, Piwnica-Worms H, Canman CE. Threonine 68 phosphorylation by ataxia telangiectasia mutated is required for efficient activation of Chk2 in response to ionizing radiation. Cancer research **2000**;60(21):5934-6.

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

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

60. Chrisanthar R, Knappskog S, Lokkevik E, Anker G, Ostenstad B, Lundgren S*, et al.* CHEK2 mutations affecting kinase activity together with mutations in TP53 indicate a functional pathway associated with resistance to epirubicin in primary breast cancer. PloS one **2008**;3(8):e3062 doi 10.1371/journal.pone.0003062.

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

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

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

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

SUPPLEMENTARY INFORMATION

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

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

The supplementary information contains*:

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

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

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

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

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

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

Supplementary Table S2. Continued

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

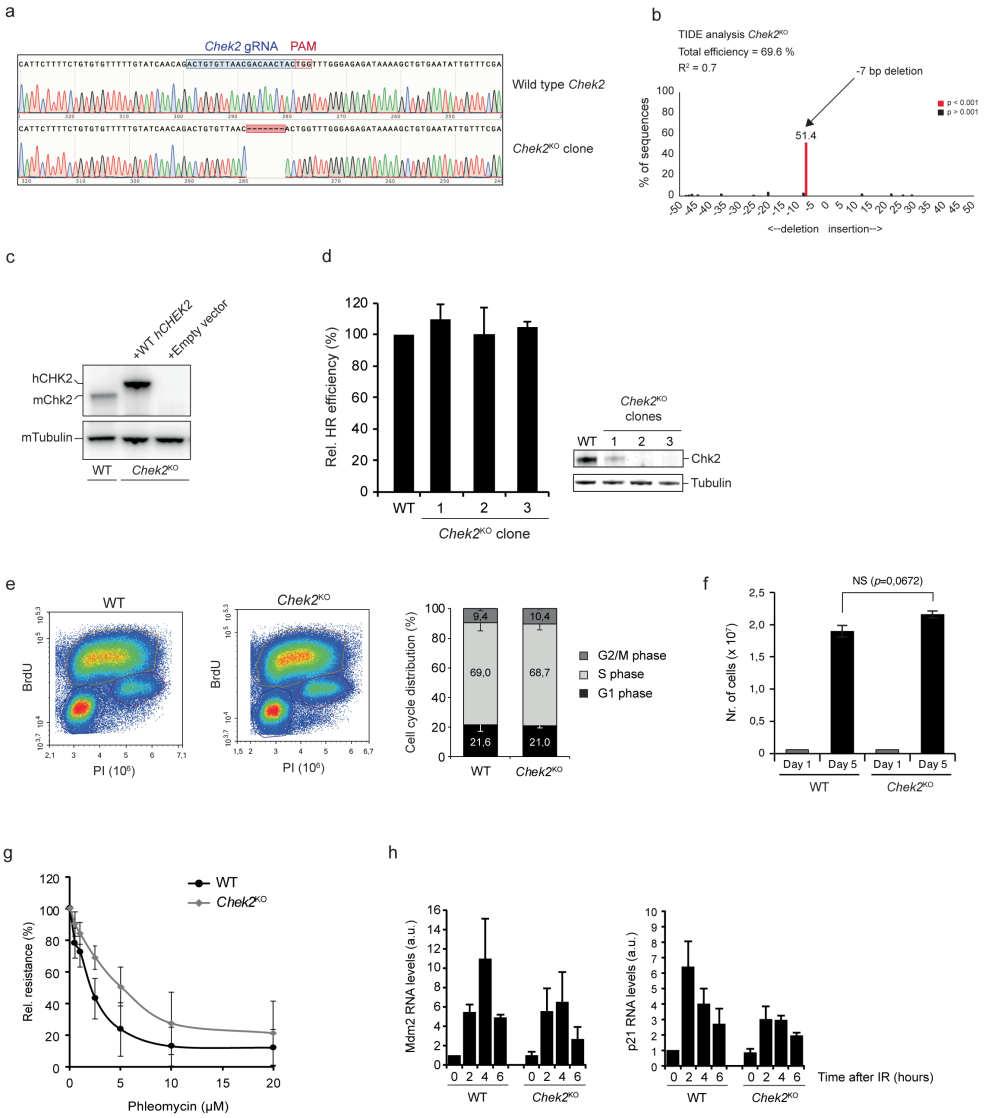
Only predictions from SpliceAI are shown in this table. Predictions using Alamut (i.e., from four algorithms; SpliceSiteFinder-like, MaxEntScan, GeneSplicer and NNSPLICE) are available in the online version of this table. SpliceAI scores range from 0-1 and can be interpreted as the probability that the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

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

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

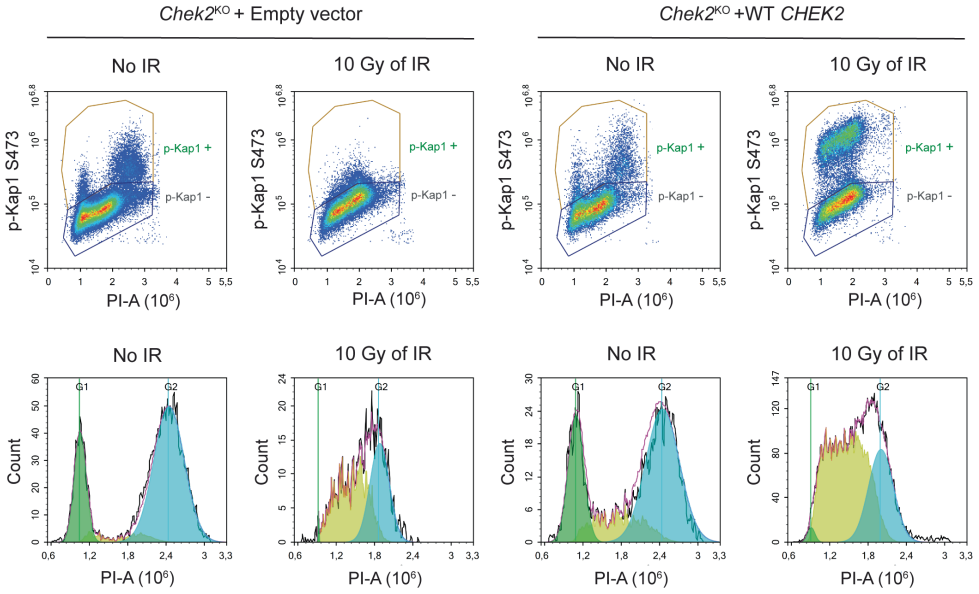
Supplementary Table S3. Continued

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

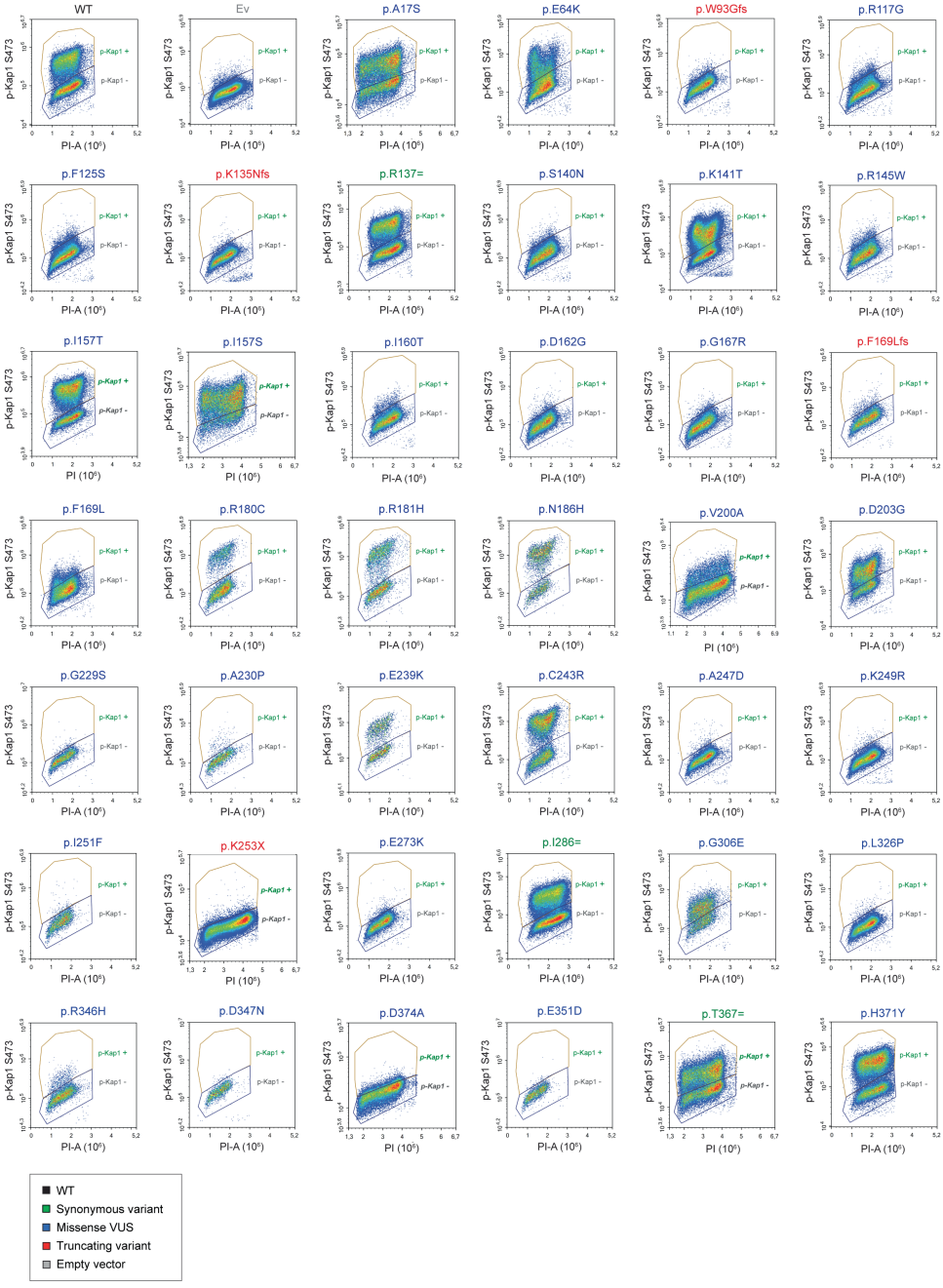


Supplementary Figure S1. Validation and functional analysis of *Chek2*^{KO} mES cells. **a** Sequence alignment of a fragment of exon 3 of the *Chek2* gene showing a -7 bp deletion. **b** TIDE analysis confirming the -7 bp deletion in exon 3 of the *Chek2* gene. **c** Western blot analysis confirming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

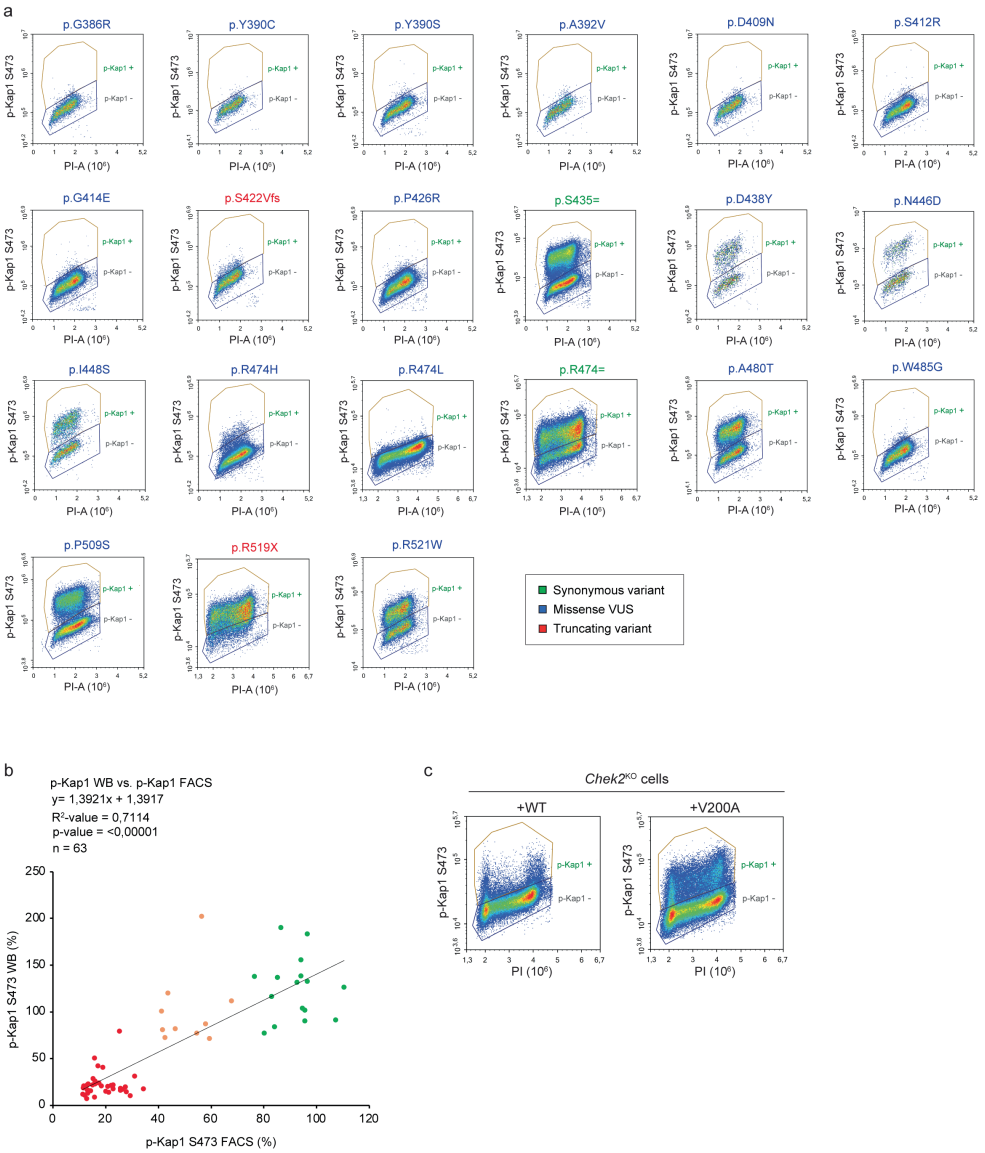
The bar graph represents the mean percentage of cell cycle phase distributions from 2 independent measurements. **f** Analysis of the proliferation rate of WT and *Chek2*^{KO} mES cells. On day 1, 0.5x10^{6} cells were seeded for both conditions and on day 5, cell growth was assessed by cell counting.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



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

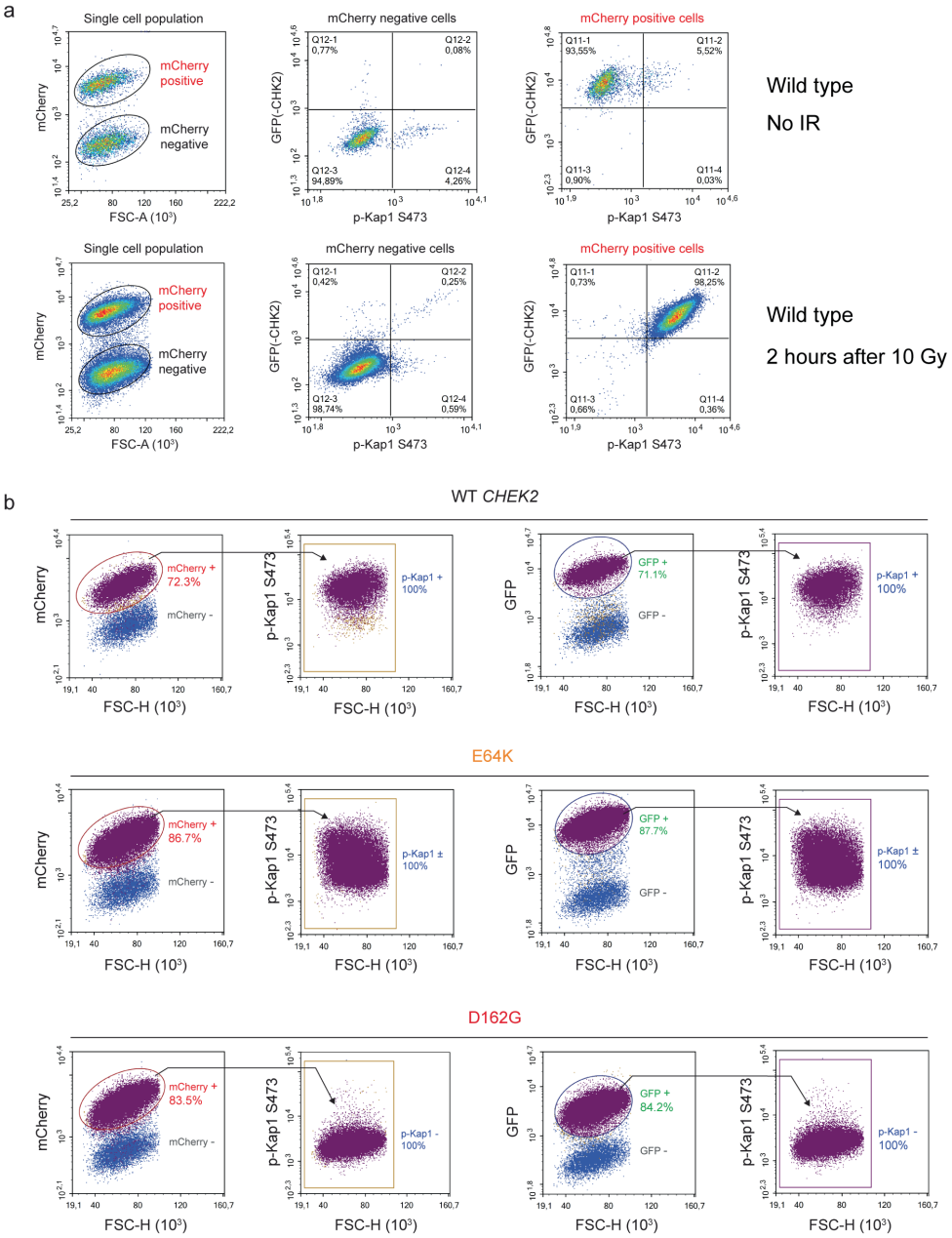


Supplementary Figure S3. Human *CHEK2* variants and their effect on CHK2's kinase activity toward Kap1 p.S473. Quantitative FACS-based analysis, 2 hours after IR, of Kap1 p.S473 phosphorylation in *Chek2*^{KO} mES cells complemented with the indicated conditions. Variants/conditions are categorized by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



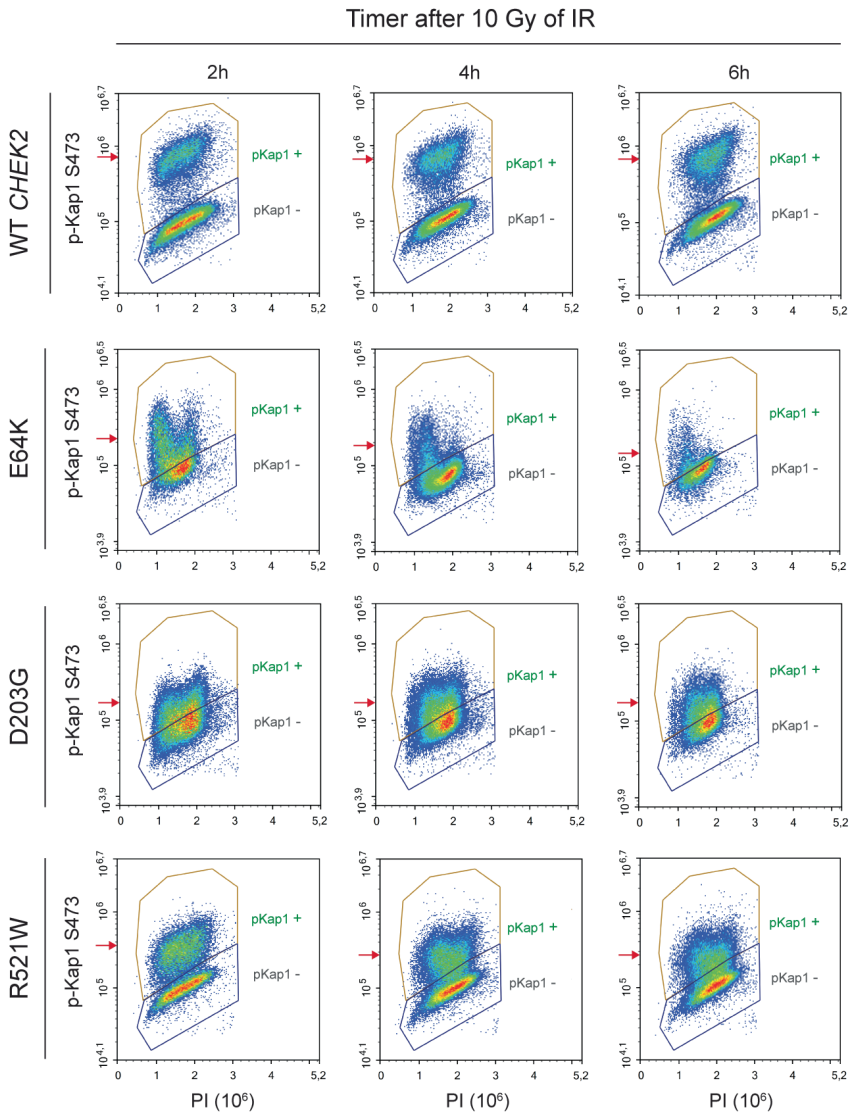
Supplementary Figure S4. Human *CHEK2* variants and their effect on CHK2's kinase activity toward Kap1 p.S473. **a** Quantitative FACS-based analysis, 2 hours after IR, of Kap1 p.S473 phosphorylation in *Chek2*^{KO} mES cells complemented with the indicated conditions. Variants/conditions are categorized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

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



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

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

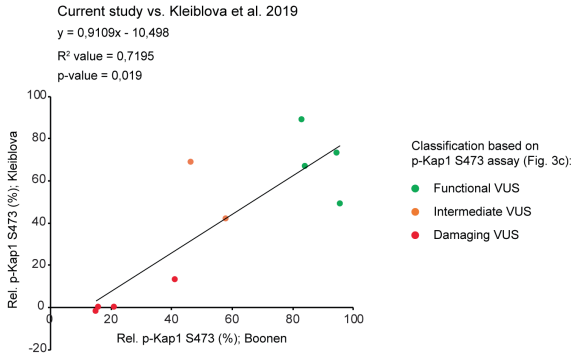


Supplementary Figure S6. CHK2 kinase activity in time after DNA damage induction. Quantitative FACS-based analysis of Kap1 p.S473 phosphorylation in *Chek2*^{KO} mES cells complemented with the indicated conditions. Cells were fixed and measured at the indicated times after IR. The red arrows indicate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

a

10 overlapping variants

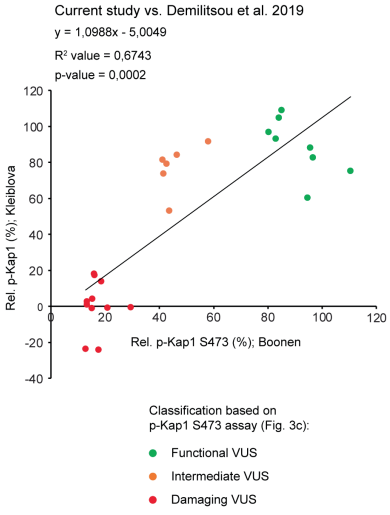
CHEK2 variants		Current study	Kleiblova et al., 2019	
Nt position	Aa position	p-Kap1 intensity (%)	p-Kap1 microscopy (%)	In vitro
c.190G>A	p.E64K	41	13	neutral
c.470 T>C	p.I157T	96	49	neutral
c.538C>T	p.R180C	84	67	neutral
c.542G>A	p.R181H	83	89	neutral
c.715G>A	p.E239K	58	42	intermediate
c.1037G>A	p.R346H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



b

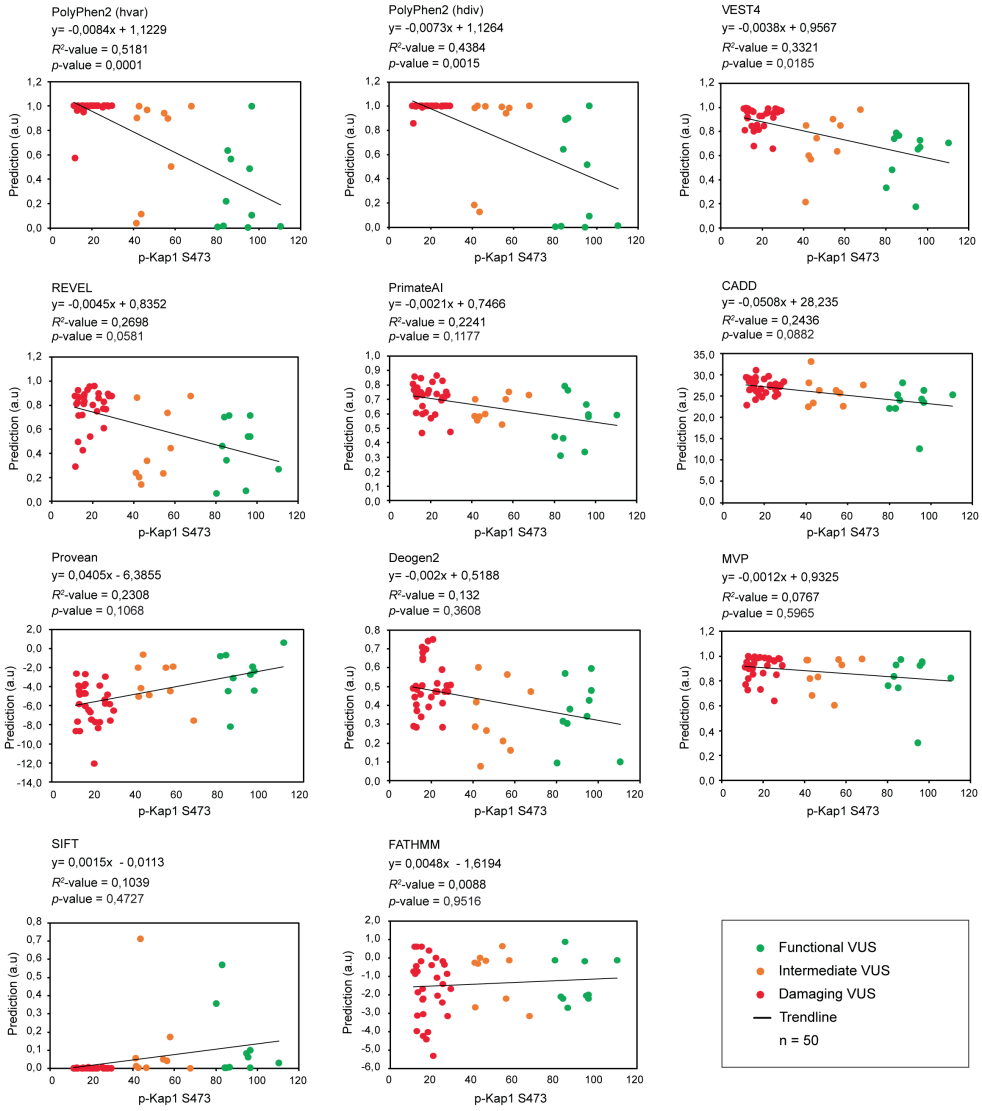
25 overlapping variants

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



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

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



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

SUPPLEMENTARY REFERENCES

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

CHAPTER 7

Future perspectives

7

FUTURE PERSPECTIVES

According to ClinVar (as of October 2022), a clinically oriented database for genetic variants, around 75% of all germline variants in the coding sequences of *PALB2* and *CHEK2* are missense variants, whereas 17% are frameshift and 8% are nonsense. Almost all (94%) *PALB2* (n=1987) and *CHEK2* (n=1284) missense variants are classified as variants of uncertain significance (VUS), which constitute variants that cannot be used for clinical decision-making or cancer risk assessment due to insufficient available evidence that can be used for clinical interpretation. Potential biochemical and structural alterations resulting from missense VUS are often extremely challenging to predict, meaning that they cannot be used for clinical interpretation of these variants. In contrast, the concerted efforts to functionally characterize numerous *PALB2* and *CHEK2* missense VUS in distinct functional assays on a variant-by-variant basis (as reviewed in **Chapters 2** and **3**), represent milestones for clinical interpretation and clinical management of *PALB2* and *CHEK2* VUS carriers. However, some challenges still remain, such as addressing the functional effects of an overwhelming number of VUS in these genes that are yet to be functionally characterized. This also includes addressing the functional impact of variants on RNA splicing as many missense or synonymous variants that do not affect protein function in many of the cDNA-based complementation systems that have been used for functional analysis of variants, may still have a negative impact on protein function due to the potential introduction of cryptic splice sites (1). A future outlook to address these challenges, as well as the challenge of implementing functional evidence in clinical variant interpretation, is provided below.

Characterizing functional impact of genetic variants at scale

Sequencing hereditary breast cancer susceptibility genes has proven to be a powerful diagnostic tool to identify individuals at increased risk for breast cancer (2). However, each individual's genome contains millions of sites where his or her DNA differs from the reference sequence. Despite major advances in cataloging and this overwhelming number of genomic variants (e.g., in ClinVar), current clinical and functional understanding of most of the identified variants is insufficient. Even with the ongoing efforts in experimentally measuring the functional consequences of variants in *PALB2* and *CHEK2*, most missense variants in these genes are still classified as VUS (3,4). This is mainly because variant-by-variant assays (as those discussed in this thesis) are too time and resource intensive to keep up with the number of identified variants. Also, functional assays are generally performed after a variant is identified in a carrier. Results may then become available years later, when they may no longer be relevant for deciding on preventive therapeutic strategies (at least for the carrier in which the variant was identified first). Alternatively, to keep up with the large number of genetic variants

that are being detected during genetic testing, high-throughput systems can be applied to measure the functional consequences of all possible variants in disease-relevant loci, for a variety of molecular and cellular phenotypes, simultaneously. The functional data obtained by high-throughput approaches can then be presented in the form of comprehensive atlases that do not only facilitate interpretation of variants that have already been identified in carriers, but also variants that are yet to be identified during genetic testing.

For *PALB2*, we employed a cDNA-based complementation system, using a variant library for the Coiled-Coil (CC) region of *PALB2*, to assess sensitivity to PARPi treatment in a high-throughput manner (**Chapter 5**). Using this strategy, we functionally assessed 91.1% of all possible missense variants in this region. Importantly, this strategy can be extended to the WD40 domain of *PALB2*, or other regions, or even *CHEK2* (with phospho-Kap1 as a readout, **Chapter 6**), with the use of additional variant libraries. An alternative approach to our high-throughput assay may be to introduce variants endogenously. For instance, a saturation genome editing technique that relies on CRISPR/Cas9-induced DNA breaks in combination with the use of repair templates (each containing a distinct variant), was reported for *BRCA1* (5). Using a similar approach for *PALB2*, introduction of a damaging *PALB2* variant in the human haploid cell line HAP1 (6) will result in cell death since this is an essential gene (5). Thus, after introduction of a large number of variants in this cell line, cells expressing damaging *PALB2* variants should be depleted from the population and cell survival can be used as a functional readout in a high-throughput manner. In contrast, HAP1 cells expressing damaging *CHEK2* variants may gain a growth advantage, as was also evident in a recent base-editing screen in which variants were introduced endogenously by direct modification of bases with a nuclease-deficient Cas9 tethered to the cytosine deaminase APOBEC1 (7). Importantly, when variants are introduced endogenously, their effects on multiple layers of gene function (such as RNA splicing and stability, and protein stability and function), can be studied simultaneously. Another option to address functional effects of variants *en masse* is by massively parallel sequencing (VAMP-seq), which is an experimental strategy that can measure the effects of thousands of missense variants on intracellular abundance simultaneously (8). In this approach, a mixed population of cells is generated where each cell expresses one protein variant fused to a fluorescent tag. Cells can be sorted by flow cytometry based on their levels of fluorescence, which in turn corresponds to a certain degree of variant protein stability. As many variants in the WD40 domain of *PALB2*, or throughout the entire *CHEK2* sequence, have been shown to affect protein stability (9-11), such a method may enable the identification of many damaging variants. Collectively, large-scale functional data resulting from the above-mentioned approaches can eventually result in lookup tables that will aid in a more accurate ascertainment of the pathogenicity of many genetic variants.

Despite the promising utility of high-throughput assays in large scale variant interpretation, it should be noted that they can produce noisy data (12), hampering their use for clinical classification. This issue may be addressed by performing, 1) high numbers of replicate experiments, 2) cross validation with clinical data or other (high-throughput) functional studies, and 3) single variant assays for proper validation of functional effects. In addition, to further improve the clinical utility of high-throughput assays for missense variants, results can be cross-correlated with *in silico* prediction tools such as the splice predictor SpliceAI (13), which is generally valued for its accuracy and can easily provide evidence for variant interpretation at a large scale (14,15). Such a correlation could provide important insights into which missense variants affect protein function due to effects on RNA splicing or due the amino acid substitution it causes.

The possibility of examining a specific phenotype of interest at scale, such as PALB2's function in HR, in high-throughput functional assays, is a prerequisite for the ability to test large numbers of variants in a gene. However, despite the recent development of different types of high-throughput functional assays, many disease-associated genes still remain beyond reach owing to a lack of assays with a suitable read-out. Consequently, an important question is: what do we need in order to apply high-throughput functional assays to all disease-associated genes in the genome? Alternative to the aforementioned functional readouts, recent advances in microscopy-based cell sorting now allows for high-throughput examination of visual cellular phenotypes as a result of genomic variation (i.e., Visual Cell Sorting) (16). For example, this technique enables the sorting of hundreds of thousands of cells according to the nuclear localization of a fluorescently tagged protein (variant) (16). As PALB2 localizes to the nucleus in order to perform its DNA repair function, and mis-localization in the cytoplasm has been associated with impaired PALB2 protein function (11,17,18), this technique may enable the identification of damaging PALB2 variants at a large scale. This is only one example of a technique that will strongly expand the repertoire of disease-associated genes for which genetic variants can be functionally characterized. Based on this, it may be expected that high-throughput assays will soon be further adapted and optimized, allowing for an expansion of the repertoire of genes for which large-scale variant analysis is desired.

Functional analysis of splice variants

Splicing of precursor messenger RNA (pre-mRNA) is an extremely complex process, and the clinical interpretation of variants that affect splicing can take into account predictions from both computational algorithms, as well as experimental data. When these variants occur at canonical splice sites, i.e., at the 'GT' splice donor and the 'AG' splice acceptor site, they are often easy to classify. This is because predictive models, nowadays based on deep learning, perform reasonably well (13). For instance, these variants are generally considered pathogenic

when loss of function of a gene (e.g., due to expected skipping of an entire exon that may encode a region essential for protein function), is known to be causative of disease (19). However, there are many caveats that must be considered for these types of variants, as there are several scenarios in which a functional protein can be produced despite the presence of a variant in a canonical splice site. Therefore, these variants can be most problematic for clinical interpretation

Splicing assays that assess the impact of variants at the mRNA level can be highly informative and can include direct analysis of RNA, or in vitro minigene splicing assays (20). These assays have been shown to be useful for the interpretation of splice variants occurring at canonical splice sites, in coding sequences, or even in deeper intronic regions (21). However, unlike a functional readout for protein function itself, an effect on splicing (e.g., exon skipping, or intron retention) does not necessarily translate to an impact on protein function. In general, aberrant splicing can result in multiple outcomes with respect to mRNA fate and the protein-reading frame. Although it is often assumed that abnormally spliced transcripts resulting in a premature stop codon will undergo nonsense-mediated decay, this is not always the case as exemplified by normal protein levels observed for the Fanconi anemia-associated *PALB2* p.Y551X truncating variant (22). Consequently, some abnormal transcripts can lead to expression of a truncated protein with or without functional consequences (23). Thus, accurate clinical classification of variants that affect RNA splicing requires that alternative transcripts are identified, quantified, and functionally characterized.

With regards to clinical interpretation, a functional evaluation of genetic variants that includes potential effects on RNA processing, is most valuable. An advantage of the haploid HAP1 cell-based system where genome editing can be employed to introduce variants at endogenous loci (5,7), is that the effect of variants on regulatory mechanisms such as splicing can be included (24). Alternatively, complementation with a bacterial artificial chromosome (BAC) containing the complete human gene of interest (GOI), can also allow for evaluation of any type of variant. This method was, for instance, used to evaluate numerous *BRCA2* variants for their effects on splicing and their capacity to express functional BRCA2 protein isoforms after loss of the endogenous gene in mES cells. Importantly, multiple alternative transcripts encoding (partially) functional protein isoforms were identified and their altered expression attenuated the functional effects of several predicted *BRCA2* loss-of-function canonical splice variants (25). Additionally, several *BRCA2* nonsense variants in exon 12, that were initially assumed to be pathogenic, have been shown to result in enhanced expression of an alternative transcript lacking exon 12, which encodes a (partially) functional protein isoform (26). Consequently, these and other assumed loss-of-function variants in exon 12 of *BRCA2*, constitute variants for which further studies are required to estimate their associated cancer risk. These findings highlight the need of examining the effects of genetic variants on RNA

splicing and protein function. To date, however, it has been difficult to identify all the different transcripts that are expressed due to a genetic variant. Moreover, the quantification of all these distinct transcripts is extremely challenging. The advent of PacBio-based Next Generation Sequencing, however, may be able to provide a complete RNA transcript profile as it allows for the analysis of long reads up to 25kb (27). For many genes, all RNA transcripts ranging from the first to the last exon can then be captured in a quantitative manner and linked to a functional phenotype. Overall, such techniques will result in more detailed understanding of how variants can affect RNA splicing, and also a better clinical classification of these variants.

The use of functional data for clinical interpretation of variants

Most variants identified in the breast cancer susceptibility genes are exceedingly rare and it will require extremely large case-control association studies (i.e., >1 million individuals) to accurately quantify cancer risk for specific variants. Validated functional assays, are considered by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) guidelines, as strong evidence for or against the pathogenicity of rare missense variants (28,29). Accordingly, a strong concordance was observed between two high-throughput functional studies for *BRCA1* and ClinVar classifications of pathogenicity for variants with expert panel evaluations (5,30). This supports the claim that functional characterization of genetic variants is extremely useful for clinical interpretation of variants and assessment of cancer risk. Furthermore, as shown for both *PALB2* and *CHEK2* (**Chapter 5** and **Chapter 6,** respectively), results from functional assays quantitatively correlate with the degree of breast cancer risk, as calculated based on variant frequency data from the Breast Cancer Association Consortium (BCAC) (31). That is, the degree of breast cancer risk that is associated with a certain level of PALB2 or CHK2 protein function, was established using a burden-type association analysis. In this analysis, variants are grouped based on similar impact on protein function and joint frequencies of these variants in cases and controls are used to derive odds ratios per variant group (representing a level of protein functionality). However, as especially damaging *PALB2* variants are extremely rare in occurrence, it is (for this variant group particularly) difficult to obtain high enough case-control frequencies and establish an odds ratio with a narrow confidence interval that is statistically meaningful (e.g., p-value <0.01). Thus, to make these associations more conclusive, and the burden-type association analysis more valuable, high-throughput functional assays need to be performed for (nearly) all *PALB2* and *CHEK2*, variants. Even then, it may be challenging to identify enough damaging *PALB2* missense variants to warrant an accurate association with breast cancer risk. Accordingly, there is also an urgent need for more clinical data from large case control association studies, such as that from BCAC (31), that can ultimately be combined with other large case control studies, such as for instance CARRIERS (32), in order to improve

our understanding of the quantitative relationship between PALB2 protein function and cancer risk.

Although variants in *CHEK2* occur more frequently in the general population relative to variants in *PALB2* (33), the fact that it constitutes gene that is associated with moderate risk (~2 fold increased) excludes the use of genetic approaches such as co-segregation analysis to determine the pathogenicity for variants as is done in genes with high penetrance like *BRCA1* or *BRCA2*. To date, there are no *CHEK2* missense variants that are classified as benign or pathogenic based on clinical data. As a consequence, setting a functional threshold for benignity is very complicated. In contrast, although known pathogenic missense variants cannot be used, the residual functionality seen for truncating variants can be used to calibrate a functional threshold for pathogenicity. Similar to *PALB2*, it is pivotal that more *CHEK2* missense variants are functionally characterized, with for instance, high-throughput approaches. Ultimately, it may even be possible to generate a 'continuous risk model' that allows for the calculation of a variant-specific risk. For example, based on the data presented in **Chapter 6**, we could reason that the damaging *CHEK2* missense variants, for which the functional impact is similar to that of the truncating variants, are associated with a similar breast cancer risk. Relative to that, the average decrease in protein function that we observed for intermediate missense variants is 60%, while the average decrease of the functional variants is 10%. Because we associated these variant groups with odds ratios in the burden type association analysis, we can also deduct a simplified continuous risk model from this data (Fig. 1). Assuming that the odds ratio is a function of the functional score (i.e., a decrease in CHK2 protein function inversely correlates with breast cancer risk), we can then for example calculate that *CHEK2* p.D203G, for which we could not calculate an odds ratio specifically and which showed a decrease in protein function of ~50%, associates with an odds ratio of ~1.55. This continuous risk model is currently based on only 44 out of the 388 *CHEK2* missense variants that were identified the BCAC studies (31) and that were functionally characterized as functional, intermediate or damaging in **Chapter 6**. Several relatively frequent *CHEK2* missense variants for which cancer risk estimates are already available, such as those mentioned in **Chapter 3** (Table 2), could be added as individual datapoints (i.e., not being part of the functional, intermediate or damaging variant groups) to improve the linear regression analysis. In addition, high-throughput functional analysis of *CHEK2* missense variants should allow for functional characterization of nearly all 388 *CHEK2* missense variants and should therefore results in an even more accurate estimation of the correlation between residual functionality and associated cancer risk.

With an optimistic view to the future, it is foreseeable that additional high-throughput functional and clinical studies will ultimately result in the establishment of a quantitative relationship between protein function (e.g., functional, intermediate and damaging), and the

degree of associated cancer risk. This will be a major step towards the use of functional data in personalized risk prediction and clinical decision making.

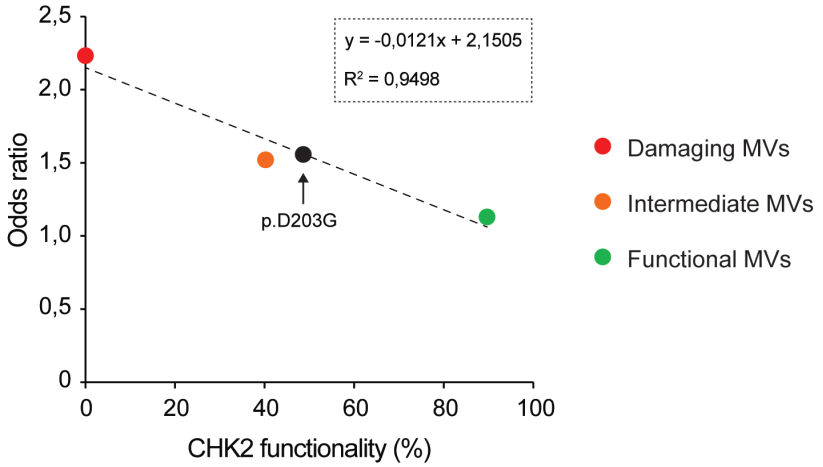


Figure 1. Quantitative relationship between CHK2 function and cancer risk. The functional defect of damaging missense variants (MVs) compares to that of truncating *CHEK2* variants and the average impact of this variant group on Kap1 S474 phosphorylation is therefore set to a 100% decrease in CHK2 protein function. The average impact of the functional and intermediate variant groups on Kap1 S474 phosphorylation are presented relative to that of the damaging *CHEK2* variant group. In **Chapter 6**, odds ratios were calculated for the three *CHEK2* variant groups (i.e., functional, intermediate and damaging). The dotted line represents the correlation between CHK2 function and associated breast cancer risk. Once the impact on CHK2 function has been established for a VUS, the OR of the variant can be estimated.

REFERENCES

1. Lee M, Roos P, Sharma N, Atalar M, Evans TA, Pellicore MJ*, et al.* Systematic Computational Identification of Variants That Activate Exonic and Intronic Cryptic Splice Sites. Am J Hum Genet **2017**;100(5):751-65 doi 10.1016/j.ajhg.2017.04.001.
2. Strande NT, Berg JS. Defining the Clinical Value of a Genomic Diagnosis in the Era of Next-Generation Sequencing. Annu Rev Genomics Hum Genet **2016**;17:303-32 doi 10.1146/annurev-genom-083115-022348.
3. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S*, et al.* ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res **2016**;44(D1):D862-8 doi 10.1093/nar/gkv1222.
4. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM*, et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res **2014**;42(Database issue):D980-5 doi 10.1093/nar/gkt1113.
5. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M*, et al.* Accurate classification of BRCA1 variants with saturation genome editing. Nature **2018**;562(7726):217-22 doi 10.1038/s41586-018-0461-z.
6. Blomen VA, Majek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, Staring J*, et al.* Gene essentiality and synthetic lethality in haploid human cells. Science **2015**;350(6264):1092-6 doi 10.1126/science.aac7557.
7. Cuella-Martin R, Hayward SB, Fan X, Chen X, Huang JW, Taglialatela A*, et al.* Functional interrogation of DNA damage response variants with base editing screens. Cell **2021**;184(4):1081-97 e19 doi 10.1016/j.cell.2021.01.041.
8. Matreyek KA, Starita LM, Stephany JJ, Martin B, Chiasson MA, Gray VE*, et al.* Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat Genet **2018**;50(6):874-82 doi 10.1038/s41588-018-0122-z.
9. Boonen R, Wiegant WW, Celosse N, Vroling B, Heijl S, Kote-Jarai Z*, et al.* Functional Analysis Identifies Damaging CHEK2 Missense Variants Associated with Increased Cancer Risk. Cancer Res **2022**;82(4):615-31 doi 10.1158/0008-5472.CAN-21-1845.
10. Boonen R, Rodrigue A, Stoepker C, Wiegant WW, Vroling B, Sharma M*, et al.* Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. Nat Commun **2019**;10(1):5296 doi 10.1038/s41467-019-13194-2.
11. Ng PS, Boonen RA, Wijaya E, Chong CE, Sharma M, Knaup S*, et al.* Characterisation of protein-truncating and missense variants in PALB2 in 15 768 women from Malaysia and Singapore. J Med Genet **2021** doi 10.1136/jmedgenet-2020-107471.

12. Starita LM, Ahituv N, Dunham MJ, Kitzman JO, Roth FP, Seelig G*, et al.* Variant Interpretation: Functional Assays to the Rescue. Am J Hum Genet **2017**;101(3):315-25 doi 10.1016/j.ajhg.2017.07.014.
13. Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI*, et al.* Predicting Splicing from Primary Sequence with Deep Learning. Cell **2019**;176(3):535-48 e24 doi 10.1016/j.cell.2018.12.015.
14. Strauch Y, Lord J, Niranjan M, Baralle D. CI-SpliceAI-Improving machine learning predictions of disease causing splicing variants using curated alternative splice sites. PLoS One **2022**;17(6):e0269159 doi 10.1371/journal.pone.0269159.
15. Ha C, Kim JW, Jang JH. Performance Evaluation of SpliceAI for the Prediction of Splicing of NF1 Variants. Genes (Basel) **2021**;12(9) doi 10.3390/genes12091308.
16. Hasle N, Cooke A, Srivatsan S, Huang H, Stephany JJ, Krieger Z*, et al.* High-throughput, microscope-based sorting to dissect cellular heterogeneity. Mol Syst Biol **2020**;16(6):e9442 doi 10.15252/msb.20209442.
17. Rodrigue A, Margaillan G, Torres Gomes T, Coulombe Y, Montalban G, da Costa ESCS*, et al.* A global functional analysis of missense mutations reveals two major hotspots in the PALB2 tumor suppressor. Nucleic Acids Res **2019**;47(20):10662-77 doi 10.1093/nar/gkz780.
18. Wiltshire T, Ducy M, Foo TK, Hu C, Lee KY, Belur Nagaraj A*, et al.* Functional characterization of 84 PALB2 variants of uncertain significance. Genet Med **2020**;22(3):622-32 doi 10.1038/s41436-019-0682-z.
19. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J*, et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med **2015**;17(5):405-24 doi 10.1038/gim.2015.30.
20. Gaildrat P, Killian A, Martins A, Tournier I, Frebourg T, Tosi M. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. Methods Mol Biol **2010**;653:249-57 doi 10.1007/978-1-60761-759-4_15.
21. Lopez-Perolio I, Leman R, Behar R, Lattimore V, Pearson JF, Castera L*, et al.* Alternative splicing and ACMG-AMP-2015-based classification of PALB2 genetic variants: an ENIGMA report. J Med Genet **2019**;56(7):453-60 doi 10.1136/jmedgenet-2018-105834.
22. Xia B, Dorsman JC, Ameziane N, de Vries Y, Rooimans MA, Sheng Q*, et al.* Fanconi anemia is associated with a defect in the BRCA2 partner PALB2. Nat Genet **2007**;39(2):159-61 doi 10.1038/ng1942.

23. Abou Tayoun AN, Pesaran T, DiStefano MT, Oza A, Rehm HL, Biesecker LG*, et al.* Recommendations for interpreting the loss of function PVS1 ACMG/AMP variant criterion. Hum Mutat **2018**;39(11):1517-24 doi 10.1002/humu.23626.
24. Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J. Saturation editing of genomic regions by multiplex homology-directed repair. Nature **2014**;513(7516):120-3 doi 10.1038/nature13695.
25. Mesman RLS, Calleja F, de la Hoya M, Devilee P, van Asperen CJ, Vrieling H*, et al.* Alternative mRNA splicing can attenuate the pathogenicity of presumed loss-of-function variants in BRCA2. Genet Med **2020**;22(8):1355-65 doi 10.1038/s41436-020-0814-5.
26. Meulemans L, Mesman RLS, Caputo SM, Krieger S, Guillaud-Bataille M, Caux-Moncoutier V*, et al.* Skipping Nonsense to Maintain Function: The Paradigm of BRCA2 Exon 12. Cancer Res **2020**;80(7):1374-86 doi 10.1158/0008-5472.CAN-19-2491.
27. Hon T, Mars K, Young G, Tsai YC, Karalius JW, Landolin JM*, et al.* Highly accurate long-read HiFi sequencing data for five complex genomes. Sci Data **2020**;7(1):399 doi 10.1038/s41597-020-00743-4.
28. Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD*, et al.* Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med **2019**;12(1):3 doi 10.1186/s13073-019-0690-2.
29. Brnich SE, Rivera-Munoz EA, Berg JS. Quantifying the potential of functional evidence to reclassify variants of uncertain significance in the categorical and Bayesian interpretation frameworks. Hum Mutat **2018**;39(11):1531-41 doi 10.1002/humu.23609.
30. Starita LM, Islam MM, Banerjee T, Adamovich AI, Gullingsrud J, Fields S*, et al.* A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution Variants on Protein Function. Am J Hum Genet **2018**;103(4):498-508 doi 10.1016/j.ajhg.2018.07.016.
31. Breast Cancer Association C, Dorling L, Carvalho S, Allen J, Gonzalez-Neira A, Luccarini C*, et al.* Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. N Engl J Med **2021**;384(5):428-39 doi 10.1056/NEJMoa1913948.
32. Hu C, Hart SN, Gnanaolivu R, Huang H, Lee KY, Na J*, et al.* A Population-Based Study of Genes Previously Implicated in Breast Cancer. N Engl J Med **2021**;384(5):440-51 doi 10.1056/NEJMoa2005936.
33. Woodward ER, van Veen EM, Evans DG. From BRCA1 to Polygenic Risk Scores: Mutation-Associated Risks in Breast Cancer-Related Genes. Breast Care (Basel) **2021**;16(3):202-13 doi 10.1159/000515319.

CHAPTER 7

8

Nederlandse samenvatting

Curriculum Vitae

Publications

Dankwoord

NEDERLANDSE SAMENVATTING

Borstkanker is wereldwijd de meest voorkomende oorzaak van sterfte bij vrouwen. De ontwikkeling van borstkanker ontstaat in veel gevallen waarschijnlijk door een combinatie van risico factoren die afhankelijk kunnen zijn van gedrag, afkomst en/of omgeving. Bij ruwweg 10% van de vrouwelijke borstkanker patiënten is er echter sprake van een erfelijke/genetische aanleg om deze vorm van kanker te ontwikkelen. Binnen families waar borstkanker veel voorkomt, en/of op jonge leeftijd is ontstaan, kunnen vrouwen doorverwezen worden naar een klinisch geneticus voor een erfelijkheidsonderzoek. Er wordt dan onderzoek gedaan naar een klein aantal genen, waarvan de meeste betrokken zijn bij de 'DNA damage response' (een reactie binnen een cel die plaats vindt na het detecteren van DNA-schade). Deze reactie is enorm belangrijk voor een cel, aangezien deze nodig is om de genomische stabiliteit te bewaken. Een afwijking in deze reactie, bijvoorbeeld als gevolg van een genetische verandering, kan er dan ook voor zorgen dat iemand een sterk verhoogd risico heeft om borstkanker te ontwikkelen.

Bij een erfelijkheidsonderzoek zijn er een aantal uitslagen mogelijk. Het kan zo zijn dat er geen genetische veranderingen (ofwel varianten) gevonden worden, of varianten waarvan bekend is dat ze benigne zijn (niet ziekmakend). In dergelijke gevallen lijkt de persoon geen verhoogd risico op borstkanker te lopen, of dit te kunnen overdragen aan eventuele kinderen (althans op basis van de onderzochte genen). Daartegenover is het mogelijk dat er een duidelijke pathogene (ofwel ziekmakende) variant gevonden wordt. In deze gevallen worden de dragers hiervan zo adequaat mogelijk geadviseerd over, bijvoorbeeld, de preventieve maatregelen die genomen kunnen worden. Tot slot is het mogelijk dat er varianten gevonden worden waarvan het effect onzeker is. Deze varianten worden ook wel 'variants of uncertain significance' (VUS) genoemd, en voor dragers hiervan is het dus compleet onduidelijk of ze een verhoogd risico lopen om borstkanker te ontwikkelen en of er bijvoorbeeld ingrijpende maatregelen, zoals risico verlagende chirurgische ingrepen, aanbevolen moeten worden. Als gevolg hiervan kunnen deze varianten voor veel stress zorgen bij dragers, en frustratie bij klinisch genetici.

Om inzicht te krijgen in de pathogeniciteit van een VUS kan er gebruik gemaakt worden van klinische gegevens, bijvoorbeeld data die laat zien of de variant binnen een familie wel of niet segregeert met het voorkomen van kanker. Een dergelijke waarneming moet echter statistisch onderbouwd worden om een toevalsbevinding uit te sluiten en daarvoor zijn in het algemeen te weinig families met dezelfde variant beschikbaar. In dat soort gevallen is er dus een ander soort onderzoek nodig om deze VUSsen te interpreteren.

Voor de meeste genen zorgt een pathogene variant voor verlies van eiwitfunctie. Dus een functioneel onderzoek, waarbij er gekeken wordt of een variant effect heeft op de functie van het geproduceerde eiwit in de cel, kan de uitkomst bieden voor het inschatten van de

pathogeniciteit van veel zeldzame VUSsen. In dit proefschrift presenteren we de ontwikkeling en toepassing van een dergelijke functionele analyse, waarbij de focus ligt op twee 'DNA damage response' genen die beide geassocieerd zijn met borstkanker; namelijk *PALB2* (een hoog risico gen voor borstkanker) en *CHEK2* (een gematigd verhoogd risico gen voor borstkanker).

Het PALB2 eiwit heeft een belangrijke functie in een specifieke vorm van DNA-schade herstel, namelijk de homologe recombinatie (HR). HR zorgt ervoor dat dubbelstrengs DNA-breuken op een foutloze manier gerepareerd worden. Wanneer de functie van het PALB2 eiwit verstoord wordt door de aanwezigheid van een pathogene variant, kunnen dubbelstrengs DNA-breuken niet goed hersteld worden en kan er dus genomische instabiliteit optreden. Het CHK2 eiwit, geproduceerd door het *CHEK2* gen, bewaakt de genomische stabiliteit van cellen op een andere manier, namelijk door celdelingen een halt toe te roepen wanneer er DNA-schade is gedetecteerd. Dit geeft een cel de tijd om de DNA-schade te repareren voordat deze definitief de celdeling afrondt. Daarnaast kan CHK2 in sommige gevallen ervoor zorgen dat een cel in apoptose gaat (geprogrammeerde celdood) wanneer de DNA-schade onherstelbaar is of in te grote mate aanwezig is in het genoom van de cel. Een verstoring in de functie van het CHK2 eiwit (als gevolg van een pathogene *CHEK2* variant), kan dus op een andere manier leiden tot genomische instabiliteit. Deze genomische instabiliteit, die dus kan optreden als gevolg van een defect in PALB2 of CHK2 eiwit functie, kan uiteindelijk resulteren in ongecontroleerde celdelingen en een verhoogd risico om borstkanker te ontwikkelen. In het geval van een defect in PALB2 eiwit functie (als gevolg van een pathogene variant), gaat het dan om een sterk verhoogd risico. In het geval van een defect in CHK2 eiwit functie (als gevolg van een pathogene variant), gaat het om een gematigd verhoogd risico.

Voor het interpreteren van VUSsen in zowel het *PALB2* als het *CHEK2* gen, hebben wij functionele *in vitro* testsystemen opgezet waarvoor we muis embryonale stamcellen (mESC) gebruiken. In deze cellen is het muis gen voor *Palb2* of *Chek2* onbruikbaar gemaakt en kan, respectievelijk, het humane *PALB2* of *CHEK2* gen (eventueel inclusief variant) geïntroduceerd worden om dit verlies te compenseren. Voor *PALB2* kunnen we dan (voornamelijk) naar de efficiëntie in HR kijken, en het effect van *PALB2* varianten daarop bestuderen. Voor *CHEK2* varianten daarentegen, kijken we voornamelijk naar de fosforylatie van een doeleiwit van CHK2, i.e., Kap1, die op het serine 473 residu specifiek door CHK2 gefosforyleerd wordt na de detectie van DNA-schade (bijvoorbeeld na röntgenbestraling van cellen). De uitlezing van beide testsystemen (i.e., de mate van DNA-schadeherstel, dan wel de mate van Kap1-fosforylering) zegt iets over de functionaliteit van een *PALB2* of *CHEK2* VUS en kan op een kwantitatieve of semi-kwantitatieve manier gebruikt worden bij het inschatten of een VUS pathogeen of benigne is.

In de **Hoofdstukken 2 en 3** worden twee literatuurstudies gepresenteerd voor, respectievelijk, *PALB2* en *CHEK2*, waarin verscheidene gepubliceerde functionele analyses voor beide genen beschreven worden en op een kritische manier met elkaar worden vergeleken. In de opvolgende onderzoek hoofdstukken wordt in **Hoofdstuk 4** het testsysteem voor *PALB2* gevalideerd door het gebruik van 9 bekende functionele varianten en 12 bekende pathogene (niet functionele) varianten. Vervolgens wordt de functionaliteit van 48 *PALB2* VUSsen geanalyseerd, wat duidelijk laat zien dat er 5 VUSsen, functioneel gezien net zo schadelijk zijn als de 12 pathogene controle varianten. Tevens zien we dat de VUSsen die een functioneel defect laten zien, zich bevinden in het coiled-coil (CC) domein en het WD40 domein van het *PALB2* gen. De schadelijke VUSsen in het CC domein verstoren de interactie tussen het PALB2 eiwit en het BRCA1 eiwit, en daarmee als gevolg de HR-efficiëntie. Schadelijke VUSsen in het WD40 domein daarentegen, verstoren de stabiliteit van het PALB2 eiwit en als gevolg dus ook HR. In **Hoofdstuk 5** worden deze mechanistische bevindingen verder bevestigd met de analyse een nieuwe groep *PALB2* varianten, waarbij tevens nog een aantal schadelijke *PALB2* VUSsen worden geïdentificeerd. Zo laten de resultaten in dit hoofdstuk ook zien dat volledige verwijdering van andere functionele domeinen dan het CC- of WD40-domein (i.e., het ChAM of MRG15 domein) van *PALB2*, geen effect heeft op de functie van het PALB2 eiwit in HR. Hierdoor is het dus aannemelijk dat VUSsen die zich in deze domeinen bevinden, minder snel een effect zullen hebben op de functie van PALB2 binnen HR. Daarnaast wordt er een methode gepresenteerd die de functionele karakterisatie van honderden *PALB2* varianten binnen een enkel experiment mogelijk maakt, wat o.a. van groot belang zal zijn voor de interpretatie van het grote aantal *PALB2* varianten dat momenteel (nog) als VUS geclassificeerd is. Tot slot, worden in hoofdstuk 5 onze functionele resultaten geassocieerd met borstkankerrisico middels een ‘burden-type’ associatie analyse. Hiervoor wordt gebruik gemaakt van de gegevens uit een grote ‘case-control’ studie van het Borstkanker Associatie Consortium (BCAC), die voor sommige *PALB2* varianten heeft laten zien hoe vaak de variant gevonden is in een borstkanker patiënt (i.e., case) en/of in een gezonde controle (i.e., control). Deze ‘burden-type’ associatie analyse laat duidelijk zien dat een verlaagde HR-efficiëntie inderdaad correleert met een verhoogd borstkankerrisico. Deze bevinding vergroot de klinische waarde van de functionele analyse van *PALB2* VUSsen, omdat de varianten die schadelijk zijn met betrekking tot eiwit functie (en al gevolg het risico voor borstkanker verhogen), direct vertaald kunnen worden naar klinisch handelen, vergelijkbaar met hoe het nu voor pathogene varianten gebeurt.

Het onderzoek in **Hoofdstuk 6** is gericht op de validatie van het functionele testsysteem voor *CHEK2* varianten en vervolgens het functioneel testen van 50 *CHEK2* VUSsen. Voor de validatie gebruiken we hier 6 functionele controle varianten en 6 bekende pathogene varianten. Voor deze varianten zien we een duidelijke scheiding in het vermogen

van het CHK2 eiwit om een downstream target, Kap1, te fosforyleren. Terwijl de functionele varianten, als reactie op het voordoen van DNA schade, prima in staat zijn om Kap1 te fosforyleren op serine 473, zijn de pathogene varianten hiertoe niet meer in staat. Wat betreft de 50 *CHEK2* VUSsen zien we vervolgens dat er maar liefst 31 net zo schadelijk zijn als de bekende pathogene varianten en dat er 9 een middelmatig functioneel defect laten zien. Al deze VUSsen lijken verspreid te liggen over de hele sequentie van het *CHEK2* gen (in tegenstelling dus tot *PALB2* waar ze alleen in het CC- en WD40-domein zaten) en hebben voornamelijk een effect op de stabiliteit van het CHK2 eiwit. Een klein aantal *CHEK2* VUSsen lijkt echter schadelijk te zijn door een defect te veroorzaken in de autofosforylatie van CHK2 en/of mogelijk de binding van ATP. Tot slot associëren we in dit hoofdstuk ook de functionele resultaten met borstkankerrisico middels een 'burden-type' associatie analyse, wat laat zien dat de schadelijke *CHEK2* VUSsen met eenzelfde mate van risico geassocieerd zijn als de bekende pathogene *CHEK2* varianten; namelijk een gematigd verhoogd risico voor borstkanker. De VUSsen die resulteren in een middelmatig functioneel defect lijken geassocieerd te zijn met een iets lager maar nog steeds significant verhoogd risico, terwijl de functionele VUSsen niet geassocieerd zijn met een verhoogd risico. Deze associatie analyse is een verdere validatie van ons test systeem en benadrukt wederom het belang van de functionele analyse van VUSsen. Deze informatie blijkt namelijk duidelijk van waarde te zijn voor het interpreteren van deze varianten en het bepalen of deze varianten pathogeen of benigne zijn.

Kortom, dit proefschrift laat zien dat we robuuste testsystemen hebben ontwikkeld die het mogelijk maken om de functionaliteit van *PALB2* en *CHEK2* varianten te bepalen. Deze functionaliteit correleert met de mate van borstkankerrisico en kan dus van enorme waarde zijn bij het bepalen van welke VUSsen pathogeen of benigne zijn. Wanneer dit op een betrouwbare manier bepaald kan worden, kunnen dragers van deze varianten dus beter geadviseerd worden met betrekking tot preventieve maatregelen, of in sommige gevallen (wanneer het al een drager met borstkanker betreft) over de behandelmogelijkheden. Zo kan hopelijk de zorg voor deze VUS-dragers uiteindelijk verbeterd worden.

CURRICULUM VITAE

Ricky Antonius Cornelius Martin Boonen werd geboren op 11 september 1983 in Boxmeer. In het jaar 2000 behaalde hij zijn Havo diploma bij de Scholengemeenschap Stevensbeek. Na een jaar werkervaring en oriënteren, begon hij in 2002 met de studie Biologie en Medisch Laboratoriumonderzoek aan de Fontys Hogeschool in Eindhoven, waarvoor hij na één jaar zijn propedeuse behaalde. Met de gewekte interesse op dit gebied begon hij vervolgens in 2003 aan de universitaire opleiding Biomedische Wetenschappen in Utrecht, waarvoor hij in 2009 zijn bachelor diploma behaalde. Kort na het starten van deze studie vond hij in 2004 het Hubrecht Instituut, waar hij de mogelijkheid kreeg om één dag in de week als analist werkervaring op te doen in de onderzoeksgroep van dr. Danica Zivkovic. Hier deed hij onder andere ervaring op met de zebravis als modelsysteem en werkte hij aan de rol van de Wnt pathway in de ontwikkeling van spiercellen. Tevens schreef hij hier onder begeleiding van dr. Danica Zivkovic zijn bachelor scriptie over de betrokkenheid van de Wnt pathway in de ontwikkeling van de ziekte van Alzheimer, die in 2008 gepubliceerd werd als review artikel in het tijdschrift 'Ageing Research Reviews'. Na deze werkervaring, kreeg hij in 2011 weer de mogelijkheid om als analist te werken bij het Hubrecht Instituut, ditmaal voltijd en in de onderzoeksgroep van dr. Puck Knipscheer. Hier werkte hij met behulp van 'Xenopus egg' extracten aan de mechanismen die een rol spelen bij 'DNA interstrand crosslink repair' en wist hij hierover in 2014 een gedeeld eerste auteurs artikel te publiceren in het tijdschrift 'Molecular Cell'. In 2012 startte hij alsnog met een masterstudie, namelijk 'Cancer Stem Cells and Developmental Biology', en combineerde deze met zijn baan als analist die voor deze studie ook zijn eerste master stage is geworden. Voor zijn tweede stage ging hij een half jaar naar New York, om daar in de onderzoeksgroep prof. dr. Karen Duff, bij 'Columbia University' te werken aan de pathogene mechanismen die een rol spelen bij de ontwikkeling van de ziekte van Alzheimer. In 2014 behaalde hij zijn masterdiploma en in het opvolgende jaar vond hij de mogelijkheid om als PhD-kandidaat in de groep van prof. dr. Haico van Attikum, bij het Leids Universitair Medisch Centrum, te werken aan de functionele karakterisatie van varianten in de genen *PALB2* en *CHEK2*, die beide geassocieerd zijn met borstkanker. In 2020 zette hij dit werk voort als postdoctoraal onderzoeker onder de toegevoegde begeleiding van Maaike Vreeswijk. In hetzelfde jaar trouwde hij ook met Tamara Abma en werd kort daarna zijn dochtertje Isabella Hazel Jo geboren. Twee jaar later kwam ook zijn zoontje Odin James ter wereld en begon hij rond dezelfde tijd met een nieuw werk/leertraject als klinisch moleculair bioloog in de pathologie in opleiding, bij Eurofins PAMM in de regio Eindhoven.

PUBLICATIONS

***CHEK2* variants: linking functional impact to cancer risk**

Boonen RACM, Vreeswijk MPG, van Attikum H.

Trends in Cancer. 2022 Sep;8(9):759-770. doi: 10.1016/j.trecan.2022.04.009. Epub 2022 May 25. Review.

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

Boonen RACM, Wiegant WW, Celosse N, Vroling B, Heijl S, Kote-Jarai Z, Mijuskovic M, Cristea S, Solleveld-Westerink N, van Wezel T, Beerenwinkel N, Eeles R, Devilee P, Vreeswijk MPG, Marra G, van Attikum H.

Cancer Research. 2022 Feb 15;82(4):615-631. doi: 10.1158/0008-5472.CAN-21-1845.

Functional characterization *PALB2* variants of uncertain significance: toward cancer risk and therapy response prediction

Boonen RACM, Vreeswijk MPG, van Attikum H.

Frontiers in Molecular Biosciences. 2020 Sep 16;7:169. doi: 10.3389/fmolb.2020.00169. eCollection 2020. Review.

Characterisation of protein-truncating and missense variants in *PALB2* in 15,786 woman from Malaysia and Singapore

Ng PS, Boonen RA, Wijaya E, Chong CE, Sharma M, Knaup S, Mariapun S, Ho WK, Lim J, Yoon SY, Mohd Taib NA, See MH, Li J, Lim SH, Tan EY, Tan BK, Tan SM, Tan VK, van Dam RM, Rahmat K, Yip CH, Carvalho S, Luccarini C, Baynes C, Dunning AM, Antoniou A, van Attikum H, Easton DF, Hartman M, Teo SH.

Journal of Medical Genetics. 2022 May;59(5):481-491. doi: 10.1136/jmedgenet-2020-107471. Epub 2021 Apr 2.

Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene *PALB2*

Boonen RACM, Rodrigue A, Stoepker C, Wiegant WW, Vroling B, Sharma M, Rother MB, Celosse N, Vreeswijk MPG, Couch F, Simard J, Devilee P, Masson JY, van Attikum H.

Nature Communications. 2019 Nov 22;10(1):5296. doi: 10.1038/s41467-019-13194-2.

The role of SLX4 and its associated nucleases in DNA interstrand crosslink repair
Hoogenboom WS, Boonen RACM, Knipscheer P.
Nucleic Acids Research. 2019 Mar 18;47(5):2377-2388. doi: 10.1093/nar/gky1276.

Recruitment and positioning determine the specific role of XPFERCC1 in ICL repair
Klein Douwel D, Hoogenboom WS, Boonen RA, Knipscheer P.
The EMBO Journal. 2017 Jul 14;36(14):2034-2046. doi: 10.15252/embj.201695223. Epub 2017 Mar 14.

A PALB2-interacting domain in RNF168 couples homologous recombination to DNA break-induced chromatin ubiquitylation
Luijsterburg MS, Typas D, Caron MC, Wiegant WW, van den Heuvel D, Boonen RA, Couturier AM, Mullenders LH, Masson JY, van Attikum H.
eLife. 2017 Feb 27;6:e20922. doi: 10.7554/eLife.20922.

Neuronal activity enhances tau propagation and tau pathology in vivo
Wu JW, Hussaini SA, Bastille IM, Rodriguez GA, Mrejeru A, Rilett K, Sanders DW, Cook C, Fu H, Boonen RA, Herman M, Nahmani E, Emrani S, Figueroa YH, Diamond MI, Clelland CL, Wray S, Duff KE.
Nature Neuroscience. 2016 Aug;19(8):1085-92. doi: 10.1038/nn.4328. Epub 2016 Jun 20.

Functional analysis of missense variants in the putative breast cancer susceptibility gene *XRCC2*
Hilbers FS, Luijsterburg MS, Wiegant WW, Meijers CM, Völker-Albert M, Boonen RA, van Asperen CJ, Devilee P, van Attikum H.
Human Mutation. 2016 Sep;37(9):914-25. doi: 10.1002/humu.23019. Epub 2016 Jun 17.

XPF-ERCC1 acts in unhooking DNA interstrand crosslinks in cooperation with FANCD2 and FANCP/SLX4
Boonen RA*, Klein Douwel D*, Long DT, Szypowska AA, Räschle M, Walter JC, Knipscheer P.
Molecular Cell. 2014 May 8;54(3):460-71. doi: 10.1016/j.molcel.2014.03.015. Epub 2014 Apr 10.
**Shared first authorship*

Regulation of slow and fast muscle myofibrillogenesis by Wnt/beta-catenin and myostatin signaling

Tee JM, van Rooijen C, Boonen R, Zivkovic D.

PLOS ONE. 2009 Jun 11;4(6):e5880. doi: 10.1371/journal.pone.0005880.

Wnt signaling in Alzheimer's disease: Up or down, that is the question

Boonen RA, van Tijn P, Zivkovic D.

Ageing Research Reviews. 2009 Apr;8(2):71-82. doi: 10.1016/j.arr.2008.11.003. Epub 2008 Dec 3. Review

DANKWOORD

While doing a PhD was at times overwhelming, it was mostly an unforgettable experience that involved the support of many people. Here I would like to express my gratitude to all of you.

Te beginnen met mijn supervisors en collega's.

Haico, ik ben blij dat ik zoveel heb mogen leren van jouw begeleiding en adviezen, die hebben geleid tot een aantal mooie publicaties en prachtige ervaringen. Ik wil je graag bedanken voor de fijne samenwerking en de ontwikkelingen die ik in jouw lab heb mogen doormaken.

Mijn copromotoren Peter Devilee en Maaike Vreeswijk, Ik wil jullie graag bedanken voor jullie enorme betrokkenheid bij alle projecten waar ik aan heb mogen werken. Ik ben ontzettend dankbaar dat ik van jullie beide zoveel heb mogen leren op het gebied van klinisch wetenschappelijk onderzoek.

Ook mijn voorgaande werkgevers Dana Zivkovic en Puck Knipscheer wil ik graag bedanken voor het vertrouwen dat jullie altijd in mij hebben gehad en de steun die ik van jullie beide heb gekregen. Zonder jullie had ik nooit een PhD traject aan kunnen gaan.

This thesis would not have reached its final form without the help of many current and former lab members. Sabine, Wouter, Magda, Chantal, Milan, Nandi and Aarzo. I greatly appreciate all of your contributions to this work, as well as the pleasant collaboration with all of you. Especially Sabine and Wouter, whose contributions to this work were massive.

Jenny, I want to thank you for all the great conversations and for simply being a great friend and colleague. I hope I will see more of you in times to come.

Bert, bedankt voor je luisterend oor. Het was ontzettend fijn dat ik met jou altijd werk gerelateerde en persoonlijke ervaringen kon delen.

Romy and Veronica, thank you for all the laughs during the more recent times. Romy, I will dearly miss our conversations about photography, as well as working together with you in the VUS field. I will always consider you a great friend.

Also, thanks to all current or former members of the 'van Attikum lab', Anton, Kees, Fenna, Sophie, Jolanthe, Leonie, Amandine, Pierre, Sylvie, Martijn, Madelon, Mireille, for all your incredible support.

Joost, I greatly appreciate all your advice regarding the mESC system and techniques such as CRISPR/Cas, which was of course invaluable to this thesis.

I want to thank everyone from the Human Genetics department, especially Nuria for the great and pleasant working environment.

Dan nu mijn dierbare vrienden.

Bart en Bram, het blijft fijn dat ik met jullie te alle tijden alles zou kunnen bespreken. Ik wil jullie beide bedanken voor alle steun in de afgelopen jaren en ik hoop dat ik jullie altijd mijn vrienden mag noemen.

Iedereen die ik binnen MAB heb mogen leren kennen; Oscar, Niek Joanknecht, Sjoerd, Salam, Maarten, Niek Sonneveld, Frank, Mathieu, Wilbert, Rodi, Fedor, Wouter, Thomas, Jelmer, Ischa, Gerson en Ouren. Ik wil jullie allemaal ontzettend bedanken voor jullie vriendschap en alle onvergetelijk momenten. Deze hebben altijd veel voor me betekend en ervoor gezorgd dat ik alle zorgen even opzij kon zetten.

Tot slot mijn lieve familie.

Kim, ik weet dat ik je veel te weinig heb gezien de afgelopen jaren en ik hoop dat daar (met deze mijlpaal achter de rug) snel verandering in gaat komen. Ook al laat ik het niet altijd merken, weet dat je ontzettend veel voor mij betekent.

Mam, om redenen die ik hier niet zal noemen ben jij uitgegroeid tot een ontzettend sterk persoon. Ik kan alleen maar hopen dat daarvan een beetje is overgedragen op mij. Door jou steun wist ik in ieder geval altijd mijn rug recht te houden in lastige tijden.

Pap, je bent een prachtmens en ik zal altijd tegen je op kijken. Vandaar ook deze quote; “Every father should remember that one day his son will follow his example, not his advice” (Charles F. Kettering).

Joke en Frans, ik ben blij dat ik jullie mijn schoonouders mag noemen en dank jullie graag voor alle steun die wij (Tamara en ik) de afgelopen jaren van jullie hebben gekregen. Salam en Amanda, een erg speciale dankbaarheid gaat uit naar jullie. Zonder jullie had ik namelijk mijn vrouw niet gekend en had ik dit proefschrift niet met liefde kunnen opdragen aan mijn dierbare kinderen Isabella en Odin.

Lieve Tamara en mama van Isabella en Odin. Wat ben ik toch blij dat ik jou mijn vrouw mag noemen. We hebben geen makkelijke start gehad samen, maar zonder die ervaring had ik me misschien nooit gerealiseerd hoeveel jij me waard bent. Daar zijn geen woorden voor. Zelfs met de vermoeidheid van twee jonge kinderen, geniet ik nog elke dag van jou. Ik draag dit boekje dan wel op aan Isabella en Odin, maar zonder jou had ik hier niet gestaan met zoveel rijkdom om me heen.