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

Boonen, R.A.C.M.

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

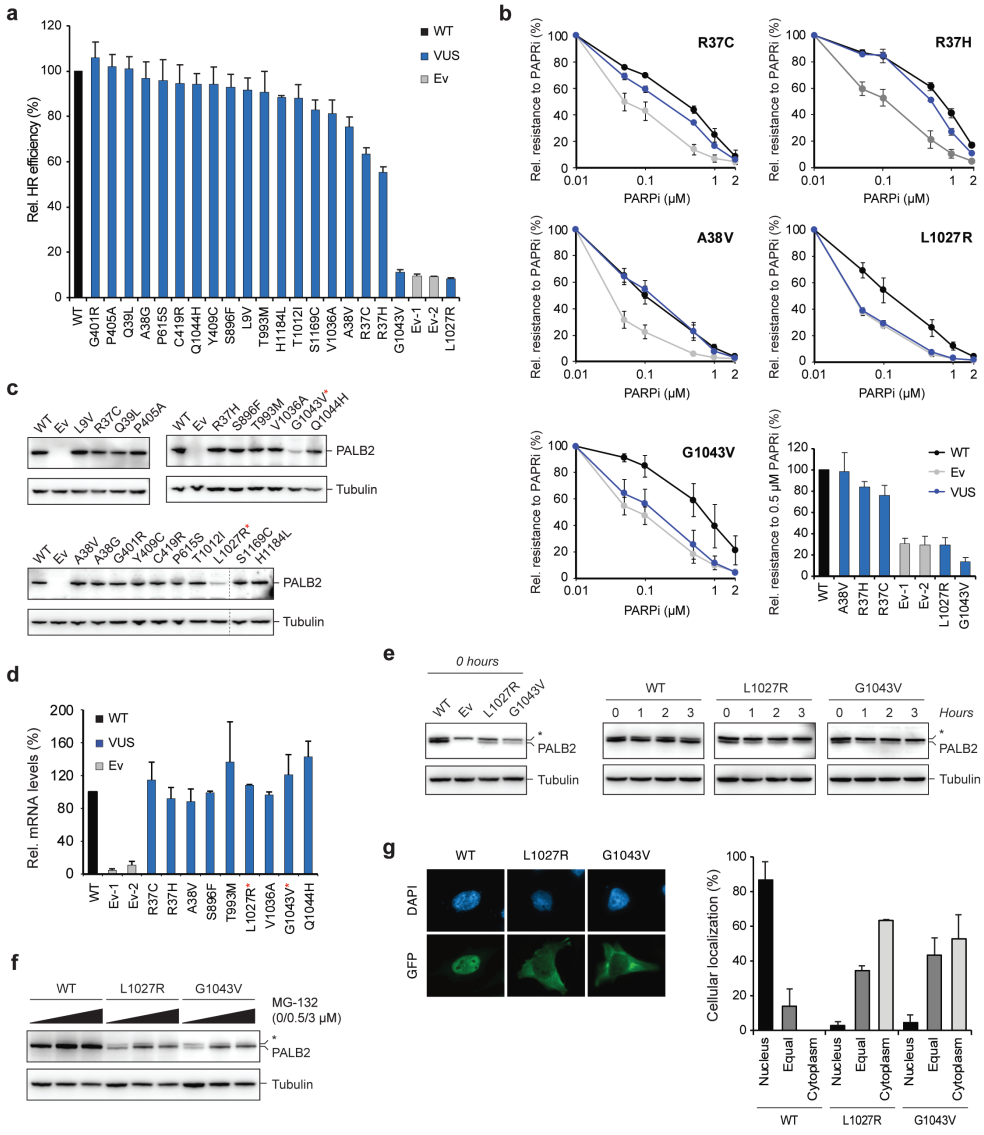

Figure 1. Functional analysis of *PALB2* missense variants from an Asian cohort. **a** HR assay (DR-GFP) in *Trp53*^{KO}/*PALB2*^{KO} mES cells expressing the indicated *PALB2* variants (or an empty vector, Ev). Normalized values are plotted with the wild type (WT) condition set to 100% (absolute HR 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

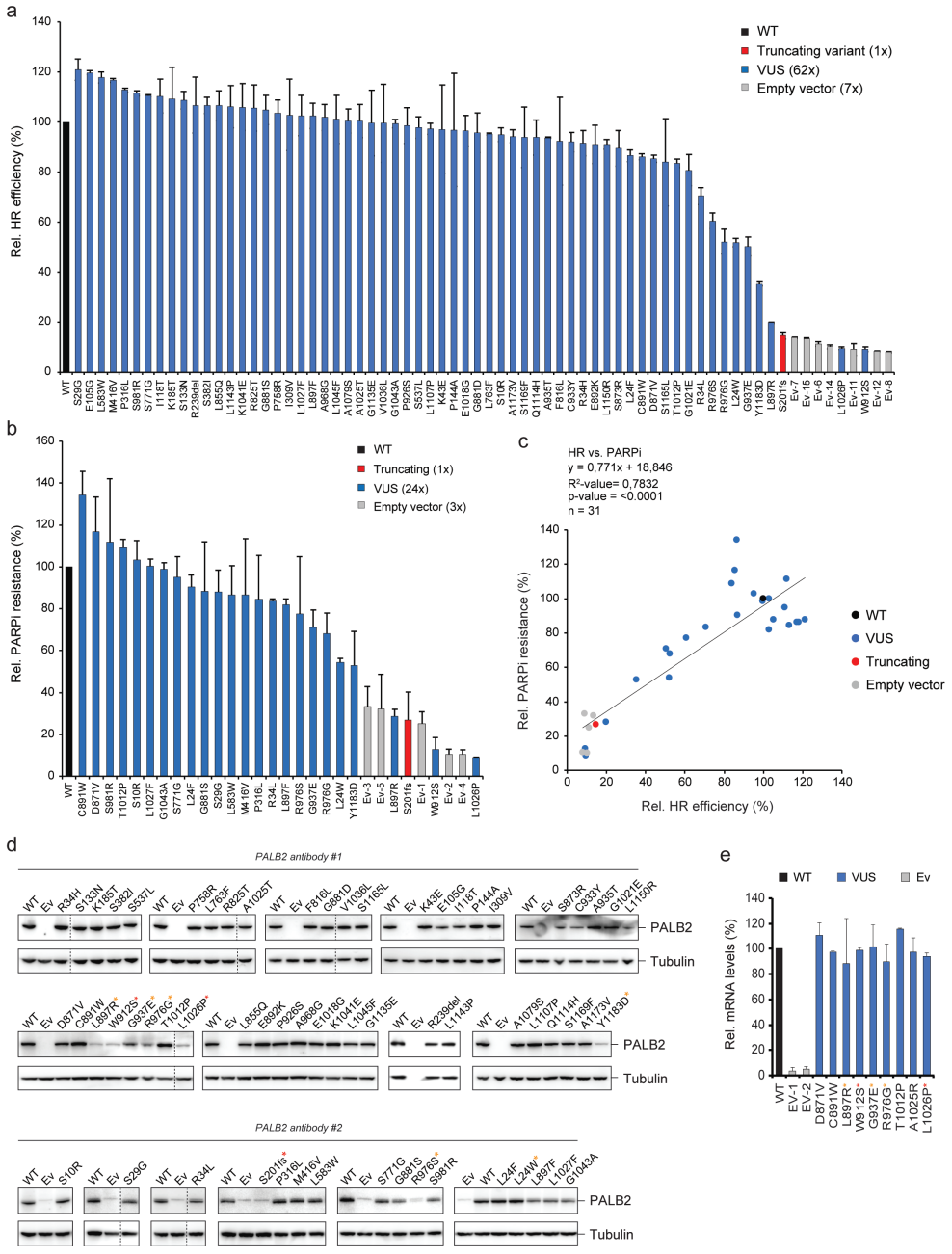

Figure 2. Functional analysis of *PALB2* missense variants identified in 44 BCAC studies. **a** HR assay as in (Fig. 1a). **b** Proliferation-based PARP inhibitor (PARPi) sensitivity assay as in (Fig. 1b). **c** Scatter plot showing the correlation between HR efficiencies and PARPi sensitivity for variants 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

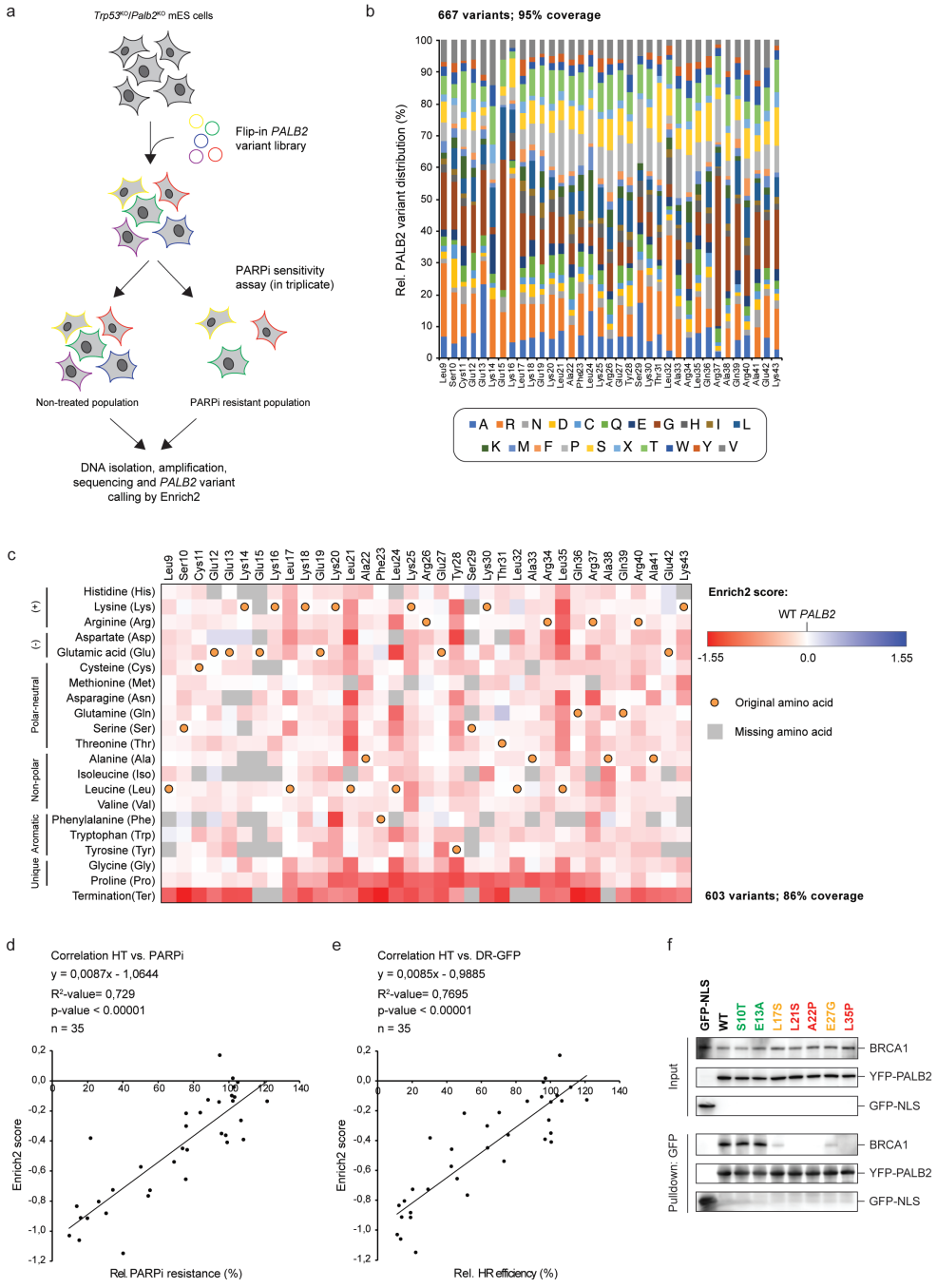

Figure 3. High-throughput analysis of *PALB2* variants in the CC domain. **a** Schematic flow of the high-throughput functional analysis employed in this study. **b** Bar graph showing the variant diversity distribution of the CC-variant library containing 667 distinct *PALB2* variants. **c** Amino acid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

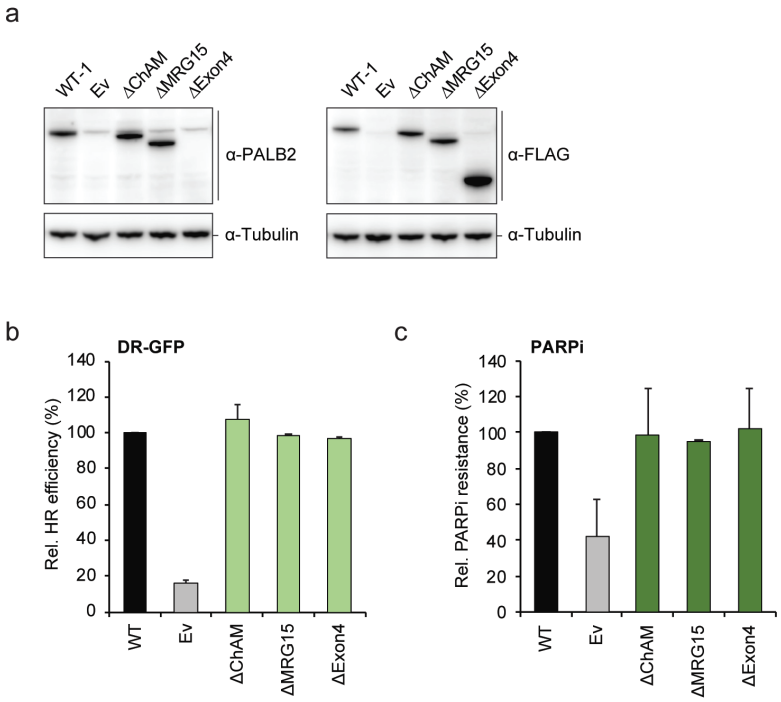

Figure 4. Functional analysis of *PALB2* deletion variants. **a** Western blot analysis for the expression of three *PALB2* domain deletion variants as indicated. Tubulin is a loading control. **b** HR assay as in (Fig. 1a). **c** Proliferation-based PARP inhibitor (PARPi) sensitivity assay as in (Fig.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

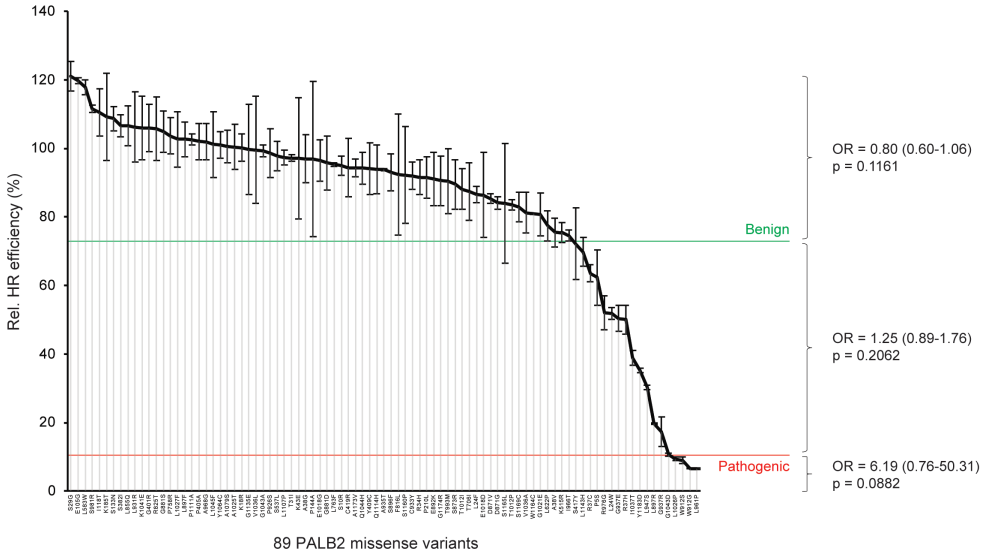

Figure 5. Association of PALB2 HR efficiency with breast cancer risk. Bar graph showing results from HR assays (DR-GFP) in *Trp53*^{KO}/*PALB2*^{KO} mES cells complemented with 89 distinct human *PALB2* variants. Previously published results [3], as well as those from (Fig. 1a) and (Fig. 2a) are shown. 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.