The genetic etiology of familial breast cancer: Assessing the role of rare genetic variation using next generation sequencing

Hilbers, F.S.M.

Chapter 4

Functional analysis of missense variants in the putative breast cancer susceptibility gene *XRCC2*

Florentine S. Hilbers, Martijn S. Luijsterburg, Wouter W. Wiegant, Caro M. Meijers, Moritz Völker-Albert, Rick A. Boonen, Christi J. van Asperen, Peter Devilee, Haico van Attikum

Hum Mutat. 2016 Sep; 37(9): 914-25

Abstract

XRCC2 genetic variants have been associated with breast cancer susceptibility. However, association studies have been complicated because *XRCC2* variants are extremely rare and consist mainly of amino acid substitutions whose grouping is sensitive to misclassification by the predictive algorithms. We therefore functionally characterized variants in *XRCC2* by testing their ability to restore XRCC2-DNA repair deficient phenotypes using a cDNA-based complementation approach. While the protein-truncating variants p.Leu117fs, p.Arg215* and p.Cys217* were unable to restore XRCC2 deficiency, 19 out of 23 missense variants showed no or just a minor (<25%) reduction in XRCC2 function. The remaining 4 (p.Cys120Tyr, p.Arg91Trp, p.Leu133Pro and p.Ile95Lx) had a moderate effect. Overall, measured functional effects correlated poorly with those predicted by in silico analysis. After regrouping variants from published case-control studies based on the functional effect found in this study and reanalysis of the prevalence data, there was no longer evidence for an association with breast cancer. This suggests that if breast cancer susceptibility alleles of *XRCC2* exist, they are likely restricted to protein-truncating variants and a minority of missense changes. Our study emphasizes the use of functional analyses of missense variants to support variant classification in association studies.

Introduction

One of the most important risk factors for breast cancer is having family members affected with the disease.[1] Mutations in two of the most well-characterized susceptibility genes, *BRCA1* (OMIM 113705) and *BRCA2* (OMIM 600185), explain about 17% of this familial relative risk.[2–4] Although a number of additional moderate and high-risk genes have been identified,[5] these explain at most another 5%. [4] Any yet unidentified moderate or high-risk breast cancer gene is likely to have extremely low mutation frequencies, hampering its discovery. Next-generation sequencing has provided the possibility to detect these rare variants on a genome-wide scale. However, due to the very low allele frequencies of these variants, obtaining conclusive evidence for an association with breast cancer has proven a considerable challenge.

XRCC2 (OMIM 600375) was among the first genes in which variants were reported to be associated with breast cancer risk by a study employing next-generation sequencing.[6] However, variants in *XRCC2* are very rare, with about 0.5% of healthy individuals carrying one (with the exception of the common single nucleotide polymorphism (SNP) rs3218536), and most have been detected only once. Nonetheless, by genotyping familial or early onset cases and healthy controls a significant association with breast cancer was detected.[6] *XRCC2* was a very interesting candidate breast cancer gene for several reasons. Firstly, it has been identified as a potential Fanconi anemia gene in a Saudi Arabian patient from consanguineous parents.[7] Currently, for at least four of the sixteen known Fanconi anemia genes, it has been shown that, while two mutated alleles cause Fanconi anemia, a single affected allele increases the risk of breast cancer.[8–12] Secondly, *XRCC2*, like many other confirmed breast cancer genes, encodes a protein involved in the repair of DNA double-strand breaks (DSB) via homologous recombination (HR). XRCC2 forms a complex with several members of the RAD51 protein family, including RAD51B, RAD51C and RAD51D. This complex is required for the localization of RAD51, a key mediator of HR, to DSB sites.[13] During HR, the ends of a DSB are resected to form single-strand DNAs, which are coated with RPA. RAD51 replaces RPA and subsequently initiates the search for a homologous sequence to enable repair of the DSB.[14]

Although *XRCC2* is a promising candidate breast cancer gene, a second, larger case-control study initiated by our lab, did not confirm its association with breast cancer since an equal percentage of variants in familial cases and healthy controls was found.[15] However, both case-control studies lacked the power to examine associations with individual variants or even with variant subtypes, and therefore protein-truncating variants and missense variants predicted to be damaging by in silico analysis were pooled in order to calculate the association with breast cancer. Given the very low number of variants detected, misclassification of only a few variants can already strongly influence the results of an association analysis, which might explain the discrepancy between the two studies.

Here we aimed to functionally characterize the *XRCC2* non-synonymous coding variants by testing their ability to complement the DNA repair phenotype of XRCC2-deficient cells. Remarkably, in our functional assays the majority of missense variants found in the population did not affect XRCC2's function in DNA repair, although in silico analyses predicted several of them to be damaging. When we re-classified the *XRCC2* variants based on these functional effects and re-analyzed the data from the two previously published case-controls studies, evidence for an association with breast cancer disappeared. These findings illustrate how functional analyses may aid the interpretation of rare genetic variation in the context of disease susceptibility.

Materials and Methods

Variant selection and in silico analysis

We used four different prediction algorithms to estimate how likely variants from breast cancer cases and healthy controls[6,15] as well as a Fanconi Anemia case[7] would affect XRCC2 protein function. PolyPhen2 (v2.2.2r398),[16] SIFT (JCVI-SIFT v.1.03, cut-off 0.05)[17] and AlignGVGD (depth until Branchiostoma floridae (Lancelet), cut-off C45)[18] predict the effect of a single amino acid change, while CADD (v1.2, cut-off 20)[19] can also provide predictions for small insertions or deletions and variants that result in a premature stop codon. All variants mentioned in this paper are described based on DNA sequence NM_005431.1 and protein sequence NP_005422.1. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1.

We used five different splice site prediction algorithms to predict the effect of the selected variants on splicing: SpliceSiteFinder (cut-off 70), MaxEntScan (cut-off 0),[20] Splice Site Prediction by Neural Network (cut-off 0.4),[21] GeneSplicer (cut-off 0),[22] and Human Splicing Finder (cut-off 65).[23] Prediction algorithms were run by the integrated software package Alamut (version 4.7.1) using default settings. An effect on splicing was considered when more than 10% difference in at least two algorithms was observed[24].

Cell lines and culture conditions

All cell lines were grown in DMEM-F12 (Gibco) containing 10% fetal bovine serum (Bodinco), Glutamax (Gibco, 1x), HEPES (Gibco, 10mM), Na-pyruvate (Lonza, 1mM) and penicillin/streptomycin (Gibco, 100 units/mL /100ug/mL). *irs1* cells were kindly provided by John Thacker.[25,26] *irs1* cells carrying the sister chromatid reporter (SCR) (irs-SCR) were a kind gift from Ralph Scully.[27,28] HEK293 cells containing a DR-GFP reporter were obtained from Maria Jasin.[29]

siRNAs, plasmids and site-directed mutagenesis

An mCherry-XRCC2 expression vector was created by inserting the human *XRCC2* cDNA sequence (OriGene, RG208330) into mCherry-C1. All selected variants were introduced into this vector using two-step site directed mutagenesis or, in case of variants in regions with high A/T content, by mutagenic overlap PCR followed by cloning. For both methods we made use of complementary primers containing the desired variant and an approximately 20bp sequence on both sites. The presence of the variants and the lack of additional mutations in the cDNA region were confirmed with Sanger sequencing (data not shown). An siRNA-resistant version of mCherry-XRCC2 was generated by mutating all wobble bases in the siRNA-targeting sequence using mutagenic overlap PCR such that the amino acid sequence of the XRCC2 protein would not change. Variants of interest were introduced in this vector using the aforementioned methods. In addition, wildtype and p.Arg17*-containing *XRCC2* cDNAs were amplified by PCR to remove the stop codon and introduce a C-terminal FLAG tag. The PCR products were subsequently inserted into mCherry-C1. All primer and siRNA sequences can be found in Supplementary Material, Tables S1 and S2.

Western blot analysis

Cell pellets we lysed in RIPA buffer and subsequently incubated at 95°C for 10 minutes in Laemmli buffer. Equal amounts of each sample were separated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis and transferred to PVDF membranes (Millipore).

Membranes were probed with primary antibodies to XRCC2 (N-20, 73278, Santa Cruz), mCherry (ab125096, Abcam) or FLAG (#F1804, Sigma-Aldrich), and appropriate secondary antibodies (LI-COR Biosciences) followed by protein detection using the Odyssey imaging scanning system (LI-COR Biosciences).

RAD51 foci formation after MMC treatment in *irs1*

The localization of RAD51 to DSBs was examined by assessing RAD51 foci formation after mitomycin C (MMC) treatment. On day one *irs1* cells were seeded in 12-wells plates containing 12mm coverslips at 30.000 cells /well. On day two the cells were transfected with 500ng mCherry or mCherry-XRCC2 expression vector using Effectene Transfection Reagent (Qiagen) according to the manufactor's protocol. On day three the medium containing the transfection reagents was removed and the cells were incubated for one hour with medium containing 6μM MMC. Cells were then incubated with fresh medium lacking MMC and six hours later fixed using 4% formaldehyde and permeabilized with 0.5% Triton X-100. Subsequently, the coverslips were incubated with 100mM glycine for 10 minutes and phosphate buffered saline containing 0.5% bovine serum albumin and 0.05% Tween 20 for 10 minutes. Finally, coverslips were incubated with anti-RAD51 antibody (H-92, Santa Cruz) at a 1:100 dilution for one hour, followed by incubation with goat anti-rabbit-488 (A-11034, Life Technologies) at 1:1000 and DAPI at 0.1 μg/ml for one hour. The fraction of mCherry-positive cells containing ≥5 RAD51 foci was scored manually. For each variant the experiment was repeated at least three times. Between 100 and 300 cells were scored in each experiment.

Sister chromatid recombination reporter assay in *irs1*

This sister chromatid recombination (SCR) reporter assay was essentially performed as described previously.[27] Briefly, *irs1* cells containing the SCR reporter were seeded at day one at 30.000 cells/well in a 12-well plate format. On day two the cells were transfected with 200ng of mCherry or mCherry-XRCC2 expression vector in combination with 600ng of either a I-SceI expression vector (pCBASce) or a control vector (pCAGGS).[29] Three days later the fraction of mCherry- and GFP-positive cells was determined on a LSRII flow cytometer (BD Biosciences). Data analysis was performed using Cytobank.[30] HR efficiencies were calculated as GFP-to-mCherry ratios. For each variant the experiment was repeated at least three times.

Microscopic analysis of fixed cells

Images of fixed samples were acquired on a Zeiss AxioImager M2 widefield fluorescence microscope equipped with 40x, 63x and 100x Planar apochromatic (1.4 numerical aperture) oil-immersion objectives (Zeiss) and an HXP 120 metal-halide lamp used for excitation. Fluorescent probes were detected using the following filters: DAPI (excitation filter: 350/50 nm, dichroic mirror: 400 nm, emission filter: 460/50 nm), GFP/Alexa 488 (excitation filter: 470/40 nm, dichroic mirror: 495 nm, emission filter: 525/50 nm) and mCherry (excitation filter: 560/40 nm, dichroic mirror: 585 nm, emission filter: 630/75 nm). Images were recorded using ZEN 2012 software (Zeiss).

DR-GFP reporter assay in HEK293

HEK293 cells carrying the DR-GFP reporter[29] were used to measure HR efficiencies essentially as described.[31] Briefly, on day one cells were seeded in 12-well plates at 225.000 cells/well. Cells were then transfected the same day and the second day with siRNA against *XRCC2* or a

control siRNA (Sigma, Universal Negative Control #1) at 20nM using Lipofectamine RNAiMax (Life Technologies) according to the manufacturer's protocol. At day three the cells were co-transfected with 500ng mCherry or siRNA-resistant mCherry-XRCC2 expression vector and 4µg of either an I-SceI expression vector (pCBASce) or a control vector (pCAGGS) using Lipofectamine 2000 (Life Technologies) (Pierce et al., 1999). Two days later cells were analyzed on a LSRII flow cytometer. Data analysis was performed using Cytobank.[30] HR efficiencies were calculated as GFP-to-mCherry ratios. For each variant the experiment was repeated at least three times. For cell cycle analysis, on day five cells were pulse labeled with 1µM EdU for two hours and then fixed and stained using the Click-iT EdU Flow Cytometry Kit (Invitrogen) according to protocol using an Alexa Fluor 488 dye. Knockdown of XRCC2 was validated using reverse-transcriptase quantitative PCR (RT-qPCR) as described in [32] using primer sets that either detect mRNA from the endogenous *XRCC2* or siRNA-resistant *XRCC2* cDNA (Table S1).

Table 1. In silico predictions and measured functional effect for all *XRCC2* genetic variants

DNA (NM_005431.1)	Protein (NP_005422.1)	Study	Poly-Phen-2	SIFT[a]	Align-GVGD[b]	CADD[c]	RAD51 foci - Hamster	SCR - Hamster	DR-GFP - Human
c.46 G>T [d]	p.Ala16Ser	(5)	Possibly damaging	Tolerated	C0	24.300	81%	100%	NA
c.49C>T	p.Arg17*	(5)	-	-	-	26.600	66%	111%	77%
c.140A>G	p.His47Arg	(15)	Possibly damaging	Tolerated	C0	24.000	90%	89%	NA
c.181C>A	p.Leu61Ile	(5)	Possibly damaging	Damaging	C0	22.600	83%	96%	NA
c.223G>C	p.Glu75Gln	(15)	Possibly damaging	Tolerated	C0	22.100	88%	89%	NA
c.247dup[e]	p.Thr83fs	-	-	-	-	29.000	11%	7%	48%
c.271C>T	p.Arg91Trp	(5)	Probably damaging	Damaging	C65	27.400	71%	66%	62%
c.283A>C [f]	p.Ile95Leu	(5)	Benign	Damaging	C0	0.001	77%	91%	70%
c.283A>G	p.Ile95Val	(5,15)	Benign	Tolerated	C0	0.002	90%	95%	NA
c.350delT	p.Leu117fs	(15)	-	-	-	29.400	51%	81%	56%
c.353T>C [g]	p.Val118Ala	(15)	Benign	Tolerated	C0	4.263	81%	101%	NA
c.359G>A	p.Cys120Tyr	(15)	Probably damaging	Damaging	C65	24.900	77%	70%	60%
c.398T>C	p.Leu133Pro	(15)	Probably damaging	Damaging	C65	24.000	62%	82%	67%
c.490G>C	p.Glu164Gln	(15)	Probably damaging	Damaging	C0	25.000	96%	75%	NA
c.509A>C	p.Glu170Ala	(15)	Probably damaging	Damaging	C65	27.400	101%	98%	NA
c.562C>T	p.Arg188Cys	(15)	Probably damaging	Damaging	C15	29.300	73%	102%	81%
c.563G>A [h]	p.Arg188His	(5,15)	Benign	Tolerated	C0	18.980	88%	86%	97%
c.581C>T	p.Thr194Met	(15)	Probably damaging	Damaging	C15	27.000	82%	92%	NA
c.595A>C	p.Met199Ile	(15)	Benign	Tolerated	C0	9.678	69%	97%	99%
c.620A>G [i]	p.Glu207Gly	(15)	Benign	Tolerated	C0	14.470	91%	84%	82%
c.643C>T	p.Arg215*	(6)	-	-	-	34.000	41%	79%	42%
c.651_652delTG	p.Cys217*	(5)	-	-	-	24.400	27%	61%	31%
c.659A>T	p. Asp220Val	(15)	Benign	Damaging	C0	13.880	86%	96%	NA
c.693G>T	p.Trp231Cys	(5)	Probably damaging	Damaging	C65	32.000	88%	88%	NA
c.714G>C	p.Arg238Ser	(15)	Probably damaging	Tolerated	C65	18.570	77%	94%	NA
c.742C>G [j]	p.Gln248Glu	(15)	Benign	Tolerated	C0	0.001	94%	92%	NA
c.772C>T	p.Arg258Cys	(15)	Benign	Tolerated	C0	22.500	71%	88%	108%
c.808T>G [k]	p.Phe270Val	(5,15)	Probably damaging	Damaging	C45	24.900	76%	85%	NA

All non-synonymous coding *XRCC2* variants from Park et al. (2012) and Hilbers et al. (2012) were selected for functional testing and in silico analysis by Polyphen-2, SIFT, AlignGVGD and CADD. In addition, a variant described by Shamseldin et al. (2012), detected in a Fanconi Anemia case and a truncating variant, p.Thr83fs reported by Douglas Easton (personal communication) were analyzed. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1. [a] based on a 0.05 cut-off, [b] AlignGVGD depth until Branchiostoma floridae (Lancelet) Class 0 is the least likely to be damaging, Class C65 is the most likely to be damaging, [c] PHRED scaled C-score, a high score indicates a high change for the variant to be deleterious, a score of 20 is commonly used as a cut-off, [d] rs4987090, [e] reported by Douglas Easton (personal communication), [f] rs140214637, [g] rs185815454, [h] rs3218536, [i] rs61762969, [j] rs190900560 [k] rs145085742 .

Results

In silico prediction tools yield conflicting results

We selected all non-synonymous coding variants from the two major studies assessing the association of rare genetic variants in *XRCC2* and breast cancer risk for functional analysis in this study (Table 1).[6,15] Among these variants was p.Arg188His (rs3218536), the only variant in *XRCC2* common enough to directly establish its association with breast cancer. However, two meta-analyses, both including over 30,000 cases and controls, did not find association with breast cancer risk for this SNP.33,34 In addition we included a variant found in a Fanconi Anemia case (c.643C>T, p.Arg215*).[7] These variants comprised twenty-three amino acid changes, two deletions and two premature stop codons (Fig. 1A). In silico prediction algorithms PolyPhen-2,[16] SIFT,[17] AlignGVGD[18] and CADD[19] were used to predict if these variants would affect XRCC2 protein function. Polyphen and SIFT both use multiple sequence alignment to assess the effect of amino acid changes on protein function. Polyphen also takes into account information on protein domains and three-dimensional structures. AlignGVGD combines Grantham difference of amino acids with ortholog multiple sequence alignment, whereas CADD combines 63 different annotations, including PolyPhen, SIFT and Grantham differences of amino acids, to predict the effect of amino acid changes on protein function. The number of amino acid changes predicted to be "damaging" based on these prediction tools ranged from seven (AlignGVGD) to fourteen (CADD). Six variants, p.Arg91Trp, p.Cys120Tyr, p.Leu133Pro, p.Glu170Ala, p.Trp231Cys and p.Phe270Val (26%), were classified as damaging by all prediction algorithms. For eleven variants (48%) the algorithms yielded conflicting results and for six variants (26%) all agreed on them being benign. The level of disparity between the in silico tools demonstrates the need for more conclusive data on the functional effects of these variants.

Few XRCC2 variants affect RAD51 foci formation in hamster cells

To assess the effect of the missense variants on *XRCC2* function, we introduced all variants in a human XRCC2 cDNA expression vector and assessed to which extent these variants could complement the repair defect of XRCC2-deficient *irs1* hamster cells (Fig. 1B). One of the main phenotypic characteristics of *irs1* cells is that the accumulation of RAD51 into distinct foci containing DSBs is strongly reduced.[26,35] Indeed, we could confirm that in response to treatment with the DSB-inducing agent Mytomycin C (MMC) RAD51 foci formation was almost completely absent in *irs1* cells that were transfected with an mCherry expression vector (Fig. 1C). In contrast, the expression of mCherry-tagged human *XRCC2* restored RAD51 recruitment to sites of DNA damage in *irs1* cells. This is consistent with the results of previous studies showing that human *XRCC2* expression can complement the RAD51 foci formation defect in *irs1*.[26,35]

To assess the functionality of the selected *XRCC2* missense variants, we generated a collection of mCherry-tagged XRCC2 fusions containing these variants and measured their ability to restore RAD51 foci formation following MMC exposure in *irs1* cells (Fig. 1B). We found most missense variants, including the common SNP p.R188H for which no association with breast cancer risk was found, to rescue RAD51 foci formation to more than 75% of the levels seen for wild-type XRCC2 (Fig. 1D). Only five out of twenty-three variants were able to rescue RAD51 foci formation to levels below 75% (p.Arg91Trp, p.Leu133Pro, p.Arg188Cys, p.Met199Ile

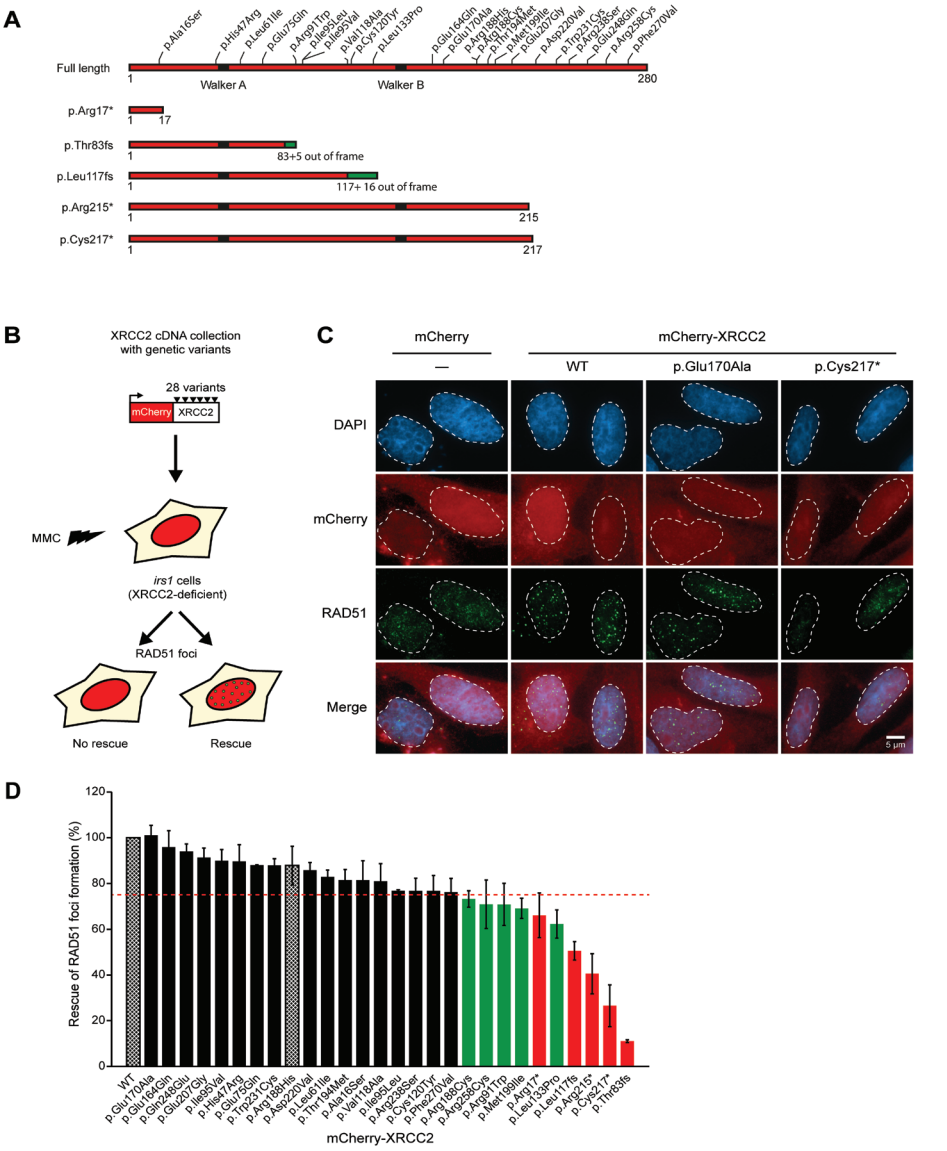

Figure 1. Functional analysis of the effect of *XRCC2* variants on RAD51 foci formation in hamster cells.
(A) Schematic overview of the *XRCC2* variants. Red indicates the in-frame *XRCC2* sequence, while green represents out-of-frame amino acids. (B) Workflow for the functional analysis of the effect of *XRCC2* variants on RAD51 foci formation in XRCC2-deficient *irs1* hamster cells. A collection of mCherry-tagged *XRCC2* cDNAs containing selected genetic variants was introduced in *irs1* cells. Following treatment of the cells with MMC, RAD51 foci formation was monitored to determine whether the *XRCC2* variants would rescue the defect in RAD51 foci formation of the *irs1* cells. (C) Examples of RAD51 foci formation in *irs1* cells expressing mCherry, mCherry-XRCC2 (wildtype; WT), mCherry-XRCC2 carrying a functional variant (p.E170A), and mCherry-XRCC2 carrying a protein-truncating variant (p.C217*). (D) Effect of *XRCC2* variants on the rescue of RAD51 foci formation in *irs1* cells. The percentage of cells with >5 MMC-induced RAD51 foci was scored. Cells expressing mCherry-XRCC2 (WT) served as a reference, which was set to 100%. The checked bars indicate wildtype *XRCC2* and the neutral common polymorphism p.R188H. Green bars indicate missense variants, whereas red bars indicate truncating variants with rescue levels below 75%. Error bars indicate standard error of the mean.

and p.Arg258Cys; from 62%-73%). Importantly, three protein-truncating variants, p.Leu177fs, p.Arg215* and p.Arg217*, rescued RAD51 foci formation to levels between 26% and 50%, suggesting impaired protein function. Unexpectedly, a fourth protein-truncating variant, p.Arg17*, showed a rescue of 66%. To verify that our assay was sensitive enough to detect differences in complementation for variants with a drastic effect on the protein produced, we also tested another variant, p.Thr83fs (Douglas Easton, personal communication), that would result in a truncated protein much shorter than the three giving a moderate effect. This variant rescued RAD51 foci formation to only 11%. Western blot analysis confirmed that expression of all protein-truncating variants produced a protein with a molecular weight consistent with the position of the premature stop-codon in the *XRCC2* coding sequence (Supplementary Material, Fig. S1), ruling out that the observed defects were due to impaired protein production. Thus, these results reveal that the truncating variants in *XRCC2* affect protein function, while most missense variants do not.

Few XRCC2 variants affect DSB repair via HR in hamster cells

RAD51 foci formation represents an intermediate step in the process of HR. Therefore, it is important to also establish the effect of genetic variants in *XRCC2* on the completion of DSB repair via this pathway. To this end, we employed *irs1* cells carrying a sister chromatid-recombination (SCR) reporter (*irs1*-SCR)[27] (Fig. 2A). This construct consists of two differentially mutated GFP sequences, one of which harbors an I-SceI restriction site (Fig. 2B). Expression of the I-SceI nuclease in these cells generates a DSB that is primarily repaired via HR, which results in the restoration of a functional GFP gene. Hence, the percentage of GFP-positive cells after I-SceI expression is a measure of the efficiency of HR. When we co-transfected *irs1*-SCR cells with an I-SceI expression and mCherry control vector hardly any GFP positive cells were detected, indicating that the cells are indeed incapable of repairing DSBs via HR. However, when mCherry-tagged wild-type *XRCC2* was introduced, a 16-fold increase in GFP-positive cells compared to mCherry alone was observed (Fig. 2C), comparable with previous results.[28] This also indicates that under our experimental conditions, HR can be rescued with an efficiency that allows us to assess the functionality of *XRCC2* variants.

We subsequently introduced our collection of mCherry-tagged *XRCC2* expression vectors containing the selected variants into the *irs1*-SCR cells and measured HR by flow cytometry. Similar to the results of our RAD51 foci analysis, we found that most missense variants, including the common SNP p.R188H, had no or only a very small effect on the ability to complement the HR repair defect in these cells (Fig. 2D). On average, the effects of these genetic variants in the SCR reporter assay were, however, smaller than those in the RAD51 foci formation assay. In fact, only three missense variants (p.Glu164Gln, p.Cys120Tyr and p.Arg91Trp) rescued to 75% or slightly less when compared to wild-type XRCC2. Surprisingly, for three of the four protein-truncating variants the effects were also smaller than those seen in the RAD51 foci analysis, ranging from 60% to 81% rescue. However, for the shortest truncated protein (p.Thr83fs), we observed a very strong defect in HR with only 7% rescue compared to wild-type, indicating that the assay is capable of identifying alleles that have impaired protein function. Interestingly, p.Arg17*, the protein-truncating variant that showed unexpectedly good rescue in the RAD51 foci assay, also fully restored HR in this assay. Overall the results from the sister chromatid reporter assay concur with those from the RAD51 foci assay and indicate that only a small minority of the missense variants affects XRCC2 function.

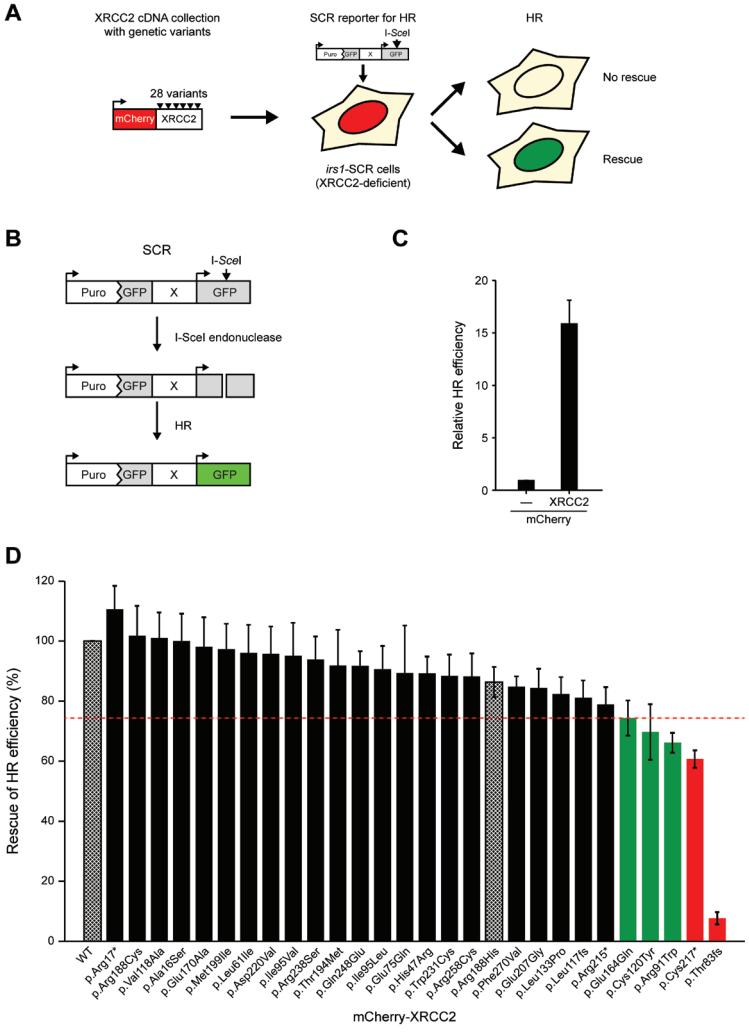

Figure 2. Functional analysis of the effect of *XRCC2* variants on HR in hamster cells.
(A) Workflow for the functional analysis of the effect of *XRCC2* variants on HR in XRCC2-deficient *irs1* hamster cells. A collection of mCherry-tagged XRCC2 cDNAs containing selected genetic variants was introduced in *irs1* cells containing the SCR reporter for HR (*irs1*-SCR cells). Following expression of the I-SceI nuclease, the fraction of GFP-expressing cells among the mCherry-positive cells was determined to assess whether the *XRCC2* variants would rescue the HR defect of *irs1*-SCR cells. (B) Schematic of the sister chromatid reporter (SCR) assay. (C) Relative HR efficiency of *irs1*-SCR cells expressing mCherry-XRCC2 compared to mCherry alone. The repair efficiency in cells expressing mCherry was set to 1. (D) Effect of *XRCC2* variants on the rescue of HR in *irs1*-SCR cells. Cells expressing mCherry-XRCC2 (WT) served as a reference, which was set to 100%. The checked bars indicate wildtype *XRCC2* and the neutral common polymorphism p.R188H. Green bars indicate missense variants, whereas red bars indicate truncating variants with rescue levels below 75%. Error bars indicate standard error of the mean.

Validation of the effect of XRCC2 variants on HR in human cells

Given our observation that only few *XRCC2* variants fail to complement the repair phenotype of *irs1* hamster cells, we next asked if these variants would show a similar effect in human cells. To this end, we used human HEK293 cells carrying the HR reporter DR-GFP to validate the results of a subset of *XRCC2* variants[29] (Fig. 3A). Like the SCR reporter in *irs1* cells, the DR-GFP reporter allows us to measure the efficiency of HR by flow cytometry as the percentage of GFP-positive cells after I-SceI expression (Fig. 3B). We used siRNA-mediated knockdown to reduce the expression of endogenous XRCC2 in cells that either expressed mCherry alone or an siRNA-resistant version of mCherry-tagged XRCC2. Western blot analysis and quantitative reverse transcription PCR (RT-qPCR) with primers that specifically recognize transcripts from the endogenous *XRCC2* gene confirmed the efficient knockdown at the protein and mRNA level, respectively (Fig. 3C and 3D). Importantly, knockdown of endogenous XRCC2 led to an ~80% reduction in HR efficiency in mCherry-expressing cells, indicating a clear defect in executing HR (Fig. 3E). Expression of mCherry-XRCC2 restored HR levels to almost 70% of that observed in control cells expressing mCherry (Fig. 3E), showing that our experimental setup allows the functional characterization of *XRCC2* variants. We selected all variants that showed <75% rescue compared to wild-type in at least one of the two *irs1* assays in addition to a subset of variants that showed full rescue in either assay and examined their ability to rescue the HR-deficient phenotype of human XRCC2-depleted cells. The protein-truncating variants p.Thr83fs, p.Leu117fs, p.Arg215* and p.Cys217* restored the HR-defect to 48%, 56%, 42% and 31%, respectively (Fig. 3F). As observed before, the truncating variant p.Arg17* showed a more moderate effect and rescued HR up to 67%. We considered the possibility that this truncating mutant showed a considerable rescue in all functional assays due to re-initiation of translation downstream of the stop codon (Supplementary Material, Fig. S2A). To study this, we expressed mCherry-XRCC2-FLAG and mCherry-XRCC2-p.Arg17*-FLAG constructs in HEK293T cells and examined the expression of mCherry- and FLAG-tagged XRCC2 products by Western blot analysis using antibodies against mCherry or FLAG (Supplementary Material, Fig. S2B). While we detected full-length mCherry-XRCC2-FLAG using either the mCherry or FLAG antibody, we detected the presence of a truncated mCherry-XRCC2 protein in mCherry-XRCC2-p.Arg17*-FLAG- expressing cells when using the mCherry antibody. We were unable to detect expression of a truncated XRCC2-FLAG protein in these cells when using the FLAG antibody (Supplementary Material, Fig. S2B). This suggests there is no re-initiation of translation downstream of the stop codon. Surprisingly, when we examined higher exposure Western blots, we detected low levels of mCherry-XRCC2-FLAG protein in mCherry-XRCC2-p.Arg17*-FLAG expressing cells. This suggests a certain degree of read-through of the stop codon introduced by the p.Arg17* mutation (Supplementary Material, Fig. S2B and S2C). Indeed, the p.Arg17* mutation introduces a TGA stop codon, which has been reported to result in translational read-through in mammalian genes.36 This may explain why we observed a considerable rescue for p.Arg17* in the functional assays. Finally, also most missense variants showed considerable rescue of the HR defect with the common polymorphism p.Arg188His even up to 97%, while some others, such as p.Cys120Tyr were less efficient in restoring HR (~60% rescue; Fig. 3F).

Given that HR is most prominent in S-phase,[37] we examined if the expression of *XRCC2* variants could impacted HR indirectly by affecting cell cycle progression. Flow cytometry analysis, however, did not show any major differences in cell cycle progression between HEK293 DR-GFP cells expressing wildtype or variant *XRCC2* (Supplementary Material, Fig. S3), indicating that the outcome of the assay was not influenced by alterations in cell cycle

distribution. Thus, overall our results in human cells confirm that truncating variants more strongly impair XRCC2's DNA repair function compared to missense variants.

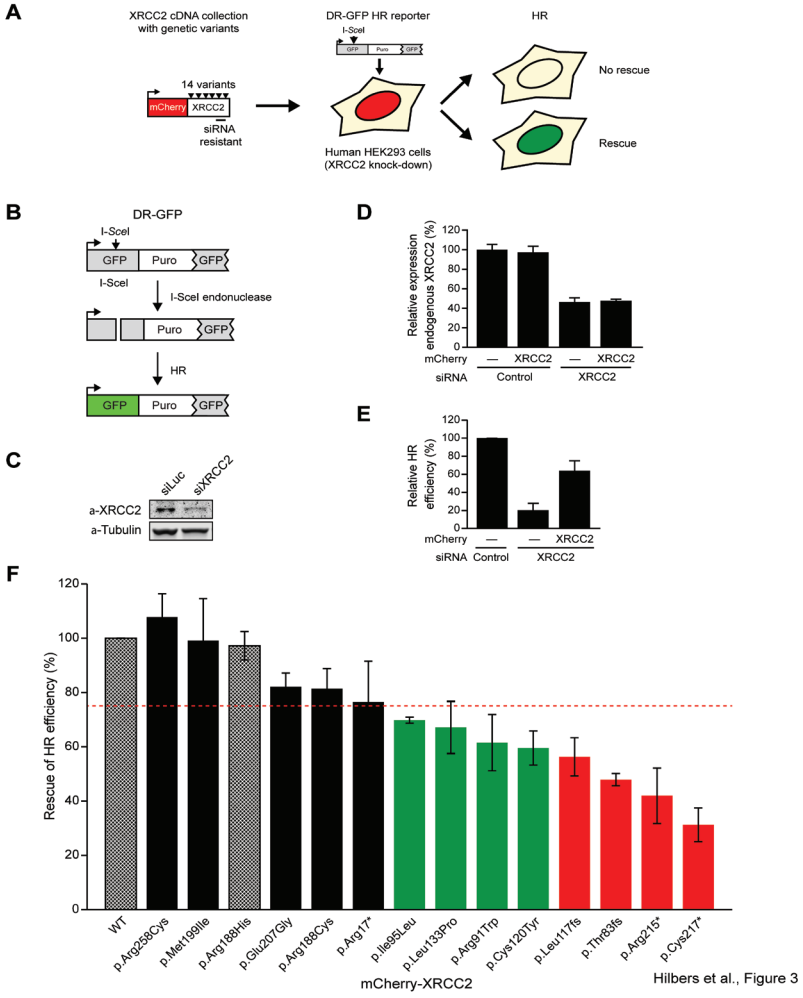

Figure 3. Functional analysis of the effect of *XRCC2* variants on HR in human cells.
(A) Workflow for the functional analysis of the effect of *XRCC2* variants on HR in human cells. A collection of mCherry-tagged XRCC2 cDNAs containing selected genetic variants was introduced in HEK293 cells containing the DR-GFP reporter for HR after knockdown of *XRCC2* by siRNA interference. Following expression of the I-SceI nuclease, the fraction of GFP-expressing cells among mCherry-positive cells was determined to assess whether the *XRCC2* variants would rescue the HR defect of the *XRCC2* knockdown cells. (B) Schematic of the DR-GFP reporter assay. (C) Validation of *XRCC2* knock-down after siRNA transfection (80 nM) of HEK293T cells using a specific XRCC2 antibody. (D) Relative expression of endogenous *XRCC2* in HEK293 cells expressing siRNA-resistant wild-type mCherry-XRCC2 cDNA or mCherry alone, following treatment with control or *XRCC2* siRNAs. Expression was determined by RT-qPCR using primers specific for the endogenous *XRCC2* transcript. The XRCC2 expression level in control cells expressing mCherry alone was set to 100%. (E) Effect of siRNA-mediated XRCC2 knockdown and subsequent re-expression of siRNA-resistant mCherry-XRCC2 cDNA on HR in HEK293 DR-GFP cells. The bars indicate the relative percentage of GFP-positive cells after DSB induction by I-SceI. The HR efficiency of mCherry-expressing cells transfected with control siRNAs were set to 100%. (F) Effect of XRCC2 variants on the rescue of HR in XRCC2-depleted HEK293 DR-GFP cells. Cells expressing mCherry-XRCC2 (WT) served as a reference, which was set to 100%. The checked bars indicate wildtype *XRCC2* and the neutral common polymorphism p.R188H. Green bars indicate missense variants, whereas red bars indicate truncating variants with rescue levels below 75%. Error bars indicate standard error of the mean.

The outcome of functional assays correlates poorly with in silico predictions

In order to assess the robustness of the functional assays, we examined the correlation between the three HR assays (RAD51 foci and GFP-based reporter assays in hamster cells and human cells) in more detail. We found moderately high Pearson correlation coefficients for these assays ranging from 0.57 to 0.72 (Fig. 4A-C). This degree of correlation likely reflects the variation within the assays and the fact that each measures different but related features. We also compared the results of the assays with the predictions of the in silico algorithms. For the prediction tools that give categorical results for amino acid changes (PolyPhen2, SIFT and AlignGVGD), there seems to be no correlation with protein function in either of the assays (Supplementary Material, Fig. S4). Likewise, we found that CADD, which assigns a continuous prediction score instead of an effect category to each variant, shows a rather poor correlation with all three assays ranging from -0.31 to -0.45 (Fig. 4D-F). Based on this we conclude that the outcome of these predictive algorithms poorly correlates with the effect of missense variants on XRCC2's function in DNA repair.

Lack of association with breast cancer upon functional reclassification of variants

Based on the results of our functional assays, we reclassified the *XRCC2* variants and re-assessed their association with breast cancer. We used the results of the common SNP p.R188H as a benchmark for neutrality. The rescue of HR found for this variant ranged between 86% to 97% when compared to wild-type in the different assays. Any variant showing a similar or better rescue of HR in these assays is unlikely to have an association with breast cancer. In contrast, no individual *XRCC2* variant has been convincingly shown to increase breast cancer risk. The only variant with a demonstrable clinical association is the p.R215* variant found in the Saudi Arabian Fanconi Anemia patient. Given the overlap between Fanconi Anemia genes and breast cancer susceptibility genes, this variant currently gives the best estimation of the range of functional effects that could be relevant for breast cancer susceptibility. The variant shows an HR rescue of 41%, 79% and 42% in the RAD51 foci formation assay, SCR assay and DR-GFP assay, respectively.

Using these extremes we defined three functional categories. One category with variants strongly affecting XRCC2 function, with <50% HR rescue in at least two out of three assays. This category contains the protein-truncating variants p.Thr83fs, p.Arg215*and p.Cys217*. A second set of variants with a moderate effect on XRCC2 function defined as 50-75% HR rescue in two out of three assays. This category contains missense variants, p.Arg91Trp, p.Ile95Leu, p.Cys120Tyr and p.Leu133Pro, and the frame-shifting variant p.Leu117fs. The last category includes variants that showed 75-100% HR rescue in at least two assays. This category includes all other missense variants and protein-truncating variant p.Arg17* (for an overview of the results in all three assays Table 1). When we reexamine the association data provided by Park et al.[6] and Hilbers et al.[15] we do not find any evidence for an association with breast cancer (Table 2). For the >50% and 25-50% reduction in complementation categories combined we find an odds ratio of 0.75 (95% CI: 0.18-3.13). It should, however, be noted that the number of individuals with a variant in the <50% rescue category is too low for a meaningful association analysis. An association with breast cancer can therefore not be excluded for these variants. In conclusion, our functional classification data suggest that it is highly unlikely that the missense variants analyzed in this study are associated with breast cancer risk.

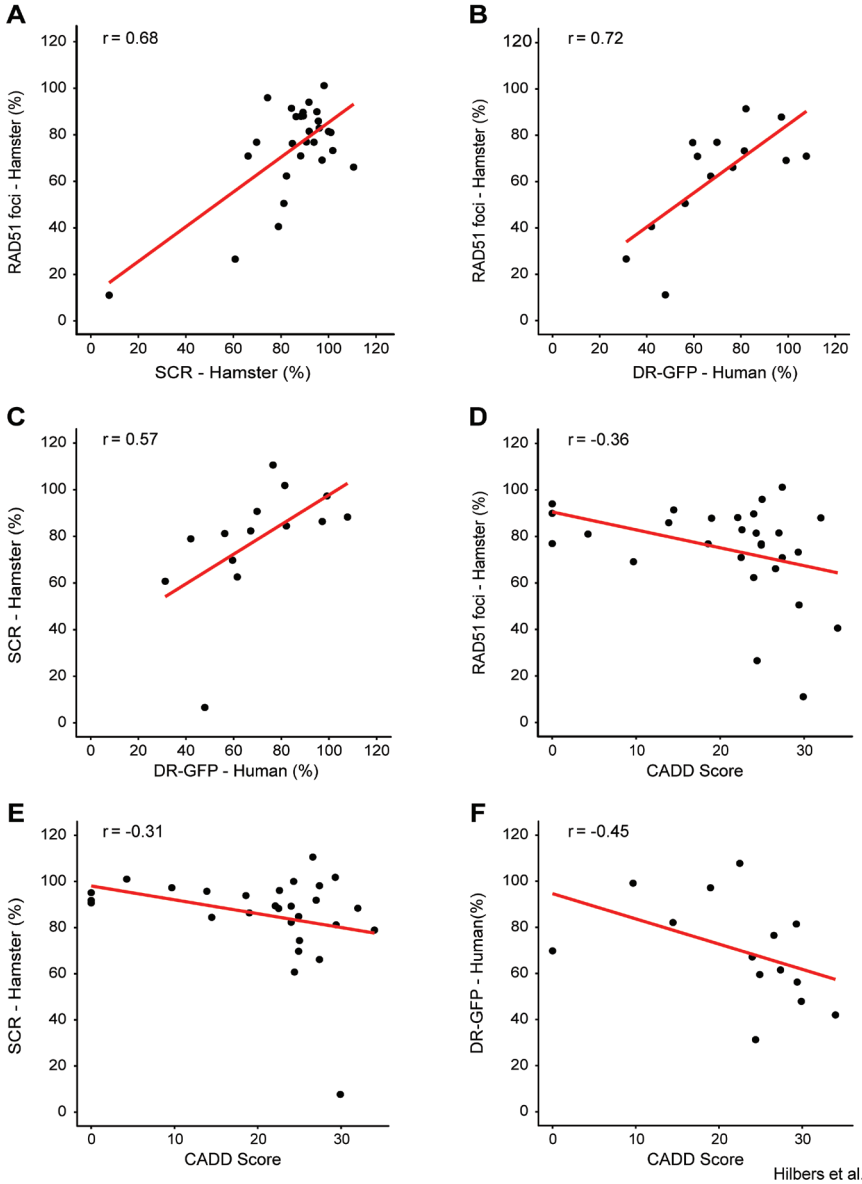

Figure 4. Correlation between the outcome of functional assays and in silico analysis.
(A) Correlation between the outcome of RAD51 foci formation and SCR assays in hamster cells, (B) RAD51 foci formation assay in hamster cells and DR-GFP assays in human cells, (C) SCR assays in hamster cells and DR-GFP assays in human cells, (D) RAD51 foci formation assays in hamster cells and CADD prediction analysis, (E) SCR assays in hamster cells and CADD prediction analysis, and (F) RAD51 foci formation assays in hamster cells and CADD prediction analysis. r indicates the Pearson correlation coefficient.

Table 2. Association analysis of functionally characterized XRCC2 variants

	Degree of variant complementation	Cases (%) (n=2147)	Controls (%) (n=1120)	
Park et al. (5)	**All**	12 (0.56)	1 (0.09)	
	75-100%	7 (0.33)	1 (0.09)	
	50-75%	3 (0.14)	0 (0)	
	<50%	2 (0.09)	0 (0)	
	<75%	5 (0.23)	0 (0)	
	Degree of variant complementation	**Cases (%) (n=3548)**	**Controls (%) (n=1435)**	
Hilbers et al. (15)	**All**	18 (0.51)	10 (0.70)	
	75-100%	18 (0.51)	7 (0.49)	
	50-75%	0 (0)	3 (0.21)	
	<50%	0 (0)	0 (0)	
	<75%	0 (0)	3 (0.21)	
	Degree of variant complementation	**Cases (%) (n=5695)**	**Controls (%) (n=2555)**	**OR (95% CI)**
Total	**All**	30 (0.53)	11 (0.43)	1.22 (0.61-2.45)
	75-100%	25 (0.44)	8 (0.31)	1.40 (0.63-3.12)
	50-75%	3 (0.05)	3 (0.12)	0.45 (0.09-2.22)
	<50%	2 (0.04)	0 (0)	NA
	<75%	5 (0.09)	3 (0.12)	0.75 (0.18-3.13)

The *XRCC2* genetic variants reported by Park et al. (2012) and Hilbers et al. (2012) were classified according to the level of complementation compared to wild-type.

Discussion

Here we report the functional analysis of all *XRCC2* genetic variants detected previously in two large case-control studies,[6,15] which assessed their association with breast cancer risk. We show that most non-synonymous variants have no, or only a small effect on XRCC2 function, with the exception of p.Arg91Trp, p.Ile95Leu, p.Cys120Tyr and p.Leu133Pro, which displayed moderate effects in our functional assays. While three protein-truncating variants had a strong effect on XRCC2 function in our assays, two other such variants, p.Arg17* and p.Leu117fs show unexpectedly good complementation of the XRCC2-deficient phenotypes. Upon reclassification of the *XRCC2* genetic variants according to the results of our functional analyses, the originally reported association with breast cancer was no longer evident.

Most prediction algorithms classify protein-truncating variants as pathogenic and this is also predicted for p.Arg17*. However, a closer inspection of this variant, which showed unexpected functionality in our assays, suggested that the introduced TGA stop codon frequently leads to translational read-through as has been reported for a number of mammalian genes.[36] Indeed, we can detect small amounts of full-length XRCC2 protein produced from this allele, which is a likely explanation for the considerable rescue of p.Arg17*

in the functional assays. However, we should note that it is unclear whether the endogenous *XRCC2* sequence harboring this mutation would produce sufficient amounts of full-length protein to support HR at normal levels. Nonetheless, these results suggest a mechanism through which protein-truncating mutations might still retain function and illustrate that our functional assays are sensitive enough to detect this phenomenon.

Two *XRCC2* variants assayed in our study have previously been studied by others. The common SNP p.Arg188His was shown by Rafii and colleagues to have a negligible effect on XRCC2 function in conferring resistance to Mitomycin C in clonogenic survival assay,[38] consistent with the results of our three functional assays. The frame-shifting variant p.Leu117fs has previously been identified in the mismatch-repair deficient uterine sarcoma cell line SKUT-1.[39] Functional analysis of this variant has shown that, although it has a reduced function compared to wild-type XRCC2, the effect is not as strong as other truncating or frame-shifting mutations in the same region.[40] In our analysis, p.Leu117fs retained approximately 55% function compared to wild-type, whereas more C-terminal variants resulting in protein truncation seem to have stronger inactivating effects. Although our findings are consistent with the previously reported results,[40] we can at this moment not explain why p.Leu117fs has this relatively mild effect.

The two main conserved domains in XRCC2 are the ATP-binding Walker A (amino acids 48-55) and Walker B (amino acids 145-149) motifs. It has been shown that mutations in the most conserved part of the Walker A motif do not affect XRCC2 function.[26,28] Moreover, one study suggested that the Walker A motif is important for the interaction between XRCC2 and RAD51D, whereas another study suggested that the RAD51D-binding region of XRCC2 is located in the N-terminus of XRCC2 (amino acids 1-42).[40] On the other hand, the C-terminal half of the XRCC2 protein (amino acids 144-280) is required for the interaction with RAD51.40 Given that the variants p.Leu117fs, p.Arg215* and p.Cys217*, which result in the expression of C-terminally truncated XRCC2 proteins, which are still partly functional in our assays, it is unlikely that the localization of RAD51 and the repair of DSBs via HR only depends on this interaction. The notion that the Walker A motif does not seem to be important for XRCC2 function might explain why all in silico prediction tools overestimated the number of missense variants affecting XRCC2 function, as these tools weigh protein conservation so heavily. XRCC2 is one of five RAD51-paralogs that have no functional redundancy.[35] Especially in the case of a paralogue, which originates from a segmental duplication and subsequently evolved into a gene with a different function, in silico predictions tools based on sequence conservation might have trouble correctly predicting functional effects. Interestingly, the four missense variants with the strongest effect in our assays are all located in between the Walker A and Walker B motif. This region has no known function at present, but might affect the three-dimensional structure of the protein.

It remains possible that because of limitations of the assays we have used, we missed certain variant effects. For example, the use of cDNA-constructs instead of a construct with an intron-exon structure, does not allow the detection of effects of variants on RNA splicing. In fact, our in silico analysis of possible effects on splicing suggested that 4 of the 28 variants examined in this study may affect splicing. Whether this prediction has relevance for the splicing of *XRCC2* mRNAs in vivo remains to be established. In addition, we are not able to detect the effect of nonsense mediated decay (NMD) on variants resulting in a premature stop-codon. *XRCC2* consists of three exons of which the third exon is by far the largest. All truncating variants except p.R17* are located in this last exon. Given that NMD is thought to only affect mRNA molecules with a premature stop codon upstream of the last exon-exon

junction, it seems highly unlikely that this process will affect any of the truncating variants other than p.R17*.[41]

Crucial to the functional analysis of variants in any gene, is a full understanding of the processes the encoded protein is involved in. In the case of XRCC2 it is widely accepted that its main function is in RAD51-dependent HR. There are however some indications that XRCC2 is also involved in other pathways. A study in Arabidopsis thaliana has implicated a role of XRCC2 in the repair of DSBs via RAD51-independent single-strand annealing, a pathway that by definition is mutagenic.[42] In addition, it has been shown that XRCC2, like many other HR proteins, is also involved in centrosome integrity and correct segregation of chromosomes.[43,44] However, it is not known what its exact role is in this process and whether it depends on interactions other than those important for HR. We can therefore not exclude that the variants that showed full complementation of the XRCC2-deficient phenotype in our assays, affect an XRCC2-regulated process that was not assessed in our study. A last limitation of our study is the use of transient transfections. For example, the use of siRNAs in the DR-GFP assay resulted in a partial knockdown of endogenous XRCC2. A stable knockdown or knockout of XRCC2, would most likely reduce inter-experiment variability and might provide a larger window for functional analysis of *XRCC2* variants.

In general, this study demonstrates the pitfalls in establishing an association between rare genetic variation and disease risk. Now that next-generation sequencing makes it possible to explore variation in large regions of the genome at reasonable costs at epidemiological scale, the number of extremely rare variants with a possible relation to disease risk will increase tremendously. Likewise, diagnostic testing of breast cancer genes other than *BRCA1* and *BRCA2* is likely to result in a steep increase in variants of uncertain significance. When no extensively sampled families are available for co-segregation analysis of individual variants, a gene-centered burden analysis is the only way to establish a disease association, but this crucially depends on their correct classification. Functional analyses as shown here for *XRCC2* variants can prove a crucial step in this process. To this end, future efforts should establish a more sensitive and robust cellular system that allows for the high-throughput testing of large numbers of variants in (putative) breast cancer susceptibility genes.

In conclusion, this study has shown that most rare *XRCC2* missense variants have little effect on XRCC2 protein function and are unlikely to be associated with familial and early onset breast cancer. In the group of *XRCC2* variants that resulted in a >25% reduction in complementation no evidence of an association with breast cancer has been found. Although we cannot exclude that the few *XRCC2* variants with a very strong effect on protein function increase breast cancer risk, these results suggests that the previously reported association with breast cancer is a false-positive finding. Unfortunately, none of the studies published to date had sufficient families in which one of the variants showing reduced complementation was found and for which DNA samples from additional family members were available for a co-segregation analysis. Much larger case-control studies are necessary to provide conclusive results on the association between the variants with >50% reduction in complementation and disease risk. Nevertheless, based on the current data, we conclude that *XRCC2* variants are unlikely to explain a significant fraction of the familial risk to breast cancer.

Funding

This work was supported by the Netherlands Organisation for Scientific Research [NWO-VENI to M.S.L.]; and the Dutch Cancer Society [KWF UL 2009-4388 to P.D., KWF UL 2014-6853 to P.D.,

C.J.v.A. and H.v.A.].

Acknowledgements

The authors would like to thank John Thacker, Ralph Scully and Maria Jasin for generously providing *irs1*, *irs1*-SCR and HEK293 DR-GFP cells, respectively, and Maaike Vreeswijk for help with the in silico splice prediction analysis.

References

1. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet. 2001;358(9291):1389-1399. doi:10.1016/S0140-6736(01)06524-2
2. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Anglian Breast Cancer Study Group. Br J Cancer. 2000;83(10):1301-1308. doi:10.1054/bjoc.2000.1407
3. Peto J, Collins N, Barfoot R, et al. Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer. J Natl Cancer Inst. 1999;91(11):943-949. doi:10.1093/jnci/91.11.943
4. Lalloo F, Evans DG. Familial breast cancer. Clin Genet. 2012;82(2):105-114. doi:10.1111/j.1399-0004.2012.01859.x
5. Easton DF, Pharoah PDP, Antoniou AC, et al. Gene-panel sequencing and the prediction of breast-cancer risk. N Engl J Med. 2015;372(23):2243-2257. doi:10.1056/NEJMsr1501341
6. Park DJ, Lesueur F, Nguyen-Dumont T, et al. Rare mutations in XRCC2 increase the risk of breast cancer. Am J Hum Genet. 2012;90(4):734-739. doi:10.1016/j.ajhg.2012.02.027
7. Shamseldin HE, Elfaki M, Alkuraya FS. Exome sequencing reveals a novel Fanconi group defined by XRCC2 mutation. J Med Genet. 2012;49(3):184-186. doi:10.1136/jmedgenet-2011-100585
8. Antoniou AC, Casadei S, Heikkinen T, et al. Breast-cancer risk in families with mutations in PALB2. N Engl J Med. 2014;371(6):497-506. doi:10.1056/NEJMoa1400382
9. D'Andrea AD. Susceptibility pathways in Fanconi's anemia and breast cancer. N Engl J Med. 2010;362(20):1909-1919. doi:10.1056/NEJMra0809889
10. Howlett NG, Taniguchi T, Olson S, et al. Biallelic inactivation of BRCA2 in Fanconi anemia. Science. 2002;297(5581):606-609. doi:10.1126/science.1073834
11. Kiiski JI, Pelttari LM, Khan S, et al. Exome sequencing identifies FANCM as a susceptibility gene for triple-negative breast cancer. Proc Natl Acad Sci USA. 2014;111(42):15172-15177. doi:10.1073/pnas.1407909111
12. Sawyer SL, Tian L, Kähkönen M, et al. Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype. Cancer Discov. 2015;5(2):135-142. doi:10.1158/2159-8290.CD-14-1156
13. Chun J, Buechelmaier ES, Powell SN. Rad51 paralog complexes BCDX2 and CX3 act at different stages in the. Mol Cell Biol. 2013;33(2):387-395. doi:10.1128/MCB.00465-12
14. Krejci L, Altmannova V, Spirek M, Zhao X. Homologous recombination and its regulation. Nucleic Acids Res. 2012;40(13):5795-5818. doi:10.1093/nar/gks270
15. Hilbers FS, Wijnen JT, Hoogerbrugge N, et al. Rare variants in XRCC2 as breast cancer susceptibility alleles. J Med Genet. 2012;49(10):618-620. doi:10.1136/jmedgenet-2012-101191
16. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit7.20. doi:10.1002/0471142905.hg0720s76
17. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31(13):3812-3814. doi:10.1093/nar/gkg509
18. Tavtigian SV, Deffenbaugh AM, Yin L, 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
19. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310-315. doi:10.1038/ng.2892
20. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004;11(2-3):377-394. doi:10.1089/1066527041410418
21. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. J

Comput Biol. 1997;4(3):311-323. doi:10.1089/cmb.1997.4.311
22. Pertea M, Lin X, Salzberg SL. GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res. 2001;29(5):1185-1190. doi:10.1093/nar/29.5.1185
23. Desmet F-O, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. 2009;37(9):e67. doi:10.1093/nar/gkp215
24. Thery JC, Krieger S, Gaildrat P, et al. Contribution of bioinformatics predictions and functional splicing assays to the interpretation of unclassified variants of the BRCA genes. Eur J Hum Genet. 2011;19(10):1052-1058. doi:10.1038/ejhg.2011.100
25. Jones NJ, Cox R, Thacker J. Isolation and cross-sensitivity of X-ray-sensitive mutants of V79-4 hamster cells. Mutat Res. 1987;183(3):279-286. doi:10.1016/0167-8817(87)90011-3
26. O'Regan P, Wilson C, Townsend S, Thacker J. XRCC2 is a nuclear RAD51-like protein required for damage-dependent RAD51 focus formation without the need for ATP binding. J Biol Chem. 2001;276(25):22148-22153. doi:10.1074/jbc.M102396200
27. Puget N, Knowlton M, Scully R. Molecular analysis of sister chromatid recombination in mammalian cells. DNA Repair (Amst). 2005;4(2):149-161. doi:10.1016/j.dnarep.2004.08.010
28. Nagaraju G, Hartlerode A, Kwok A, Chandramouly G, Scully R. XRCC2 and XRCC3 regulate the balance between short- and long-tract gene conversions between sister chromatids. Mol Cell Biol. 2009;29(15):4283-4294. doi:10.1128/MCB.01406-08
29. Pierce AJ, Johnson RD, Thompson LH, Jasin M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. Genes Dev. 1999;13(20):2633-2638. doi:10.1101/gad.13.20.2633
30. Kotecha N, Krutzik PO, Irish JM. Web-based analysis and publication of flow cytometry experiments. Curr Protoc Cytom. 2010;Chapter 10:Unit10.17. doi:10.1002/0471142956.cy1017s53
31. Smeenk G, Wiegant WW, Marteijn JA, et al. Poly(ADP-ribosyl)ation links the chromatin remodeler SMARCA5/SNF2H to. J Cell Sci. 2013;126(Pt 4):889-903. doi:10.1242/jcs.109413
32. Typas D, Luijsterburg MS, Wiegant WW, et al. The de-ubiquitylating enzymes USP26 and USP37 regulate homologous recombination by counteracting RAP80. Nucleic Acids Res. 2015;43(14):6919-6933. doi:10.1093/nar/gkv613
33. He Y, Zhang Y, Jin C, et al. Impact of XRCC2 Arg188His polymorphism on cancer susceptibility: a meta-analysis. PLoS One. 2014;9(3):e91202. doi:10.1371/journal.pone.0091202
34. Yu K-D, Chen A-X, Qiu L-X, Fan L, Yang C, Shao Z-M. XRCC2 Arg188His polymorphism is not directly associated with breast cancer risk: evidence from 37,369 subjects. Breast Cancer Res Treat. 2010;123(1):219-225. doi:10.1007/s10549-010-0753-y
35. Takata M, Sasaki MS, Tachiiri S, et al. Chromosome instability and defective recombinational repair in knockout mutants of the five Rad51 paralogs. Mol Cell Biol. 2001;21(8):2858-2866. doi:10.1128/MCB.21.8.2858-2866.2001
36. Loughran G, Chou M-Y, Ivanov IP, et al. Evidence of efficient stop codon readthrough in four mammalian genes. Nucleic Acids Res. 2014;42(14):8928-8938. doi:10.1093/nar/gku608
37. Karanam K, Kafri R, Loewer A, Lahav G. Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. Mol Cell. 2012;47(2):320-329. doi:10.1016/j.molcel.2012.05.052
38. Rafii S, O'Regan P, Xinarianos G, et al. A potential role for the XRCC2 R188H polymorphic site in DNA-damage repair and breast cancer. Hum Mol Genet. 2002;11(12):1433-1438. doi:10.1093/hmg/11.12.1433
39. Mohindra A, Hays LE, Phillips EN, Preston BD, Helleday T, Meuth M. Defects in homologous recombination repair in mismatch-repair-deficient tumour cell lines. Hum Mol Genet. 2002;11(18):2189-2200. doi:10.1093/hmg/11.18.2189
40. Tambini CE, Spink KG, Ross CJ, Hill MA, Thacker J. The importance of XRCC2 in RAD51-related DNA damage repair. DNA Repair (Amst). 2010;9(5):517-525. doi:10.1016/j.

dnarep.2010.01.016
41. Schoenberg DR, Maquat LE. Regulation of cytoplasmic mRNA decay. Nat Rev Genet. 2012;13(4):246-259. doi:10.1038/nrg3160
42. Serra H, Da Ines O, Degroote F, Gallego ME, White CI. Roles of XRCC2, RAD51B and RAD51D in RAD51-independent SSA recombination. PLoS Genet. 2013;9(11):e1003971. doi:10.1371/journal.pgen.1003971
43. Cappelli E, Townsend S, Griffin C, Thacker J. Homologous recombination proteins are associated with centrosomes and are required for mitotic stability. Exp Cell Res. 2011;317(8):1203-1213. doi:10.1016/j.yexcr.2011.01.021
44. Griffin CS, Simpson PJ, Wilson CR, Thacker J. Mammalian recombination-repair genes XRCC2 and XRCC3 promote correct chromosome segregation. Nat Cell Biol. 2000;2(10):757-761. doi:10.1038/35036399

Supplementary data

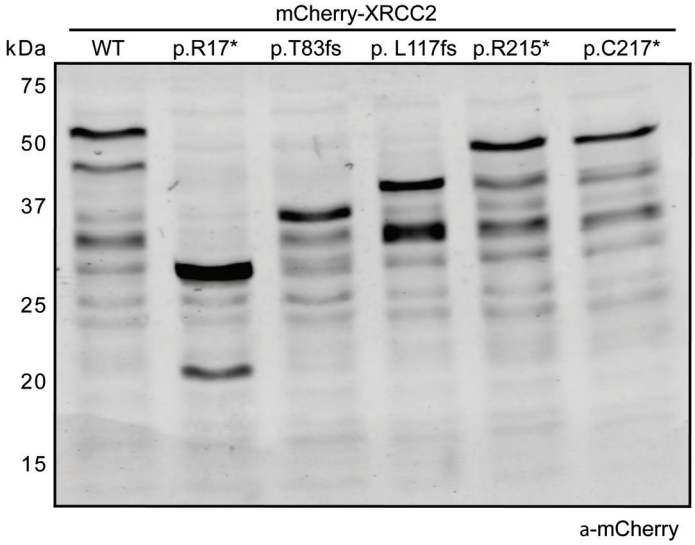

Figure S1. Western blot analysis of selected *XRCC2* variants.
Wild-type (WT) mCherry-XRCC2 and mCherry-XRCC2 carrying different truncating variants were expressed in HEK293 cells. Cell lysates were prepared and subjected to Western blot (WB) analysis using anti-mCherry antibody.

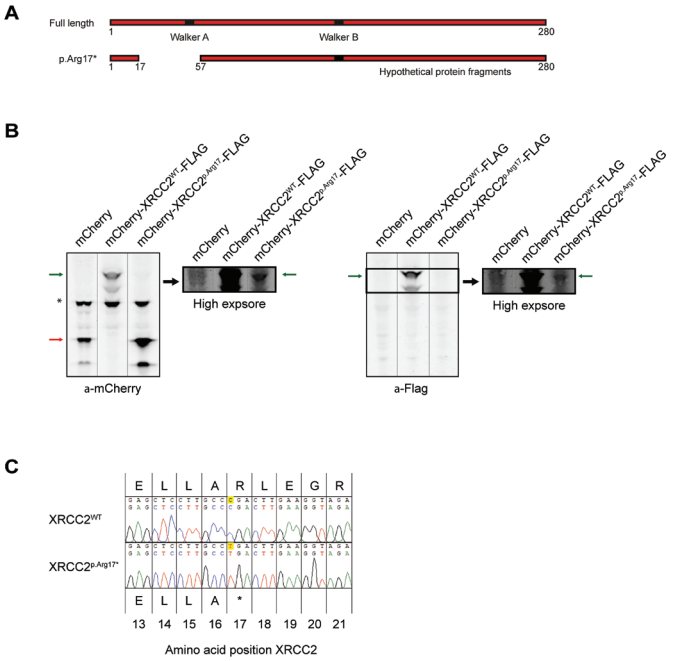

Figure S2.
(A) Schematic overview of *XRCC2* wildtype and XRCC2-p.Arg17* mutant proteins. The hypothetical protein generated by re-initiation of translation downstream of the stop codon introduced by the p.Arg17* mutation is shown. (B) Western blot analysis of HEK293T cells expressing mCherry, mCherry-XRCC2-FLAG or mCherry-XRCC2p.Arg17-FLAG. The star (*) indicates a non-specific band detected with the mCherry antibody. The red arrow indicates the position of mCherry protein. The green arrow indicates the position of the mCherry-XRCC2-FLAG fusion protein. (C) Mutation detection analysis by sequencing of mCherry-XRCC2-FLAG and mCherry-XRCC2p.Arg17-FLAG constructs. Amino acids and corresponding DNA sequences in *XRCC2* are shown.

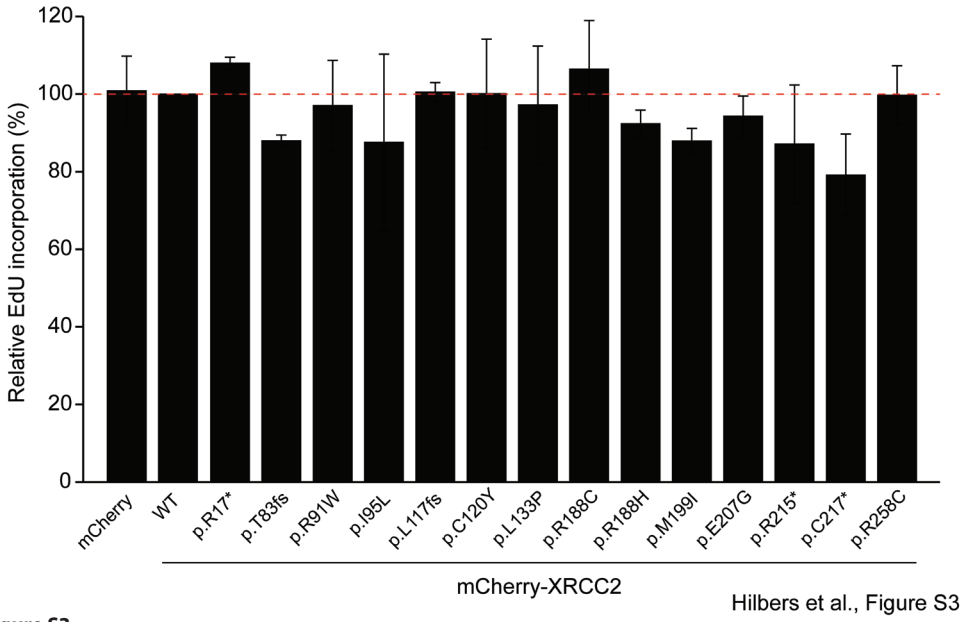

Figure S3.
Cell cycle analysis of HEK293 DR-GFP cells expressing selected *XRCC2* variants. *XRCC2* knockdown cells expressing I-SceI and either mCherry alone, wild-type (WT) mCherry-XRCC2 or different mCherry-XRCC2 variants were labeled with EdU and examined by flow cytometry to determine the fraction of cells in early, mid and late S phase. The fraction of EdU-positive cells for each sample is presented relative to that for the mCherry-XRCC2 sample, which was set to 100%.

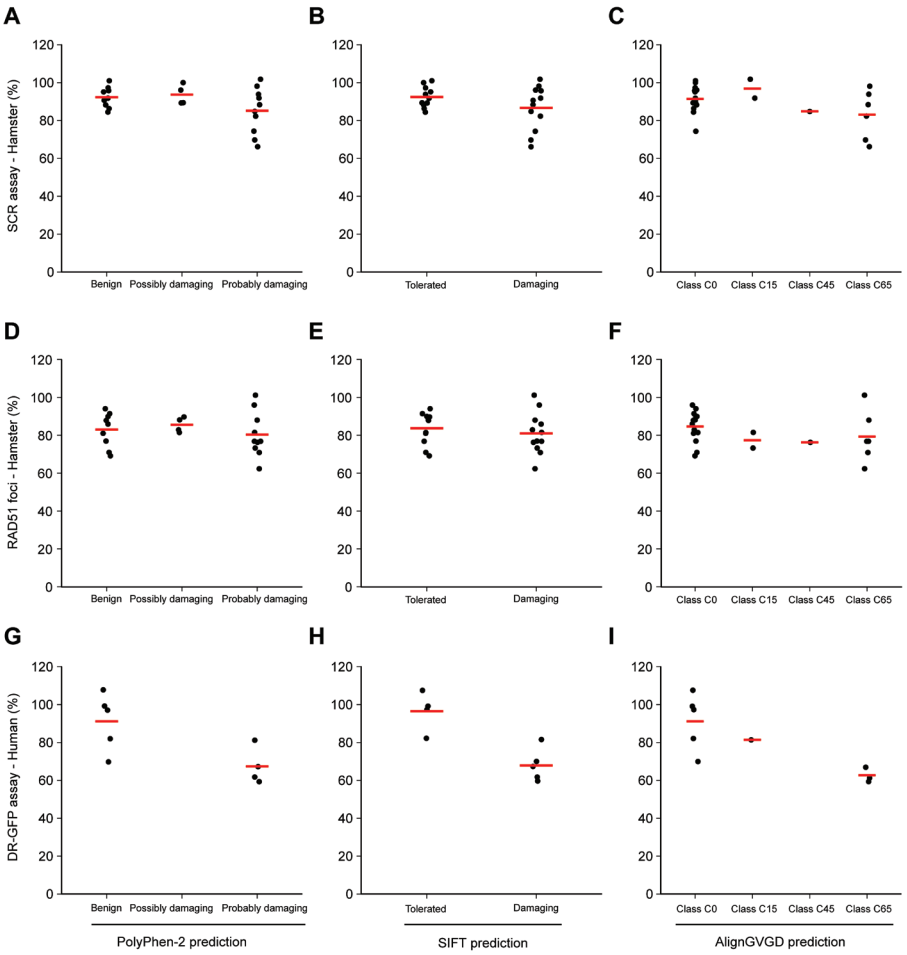

Figure S4.
Correlation between the outcome of functional assays and in silico predictions. The correlation between the functional effects of all *XRCC2* missense variants as found in the SCR assay in hamster cells and the predictions from in silico analysis by (A) PolyPhen-2, (B) SIFT and (C) AlignGVGD are shown. (D-F) As in A-C, except that the functional effects of *XRCC2* variants on RAD51 foci formation in hamster cells were compared. (G-I) As in A-C, except that the functional effects of *XRCC2* variants on HR in the DR-GFP assay in HEK293 cells were compared.

Table S1. Primers used for PCR, site-directed mutagenesis and RT-qPCR

Description		Forward primer	Reverse primer
SDM[a]	c.46 G>T	GGACCGAGCTCCTTTCCCGACTTGAAGG	CCTTCAAGTCGGGAAAGGAGCTCGGTCC
	c.49C>T	CCGAGCTCCTTGCCTGACTTGAAGGTAGAAG	CTTCTACCTTCAAGTCAGGCAAGGAGCTCGG
	c.140A>G	GGTGATATTCTTGAATTTCGTGGCCCAGAAGGAACAGG	CCTGTTCCTTCTGGGCCACGAAATTCAAGAATATCACC
	c.181C>A	CAGAAATGCTTTATCACATAACAGCACGATGTATACTTCCC	GGGAAGTATACATCGTGCTGTTATGTGATAAAGCATTTCTG
	c.223G>C	CCAAATCAGAAGGTGGCCTGCAAGTAGAAGTCTTATTTATT	AATAAATAAGACTTCTACTTGCAGGCCACCTTCTGATTTGG
	c.247dup	CCTGGAAGTAGAAGTCTTATTTATTGATAACAGATTACCACTTTGATATGCTCC	GGAGCATATCAAAGTGGTAATCTGTTATCAATAAATAAGACTTCTACTTCCAGG
	c.271C>T	CCACTTTGATATGCTCTGGCTAGTTACAATTCTTGAGC	GCTCAAGAATTGTAACTAGCCAGAGCATATCAAAGTGG
	c.283A>C	GCTCCGGCTAGTTACACTTCTTGAGCACAGACTATCCC	GGGATAGTCTGTGCTCAAGAAGTGTAACTAGCCGGAGC
	c.283A>G	GCTCCGGCTAGTTACAGTTCTTGAGCACAGACTATCCC	GGGATAGTCTGTGCTCAAGAACTGTAACTAGCCGGAGC
	c.350delT	TCAAATACTGCCTGGGAAGATTTTTTTGGTGTACTGCAGTAGTAGCA	TGCTACTACTGCAGTACACCAAAAAAATCTTCCCAGGCAGTATTTGA
	c.353T>C	GCCTGGGAAGATTTTTTTTGGCGTACTGCAGTAGTAGCACCC	GGGTGCTACTACTGCAGTACGCCAAAAAAAATCTTCCCAGGC
	c.359G>A	GGGAAGATTTTTTTTGGTGTACTACAGTAGTAGCACCCACTTAC	GTAAGTGGGTGCTACTACTGTAGTACACCAAAAAAAATCTTCCC
	c.398T>C	CCCACTTACTTCTTACACTTTACTCACCAGAAAGTATGTTTTGTAGTCACCC	GGGTGACTACAAAACATACTTTCTGGTGAGTAAAGTGTAAGAAGTAAGTGGG
	c.490G>C	CCGCGTCAATGGAGGACAAAGTGTGAACTTACAGG	CCTGTAAGTTCACACTTTGTCCTCCATTGACGCGG
	c.509A>C	GGAGGAGAAAGTGTGAACTTACAGGCGTCTACTCTGAGG	CCTCAGAGTAGACGCCTGTAAGTTCACACTTTCTCCTCC
	c.562C>T	GCTTGTAAATGACTATTGCCTGGTTCTTTTTGCAACGACAC	GTGTCGTTGCAAAAAGAACCAGGCAATAGTCATTTACAAGC
	c.563G>A	GCTTGTAAATGACTATCACCTGGTTCTTTTTGCAACGACAC	GTGTCGTTGCAAAAAGAACCAGGTGATAGTCATTTACAAGC
	c.581C>T	CGCCTGGTTCTTTTTGCAATGACACAAACTATAATGCAG	CTGCATTATAGTTTGTGTCATTGCAAAAAGAACCAGGCG
	c.595A>C	GCAACGACACAAACTATACTGCAGAAAGCCTCGAGCTCA	TGAGCTCGAGGCTTTCTGCAGTATAGTTTGTGTCGTTGC
	c.620A>G	GAAAGCCTCGAGCTCATCAGGAGAACCTTCTCATGCCTCTCG	CGAGAGGCATGAGAAGGTTCTCCTGATGAGCTCGAGGCTTTC
	c.643C>T	AACCTTCTCATGCCTCTCGATGACTGTGTGATGTGGACATAG	CTATGTCCACATCACACAGTCATCGAGAGGCATGAGAAGGTT
	c.651_652delTG	GCCTCTCGACGACTGTGATGTGGACATAG	CTATGTCCACATCACAGTCGTCGAGAGGC
	c.659A>T	GCCTCTCGACGACTGTGTGATGTGGTCATAGACTACAGACC	GGTCTGTAGTCTATGACCACATCACACAGTCGTCGAGAGGC
	c.693G>T	CAGTTGCTGACATGCCTTACAG	CTGTAAGGCATGTCAGCAACTG
	c.714G>C	GGCAGCAACTGGTGAAGCACAGCATGTTTTTCTCC	GGAGAAAAACATGCTGTGCTTCACCAGTTGCTGCC
	c.742C>G	GGATGTTTTTCTCCAAACAAGATGATTCTGAAAGCAGCAACC	GGTTGCTGCTTTCAGAATCATCTTGTTTGGAGAAAAACATCC
	c.772C>T	ATTTTCATTAGTTTCATGTTGTTTAAAAAGTA	TACTTTTTAAACAACATGAAACTAATGAAAAT
	c.808T>G	CTCCAATAATAACAAAATGTTTTTTTAAACTG	CAGTTTAAAAAAACATTTTGTTATTATTGGAG
SDM siRNA resistance		GTTCTTTTCGCTACTACTCAGACAATTATGCAG	CTGCATAATTGTCTGAGTAGTAGCGAAAAGAAC
SDM siRNA res.[b]	c.562C>T	GCTTGTAAATGACTATTGCCTGGTTCTTTTCGCTACTACTC	GAGTAGTAGCGAAAAGAACCAGGCAATAGTCATTTACAAGC
	c.563G>A	GCTTGTAAATGACTATCACCTGGTTCTTTTCGCTACTACTC	GAGTAGTAGCGAAAAGAACCAGGTGATAGTCATTTACAAGC
	c.581C>T	CGCCTGGTTCTTTTCGCTATGACTCAGACAATTATGCAG	CTGCATAATTGTCTGAGTCATAGCGAAAAGAACCAGGCG
	c.595A>C.	GCTACTACTCAGACAATTCTGCAGAAAGCCTCGAGCTCA	TGAGCTCGAGGCTTTCTGCAGAATTGTCTGAGTAGTAGC
mChrerry-XRCC2 [c]		CACCATCGTGGAACAGTACG	ACCTCTACAAATGTGGTATGGCTG
mChrerry-XRCC2 siRNA res.[c]		CACCATCGTGGAACAGTACG	GCCAAAAGACGGCAATATGGTG
mCherry-XRCC2-FLAG		CTCAGATCTGCCGCCGCGATCGCCATG	ATAGAATTCTTACTTGTCATCGTCGTCCTTGTAGTCACAAAATTCAACCCCAC
RT-qPCR	endo XRCC2[d]	CGTCAATGGAGGAGAAAGTG	TGCATTATAGTTTGTGTCGTTGC
	cDNA XRCC2[d]	CGCTACTACTCAGACAATTATGC	ATCCTGTGCTTCACCAGTTG

[a] Site-directed mutagenesis (SDM). [b] Site-directed mutagenesis specific for the siRNA resistant mCherry-XRCC2 construct. [c] Primers for the amplification of the mCherry-*XRCC2* sequence in the mCherry-*XRCC2* construct, used during site-directed mutagenesis and for verification of the presence of the desired variants and the lack of additional mutations with the help of Sanger sequencing. [d] RT-qPCR primers that either detect mRNA from the endogenous XRCC2 (endo) or siRNA-resistant XRCC2 cDNA (cDNA).

Table S2. siRNAs

Target	Sequence
XRCC2	5'-UUAUAGUUUGUGUCGUUGCAA-3'
Control	Universal Negative Control #1 (Sigma, SIC001)

Table S3. In silico predictions on splicing for all *XRCC2* genetic variants

DNA (NM_005431.1)	Protein (NP_005422.1)	Study	Predicted effect on	In silico prediction by					Splicing effect likely [i]
				SpliceSiteFinder [0-100; Th. ≥ 70]	MaxEntScan [0-12; Th. ≥ 0]	NNSplice [0-1; Th. ≥ 0.4]	GeneSplicer [0-15; Th. ≥ 0]	Human Splice Finder [0-100; Th. ≥ 65]	
c.46 G>T [a]	p.Ala16Ser	(5)	Exon 2 –c.40N	AS: 78.99 = 78.99	AS: 7.24 = 7.24	AS: 0.41 => NI (-100%)	AS: 6.42 => 6.01 (-6.4%)	AS: 79.84 = 79.84	-
			Exon 2 – c.59	AS: 71.70 => 77.3 (+7.8%)	AS: 5.33 => 7.11 (+33.4)	AS: NI => 0.47 (+100%)		AS: 79.82 = 79.82	+
c.49C>T	p.Arg17*	(5)	Exon 2 –c.40N	AS: 87.99 = 78.99	AS: 7.24 = 7.24	AS: 0.41 => 0.56 (+36.2%)	AS: 6.42 => 6.04 (-5.9%)	AS: 79.84 = 79.84	-
			Exon 2 – c.59	AS: 71.70 +> 74.98 (+4.6%)	AS: 5.33 => 5.82 (+9.4%)			AS: 79.82 => 80.53 (+0.9%)	-
c.140A>G	p.His47Arg	(15)	Exon 3 – c.122N	AS: 82.42 = 82.42	AS: 7.33 = 7.33		4.1 = 4.1	86.84 = 86.84	-
			Exon 3 – c.149	AS: 77.54 => 77.80 (+0.3%)	AS: 2.5 => 2.31 (-7.7%)			84.56 => 84.68 (+0.1%)	-
			Exon 3 – c.152		AS: NI => 0.17 (+100%)			76.13 => 76.32 (+0.2%)	-
c.181C>A	p.Leu61Ile	(5)	Exon 3 – c.122N	AS: 82.42 = 82.42	AS: 7.33 = 7.33		AS: 4.1 => 4.21 (+2.7%)	AS: 86.84 = 86.84	-
			Exon 3 – c.188	AS: 79.67 => 76.18 (-4.4%)	AS: 5.95 => 3.54 (-40.5%)			AS: 83.03 => 80.45 (-3.1%)	-
c.223G>C	p.Glu75Gln	(15)	Exon 3 – c.221	DS: NI => 75.43 (+100%)					-
			Exon 3 – c.225					DS: NI => 66.58 (+100%)	-
			Exon 3 – c.227					AS: 71.3 => 71.93 (+0.9%)	-
			Exon 3 – c.230					AS: 69.94 => 67.83 (+3.7%)	-
			Exon 3 – c.233					65.46 => 67.83 (+3.6%)	-
c.247dup [b]	p.Thr83fs	-	Exon 3 – c.251	AS: 83.89 => 83.23 (-0.8%)	AS: 4.83 => 5.0 (+3.5%)	AS: 0.54 => 0.51 (-5.3%)		AS: 86.47 => 83.08 (-3.9%)	-
c.271C>T	p.Arg91Trp	(5)	Exon 3- c.272	DS: 76.35 => 76.81 (+0.6%)					-
			Exon 3- c.278		AS: 3.43 => 4.13 (+20.2%)			AS: 76.55 => 76.13 (-0.5%)	-
c.283A>C [c]	p.Ile95Leu	(5)	Exon 3 – c.292	AS: NI => 72.72 (+100%)				AS: 71.47 => 73.5 (+2.8%)	-
c.283A>G	p.Ile95Val	(5,15)	Exon 3 – c.284					AS: NI => 81.75 (+100%)	-
			Exon 3 – c.292					AS: 71.47 => 71.59 (+0.2%)	-

DNA (NM_005431.1)	Protein (NP_005422.1)	Study	Predicted effect on	In silico prediction by					Splicing effect likely [i]
				SpliceSiteFinder [0-100; Th. ≥ 70]	MaxEntScan [0-12; Th. ≥ 0]	NNSplice [0-1; Th. ≥ 0.4]	GeneSplicer [0-15; Th. ≥ 0]	Human Splice Finder [0-100; Th. ≥ 65]	
c.350delT	p.Leu117fs	(15)	Exon 3-c.363	AS: 78.97 = 78.97	AS: 6.52 => 5.74 (-11.9%)			AS: 81.2 = 81.2	-
			Exon 3 – c.401	AS: 75.64 = 75.64	AS: 6.07 = 6.07		AS: 3.25 => 3.44 (+5.6%)	AS: 81.22 = 81.22	-
c.353T>C [d]	p.Val118Ala	(15)	Exon 3-c.351					DS: 68.97 => NI (-100%)	-
			Exon 3-c.363	AS: 78.97 => 75.68 (-4.2%)	AS: 6.52 => 4.72 (-27.5%)			AS: 81.2 => 80.5 (-0.9%)	-
			Exon 3-c.401	AS: 75.64 = 75.64	AS: 6.07 =6.07		AS: 3.25 => 2.85 (-12.5%)		-
c.359G>A	p.Cys120Tyr	(15)	Exon 3 – c.363	AS: 78.97 = 78.97	AS: 6.52 => 5.8 (-11.0%)	AS: NI => 0.49 (+100%)		AS: 81.2 => 81.27 (+0.1%)	-
			Exon3 – c.366					AS: 72.3 => 72.34 (+0.1%)	-
			Exon 3-c.369					AS: 71.81 => 71.69 (-0.2%)	-
			Exon 3-c.401	AS: 75.64 = 75.64	AS: 6.07 = 6.07		AS: 3.25 => 3.42 (+5.0%)	AS: 81.22 = 81.22	-
c.398T>C	p.Leu133Pro	(15)	Exon 3 – c.403	DS: 77.38 = 77.38	DS: 6.93 = 6.93	DS: 0.53 => 0.8 (+50.4%)		DS: 79.66 = 79.66	-
			Exon 3 – c.401	AS: 75.64 => 81.82 (+8.2%)	AS: 6.07 => 6.53 (+7.7%)		AS: 3.25 =>2.79 (-14.4%)	AS: 81.22 => 88.91 (+9.5%)	-
			Exon 3-c.405					AS: 72.35 => 72.77 (+0.6%)	-
c.490G>C	p.Glu164Gln	(15)	Exon 3-c.473	AS: 80.38 = 80.38	AS: 3.79 = 3.79	AS: 0.62 => 0.57 (-7.5%)		AS: 77.04 = 77.04	-
			Exon 3 – c.491					AS: 66.64 => NI (-100%)	-
			Exon 3 – c.495					AS: NI => 65.73 (+100%)	-
			Exon 3 – c.508	AS: 80.85 = 80.85	AS: 1.61 => 3.34 (+107.7%)			AS: 85.96 = 85.96	-
c.509A>C	p.Glu170Ala	(15)	Exon 3 – c.508	AS: 80.85 = 80.85	AS: 1.61 => 1.58 (-1.6%)			AS: 85.96 => 85.89 (-0.1%)	-
			Exon 3 – c.511					AS: 69.22 => NI (-100%)	-
			Exon 3 – c.522	AS: NI => 71.6 (+100%)				AS: 80.25 = 80.25	-
c.562C>T	p.Arg188Cys	(15)	-						-
c.563G>A [e]	p.Arg188His	(5,15)	-						-
c.581C>T	p.Thr194Met	(15)	Exon 3 – c.601		AS: 1.03 => 2.34 (+127.3%)			AS: 80.46 = 80.46	-
c.595A>C	p.Met199Ile	(15)	Exon 3 – c.601	AS: NI => 72.3 (+100%)	AS: 1.03 => 2.45 (+137.8%)			AS: 80.46 => 83.87 (+4.2%)	+
			Exon 3 – c.605					AS: 68.36 => 70.84 (+3.6%)	-

DNA (NM_005431.1)	Protein (NP_005422.1)	Study	Predicted effect on	In silico prediction by					Splicing effect likely [i]
				SpliceSiteFinder [0-100; Th. ≥ 70]	MaxEntScan [0-12; Th. ≥ 0]	NNSplice [0-1; Th. ≥ 0.4]	GeneSplicer [0-15; Th. ≥ 0]	Human Splice Finder [0-100; Th. ≥ 65]	
c.620A>G [f]	p.Glu207Gly	(15)	Exon 3 – c.682		DS: 4.7 = 4.7		DS: 4.33 => 4.35 (+0.7%)	DS: 80.47 = 80.47	-
			Exon 3 – c.620					AS: 80.26 => 83.39 (+3.9%)	-
			Exon 3 – c.623					AS: 73.31 => 72.4 (-1.2%)	-
c.643C>T	p.Arg215*	(6)	Exon 3 – c.682		DS: 4.7 = 4.7		DS: 4.33 => 3.79 (-12.3%)	DS: 80.47 = 80.47	-
c.651_652delTG	p.Cys217*	(5)	Exon 3 – c.647					DS: 71.13 => 68.02 (-4.4%)	-
			Exon 3 – c.649					DS: 66.88 => NI (-100%)	-
			Exon 3 – c.682		DS: 4.7 = 4.7 AS: 3.89 => 2.75 (-29.1%)		DS: 4.33 => 4.23 (-2.2%)	DS: 80.47 = 80.47	-
			Exon 3 – c.665		AS: 3.89 => 2.75 (-29.1%)			AS: 71.89 = 71.89	-
c.659A>T	p. Asp220Val	(15)	Exon 3 – c.657					DS: NI => 73.81 (+100%)	-
			Exon 3 – c.682		DS: 4.7 = 4.7		DS: 4.33 => 3.68 (-14.9%)	DS: 80.47 = 80.47	-
			Exon 3 – c.665		AS: 3.89 => 4.93 (+26.7%)			AS: 71.89 => 74.65 (+3.8%)	-
c.693G>T	p.Trp231Cys	(5)	Exon 3 – c.682		DS: 4.7 = 4.7		DS: 4.33 => 2.94	DS: 80.47 = 80.47	-
			Exon 3 – c.687	DS: 81.88 => 89.67 (+9.5%)					-
			Exon 3 – c.691	DS: NI => 71.72 (+100%)				DS: NI => 81.93 (+100%)	+
			Exon 3 – c.697					AS: 74.15 => 74.11 (-0.0%)	-
c.714G>C	p.Arg238Ser	(15)	Exon 3 – c.682		DS: 4.7 = 4.7		DS: 4.33 => 4.21 (-2.6%)	DS: 80.47 = 80.47	-
			Exon 3 – c.714					AS: 79.2 => 75.04 (-5.3%)	-
			Exon 3 – c.734	AS: 70.48 = 70.48	AS: 3.1 => 3.88 (+25.3%)			AS: 74.9 = 74.9	-
c.742C>G [g]	p.Gln248Glu	(15)	Exon 3 – c.682		DS: 4.7 = 4.7		DS: 4.33 => 4.54 (+4.9%)	DS: 80.47 = 80.47	-
			Exon 3 –c.747					AS: 76.11 => 73.38 (-3.6%)	-
			Exon 3 – c.750					AS: 84.19 => 81.65 (-3%)	-
c.772C>T	p.Arg258Cys	(15)	Exon 3 – c.786	AS: NI => 70.24 (+100%)	AS: 2.36 => 2.79 (+18.4%)			AS: 73.9 = 73.9	+

DNA (NM_005431.1)	Protein (NP_005422.1)	Study	Predicted effect on	In silico prediction by					Splicing effect likely [i]
				SpliceSiteFinder [0-100; Th. ≥ 70]	MaxEntScan [0-12; Th. ≥ 0]	NNSplice [0-1; Th. ≥ 0.4]	GeneSplicer [0-15; Th. ≥ 0]	Human Splice Finder [0-100; Th. ≥ 65]	
c.808T>G [h]	p.Phe270Val	(5,15)	Exon 3 – c.807					DS: NI => 65.06 (+100%)	-
			Exon 3 – c.821	AS: 84.42 => 78.83 (-6.6%)				AS: 78.15 = 78.15	-

All non-synonymous coding *XRCC2* variants from Park et al. (5) and Hilbers et al. (15) were selected for in silico analysis by SpliceSiteFinder, MaxEntScan, NNSplice, GeneSplicer and Human Splice Finder using the integrated Alamut software package. In addition, a variant described by Shamseldin et al. (2012), detected in a Fanconi Anemia case and a truncating variant, p.Thr83fs reported by Douglas Easton (personal communication) were analyzed. Nucleotide numbering uses +1 as the A of the ATG translation initiation codon in the reference sequence, with the initiation codon as codon 1. [a] rs4987090, [b] reported by Douglas Easton (personal communication), [c] rs140214637, [d] rs185815454, [e] rs3218536, [f] rs61762969, [g] rs190900560 [h] rs145085742. [i] a variation of more than 10 % in at least two algorithms was considered as having an effect on splicing (Théry et al., 2011). The scores indicate the values for splice donor (SD) or splice acceptor (SA) sites, respectively. Changes relative to wild-type sequences are indicated in %. Th. = threshold, NI = not identified.