Clinical *BRCA1/2* reversion analysis identifies hotspot mutations and predicted neoantigens associated with therapy resistance

Stephen J. Pettitt[1,2]*, Jessica Frankum[1,2], Marco Punta[3,6], Stefano Lise[3], John Alexander[2], Yi Chen[4], Timothy A. Yap[5], Syed Haider[2], Andrew N. J. Tutt[2] and Christopher J. Lord[1,2]*

[1]The CRUK Gene Function Laboratory, [2]Breast Cancer Now Toby Robins Research Centre, [3]Centre for Evolution and Cancer and [4]Scientific Computing Team, The Institute of Cancer Research, London, SW3 6JB, UK. [5]The University of Texas MD Anderson Cancer Center, 1400 Holcombe Boulevard. Houston TX 77030, USA. [6]Current address: Center for Omics Sciences, IRCCS San Raffaele Hospital, Milan, Italy

*Corresponding authors:
Stephen Pettitt, Email: Stephen.Pettitt@icr.ac.uk, Tel: +44 20 7153 5053
Christopher Lord, Email: Chris.Lord@icr.ac.uk, Tel: +44 20 7153 5190

Keywords: *BRCA1*, *BRCA2*, homologous recombination, reversion mutation, PARP inhibitor, platinum

Conflict of Interest Statement**:** A.N.J.T. reports honoraria from Pfizer, Vertex, Prime Oncology, Artios, honoraria and stock in InBioMotion, honoraria and financial support for research from AstraZeneca, Medivation, Myriad Genetics, Merck Serono. Travel expenses covered by AstraZeneca for any trial related meetings or trial commitments abroad. A.N.J.T. reports benefits from ICR's Inventors Scheme associated with patents for use of PARP inhibitors in BRCA1/2 associated cancers, paid into his ICR honorary account for use on ICR research. C.J.L. makes the following disclosures: received research funding from: AstraZeneca, Merck KGaA, Artios. Received consultancy, SAB membership or honoraria payments from: Syncona, Sun Pharma, GLG, Merck KGaA, Vertex, AstraZeneca, Tango, 3rd Rock, Ono Pharma, Artios. Has stock in: Tango, Ovibio. C.J.L. is also a named inventor on patents describing the use of DNA repair inhibitors and stands to gain from the development as part of the ICR "Rewards to Inventors" scheme. T.A.Y. reports grants, personal fees, and non-financial support from AstraZeneca during the conduct of the study; grants, personal fees, and non-financial support from AstraZeneca outside the submitted work; and Research support (to Institution): Artios, Bayer, Clovis, Constellation, Cyteir, Eli Lilly, EMD Serono, Forbius, F-Star, GlaxoSmithKline, Genentech, ImmuneSensor, Ipsen, Jounce, Karyopharm, Kyowa, Merck, Novartis, Pfizer, Ribon Therapeutics, Regeneron, Repare, Sanofi, Scholar Rock, Seattle Genetics, Tesaro, and Vertex Pharmaceuticals. T.A.Y. reports the following consultancies: Almac, Aduro, Atrin, Axiom, Bayer, Bristol Myers Squibb, Calithera, Clovis, Cybrexa, EMD Serono, F-Star, Guidepoint, Ignyta, I-Mab, Jansen, Merck, Pfizer, Repare, Roche, Rubius, Schrodinger, Seattle Genetics, Varian and Zai Labs.

Abstract

Reversion mutations in *BRCA1* or *BRCA2* are associated with resistance to PARP inhibitors and platinum. To better understand the nature of these mutations, we collated, codified and analysed over 300 reversions. This identified reversion “hotspots” and “deserts” in the N- and C-terminal regions (respectively) of *BRCA2*, suggesting that pathogenic mutations in these regions may be at higher or lower risk of reversion. Missense and splice-site pathogenic mutations in *BRCA1/2* also appeared less likely to revert than truncating mutations. Most reversions were <100 bp deletions. Although many deletions exhibited microhomology, this was not universal, suggesting that multiple DNA repair processes cause reversion. Finally, we found that many reversions were predicted to encode immunogenic neopeptides, suggesting a route to the treatment of reverted disease. As well as providing a freely-available database for the collation of future reversion cases, these observations have implications for how drug resistance might be managed in *BRCA*-mutant cancers.

Statement of Significance

Reversion mutations in *BRCA* genes are a major cause of clinical platinum and PARP inhibitor resistance. This analysis of all reported clinical reversions suggests that the position of *BRCA2* mutations affects the risk of reversion. Many reversions are also predicted to encode tumour neoantigens, providing a potential route to targeting resistance.

Introduction

Defects in genes that control homologous recombination (HR) DNA repair, such as *BRCA1, BRCA2, RAD51C, RAD51D* and *PALB2,* are common in cancer and are enriched in high grade serous ovarian cancers (HGSOC (1)), triple-negative breast cancer (TNBC (2,3)) castrate resistant metastatic prostate cancer (4) and pancreatic cancer (5-7). Following the pre-clinical identification of synthetic lethality between *BRCA1/2*-mutation and PARP inhibitors (PARPi) (8,9), a number of clinical trials demonstrated that PARPi, as well as platinum, are effective in patients with either germ-line or somatic HR gene mutations, leading to the approval of four different PARPi for the treatment of HR-defective breast or ovarian cancers, and the increased use of platinum in a similar clinical context (8,10-12).

Despite the clinical effectiveness of PARPi and platinum, drug resistance is a growing clinical problem, especially in those with advanced disease (8). The causes of drug resistance in HR-defective cancers are not fully understood, but the observation that platinum resistance in HGSOC is predictive of a poor response to PARP inhibitors (13), suggests that clinical platinum resistance can often result in cross-resistance to PARPi. One potential explanation for PARPi/platinum cross-resistance is that tumor cells have restored HR. In *BRCA1, BRCA2, PALB2, RAD51C* or *RAD51D* mutant cancers, this occurs via reversion mutations that restore the native reading frame of each gene (14,15) (Figure 1A). When first seen in HR genes, true reversions (i.e. to wild type sequence) as well as second site reversions were identified (14,15) (Figure 1A). In many cases the second site reversions were intragenic

deletions, all of which were flanked by short regions (1-6 bp) of DNA sequence microhomology or accompanied by an insertion (14,15). This microhomology-associated DNA sequence “scar” suggested that DNA repair processes that utilise regions of microhomology to repair DSBs, such as microhomology end joining (MMEJ) or single strand annealing (SSA, (16,17)), could be the predominant cause of reversion.

Although reversion mutations have been associated with clinical PARPi and/or platinum resistance their description has been limited to individual case reports or studies of relatively small clinical cohorts where low numbers make it difficult to detect any recurring patterns with confidence (Supplementary Table 1). Therefore, in order to better understand clinical reversions, and to stimulate and enable the research community to report and analyse such events, we collated, codified and analysed over 300 HR gene reversion events described in the literature and show how by analysing the sequence context of each of these reversions, some insight can be gained as to their origin and nature.

Results

Collation, review and codification of cases of HR-gene reversion mutation

In order to collate all of the available data on HR-gene reversions associated with PARPi or platinum resistance (Figure 1A), we searched the literature (see Methods) up until March 13 2020, identifying 29 publications which described 308 reversion mutation events from a total of 91 patients (Supplementary Table 2). The majority of patient-derived reversion mutations were in *BRCA1* (n = 90, 29%) or *BRCA2* (n = 211, 68%). We also included relevant studies identifying reversion mutations in tumor cell lines and patient-derived xenografts (PDX). The number of cases of PARPi or platinum resistance that are not explained by reversion mutations is difficult to determine, as there will be many unreported cases where a reversion is not detected, not investigated or cannot be ruled out. Across all the studies that we collated, we identified a total of 96 cases (either cell line clones or patients with recurrent or platinum/PARPi resistant cancer) where the presence of reversion mutations was assessed but not detected (Supplementary Table 3).

Differences in nomenclature and annotation exist between publications. This often arises from the use of historical mutation nomenclature for *BRCA1/2*, and/or the varied use of either transcript-based or coding sequence (CDS)-based numbering across different studies. In addition, the nucleotide-based annotation of microhomologies at reversion deletions lacks a standard definition. Given this, we reannotated and codified all published reversion mutations, both in terms of nucleotide change and microhomology use (see

Methods and Figure 1B). In addition, we reviewed the clinical information provided for all reported cases. We collated all of this information as a singular, freely accessible, database (http://reversions.icr.ac.uk).

In terms of disease subtype, the largest number of revertant cases were from patients with ovarian cancer (56 patients with 125 reversion events; Figure 1C, D). Rather than reflecting a greater propensity for ovarian cancers to exhibit reversion mutations, the number of ovarian cancers in the collated dataset might reflect the longer period over which PARPi and platinum treatments have been in routine use in this disease. Most of the patients in the study had pathogenic mutations in *BRCA1* (39 patients with 29 mutations), or *BRCA2* (51 patients with 44 mutations) with one each for *PALB2*, *RAD51C* and *RAD51D* (Figure 1C). For the majority (84%) of patients, the pathogenic HR gene mutation was a confirmed germline mutation. Two patients (Lin 2018 SubjectID_63 and Carneiro 2018 Patient 1 in the database) had two different pathogenic alleles with reversions in each.

Reversion mutations are frequently unique events

Amongst the 91 patients we collated data from, most (68/91, 75%) had unique pathogenic mutations (Figure 1E, annotated as “single-patient mutations” and Supplementary Figure 1). There were eight pathogenic mutations represented by multiple patients in the dataset, including common founder mutations such as *BRCA2*:c.6174delT (c.5946delT in our codified annotation, five patients in the dataset) and *BRCA1*:c.185delAG (c.68_69delAG, six patients in the dataset; Figure 1E, Supplementary

Figure 2A). Even where patients had the same founder pathogenic mutation, the DNA sequences of the reversion mutations that emerged in these patients were all unique, with the exception of true reversions to wild-type and two cases of reversion of the *BRCA1*:c.5266dupC founder mutation (Supplementary Figure 2B), suggesting that there is not a strong propensity for any particular reversion mutation to arise from a particular pathogenic mutation (Figure 1E, Supplementary Figure 1). True wild-type reversions were recurrently observed for the *BRCA1:*c.68_69delAG (n = 3) and *BRCA2*: c.5946delT (n = 2) pathogenic mutations (Figure 1F, Supplementary Figure 2C).

For each of these common founder mutations, we noted that the reversions that emerged in these patients were generally localised to the 3' flanking sequence of the original pathogenic mutation (transcriptionally downstream, Figure 1F, Supplementary Figure 2B, C). Several other sites in both *BRCA1* and *BRCA2* exhibited a predominant directionality in the deletion reversions that were associated with them (e.g. *BRCA2*:c.7355delA, Figure 2A, B). However, other pathogenic mutations in *BRCA1* or *BRCA2* had reversion deletions that occurred on either side of the pathogenic mutation, suggesting that this was not a universal property, but specific to certain pathogenic mutations (Figure 2A, Supplementary Figure 2D-G).

One possible explanation for the directionality of some reversion mutations is that there is critical amino acid sequence encoded by the DNA upstream of the pathogenic mutation that cannot be disrupted if a productive reversion

allele is to be formed. However, we did not find any evidence for particular evolutionary conservation of the amino acid residues immediately upstream of the pathogenic mutation, as assessed by Conservation Score (see Methods, Figure 2B).

Reversion mutations in *BRCA2* exhibit position dependence

Although the reversion events that emerged in patients with the same founder pathogenic mutations tended to be unique, we assessed whether the propensity of a pathogenic mutant allele to acquire reversion mutations might depend on its position in either *BRCA1* or *BRCA2*. To do this, we compared the CDS positions of pathogenic *BRCA*-gene mutations known to revert (i.e. those in our reversion dataset) to the CDS positions of pathogenic *BRCA*-gene mutations in a set of clinical sequencing studies ("Incidence" dataset, see Methods, Supplementary Tables 4 and 5) covering ovarian, breast, pancreatic and prostate cancers – the predominant tumor types in our reversion dataset. In the case of *BRCA1* mutations, the pathogenic mutations in the reversion dataset were distributed throughout the *BRCA1* coding sequence, suggesting that reversion mutation is a possible resistance mechanism for pathogenic mutations at most positions (Figure 2C) and their distribution was not significantly different from the distribution of *BRCA1* mutations in the Incidence dataset (Figure 2D, p = 0.23, two-sided Kolmogorov-Smirnov test).

In contrast to *BRCA1*, the position distribution of *BRCA2* pathogenic mutations that reverted differed from the distribution in the Incidence data (Figure 2C, p = 0.023, Kolmogorov-Smirnov test). Despite pathogenic

truncating mutations in the C-terminal region of *BRCA2* being relatively common in large-scale tumor sequencing studies (22% of the pathogenic mutations in the Incidence dataset occurred in exon 16 onwards (CDS position 7617) Figure 2D), reversions of pathogenic mutations in this region were rare (Figure 2C; four reversions from four patients, 7.8%, $P < 0.015$, permutation test). All but one of the reversions in this "desert" region were true reversions to wild-type (n = 2), or missense mutations (n = 1) rather than deletions (only one deletion observed, Supplementary Figure 3). This might suggest that pathogenic mutations in the C-terminal coding sequence of *BRCA2* are less able to be productively reverted by second site mutations, particularly deletions, possibly because the surrounding sequence is important for HR function. This hypothesis is consistent with the known importance of the C terminus for HR function (18) and the high degree of amino acid sequence conservation in this region (Figure 2B). This region of *BRCA2* encodes the oligonucleotide/oligosaccharide binding (OB) folds, the nuclear localisation signal (NLS) and TR2 domains (Figure 2E). Although loss of the TR2 domain only causes a moderate defect in homologous recombination deficiency(19-21), studies in *BRCA2* mutant tumour cell lines with PARP inhibitor resistance indicated that reversion alleles that cause PARPi resistance all encode the TR2 domain even where they delete multiple C-terminal exons, suggesting that it is required for PARP inhibitor resistance ((14,15), Supplementary Figure 4A, B).

In contrast to the reversion "desert" at the C-terminus of *BRCA2*, we noted a large number of reversion mutations in the N-terminal c.750-775 region (61

reversions in total from four patients in four separate studies, Figure 2A, Supplementary Figure 5). These reversions were identified by ctDNA sequencing, which might be more effective in identifying more reversion events per patient than, for example, the bulk sequencing of tumor cells from a solid tumor biopsy (22). However, these mutations originated from four different patients, and this region of *BRCA2* did not show a high frequency of pathogenic mutations in the Incidence dataset (Figure 2D). This suggested that *BRCA2* mutations in this region might show a greater propensity to acquire reversions and/or better tolerate the local disruption of the coding sequence in the reverted *BRCA2* allele, although more data will be required to confirm this. Consistent with this hypothesis, the c.750-775 region is not a highly-conserved region of BRCA2 compared to the C-terminus of the protein (Figure 2B).

Reversion of pathogenic missense mutations is rare

Multiple types of known pathogenic *BRCA1* and *BRCA2* mutation exist, including frameshift or nonsense mutations, as well as well-characterised missense and splice site mutations (23-26). We therefore investigated whether the propensity of a *BRCA*-gene mutation to acquire reversion mutations might depend on the nature of the pathogenic mutation. Of the 74 *BRCA1/2* pathogenic mutations in our reversion dataset, 49 were present in the BRCA Exchange database of reported mutations (23). All of these 49 mutations were classified as pathogenic by the ENIGMA (27) or ClinVar (26) criteria. All remaining mutations (n = 25) without an entry in the BRCA

Exchange database were frameshift or nonsense mutations and therefore predicted to be pathogenic.

Interestingly, we noted very few missense pathogenic mutations in the set of reported reversions. For example, in the Incidence tumour sequencing datasets used previously, we found that (40/849, 4.7%) of these pathogenic *BRCA1/2* mutations were missense variants; conversely in the reversion dataset, only a single patient with a pathogenic missense mutation (*BRCA1:*p.C61S missense mutation, known to be pathogenic) was present (1/91, 1.1%, Figure 2F). We also noted a patient with a *BRCA1* p.M1I pathogenic mutation, which would result in loss of the translation start site. In each of these cases, the reversion seen was a true reversion to wild-type. Moreover, there were no splice-site pathogenic mutations among the reversion cases, despite such mutations constituting 7.3% of Incidence mutations. Splice site mutations affect nucleotides critical for correct splicing; similarly, pathogenic missense mutations, by definition, affect amino acid residues that are critical for function. Thus, these classes of pathogenic may be under similar constraints when it comes to reversion, and in particular are unlikely to be reverted productively by a deletion. The single missense mutation in the reversion dataset was not a statistically significant underrepresentation compared to the Incidence data (P = 0.08, Fisher's exact test); however the absence of reversions in splice site mutations, or splice and missense mutations considered as a combined category, was significant (P = 0.001 and P = 0.0002 respectively, Fisher's exact test, Figure 2F).

A similar observation has been previously made in an analysis of the ARIEL2 clinical trial assessing the efficacy of the PARPi, rucaparib, in relapsed, platinum-sensitive high-grade ovarian carcinomas; out of a cohort of 112 patients, four had *BRCA*-gene missense mutations and ten possessed splice-site mutations. No reversions were found in any of these 14 patients, five of which were platinum resistant or refractory at the start of the study (13).

Microhomology use in reversions is frequent but not universal

When *BRCA*2 reversion mutations were originally identified in cultured tumor cell lines, each of the deletion-mediated second site reversion events was characterised by the presence of DNA sequence microhomology at the ends of deleted regions (14,15,28). This suggested that DNA repair processes that exploit regions of microhomology to repair DSBs could be responsible for the reversion events. From a mechanistic perspective, the loss of homologous recombination is known to cause increased use of MMEJ (29), suggesting that the microhomology-characterised reversions could even be a downstream effect of the loss of HR (14). Inhibition of DNA polymerase theta, which is involved in MMEJ, has been proposed as a strategy to target HR-defective cancers via their increased reliance on MMEJ (30-32). In subsequent reports of HR-gene reversion in patients, microhomology was also a frequent feature of reversions mediated by deletion, an observation that extended beyond *BRCA1* or *BRCA2* reversion, to reversion events in *PALB2*, *RAD51C* and *RAD51D* (14,15,22,33-38). Therefore, to better understand the aetiology of reversion mutations, we assessed the use of microhomology for the reversion events in our dataset. Such events can be recognised via their

ambiguous alignments to the reference sequence, as the bases immediately adjacent to the deletion can be aligned equally well at either side of the deletion (Figure 3A, alignment 1 and 2). Surprisingly, when we systematically assessed all of the reported reversion events, the use of microhomology mediated deletions was clearly not universal. Only 56% (159 of 283 with sequence information) of the reversion cases across the whole dataset were deletions that had evidence of microhomology. In cases of *BRCA1* reversion, only 47% of all reversions (including those not mediated by deletions) were deletions with evidence of microhomology use; for *BRCA2* reversions, 60% showed microhomology use (Figure 3B).

Overall, 71% of the *BRCA1* reversions were mediated by deletions compared to 88% for *BRCA2* (categories “deletion” and “microhomology deletion” in Figure 3B). Therefore, *BRCA1* mutant cells may use a wider range of pathways of DNA repair that lead to substitution or true wild-type reversions compared to *BRCA2*, where most events are deletion-mediated (Figure 3B). When considering only reversions mediated by deletion, the fraction for which microhomology was present was similar between *BRCA1* (67%) and *BRCA2* (68%), but still approximately one third of deletions in each case did not exhibit microhomology (Figure 3C). Taken at face value, this suggested that DNA repair or mutagenic processes that do not utilise regions of DNA microhomology could also play a major role in the formation of reversion deletion mutations in patients. There was no clear position effect on the type of reversions (Supplementary Figure 6A, B), and deletions could revert by insertion and *vice versa* (Supplementary Figure 6C).

Characteristics of reversion mutations indicate strong selective pressure for close to full-length proteins

BRCA2 reversion mutations identified in cell line models were often large intragenic deletions (> 50 kb in some cases) that removed large segments of the coding sequence despite restoring the open reading frame of the gene and leading to expression of the C-terminal NLS and OB/TR2 domains (14). This might suggest that much of the *BRCA2* coding sequence is dispensable for tolerance of PARPi or platinum, at least in cultured cells. In aggregate, deletions have been observed from CDS position 4203 to 9682, but reverted proteins retain the N-terminal PALB2 binding region, some of the BRC repeats and the C-terminal TR2 domain (Supplementary Figure 4B). For *BRCA1*, cell line-based studies suggest that much of the protein coded for by exon 11 (1142 amino acids, 60% of the coding sequence) is dispensable for therapy resistance (39) – this is supported by the observation of potential reversion mutations in the splice donor of exon 11 in two cases (38,40) that may cause skipping of exon 11 and the pathogenic mutation. However, and in contrast to the observations in pre-clinical models (14), the intragenic deletions seen in clinical reversion cases ranged from 1 to 2541 base pairs (in cDNA coordinates), with most deletions being less than 50 bp and contained within a single exon (Figure 3D, Supplementary Figure 4). Therefore, while cells in culture appeared able to tolerate, for example, the loss of thousands of bases and multiple exons of *BRCA2* coding sequence, this does not appear to be recapitulated clinically. This may reflect a greater requirement or fitness advantage for tumor cells with near-full length BRCA1 or BRCA2 proteins. It

should be noted here that some NGS technologies or variant calling pipelines may not be optimised to detect large intragenic deletions or fusion events.

Interestingly, deletion size was generally larger in reversion mutations that displayed evidence of microhomology use, an observation that appeared to be limited to reversion mutations occurring in *BRCA2*-mutant tumors (*BRCA1*, $P = 0.97$; *BRCA2*, $P = 0.0105$; Wilcoxon rank sum test, Figure 3D) perhaps reflecting a greater extent of end resection and microhomology search in *BRCA2* mutant tumors than in *BRCA1* mutant tumors. One reason for the increased deletion size in *BRCA2* reversion mutations with microhomology could be that longer regions of microhomology are required for DNA end joining in this context. Longer regions of microhomology would be expected to occur less frequently, resulting in increased DNA resection length during microhomology searching. Consistent with this hypothesis, *BRCA2* reversion mutations did indeed exhibit longer regions of microhomology on average, peaking at 2-3 nt, when compared with *BRCA1* reversion events (which predominantly utilised 1 bp of microhomology on each side of the reversion deletion, Figure 3E). A general consensus of opinion is that whilst canonical NHEJ utilises either no DNA sequence microhomology or very short regions (1-3 bp) to repair DNA, MMEJ and SSA exploit somewhat longer regions (2-20 bp and >15 bp, respectively (16,17)). Taken at face value, this might suggest that differences in DNA repair pathway usage could explain the differences in microhomology length associated with *BRCA1 vs. BRCA2* reversion deletions.

Proximity of reversion mutations to original truncating mutation suggests that many revertant proteins will constitute neoantigens

Compensatory frameshift reversions that do not restore the same codon as the original mutation (i.e. second site reversions) will introduce out-of-frame stretches of novel amino acid sequence in the revertant protein that are not encoded by the wild-type allele and may not be stably expressed from the pathogenic allele. Overall, 50% of reversions restoring the reading frame occurred at a distance of at least 7 bp from the pathogenic mutation, ranging up to 105 bp (Supplementary Figure 7A, B). This is consistent with the range of distances to out-of-frame stop codons, beyond which a reversion would not restore the reading frame (Supplementary Figure 7C). Thus, most revertant proteins will contain some out-of-frame sequence of 2-30 amino acids, or at least a novel breakpoint amino acid junction. These amino acid sequences may not have previously been visible to the host immune system and could constitute neoantigens; this in turn could provide an opportunity to therapeutically target tumor cells presenting these candidate neoantigens, using approaches such as CAR-T cell therapies, immune checkpoint inhibitors or anticancer vaccines.

To assess this possibility, we first estimated, using the NetMHCpan-4.0 algorithm (41), how frequently in the general population neopeptides derived from the out-of-frame sequence following pathogenic mutations were predicted to be presented by HLA class I complexes. We found that for many pathogenic mutations, including common founder mutations such as *BRCA2*:c.5946delT, *BRCA1*:c.68_69delAG and *BRCA1*:c.5266dupC, the

associated neoantigens were likely to be presented in a sizable fraction of the population (Figure 4A, Supplementary Table 6). Out-of-frame neopeptides can be shared to some extent by revertant sequences arising from the same pathogenic mutation and different downstream reversions. For example, reversions observed downstream of the *BRCA2*:c.5946delT pathogenic mutation retain 3-15 amino acids of the original out-of-frame pathogenic sequence before the reading frame is restored (Figure 4B). Neopeptides associated with the first 7 amino acids of the pathogenic out-of-frame sequence and shared by 3 out of 10 revertant alleles were predicted to be presented by the MHC in at least 84% of individuals (based on a set of 1,261 individuals whose HLA alleles are known, see *Methods*) making them potential tumor antigens (Figure 4C). This increased to 96% of individuals when considering a longer out-of-frame sequence (RENLSRYQMLHYKTQ) also shared by the same 3 revertant cases (Supplementary Figure 8A).

In general, we observed that revertant sequences were associated with sets of neopeptides that, as a whole, were predicted likely to be presented by a considerable fraction of the general population (median 52%, IQR 23-76; Figure 4D, Supplementary Table 7) and this was also true when considering only neopeptides that were not potentially produced by the pathogenic allele (median 44%, Supplementary Figure 8B). This raises the possibility that tumors with some revertant alleles may be targetable with immunotherapies that either relieve immune suppression or those that exploit the introduction of T cell clones that recognise specific neoepitopes. For some pathogenic mutations it may be possible to vaccinate against the peptides predicted to be

presented in revertant alleles, or exploit these as antigens for other immunotherapies, as a route to delay or prevent the emergence of therapy-resistant disease.

Discussion

Here, we show that by collating, codifying and analysing over 300 HR-gene reversion mutations, a number of principles can be established. These include the unique nature of most reversions, positional “hotspots” and “deserts” in the N- and C-terminal coding regions of *BRCA2*, the paucity of missense and splice-site pathogenic mutations leading to reversions, and differences in microhomology use in *BRCA1* compared to *BRCA2*-related reversions. Finally, we found that many reverted alleles were predicted to encode highly immunogenic neo-peptides, suggesting a route to treatment of reverted disease. We believe that by generating, analysing and expanding the reversion dataset, additional principles that govern how therapy resistance emerges in HR-defective cancers could be established.

One observation we noted was that the clinical reversion mutations seem to have a more restricted spectrum (< 100 bp deletions, close to the pathogenic mutation; Figure 2A, Figure 3D, Supplementary Figure 7) compared to those previously seen in cell line and PDX studies, where large deletions predominate (14,15,42). Although some ascertainment bias in the detection of clinical reversions cannot be eliminated, it seems that the types of reversions seen in patients are more likely to preserve the majority of the coding sequence than those seen in preclinical models. Furthermore, in contrast to the ubiquitous microhomology at deletions in cell line studies, we found that microhomology usage in clinical reversions was not universal (67% of the deletion-mediated reversion mutations exhibiting microhomology, Figure 3C). This suggests that multiple DNA repair processes might drive reversion,

implying that the design of therapeutic interventions that limit reversions might be more complex than originally thought. Tumor sequencing studies have assessed microhomology usage in somatic deletion mutations at a genome-wide level, finding, for example, that ≈40% of deletions (IQR, 30-50) showed microhomology in *BRCA1/2* mutant breast cancers, compared to ≈20% in *BRCA* wild-type (43). Thus, the frequency of microhomology-associated *BRCA*-gene reversions is at the upper end of what might be expected at the genome-wide level in *BRCA*-gene mutant cancers, but still lower than that seen for reversions isolated from cell line models.

The observation of a possible hotspot for secondary mutations around position c.750-775 in *BRCA2* has potential implications for patients with these mutations. This may indicate that patients with such mutations would be at higher risk of acquiring resistance via reversion mutations, and should be monitored more closely. Conversely, patients with missense and splice site mutations, or mutations in the *BRCA2* C-terminal desert (exon 16 onwards) may be at lower risk of developing resistance via reversion.

This study has several likely limitations and biases. There are several sources of bias in the data in terms of which tumour types have been studied, which treatments patients have received and which methods were used to detect mutations. For example, the large number of reversions in the dataset that are derived from prostate cancers is somewhat out of proportion to the number of prostate cancer patients that receive *BRCA*-targeted PARP or platinum therapy, but reflects the number of prostate cancer studies where ctDNA

sequencing has been used to detect reversions. Secondly, the identification of reversions may be impacted by the method used to detect them; whilst ctDNA sequencing is extremely sensitive and can often identify dozens of different reversion events in a single patient (thus reflecting clonal heterogeneity), singular biopsies from solid tumours often do not capture this heterogeneity and thus tend to lead to the identification of single reversions as opposed to many. Thirdly, the method of reversion mutation detection might influence the size and type of reversion detected; large, multiple exon, deletions may be more efficiently detected by RT-PCR, as in cell line studies (14,15), compared to sequence capture approaches or Sanger sequencing around the site of pathogenic mutations. In addition, a major drawback of ctDNA sequencing is that true wild type reversions are difficult to detect with confidence, due partly to the low prevalence of reversions relative to wild type or non-reverted alleles in blood DNA, but also to the low likelihood that a linked SNP is available to link the wild type reversion to the chromosome that originally bore the pathogenic mutation, either directly by being on the same sequencing read, or by inference using SNP allele frequencies (34,36,44). Thus, it is possible that the prevalence of wild type reversions is underestimated.

The mechanism by which true wild type reversions emerge is still unclear. Two possibilities are: (a) the sequence at these sites favours the specific wild type reversion event; or (b) the functional constraints on the sequence at the point of mutation are such that only a wild type reversion can restore function (36). A third possibility is that the wild type sequence is directly copied from elsewhere in the genome by a process akin to gene conversion. However,

BRCA mutant tumours generally have loss-of-heterozygosity at the pathogenic mutation, meaning that the other allele is not available as a template for gene conversion even if it were to be used, and gene conversion would likely require some BRCA1/2-dependent RAD51 function, so this seems unlikely.

As more is understood about the prevalence and nature of reversion mutations, the question of how to treat cancers that acquire drug resistance via reversion can be addressed. There are several possibilities suggested by this analysis. First, as described above, inhibiting microhomology-mediated end joining, for example by inhibiting the MMEJ DNA polymerase POLQ (30-32), may be a way of preventing the emergence of some reversions, although this might not be a completely effective approach, given the frequency of non-microhomology mediated events we observed. Targeting reverted proteins that differ from the wild type BRCA-protein might also serve some therapeutic value. For example, reverted BRCA-proteins may, because of their altered amino acid sequence, have an increased dependence on chaperones such as heat shock proteins to fold correctly, as suggested elsewhere (45). Where inserted or out-of-frame amino acid sequences are formed by reversion, these may be immunogenic. We show here that there is a high probability of presentation by the MHC across the general population for many of the revertant sequences, including at common founders such as *BRCA2*:c.5946delT (Figure 4). Thus, immunotherapies (including cancer vaccines) may also be an option for direct targeting of the revertant protein. There are other possible approaches that are not related to the revertant

protein *per se*, such as using WEE1 or ATR inhibitors, that have been empirically shown in pre-clinical models to target *BRCA*-gene mutant tumor cells even after the acquisition of reversion mutations (46), an effect likely mediated by the general replication stress that is likely to still exist in the tumor, despite reversion.

The analysis presented here demonstrates the value of codified set of secondary mutation sequences from clinical observations. We have provided this dataset online at http://reversions.icr.ac.uk along with the analysis presented in this manuscript. This will be updated as more reversion events are reported in the literature to assess whether the conclusions and hypotheses here still apply as the numbers of reported cases increase. As PARPi and platinum are now in routine clinical use for several indications, it is possible that some reversions will no longer be considered novel enough to be reported risking that these are lost from the literature. We provide a facility to directly report further cases for inclusion in the database at the web portal above, and would be happy to receive submissions from further clinical cases of resistance.

Acknowledgements

This work was funded by programme funding to CJL from Cancer Research UK (as part of CRUK Programme Funding C30061/A24439 to SP and CJL), Breast Cancer Now (as part of BCN Programme Funding to the Breast Cancer Now Toby Robins Research Centre). We also acknowledge NHS funding to the NIHR Royal Marsden Hospital Biomedical Research Centre.

MP and SL are funded by the Wellcome Trust (105104/Z/14/Z). MP and SJP are supported by the Schottlander Research Charitable Trust.

Methods

Collation, annotation and standardisation of reversion mutations

Studies for this analysis were collated by searching the PubMed database for BRCA1, BRCA2, RAD51C, RAD51D or PALB2 and “Secondary Mutation” or “Reversion”. These studies, or others referenced in these papers, describing mutations in cell lines, patients or PDX models were included (13-15,22,33-38,40,47-64). Some studies only reported mutations in cell lines (including reversions generated by CRISPR mutagenesis) and PDX (28,39,42,46,65). These are included in the database but not the analysis described in this paper. Where we identified patients whose reversion mutations were reported in multiple studies, these were only included once per reversion event. Reversions were detected by targeted sequencing of cfDNA. In one case a reversion was detected at the first cycle of the investigational regimen (olaparib combined with an AKT inhibitor, capivasertib), in the other four patients the reversion was found at the end of treatment.

To aid with the overall analysis, a single transcript was used to annotate all the mutations for a gene. For *BRCA1* and *BRCA2* we used the same reference transcripts as the ARUP and BRCA Exchange databases; for other genes we chose the longest Consensus Coding Sequence (CCDS) annotated transcript. The transcripts used for codified annotations are: *BRCA1*, NM_007294.3; *BRCA2*, NM_000059.3; *RAD51C*, NM_058216.2; *RAD51D*, NM_002878.3 and *PALB2*, NM_024675.3. Where sequence information was available in the original publication this was used to annotate the mutation, otherwise the reported annotations were checked for correspondence with the

reference transcript chosen for each gene. The original annotation in the publication is provided for cross-referencing purposes, along with patient or case identifiers where used in the published paper. If no case/patient identifiers were used in the original publication, these were constructed for the purposes of our analysis based on the study and sequentially-numbered reversion events. In the database we list both forms of annotation for the original mutation, the reversion mutations and the chromosomal location (where available). Where a chromosomal location was not annotated in the original report, we have back-calculated this from the CDS annotation using the Ensembl Variant Effect Predictor (VEP, (66)).

Once the original and reversion mutations are mapped for each case, we calculated the distance between the mutations as well as noting evidence of microhomology use. The distance between the original mutation and the reversion was measured as the shortest distance, specifically the bases between the last base of one mutation and the first base of the other. Where the reversions are deletions that span the original mutation, the distance is recorded as zero. We also annotated mutations with evidence of microhomology use (Figure 3A), requiring at least one base pair homology. Microhomology is not reported for complex mutations such as insertion-deletions.

Genomic coordinates (hg38) were retrieved using the HGVS CDS annotation on the transcripts above via the Ensembl VEP (67). In annotations of the original pathogenic mutation we aligned deletions in repetitive regions to the

3’ end of the deletion, and annotated small insertion as duplications where appropriate, in order to ensure compatibility with annotations in the BRCA exchange database. Reversion mutation alleles were annotated relative to the reference sequence, including the original pathogenic mutation where this was retained. Deletions that encompassed or were immediately adjacent to the pathogenic mutation (or an alternative valid annotation of the pathogenic mutation) were annotated as a single deletion relative to the reference sequence.

The database records reversion mutations on a “per-event” basis, an event being a single observation of a reversion mutation in a patient with a pathogenic mutation in an HR gene. Where individual patients possessed multiple, distinct, reversions (as seen in 37 (40%) of patients described in the database), each reversion was recorded as a different event. In addition, we also recorded clinical information, including, where available, information pertaining to cancer type, stage and treatment history (Figure 1B).

Mutation data from tumor sequencing studies

The reference set of *BRCA1* and *BRCA2* pathogenic mutations was assembled from several sources. Some studies were identified from published literature describing identification of *BRCA* mutations in relatively large cohorts of confirmed cases of breast, ovarian, pancreatic or prostate cancer (10,35,68-71). These mutations were curated in the same way as the reversion mutations and annotations standardised where necessary. Both germline and somatic mutations were included. All patients studied by Lin *et*

al. (13) were also included in this dataset (including the patients in which reversions were identified). *BRCA1/2* mutations were also downloaded from a series of studies available in cBioPortal (Supplementary Table 4) and filtered to retain only mutations that were classified as pathogenic or likely pathogenic by either the ENIGMA or ClinVar projects. The full set of mutations is given in Supplementary Table 5.

For comparisons with pathogenic mutations in the reversions dataset, pathogenic mutations consisting of deletion or rearrangement of entire or multiple exons were removed (there were no such mutations present in the reversion data). To assess underrepresentation of mutations in the *BRCA2* C terminus, the Incidence data were randomly sampled (n = 51; the number of patients with at least one reversion mutation in *BRCA2*) and the number of mutations falling in the desert region (CDS position > 7617) calculated. This was repeated 1000 times to calculate a *P* value for observing >= 4 mutations in this region. Fishers exact tests, Wilcoxon tests and Kolmogorov-Smirnov tests were performed in R.

Conservation analysis

Multiple sequence alignments of *BRCA1* and *BRCA2* orthologues across 11 mammalian species were downloaded from EGGNOG (72) and visualised using JalView. Sequences with large gaps relative to the human protein were removed and a consensus score generated (73).

HLA-presentation score predictions

Given a gene and a mutational event (primary or reversion), we use an in-house python script (https://github.com/GeneFunctionTeam/neopeptides/) to generate all peptides of length 8-11 amino acids associated with the mutation(s). For primary events, we generate the set *A* of all non-WT peptides associated with the primary mutation (Figure 4A); for reversions, we generate the set *B* of all non-WT peptides associated with the reversion (Figure 4D) and the set *C* of peptides in *B* that are not in *A* (i.e., unique to the revertant sequence, Supplementary Figure 8B). We then calculate the Best Rank (BR) HLA class I presentation score of the mutation with respect to each HLA allotype in a list of 195 HLA–A/-B/-C allotypes total found among 1,261 individuals from the 1000 Genomes study (74). We define the BR by predicting the eluted ligand likelihood percentile rank for each peptide associated to the mutation using the program NetMHCpan-4.0 (41) and taking the minimum elution rank among all peptides (75), excluding those with a wild-type NetMHC predicted Icore (76). We define an individual's best rank (IBR) for a mutation *m* as the minimum BR of the mutation across all HLA class I allotypes of the individual. The percentage of individuals likely to present at least one peptide associated with *m* is then calculated as the percentage of individuals for which IBR < 0.5 when considering a set of 1,261 individuals from the 1000 Genomes project (74).

Data availability

All data used in this study, along with updated analysis including any cases reported in future, are available to download from reversions.icr.ac.uk.

References

1. Cancer Genome Atlas Research Network. Integrated genomic analyses of ovarian carcinoma. Nature **2011**;474(7353):609-15 doi 10.1038/nature10166.
2. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature **2012**;490(7418):61-70 doi 10.1038/nature11412.
3. Staaf J, Glodzik D, Bosch A, Vallon-Christersson J, Reuterswärd C, Häkkinen J*, et al.* Whole-genome sequencing of triple-negative breast cancers in a population-based clinical study. Nat Med **2019**;25(10):1526-33 doi 10.1038/s41591-019-0582-4.
4. Grasso CS, Grasso CS, Wu Y-M, Wu Y-M, Robinson DR, Robinson DR*, et al.* The mutational landscape of lethal castration-resistant prostate cancer. Nature **2012**;487(7406):239-43 doi:10.1038/nature11125.
5. Bailey P, Chang DK, Nones K, Johns AL, Patch A-M, Gingras M-C*, et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. Nature **2016**;531(7592):47-52 doi:10.1038/nature16965.
6. Holter S, Borgida A, Dodd A, Grant R, Semotiuk K, Hedley D*, et al.* Germline BRCA Mutations in a Large Clinic-Based Cohort of Patients With Pancreatic Adenocarcinoma. Journal of Clinical Oncology **2015**;33(28):3124-9 doi:10.1200/JCO.2014.59.7401.
7. Waddell N, Pajic M, Patch A-M, Chang DK, Kassahn KS, Bailey P*, et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. Nature **2015**;518(7540):495-501 doi:10.1038/nature14169.
8. Lord CJ, Ashworth A. PARP inhibitors: Synthetic lethality in the clinic. Science **2017**;355(6330):1152-8 doi 10.1126/science.aam7344.
9. Lord CJ, Ashworth A. BRCAness revisited. Nat Rev Cancer **2016**;16(2):110-20 doi 10.1038/nrc.2015.21.
10. Alsop K, Fereday S, Meldrum C, deFazio A, Emmanuel C, George J*, et al.* BRCA mutation frequency and patterns of treatment response in BRCA mutation-positive women with ovarian cancer: a report from the Australian Ovarian Cancer Study Group. J Clin Oncol **2012**;30(21):2654-63 doi 10.1200/JCO.2011.39.8545.
11. Tutt A, Tovey H, Cheang MCU, Kernaghan S, Kilburn L, Gazinska P*, et al.* Carboplatin in BRCA1/2-mutated and triple-negative breast cancer BRCAness subgroups: the TNT Trial. Nat Med **2018**;24(5):628-37 doi 10.1038/s41591-018-0009-7.
12. Tutt A. Inhibited, trapped or adducted: the optimal selective synthetic lethal mix for BRCAness. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO **2018**;29(1):18-21 doi:10.1093/annonc/mdx775.
13. Lin KK, Harrell MI, Oza AM, Oaknin A, Ray-Coquard I, Tinker AV*, et al.* Reversion Mutations in Circulating Tumor DNA Predict Primary and Acquired Resistance to the PARP Inhibitor Rucaparib in High-Grade Ovarian Carcinoma. Cancer Discov **2019**;9(2):210-9 doi 10.1158/2159-8290.CD-18-0715.
14. Edwards SL, Brough R, Lord CJ, Natrajan R, Vatcheva R, Levine DA*, et al.* Resistance to therapy caused by intragenic deletion in BRCA2. Nature **2008**;451(7182):1111-5 doi 10.1038/nature06548.
15. Sakai W, Swisher EM, Karlan BY, Agarwal MK, Higgins J, Friedman C*, et al.* Secondary mutations as a mechanism of cisplatin resistance in BRCA2-mutated cancers. Nature **2008**;451(7182):1116-20 doi 10.1038/nature06633.
16. Bhargava R, Onyango DO, Stark JM. Regulation of Single-Strand Annealing and its Role in Genome Maintenance. Trends in Genetics **2016**;32(9):566-75 doi:10.1016/j.tig.2016.06.007.
17. Sinha S, Villarreal D, Shim EY, Lee SE. Risky business: Microhomology-mediated end joining. Mutat Res **2016**;788:17-24 doi:10.1016/j.mrfmmm.2015.12.005.
18. Esashi F, Galkin VE, Yu X, Egelman EH, West SC. Stabilization of RAD51 nucleoprotein filaments by the C-terminal region of BRCA2. Nat Struct Mol Biol **2007**;14(6):468-74 doi:10.1038/nsmb1245.
19. Donoho G, Brenneman MA, Cui TX, Donoviel D, Vogel H, Goodwin EH*, et al.* Deletion of Brca2 exon 27 causes hypersensitivity to DNA crosslinks, chromosomal instability, and reduced life span in mice. Genes Chromosomes Cancer **2003**;36(4):317-31 doi:10.1002/gcc.10148.

20. Kass EM, Lim PX, Helgadottir HR, Moynahan ME, Jasin M. Robust homology-directed repair within mouse mammary tissue is not specifically affected by Brca2 mutation. Nat Commun **2016**;7(1):13241-10 doi:10.1038/ncomms13241.
21. McAllister KA, Bennett LM, Houle CD, Ward T, Malphurs J, Collins NK*, et al.* Cancer susceptibility of mice with a homozygous deletion in the COOH-terminal domain of the Brca2 gene. Cancer Res **2002**;62(4):990-4 doi papers3://publication/uuid/D062F8F6-D91E-4EF8-A321-55C5BEFAC7AD.
22. Quigley D, Alumkal JJ, Wyatt AW, Kothari V, Foye A, Lloyd P*, et al.* Analysis of Circulating Cell-Free DNA Identifies Multiclonal Heterogeneity of Reversion Mutations Associated with Resistance to PARP Inhibitors. Cancer Discov **2017**;7(9):999-1005 doi 10.1158/2159-8290.CD-17-0146.
23. Cline MS, Liao RG, Parsons MT, Paten B, Alquaddoomi F, Antoniou A*, et al.* BRCA Challenge: BRCA Exchange as a global resource for variants in BRCA1 and BRCA2. PLoS Genet **2018**;14(12):e1007752-17 doi:10.1371/journal.pgen.1007752.
24. Futreal PA, Liu Q, Shattuck-Eidens D, Cochran C, Harshman K, Tavtigian S*, et al.* BRCA1 mutations in primary breast and ovarian carcinomas. Science **1994**;266(5182):120-2 doi:10.1126/science.7939630.
25. Lancaster JM, Wooster R, Mangion J, Phelan CM, Cochran C, Gumbs C*, et al.* BRCA2 mutations in primary breast and ovarian cancers. Nat Genet **1996**;13(2):238-40 doi:10.1038/ng0696-238.
26. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S*, et al.* ClinVar: improving access to variant interpretations and supporting evidence. Nucleic acids research **2017**;46(D1):D1062-D7 doi:10.1093/nar/gkx1153.
27. Spurdle AB, Healey S, Devereau A, Hogervorst FBL, Monteiro ANA, Nathanson KL*, et al.* ENIGMA-Evidence-based network for the interpretation of germline mutant alleles: An international initiative to evaluate risk and clinical significance associated with sequence variation in BRCA1 and BRCA2 genes. Hum Mutat **2011**;33(1):2-7 doi:10.1002/humu.21628.
28. Sakai W, Swisher EM, Jacquemont C, Chandramohan KV, Couch FJ, Langdon SP*, et al.* Functional restoration of BRCA2 protein by secondary BRCA2 mutations in BRCA2-mutated ovarian carcinoma. Cancer Res **2009**;69(16):6381-6 doi 10.1158/0008-5472.CAN-09-1178.
29. Yun MH, Hiom K. CtIP-BRCA1 modulates the choice of DNA double-strand-break repair pathway throughout the cell cycle. Nature **2009**;459(7245):460-3 doi:10.1038/nature07955.
30. Ceccaldi R, Liu JC, Amunugama R, Hajdu I, Primack B, Petalcorin MIR*, et al.* Homologous-recombination-deficient tumours are dependent on Polθ-mediated repair. Nature **2015**;518(7538):258-62 doi:10.1038/nature14184.
31. Higgins GS, Boulton SJ. Beyond PARP-POLθ as an anticancer target. Science **2018**;359(6381):1217-8 doi:10.1126/science.aar5149.
32. Mateos-Gomez PA, Gong F, Nair N, Miller KM, Lazzerini-Denchi E, Sfeir A. Mammalian polymerase theta promotes alternative NHEJ and suppresses recombination. Nature **2015**;518(7538):254-7 doi 10.1038/nature14157.
33. Barber LJ, Sandhu S, Chen L, Campbell J, Kozarewa I, Fenwick K*, et al.* Secondary mutations in BRCA2 associated with clinical resistance to a PARP inhibitor. J Pathol **2013**;229(3):422-9 doi 10.1002/path.4140.
34. Norquist B, Wurz KA, Pennil CC, Garcia R, Gross J, Sakai W*, et al.* Secondary somatic mutations restoring BRCA1/2 predict chemotherapy resistance in hereditary ovarian carcinomas. J Clin Oncol **2011**;29(22):3008-15 doi 10.1200/JCO.2010.34.2980.
35. Patch A-M, Christie EL, Etemadmoghadam D, Garsed DW, George J, Fereday S*, et al.* Whole-genome characterization of chemoresistant ovarian cancer. Nature **2015**;521(7553):489-94 doi 10.1038/nature14410.
36. Swisher EM, Sakai W, Karlan BY, Wurz K, Urban N, Taniguchi T. Secondary BRCA1 mutations in BRCA1-mutated ovarian carcinomas with platinum resistance. Cancer Res **2008**;68(8):2581-6 doi 10.1158/0008-5472.CAN-08-0088.
37. Goodall J, Mateo J, Yuan W, Mossop H, Porta N, Miranda S*, et al.* Circulating Cell-Free DNA to Guide Prostate Cancer Treatment with PARP Inhibition. Cancer Discov **2017**;7(9):1006-17 doi 10.1158/2159-8290.CD-17-0261.

38. Kondrashova O, Nguyen M, Shield-Artin K, Tinker AV, Teng NNH, Harrell MI*, et al.* Secondary Somatic Mutations Restoring and Associated with Acquired Resistance to the PARP Inhibitor Rucaparib in High-Grade Ovarian Carcinoma. Cancer Discov **2017**;7(9):984-98 doi 10.1158/2159-8290.CD-17-0419.
39. Wang Y, Bernhardy AJ, Cruz C, Krais JJ, Nacson J, Nicolas E*, et al.* The BRCA1-Δ11q Alternative Splice Isoform Bypasses Germline Mutations and Promotes Therapeutic Resistance to PARP Inhibition and Cisplatin. Cancer Res **2016**;76(9):2778-90 doi 10.1158/0008-5472.CAN-16-0186.
40. Waks AG, Cohen O, Kochupurakkal B, Kim D, Dunn CE, Buendia JB*, et al.* Reversion and non-reversion mechanisms of resistance to PARP inhibitor or platinum chemotherapy in BRCA1/2-mutant metastatic breast cancer. Annals of Oncology **2020**:1-22 doi:10.1016/j.annonc.2020.02.008.
41. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. J Immunol **2017**;199(9):3360-8 doi:10.4049/jimmunol.1700893.
42. Ter Brugge P, Kristel P, van der Burg E, Boon U, de Maaker M, Lips E*, et al.* Mechanisms of Therapy Resistance in Patient-Derived Xenograft Models of BRCA1-Deficient Breast Cancer. J Natl Cancer Inst **2016**;108(11) doi 10.1093/jnci/djw148.
43. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X*, et al.* HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nat Med **2017**;23(4):517-25 doi:10.1038/nm.4292.
44. Ganesan S. Tumor Suppressor Tolerance: Reversion Mutations in BRCA1 and BRCA2 and Resistance to PARP Inhibitors and Platinum. JCO Precision Oncology **2018**(2):10.1200/PO.18.00001 doi 10.1200/PO.18.00001.
45. Johnson N, Johnson SF, Yao W, Li Y-C, Choi Y-E, Bernhardy AJ*, et al.* Stabilization of mutant BRCA1 protein confers PARP inhibitor and platinum resistance. Proc Natl Acad Sci U S A **2013**;110(42):17041-6 doi:10.1073/pnas.1305170110.
46. Dréan A, Williamson CT, Brough R, Brandsma I, Menon M, Konde A*, et al.* Modeling Therapy Resistance in BRCA1/2-Mutant Cancers. Mol Cancer Ther **2017**;16(9):2022-34 doi 10.1158/1535-7163.MCT-17-0098.
47. Afghahi A, Timms KM, Vinayak S, Jensen KC, Kurian AW, Carlson RW*, et al.* Tumor BRCA1 Reversion Mutation Arising during Neoadjuvant Platinum-Based Chemotherapy in Triple-Negative Breast Cancer Is Associated with Therapy Resistance. Clin Cancer Res **2017**;23(13):3365-70 doi 10.1158/1078-0432.CCR-16-2174.
48. Banda K, Swisher EM, Wu D, Pritchard CC, Gadi VK. Somatic Reversion of Germline BRCA2 Mutation Confers Resistance to Poly(ADP-ribose) Polymerase Inhibitor Therapy. JCO Precision Oncology **2018**(2):10.1200/PO.17.00044 doi 10.1200/PO.17.00044.
49. Carneiro BA, Collier KA, Nagy RJ, Pamarthy S, Sagar V, Fairclough S*, et al.* Acquired Resistance to Poly (ADP-ribose) Polymerase Inhibitor Olaparib in BRCA2 - Associated Prostate Cancer Resulting From Biallelic BRCA2 Reversion Mutations Restores Both Germline and Somatic Loss-of-Function Mutations. JCO Precision Oncology **2018**(2):1-8 doi 10.1200/PO.17.00176.
50. Cheng HH, Salipante SJ, Nelson PS, Montgomery B, Pritchard CC. Polyclonal BRCA2 Reversion Mutations Detected in Circulating Tumor DNA After Platinum Chemotherapy in a Patient With Metastatic Prostate Cancer. JCO Precision Oncology **2018**(2):10.1200/po.17.00169 doi 10.1200/po.17.00169.
51. Christie EL, Fereday S, Doig K, Pattnaik S, Dawson S-J, Bowtell DDL. Reversion of BRCA1/2 Germline Mutations Detected in Circulating Tumor DNA From Patients With High-Grade Serous Ovarian Cancer. J Clin Oncol **2017**;35(12):1274-80 doi 10.1200/JCO.2016.70.4627.
52. Cruz C, Castroviejo-Bermejo M, Gutiérrez-Enríquez S, Llop-Guevara A, Ibrahim YH, Gris-Oliver A*, et al.* RAD51 foci as a functional biomarker of homologous recombination repair and PARP inhibitor resistance in germline BRCA-mutated breast cancer. Ann Oncol **2018**;29(5):1203-10 doi 10.1093/annonc/mdy099.
53. Gornstein EL, Sandefur S, Chung JH, Gay LM, Holmes O, Erlich RL*, et al.* BRCA2 Reversion Mutation Associated With Acquired Resistance to Olaparib in Estrogen

Receptor-positive Breast Cancer Detected by Genomic Profiling of Tissue and Liquid Biopsy. Clin Breast Cancer **2018**;18(2):184-8 doi 10.1016/j.clbc.2017.12.010.

54. Khalique S, Pettitt SJ, Kelly G, Tunariu N, Natrajan R, Banerjee S*, et al.* Longitudinal analysis of a secondary BRCA2 mutation using digital droplet PCR. J Pathol Clin Res **2020**;6(1):3-11 doi 10.1002/cjp2.146.
55. Mayor P, Gay LM, Lele S, Elvin JA. reversion mutation acquired after treatment identified by liquid biopsy. Gynecol Oncol Rep **2017**;21:57-60 doi 10.1016/j.gore.2017.06.010.
56. Meijer TG, Verkaik NS, van Deurzen CHM, Dubbink H-J, den Toom TD, Sleddens HFBM*, et al.* Direct Ex Vivo Observation of Homologous Recombination Defect Reversal After DNA-Damaging Chemotherapy in Patients With Metastatic Breast Cancer. JCO Precision Oncology **2019**(3):10.1200/PO.18.00268 doi:10.1200/PO.18.00268.
57. Patel JN, Braicu I, Timms KM, Solimeno C, Tshiaba P, Reid J*, et al.* Characterisation of homologous recombination deficiency in paired primary and recurrent high-grade serous ovarian cancer. Br J Cancer **2018**;119(9):1060-6 doi 10.1038/s41416-018-0268-6.
58. Pishvaian MJ, Biankin AV, Bailey P, Chang DK, Laheru D, Wolfgang CL*, et al.* BRCA2 secondary mutation-mediated resistance to platinum and PARP inhibitor-based therapy in pancreatic cancer. Br J Cancer **2017**;116(8):1021-6 doi 10.1038/bjc.2017.40.
59. Shroff RT, Hendifar A, McWilliams RR, Geva R, Epelbaum R, Rolfe L*, et al.* Rucaparib Monotherapy in Patients With Pancreatic Cancer and a Known Deleterious Mutation. JCO Precis Oncol **2018**;2018 doi 10.1200/PO.17.00316.
60. Simmons AD, Nguyen M, Pintus E. Polyclonal BRCA2 mutations following carboplatin treatment confer resistance to the PARP inhibitor rucaparib in a patient with mCRPC: a case report. BMC Cancer **2020**;20(1):215 doi 10.1186/s12885-020-6657-2.
61. Tao H, Liu S, Huang D, Han X, Wu X, Shao YW*, et al.* Acquired multiple secondary BRCA2 mutations upon PARPi resistance in a metastatic pancreatic cancer patient harboring a BRCA2 germline mutation. American Journal of Translational Research **2020**;12(2):612-7 doi papers3://publication/uuid/88B9F987-D5AB-4C6A-89FC-6D96673D57D6.
62. Vidula N, Rich TA, Sartor O, Yen J, Hardin A, Nance T*, et al.* Routine Plasma-Based Genotyping to Comprehensively Detect Germline, Somatic, and Reversion BRCA Mutations among Patients with Advanced Solid Tumors. Clin Cancer Res **2020**;26(11):2546-55 doi 10.1158/1078-0432.CCR-19-2933.
63. Weigelt B, Comino-Méndez I, de Bruijn I, Tian L, Meisel JL, García-Murillas I*, et al.* Diverse and Reversion Mutations in Circulating Cell-Free DNA of Therapy-Resistant Breast or Ovarian Cancer. Clin Cancer Res **2017**;23(21):6708-20 doi 10.1158/1078-0432.CCR-17-0544.
64. Yap TA, Kristeleit R, Michalarea V, Pettitt SJ, Lim JSJ, Carreira S*, et al.* Phase I trial of the poly(ADP-ribose) polymerase (PARP) inhibitor olaparib and AKT inhibitor capivasertib in patients with BRCA1/2 and non-BRCA1/2 mutant cancers. Cancer Discov **2020**:CD-20-0163-44 doi:10.1158/2159-8290.CD-20-0163.
65. Ikeda H, Matsushita M, Waisfisz Q, Kinoshita A, Oostra AB, Nieuwint AWM*, et al.* Genetic reversion in an acute myelogenous leukemia cell line from a Fanconi anemia patient with biallelic mutations in BRCA2. Cancer Res **2003**;63(10):2688-94.
66. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A*, et al.* The Ensembl Variant Effect Predictor. Genome Biol **2016**;17(1):122 doi 10.1186/s13059-016-0974-4.
67. Yates A, Akanni W, Amode MR, Barrell D, Billis K, Carvalho-Silva D*, et al.* Ensembl 2016. Nucleic acids research **2016**;44(D1):D710-6 doi:10.1093/nar/gkv1157.
68. Castro E, Goh C, Olmos D, Saunders E, Leongamornlert D, Tymrakiewicz M*, et al.* Germline BRCAMutations Are Associated With Higher Risk of Nodal Involvement, Distant Metastasis, and Poor Survival Outcomes in Prostate Cancer. Journal of Clinical Oncology **2013**;31(14):1748-57 doi:10.1200/JCO.2012.43.1882.
69. Hahnen E, Lederer B, Hauke J, Loibl S, Kröber S, Schneeweiss A*, et al.* Germline Mutation Status, Pathological Complete Response, and Disease-Free Survival in

Triple-Negative Breast Cancer. JAMA Oncol **2017**;3(10):1378-8 doi:10.1001/jamaoncol.2017.1007.

70. Song H, Cicek MS, Dicks E, Harrington P, Ramus SJ, Cunningham JM*, et al.* The contribution of deleterious germline mutations in BRCA1, BRCA2 and the mismatch repair genes to ovarian cancer in the population. Hum Mol Genet **2014**;23(17):4703-9 doi:10.1093/hmg/ddu172.
71. Tung N, Battelli C, Allen B, Kaldate R, Bhatnagar S, Bowles K*, et al.* Frequency of mutations in individuals with breast cancer referred for BRCA1and BRCA2testing using next-generation sequencing with a 25-gene panel. Cancer **2014**;121(1):25-33 doi:10.1002/cncr.29010.
72. Powell S, Forslund K, Szklarczyk D, Trachana K, Roth A, Huerta-Cepas J*, et al.* eggNOG v4.0: nested orthology inference across 3686 organisms. Nucleic acids research **2014**;42(Database issue):D231-9 doi:10.1093/nar/gkt1253.
73. Livingstone CD, Barton GJ. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. Comput Appl Biosci **1993**;9(6):745-56 doi papers3://publication/uuid/91CAB7D2-78B7-4411-B753-7C43CED9EC17.
74. Gourraud P-A, Khankhanian P, Cereb N, Yang SY, Feolo M, Maiers M*, et al.* HLA diversity in the 1000 genomes dataset. PLoS ONE **2014**;9(7):e97282 doi:10.1371/journal.pone.0097282.
75. Marty R, Kaabinejadian S, Rossell D, Slifker MJ, van de Haar J, Engin HB*, et al.* MHC-I Genotype Restricts the Oncogenic Mutational Landscape. Cell **2017**;171(6):1272-83.e15 doi:10.1016/j.cell.2017.09.050.
76. Punta M, Jennings V, Melcher A, Lise S. The immunogenic potential of recurrent cancer drug resistance mutations: an in silico study. bioRxiv **2019**:1-22 doi 10.1101/845784.

Figure Legends

Figure 1. Collation, annotation and standardisation of HR gene reversion mutations. A. Common architectures of HR gene reversion mutations associated with platinum or PARPi resistance. **B.** Workflow schematic illustrating the collation, annotation and standardisation of HR gene reversion mutations. **C.** Bar chart illustrating the primary tumor site in 91 patients with HR gene reversions described in the dataset. Patients are stratified by HR gene and by primary tumor site (see color key). **D.** Bar chart illustrating 308 reversion mutations in the dataset, stratified by HR gene and by primary tumor site. **E.** Bar chart illustrating that the majority of reversion mutations in the dataset arise from patients with different pathogenic mutations. Most patients (77%) had unique pathogenic mutations (annotated as “single-patient” mutations). Reversion cases from multiple patients with common Ashkenazi founder mutations, such as *BRCA2:*c.6174delT (c.5946delT in standardised nomenclature) and *BRCA1*:c.185delAG (c.68_69delAG), were also identified. **F.** Example of unique reversion events observed for multiple patients with a common founder mutation, *BRCA2:*c.6174delT (c.5496delT), represented on the *BRCA2* coding sequence (CDS). Two true reversions to wild-type DNA sequence were observed in two different patients. Second site reversion mutations in other patients are also shown, colored by patient. Deletions are indicated by thin black lines. Sites of insertions are shown by triangles, with the inserted bases listed to the right. Out-of-frame sequence between pathogenic and reversion mutation is shaded in grey. The position of the pathogenic c.5946delT mutation is indicated by a vertical line.

Figure 2. Directionality, hot and cold spots for reversion mutations. A. Scatter plots showing orientation (5’/upstream or 3’/downstream) of all reversions relative to original pathogenic mutation in *BRCA1* (left) or *BRCA2* (right). The start and end positions of each reversion mutation (i.e. the start and end of deleted regions) are

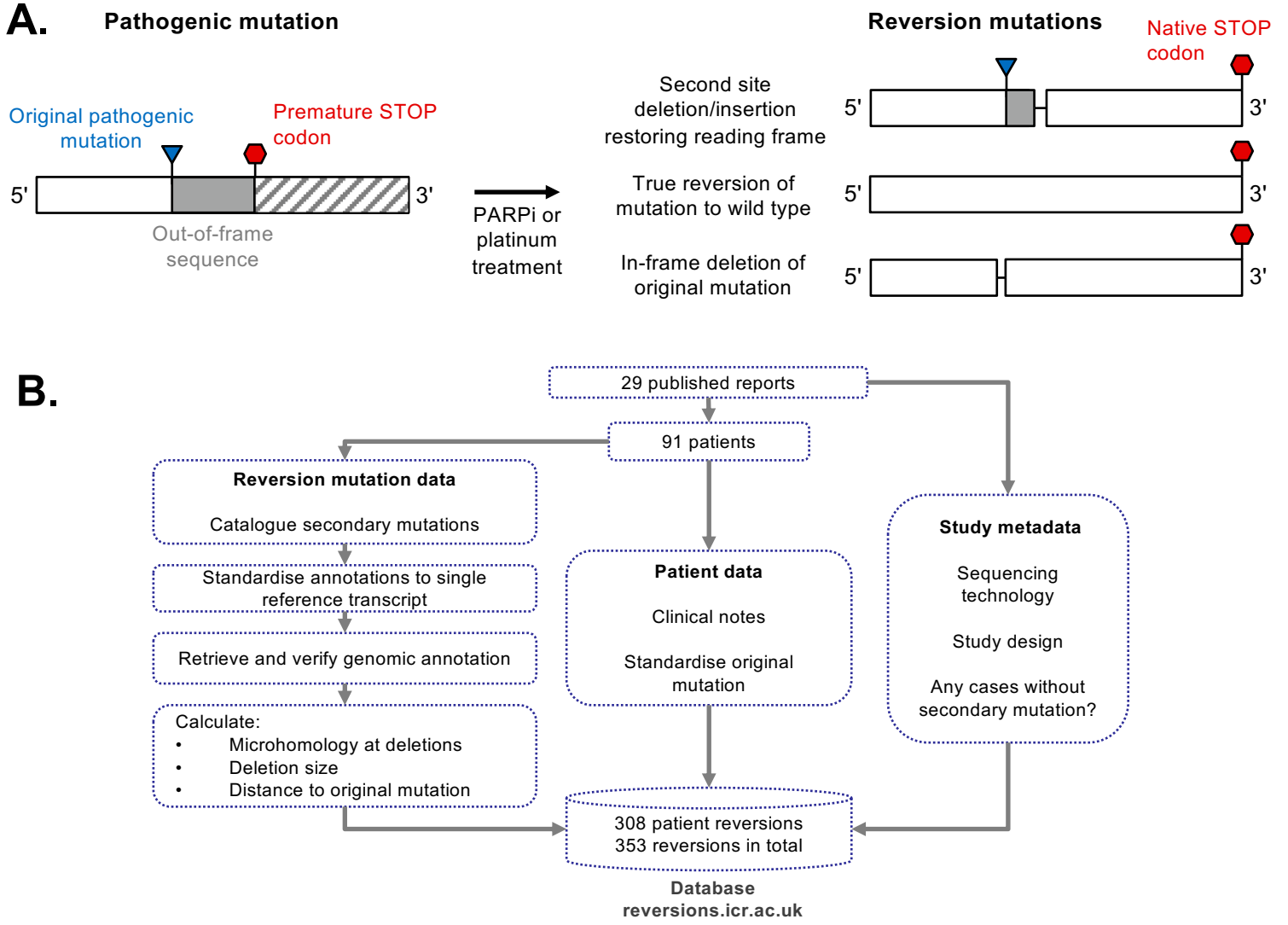
joined by lines; insertions are not shown. All positions are shown in CDS coordinates. In a few cases deletions extend beyond the plot boundaries, denoted by lines without a terminating point. For the majority of pathogenic mutations, reversion mutations do not have a directional bias and are seen both upstream and downstream of the pathogenic mutation. However, for some pathogenic mutations, e.g. *BRCA2* c.5946delT and *BRCA2*:c.7355delA, second site reversions are biased to the DNA sequence downstream of the pathogenic mutation. There is some evidence of a hotspot for reversion mutations at *BRCA2* position c.750-775 (highlighted in grey) and for a desert at the *BRCA2* C-terminus (highlighted in blue). Colors of points and lines denote different studies (colors are repeated). **B.** Conservation of amino acid sequence in BRCA1 (left) and BRCA2 (right) mapped onto CDS position for *BRCA1* and *BRCA2*, defined by conservation scores (see methods) determined by the alignment of 11 mammalian species. Notable peaks of conservation in BRCA2 are seen in the BRC region and the C-terminal OB and TR2 domains. **C.** Histogram illustrating the frequency of pathogenic mutations in the reversion dataset annotated by CDS position in *BRCA1* or *BRCA2.* Pathogenic mutations are shown in 40-bp bins. Two regions of BRCA2 are highlighted; the candidate reversion hotspot at c.750-775 (grey) and C-terminal region (blue). **D.** Histogram illustrating the frequency of pathogenic mutations in *BRCA1* or *BRCA2* in clinical studies covering breast, ovarian, pancreatic and prostate cancer ("Incidence" data, see *Methods*), plotted as in (C). The distribution of reverting mutations in *BRCA1* (shown in (C)) was not significantly different from the distribution of *BRCA1* mutations in the Incidence dataset ($P = 0.21$, two-sided Kolmogorov-Smirnov test). The frequency of reversions 3' to CDS position 7617 of *BRCA2* (exon 16 onwards) was significantly lower than expected frequency based on TCGA mutation data ($P < 0.015$, permutation test). **E.** Domain structure of BRCA1 and BRCA2 proteins annotated by CDS position. **F.** Bar chart illustrating the frequency of different pathogenic mutation types among reversions (upper) and compared to mutation types in Incidence data (lower).

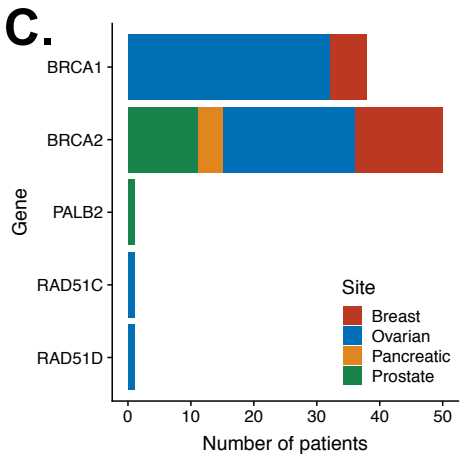
Figure 3. Microhomology usage in reversion mutations. A. Example of a reversion mutation in *BRCA2* associated with microhomology (patient 201 from Cruz *et al*.). The pathogenic G>T substitution mutation (BRCA2 c.145G>T) introduces a premature stop codon (TAA) as shown. The reversion mutation (c.145_168del24) is an in-frame deletion removing the mutated codon (shown in two different alignments). The existence of microhomology at this deletion is illustrated by the ambiguous alignment of the two nucleotides (TA) flanking it – these could be aligned equally well at either end as illustrated. **B.** Bar chart of reversion events classified by type. Reversions occurring via deletion are more frequent in *BRCA2* (88%) than in *BRCA1* (71%). **C.** Within deletion mutations, the use of microhomology occurs at a similar frequency in *BRCA1* and *BRCA2*. Reversion mutations are plotted as in (B) for deletions only. **D.** Deletion sizes are generally larger in *BRCA2* reversions (P = 0.0105, Wilcoxon rank sum test) with evidence of microhomology use. Total length of deleted sequence is shown for each reversion event, broken down by gene and presence of microhomology. **E.** *BRCA2* reversions use longer lengths of microhomology compared to *BRCA1*. Frequency distribution of length of microhomology used in *BRCA1* (red, left – mode 1 bp) compared with *BRCA2* (blue, right – mode 2 bp) plotted for all secondary deletions.

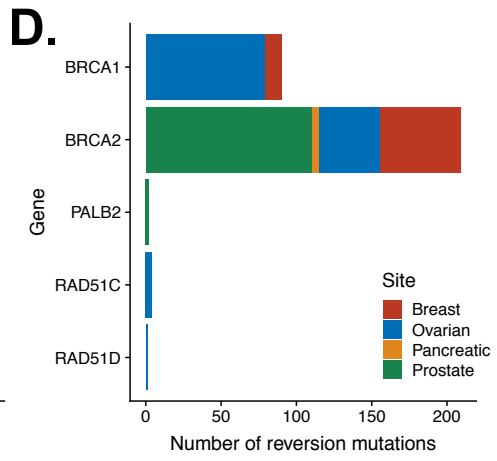
Figure 4. Prediction of HLA-mediated antigen presentation of reversion peptides. A. Percentage of individuals predicted to present at least one neopeptide from out-of-frame sequence associated with the listed pathogenic deletion mutations. This sequence will be shared with reversion mutations to some extent depending on the position of the reversion relative to the pathogenic mutation. Common founder mutations are highlighted. **B.** Predicted amino acid sequences from *BRCA2:*c.5946delT [c.6174delT] reversion events showing retention of out-of-frame sequence in many reversion alleles. The predicted protein sequence for each

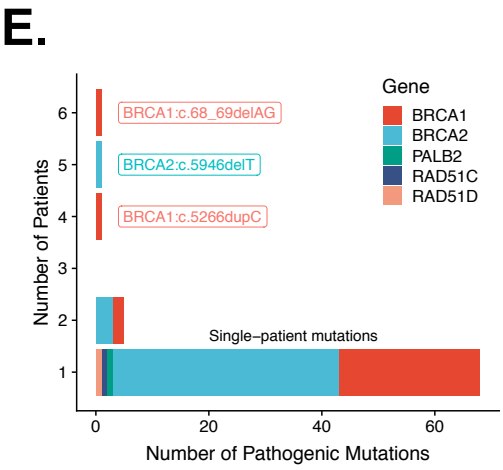
reversion observed for *BRCA2:*c.5946delT is shown compared to the wild-type (top) and predicted truncated c.5946delT protein sequence (second row). Sequences deriving from translation of out-of-frame coding sequence are shown in the yellow box. Amino acids are shaded based on their alignment to the wild type sequence. **C.** Computational prediction of HLA (HLA-A, HLA-B, HLA-C) presentation of out-of-frame protein sequences from *BRCA2* c.5946delT downstream reversions. Presentation likelihood calculated using NetMHCpan 4.0. The table shows the proportion of individuals in a set of 1,261 from the 1000 genomes project that have an HLA type predicted to present (%rank < 0.5) at least one neopeptide (length 8 to 11) associated with the indicated out-of-frame sequence (note that such neopeptides can include one or more WT amino acids upstream of the out-of-frame sequence). **D**. Percentage of individuals predicted to present at least one neopeptide for reverted protein sequences from all published cases of reversion mutations that encode neopeptides.

Figure 1.


A.
Pathogenic mutation
Original pathogenic mutation
Premature STOP codon
5'
3'
Out-of-frame sequence
PARPi or platinum treatment
Reversion mutations
Native STOP codon
Second site deletion/insertion restoring reading frame
True reversion of mutation to wild type
In-frame deletion of original mutation
B.
29 published reports
91 patients
Reversion mutation data
Catalogue secondary mutations
Standardise annotations to single reference transcript
Retrieve and verify genomic annotation
Calculate:
• Microhomology at deletions
• Deletion size
• Distance to original mutation
Patient data
Clinical notes
Standardise original mutation
Study metadata
Sequencing technology
Study design
Any cases without secondary mutation?
308 patient reversions
353 reversions in total
Database
reversions.icr.ac.uk


C.
BRCA1
BRCA2
PALB2
RAD51C
RAD51D
Gene
Site
Breast
Ovarian
Pancreatic
Prostate
0
10
20
30
40
50
Number of patients


D.
BRCA1
BRCA2
PALB2
RAD51C
RAD51D
Gene
Site
Breast
Ovarian
Pancreatic
Prostate
0
50
100
150
200
Number of reversion mutations


E.
Gene
BRCA1
BRCA2
PALB2
RAD51C
RAD51D
BRCA1:c.68_69delAG
BRCA2:c.5946delT
BRCA1:c.5266dupC
Single−patient mutations
Number of Patients
1
2
3
4
5
6
0
20
40
60
Number of Pathogenic Mutations

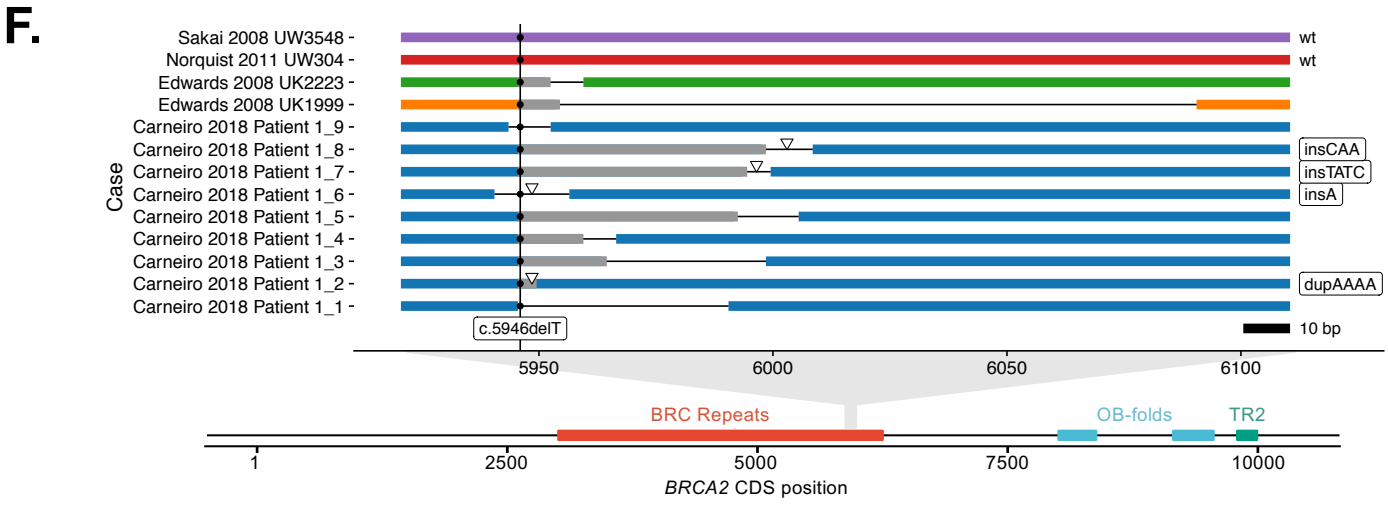

F.
Sakai 2008 UW3548
Norquist 2011 UW304
Edwards 2008 UK2223
Edwards 2008 UK1999
Carneiro 2018 Patient 1_9
Carneiro 2018 Patient 1_8
Carneiro 2018 Patient 1_7
Carneiro 2018 Patient 1_6
Carneiro 2018 Patient 1_5
Carneiro 2018 Patient 1_4
Carneiro 2018 Patient 1_3
Carneiro 2018 Patient 1_2
Carneiro 2018 Patient 1_1
Case
wt
wt
insCAA
insTATC
insA
dupAAAA
c.5946delT
10 bp
5950
6000
6050
6100
BRC Repeats
OB-folds
TR2
1
2500
5000
7500
10000
BRCA2 CDS position

Figure 2.

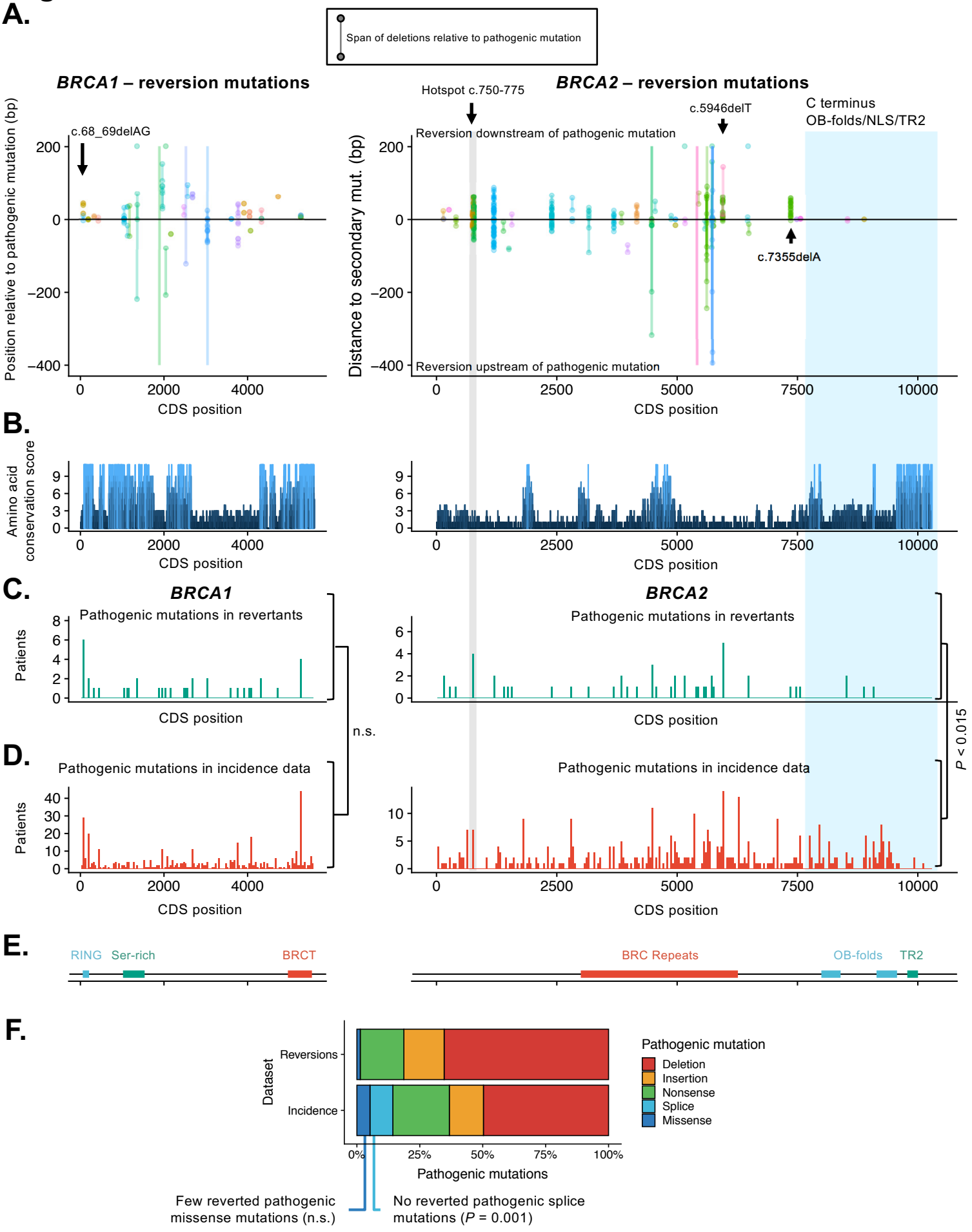
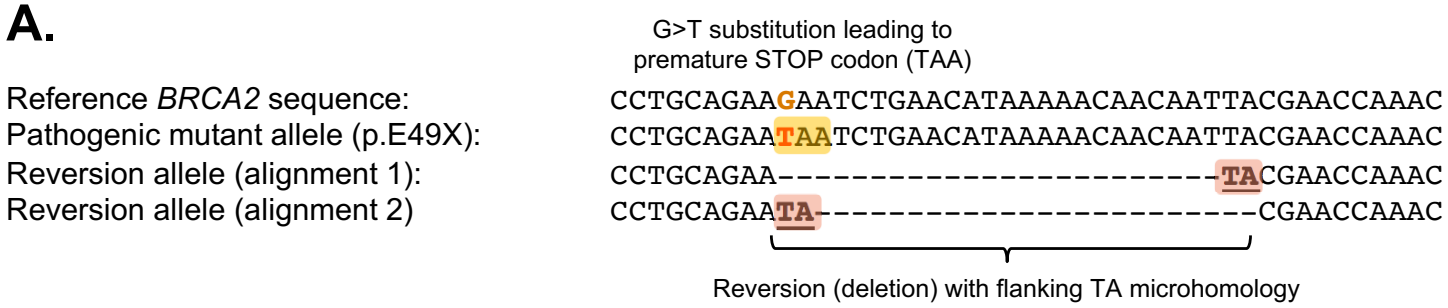
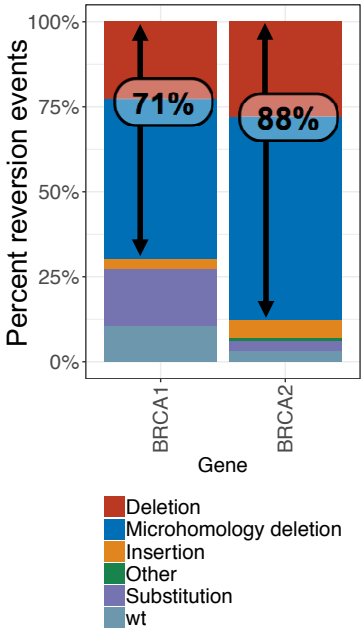

A.
Span of deletions relative to pathogenic mutation
BRCA1 – reversion mutations
BRCA2 – reversion mutations
Hotspot c.750-775
c.5946delT
C terminus
OB-folds/NLS/TR2
c.68_69delAG
c.7355delA
Reversion downstream of pathogenic mutation
Reversion upstream of pathogenic mutation
Position relative to pathogenic mutation (bp)
Distance to secondary mut. (bp)
CDS position
B.
Amino acid conservation score
C.
BRCA1
BRCA2
Pathogenic mutations in revertants
Patients
n.s.
P < 0.015
D.
Pathogenic mutations in incidence data
E.
RING
Ser-rich
BRCT
BRC Repeats
OB-folds
TR2
F.
Dataset
Reversions
Incidence
Pathogenic mutations
Pathogenic mutation
Deletion
Insertion
Nonsense
Splice
Missense
Few reverted pathogenic missense mutations (n.s.)
No reverted pathogenic splice mutations (P = 0.001)

Figure 3.

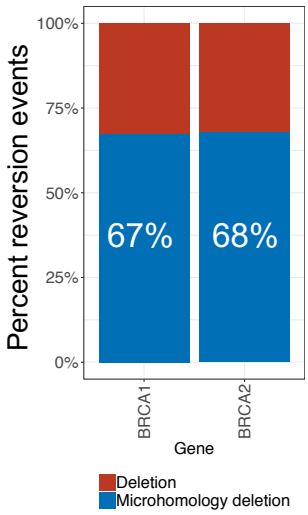
A.



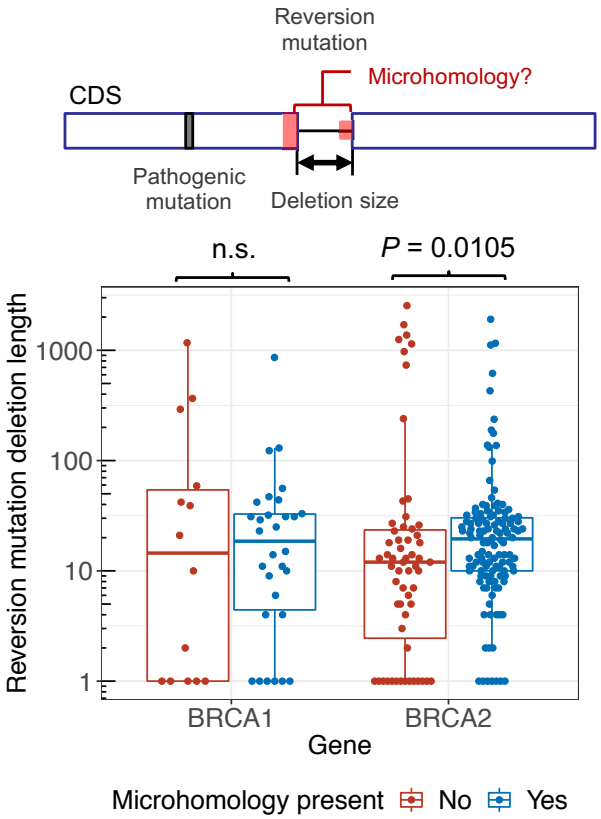
B.



C.



D.



E.

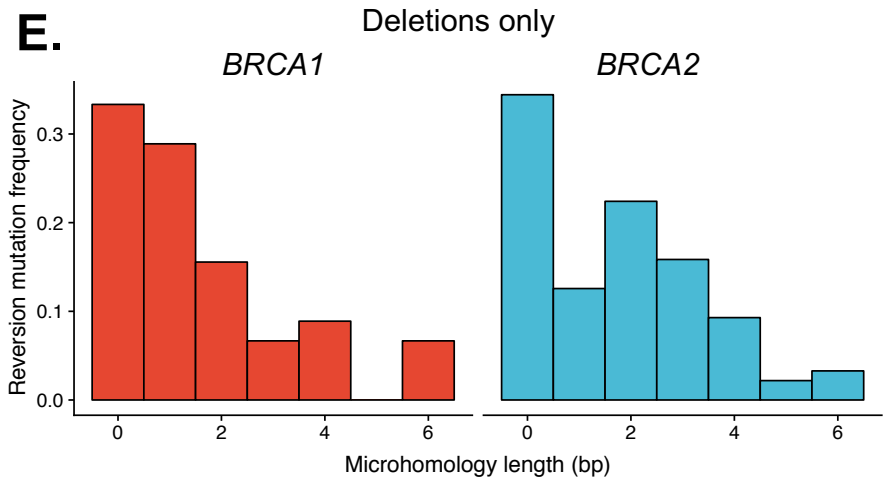
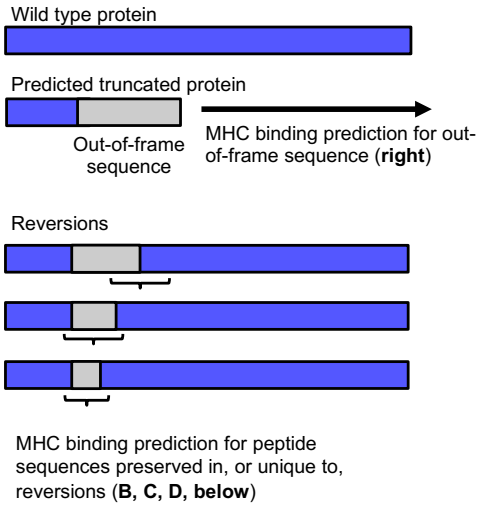
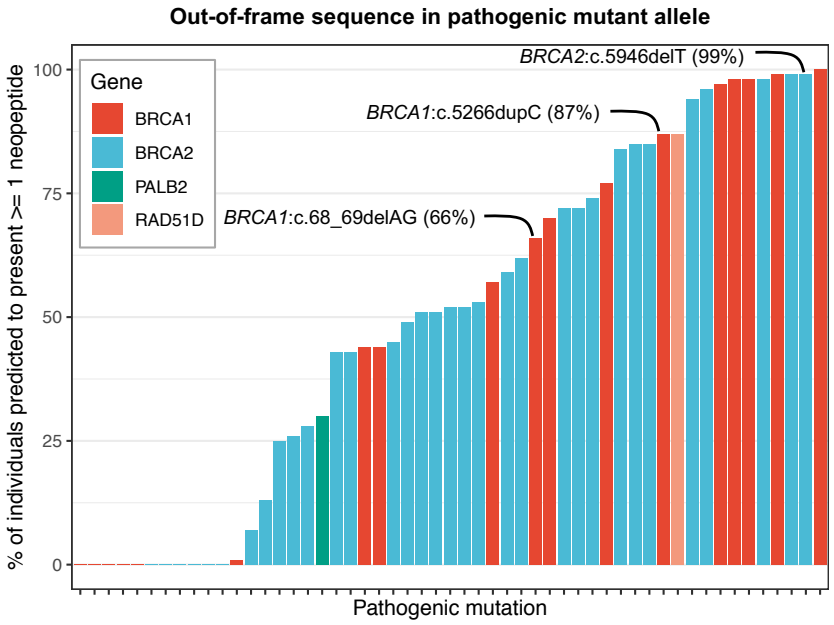


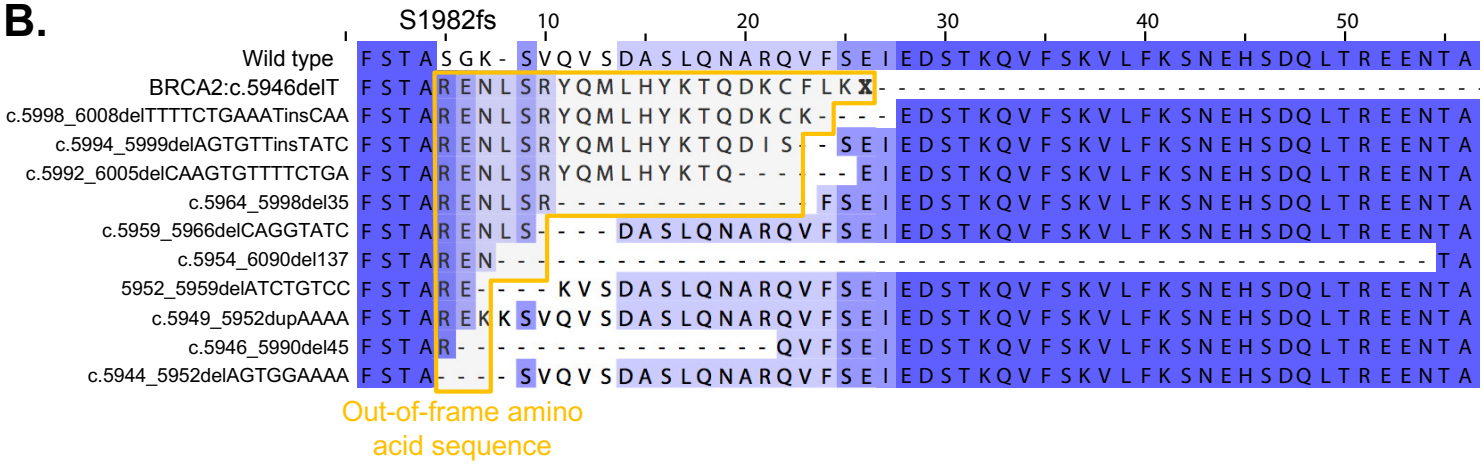
Figure 4.

A.

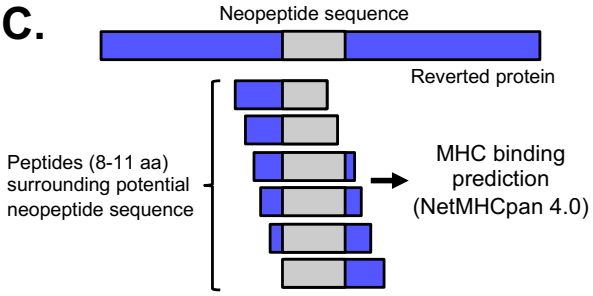




B.



C.



Neopeptide	Reversion events	Individuals presenting
REN	6	13%
RENL	5	41%
RENLS	5	41%
RENLSR	4	60%
RENLSRY	3	84%
RENLSRYQMLHYKTQ	3	96%

D.