

# Functional annotation of variants of the *BRCA2* gene via locally haploid human pluripotent stem cells

Received: 29 July 2021

Accepted: 15 June 2023

Published online: 24 July 2023

 Check for updates

Hanqin Li<sup>1,2</sup>, Rebecca Bartke<sup>1</sup>, Lei Zhao<sup>1,3</sup>, Yogendra Verma<sup>1</sup>, Anna Horacek<sup>1</sup>, Alma Rechav Ben-Natan<sup>1</sup>, Gabriella R. Pangilinan<sup>1</sup>, Netravathi Krishnappa<sup>2</sup>, Rasmus Nielsen<sup>1,3</sup> & Dirk Hockemeyer<sup>1,2,4</sup> ✉

Mutations in the *BRCA2* gene are associated with sporadic and familial cancer, cause genomic instability and sensitize cancer cells to inhibition by the poly(ADP-ribose) polymerase (PARP). Here we show that human pluripotent stem cells (hPSCs) with one copy of *BRCA2* deleted can be used to annotate variants of this gene and to test their sensitivities to PARP inhibition. By using Cas9 to edit the functional *BRCA2* allele in the locally haploid hPSCs and in fibroblasts differentiated from them, we characterized essential regions in the gene to identify permissive and loss-of-function mutations. We also used Cas9 to directly test the function of individual amino acids, including amino acids encoded by clinical *BRCA2* variants of uncertain significance, and identified alleles that are sensitive to PARP inhibitors used as a standard of care in *BRCA2*-deficient cancers. Locally haploid human pluripotent stem cells can facilitate detailed structure–function analyses of genes and the rapid functional evaluation of clinically observed mutations.

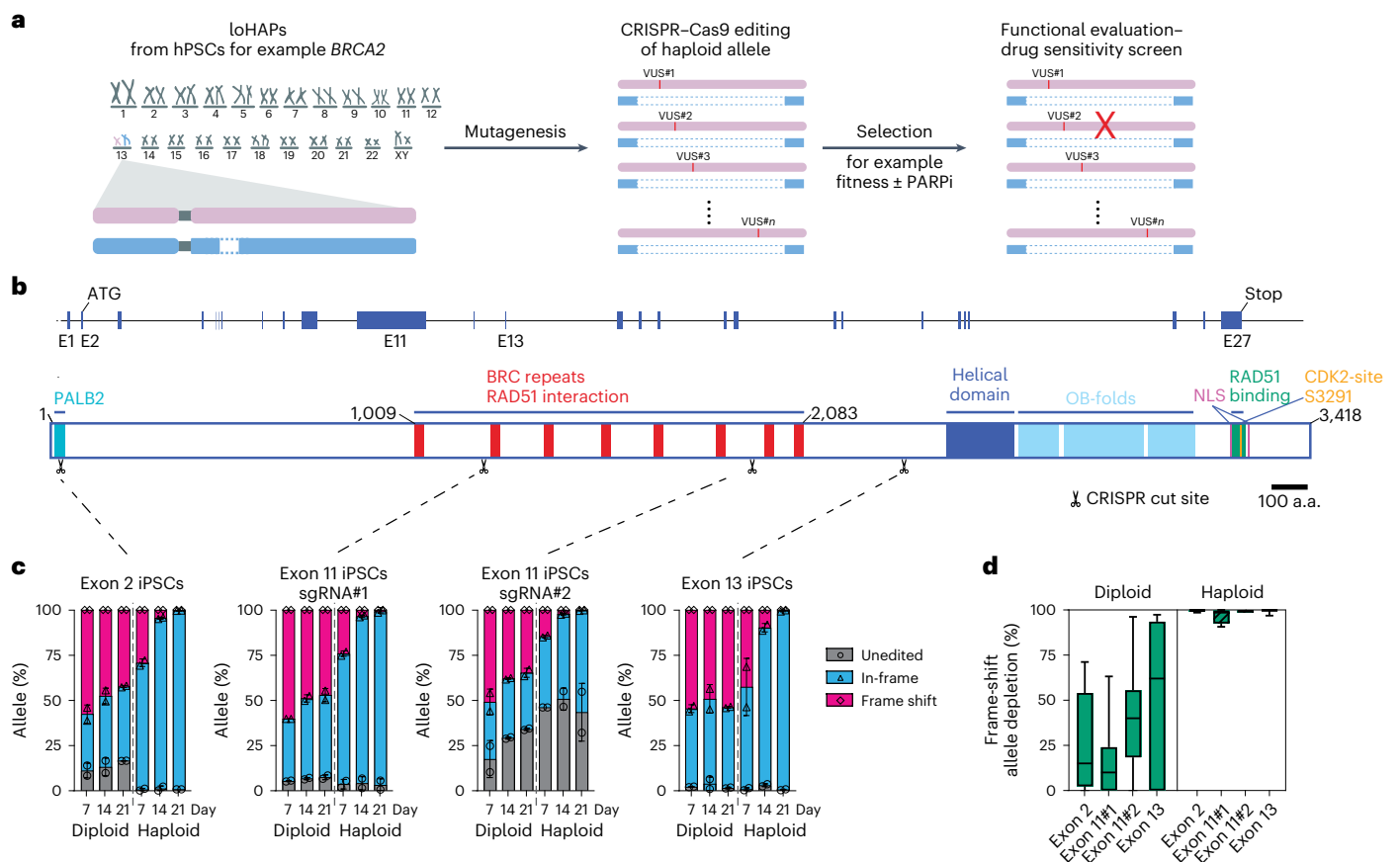
Genetic variants are key determinants of disease risk and can substantially impact diagnosis, prognosis and treatment outcomes. Classifying genetic variants of uncertain significance (VUS) in cancer genes such as *BRCA1* and *BRCA2* by PARP inhibitor sensitivity would result in clinically actionable, patient-specific information<sup>1–3</sup>. Recent advances in genome engineering have provided the necessary toolkit to introduce and evaluate disease-related variants at endogenous loci<sup>4–11</sup>. Using these tools, *BRCA* variants have previously been introduced into cancer cell lines using non-homologous end joining, homology-directed repair (HDR)<sup>12</sup>, or base editing<sup>13</sup> or prime editing<sup>14</sup> mediated mutagenesis.

A key limitation towards the implementation of these sophisticated genetic tools in the clinical interpretation of VUS is their use of cancer cell lines that inevitably harbour uncharacterized mutations in the DNA damage and repair pathways. These cell-line-specific idiosyncrasies alter cellular responses to perturbation and complicate

interpretation and generalization of the results. A clinically relevant system would allow for the assessment of a *BRCA* variant in a genetically controlled and disease-relevant primary cell system through comparison of the phenotypic impact of a specific *BRCA* mutation with its isogenic counterpart without the mutation. As outlined by the American College of Medical Genetics and Genomics/Association for Molecular Pathology interpretation guidelines, clinical variants should be evaluated using information from several orthogonal approaches in genetically diverse backgrounds that reflect the pathophysiology of patient populations<sup>15–17</sup>. For example, in the case of familial forms of *BRCA* deficiency, such a system would evaluate allele-specific drug sensitivity of a cell that underwent loss of heterozygosity compared with a patient's heterozygous non-cancer tissue<sup>18,19</sup>.

Several landmark studies<sup>20–22</sup> have shown that human haploid cells can facilitate functional annotation of genes and genetic variants due to

<sup>1</sup>Department of Molecular and Cell Biology, University of California Berkeley, Berkeley, CA, USA. <sup>2</sup>Innovative Genomics Institute, University of California, Berkeley, CA, USA. <sup>3</sup>Section for GeoGenetics, Globe Institute, University of Copenhagen, Copenhagen, Denmark. <sup>4</sup>Chan Zuckerberg Biohub, San Francisco, CA, USA. ✉e-mail: [hockemeyer@berkeley.edu](mailto:hockemeyer@berkeley.edu)



**Fig. 1 | Annotating functional protein domains in *BRCA2* using CRISPR-mediated mutagenesis in *BRCA2* loHAPs. a**, Annotation of genetic variants using loHAPs. First, one copy of a gene of interest (such as *BRCA2*) is excised from hPSCs to generate loHAPs. In a second step, the remaining allele is mutagenized (by introducing VUS) to identify alleles that are permissive and stably maintained in the cell pool. Finally, the pool of cells is challenged (for example, with a PARP inhibitor in *BRCA2* loHAPs) to identify variants sensitive to the treatment. **b**, Schematic representation of the *BRCA2* locus and the *BRCA2* protein domain structure. Shown are the *BRCA2* exons (blue boxes), the sgRNA target sites used in this study (indicated with scissors) as well as the *BRCA2* protein domain structure: PALB2 domain (turquoise), BRC repeats (red), helical domain

(dark blue), OB-folds (light blue), C-terminal RAD51 binding domain (green), two nuclear localization signals (NLS, purple) and CDK2 phosphorylation sites (orange). Scale bar, 100 amino acids (a.a.). **c**, Quantification of categorized allele frequencies (unedited, in-frame and frame-shift indels) in exons 2, 11 and 13 assayed in diploid and *BRCA2* loHAP hiPSCs at days 7, 14 and 21 after editing. Bars showing mean of two biological replicates and error bars showing s.e.m. **d**, Quantification of the relative depletion of frame-shift mutation in exons 2, 11 and 13 comparing samples collected at day 7 and 21 after editing in diploid and *BRCA2* loHAP hiPSCs. Boxes showing quartiles and whiskers showing the 10th and 90th percentile. Mutations with frequency >0.5% are plotted.

the strict genotype-phenotype relationship of single-allele alterations. Yet, haploid cells have a strong intrinsic tendency to spontaneously endoreduplicate and re-gain diploidy<sup>20–22</sup>, indicating genomic instability and an altered-physiological cellular state that compensates for an estimated 3,000 haploinsufficient genes in the human genome<sup>23</sup>. In this Article, we aim to develop a high-throughput, clinically actionable functional assay to investigate genetic variants in a genetically stable and physiology-representative human cell system.

## Results

Towards such assay, we took advantage of haploid genetics and developed a highly efficient method to generate locally haploid cells (loHAPs) from human embryonic stem cells (hESCs) and human-induced pluripotent stem cell (hiPSCs), collectively referred to as hPSCs. In loHAPs, one allele of the genomic region of interest is deleted, and genetic variants are introduced on the remaining functional allele, creating a haploid setting where mutations can be directly functionally tested (Fig. 1a). First, we designed an editing strategy that removes the entire gene of interest by excising the genomic regions between ~100 bp downstream of the transcriptional ends of the two neighbouring genes. Successful deletions are identified with polymerase

chain reactions (PCRs) detecting the presence of de novo junctions and allele-specific single-nucleotide polymorphisms (SNPs) in the deleted region (Extended Data Fig. 1a,b). To isolate loHAPs, we used a gene editing pipeline that uses a limited dilution, and next-generation sequencing (NGS)-based genotyping strategy<sup>24</sup> (Extended Data Fig. 1c) to generate clonal loHAPs for all six genes attempted with deletion sizes ranging between 64 kb and 168 kb. Our editing efficiencies ranged from 1% to 26.7%, assayed in one male iPSC cell and one female hESC line (*BRCA1*, 168 kb deletion, 2.9–24.9%; *BRCA2*, 96 kb deletion, 7.4–25.4%; *POUSF1*, 104 kb deletion, 19.9%; *HOXA* cluster, 157 kb deletion, 1%; *TERT*, 64 kb deletion, 26.7%; *FADS* cluster, 97 kb deletion, 8.6%) (Supplementary Fig. 1 and Supplementary Table 1). The notable high efficiency of loHAPs generation achieved non-virally using standard ribonucleoprotein (RNP) clustered regularly interspaced short palindromic repeats (CRISPR) delivery methods without the need for drug selection makes it a widely accessible and generalizable approach (for a detailed step-by-step protocol, see Methods). To demonstrate the power of loHAPs for functionally annotating VUS, we next focused on *BRCA2* loHAPs, which were generated in both hiPSCs and hESCs, proliferated normally and maintained pluripotency (Supplementary Fig. 2). LoHAPs remained haploid at the *BRCA2* locus without additional

karyotypic difference compared with parental cells as assayed by array comparative genomic hybridization at passages 15–20.

The *BRCA2* gene comprises 27 exons and encodes a full-length BRCA2 protein of 3,418 amino acids with previously annotated protein domains of distinct functions<sup>25,26</sup> (Fig. 1b). BRCA2 interacts at its N-terminus with PALB2 (amino acids 10–40, encoded by exons 2 and 3), contains eight BRCT repeats that can interact with RAD51, one  $\alpha$ -helical domain and three oligonucleotide/oligosaccharide-binding folds (OB-folds) involved in single-stranded DNA binding<sup>27</sup>, and harbours a functionally distinct C-terminal RAD51 interaction domain (encoded by exon 27 and mapped to amino acids 3,280–3,305) (ref. 25). A frame-shift mutation in exon 27 that disrupted this interaction has been associated with patient cancer predisposition<sup>28–31</sup>. In contrast, truncating mutations at amino acid 3326 (K3326X) have been reported to be non-pathogenic<sup>32,33</sup> or likely to have a low cancer risk<sup>34</sup>. Despite these insights, the exact truncation point at which BRCA2 function is substantially impaired, specifically which hypomorphic frame-shift mutations are sensitive to PARPi, remains largely unresolved<sup>34</sup>. About half of the 10,000 *BRCA2* variants listed in ClinVar<sup>35</sup> are annotated as VUS, distributed over the entire length of the protein, confounding their functional and therapeutic interpretation.

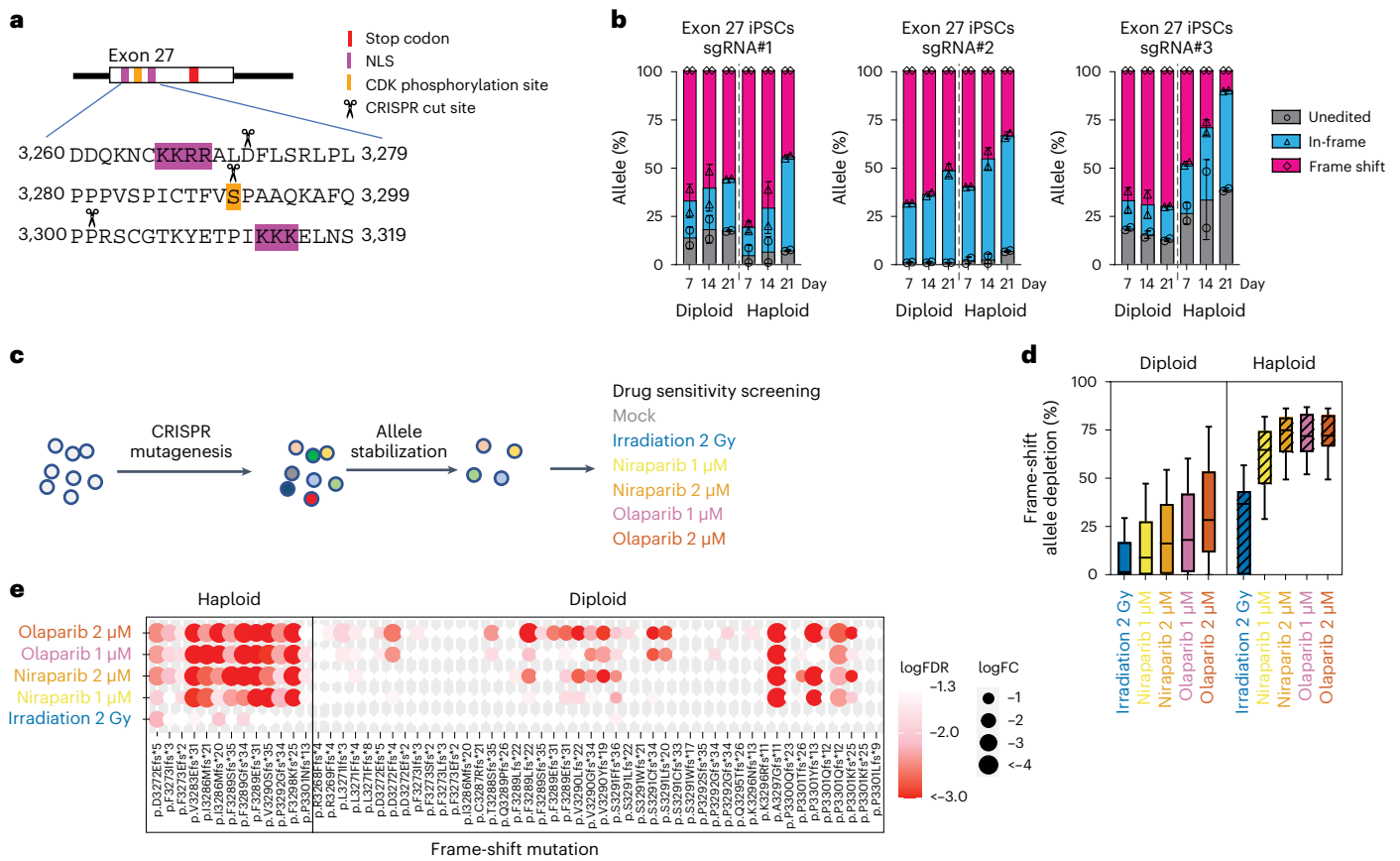
To assess the performance of our assay to test *BRCA2* genetic variants, we compared the efficacy with which we can annotate loss-of-function alleles in loHAPs compared with isogenic diploid cells. First, we introduced insertions and deletions (indels) via non-homologous end joining after CRISPR/Cas9-mediated cutting in the first coding exon, exon 2 and exons 11 and 13 with previously validated single guide RNAs (sgRNAs)<sup>36</sup> in hiPSC-based loHAPs and diploid cells (Fig. 1c). All four sgRNAs tested revealed the same pattern in these experiments: in diploid cells, the unedited allele and in-frame indels increased during 3 weeks of culture to a total of about 50%, while frame-shift mutations slightly decreased. However, this overall enrichment of in-frame indels and unedited alleles was dramatically more pronounced in *BRCA2* loHAPs, resulting in an almost complete absence of frame-shift mutations after 3 weeks of culture (Fig. 1c,d and Supplementary Table 2). These results were highly reproducible when analysing specific alleles shared between diploid and *BRCA2* loHAPs (Extended Data Fig. 2a) and were confirmed when comparing hESC-based cell lines (Extended Data Fig. 2b,c). Loss of BRCA2 is associated with a very specific cancer spectrum<sup>37,38</sup>, probably because tissues differentially rely on homologous recombination (HR) to repair DNA breaks<sup>39</sup>. Thus, we tested if our system could functionally annotate *BRCA2* variants in cell types differentiated from hPSCs. We tested this by introducing indels in exon 2 of fibroblasts differentiated from diploid or *BRCA2* loHAPs (Extended Data Fig. 2d,e), which largely recapitulated the phenotypes seen in hPSCs. Specifically, the effect size of deleterious frame-shift mutations was significantly larger in loHAPs compared with diploid fibroblasts (Extended Data Fig. 2e). However, frame-shift mutations depleted slower in fibroblasts than in stem cells, probably reflecting fibroblasts' slower proliferation rate. Together, these data demonstrate that the loHAPs system can rapidly identify loss-of-function variants in hPSCs and differentiated cell types.

A major clinical challenge is the classification of hypomorphic *BRCA2* variants that retain some BRCA2 function while still being sensitive to PARP inhibition. Thus, we next tested the hypothesis that loHAPs can be used to engineer and functionally annotate such hypomorphic alleles. To do this, we generated frame-shift mutations at three positions in exon 27 where hypomorphic alleles identified in genetically engineered mouse models<sup>40,41</sup> and VUS are prevalent (Fig. 2a). We find that frame-shift mutations were more tolerated in exon 27 than in exon 2, and the enrichment for the unedited allele and in-frame indels was reduced. The unedited and in-frame indels increased to less than 50% in diploid cells after 3 weeks (Fig. 2b), and the increase of these functional alleles was again more pronounced in loHAPs than in diploid cells. Overall, frame-shift alleles in exon 27 had reduced fitness, yet

a considerable proportion of frame-shift mutations were retained in *BRCA2* loHAP hiPSCs (Fig. 2b and Extended Data Fig. 2f), hESCs (Extended Data Fig. 2g,h) and differentiated fibroblasts (Extended Data Fig. 2i,j). To directly test the hypothesis that these remaining frame-shift alleles retained some functions yet are sensitive to PARP inhibition, we analysed the allele spectrum changes of loHAPs and diploid cells carrying frame-shift mutations in exon 27 when treated with 2 Gy of ionizing irradiation or two doses of the PARP inhibitors, niraparib and olaparib (Fig. 2c). Irradiation intensity and drug dosage were determined through dose–response curves on hPSCs (Supplementary Fig. 3a–c). Our analysis revealed that, while frame-shift mutations in exon 27 only mildly responded to irradiation, they were highly sensitive to treatment with niraparib or olaparib and depleted at all drug concentrations (Fig. 2d,e). Importantly, this depletion was much more pronounced in the *BRCA2* loHAPs than in the diploid cells (Fig. 2d). Further, analysis of individual frame-shift alleles showed that the penetrance of the effect was much more uniform in loHAPs than in diploid cells (Fig. 2e). This observation in the diploid cells is probably the consequence of a random expansion of compound heterozygotes carrying a frame-shift allele and a functional balancing allele that confound the effect of the mutation. These data suggest that *BRCA2* loHAPs can be used to reveal hypomorphic mutations that preserve residual BRCA2 function but are sensitive to PARP inhibition, providing an in vitro method to test clinical treatment options.

A primary aim of our effort is to design a rapid workflow that gives clinicians an accurate method to identify VUS affecting BRCA function and test the efficacy of treatment with PARP inhibition. Thus, we established a methodology to assay the effect of variants by introducing them through HDR-mediated CRISPR editing. We first focused on *BRCA2* exon 11, this time providing single-stranded oligonucleotides (ssODN) as an HDR template that either introduced a synonymous or non-sense mutation (Fig. 3a). HDR-mediated introduction of a synonymous mutation was highly efficient for both diploid and *BRCA2* loHAPs. Notably, the abundance of non-sense mutation was considerably lower than the synonymous mutation at day 7, suggesting that some level of selection already occurred during the first week after editing (Fig. 3b,c). Regardless of this, the relative allele frequency of the synonymous mutation increased substantially more in *BRCA2* loHAPs than in diploid cells over time (Fig. 3b), consistent with our results for a functional allele in Fig. 1. By contrast, the mutant allele frequency of the HDR-introduced non-sense mutation dramatically decreased over time, an effect that was again significantly more pronounced in *BRCA2* loHAPs than diploid cells (Fig. 3c and Extended Data Fig. 3a).

To test the robustness of HDR-mediated mutagenesis, we introduced single amino acid substitutions in *BRCA2* by performing alanine scanning experiments for the first 13 amino acids encoded by exon 2 and a non-sense substitution P9X as a complete loss-of-function control, using an ssODN pool encoding each alanine substitutions. The ssODN pools were designed to introduce synonymous PAM abrogating substitutions to prevent CRISPR recutting of alleles repaired by HDR (Fig. 3d). In our hands, HDR-based methods were highly efficient in both hESC and hiPSC (Extended Data Fig. 3b). All designer substitutions were introduced in diploid and *BRCA2* loHAPs (Extended Data Fig. 3c,d), with a frequency consistent with previous reports that demonstrate that editing efficiency decreased as the distance between designer substitutions and the cut site increased<sup>42</sup>. Like non-sense mutations in exon 11, the abundance of non-sense substitution P9X was much lower compared with the substitution P9A at day 7, suggesting that some level selection already occurs within the first week (Extended Data Fig. 3c,d). By binning in-frame and frame-shift mutations, alleles with frequency >0.05% can be reliably interpreted in these experiments (Extended Data Fig. 3e, and see Methods for detailed coverage calculations). Except for the start codon destroying M1A mutation, the relative allele frequencies were maintained for all alanine mutations over the 3 weeks, suggesting they are functional and do not cause a



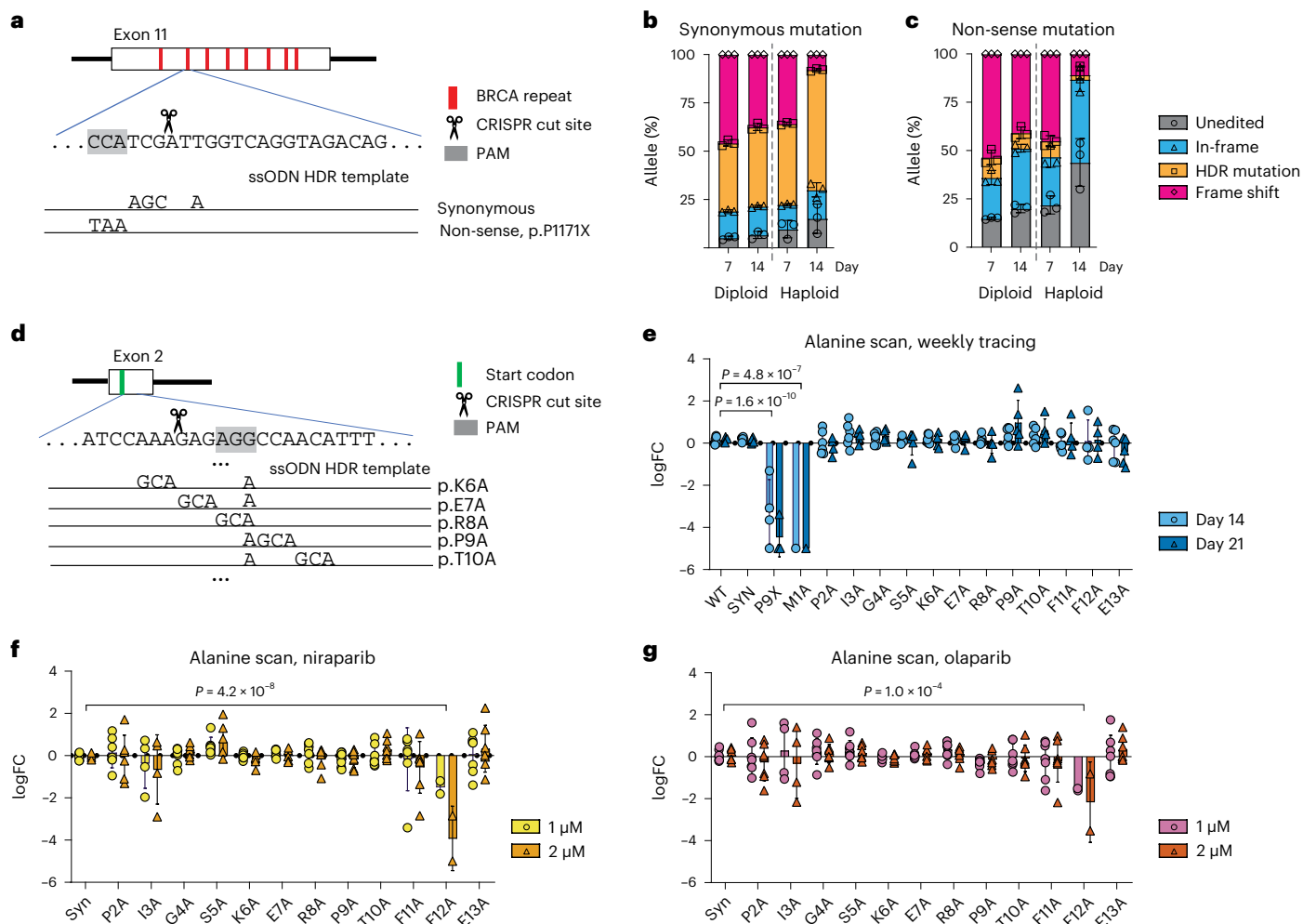
**Fig. 2 | Annotating hypomorphic mutations in *BRCA2* exon 27 using CRISPR-mediated mutagenesis in *BRCA2* loHAPs. **a**, Overview of the domain structure of *BRCA2* protein encoded by exon 27 (ref. 25). Indicated are two nuclear localization signals (NLS, purple) and a CDK phosphorylation site (orange) as well as three sgRNA cut sites used to mutagenize *BRCA2*. The *BRCA2* stop codon is indicated in red. **b**, Quantification of categorized allele frequencies (unedited, in-frame and frame-shift indels) for sgRNAs targeting exon 27 assayed in diploid and *BRCA2* loHAP hiPSCs at days 7, 14 and 21 after editing. Bars showing mean of two biological replicates and error bars showing s.e.m. **c**, Schematic of the experimental design to identify hypomorphic *BRCA2* mutations. Cells (diploid or *BRCA2* loHAPs) are mutagenized using CRISPR-mediated editing. After allele stabilization (3–4 weeks after editing), cells are exposed to 2 Gy of**

gamma radiation, 1  $\mu$ M or 2  $\mu$ M of niraparib or olaparib or remain untreated (mock). Changes in allele frequency are quantified 1 week after treatment starts. **d**, Quantification of the relative depletion of frame-shift mutations in exon 27 after drug or irradiation treatment comparing diploid and *BRCA2* loHAPs. Boxes showing the quartiles. Whiskers defined as 10th and 90th percentile. Mutations with frequency >0.5% in at least one group are plotted. **e**, Drug responses of individual *BRCA2* exon 27 frame-shift mutations comparing *BRCA2* loHAPs and diploid hiPSCs. The size of each circle represents fold change (FC) of alleles in a given treatment against mock; colour intensity represents FDR-corrected one-tailed Student's *t*-test *P* values. *n* = 6 in mock, *n* = 3 in the treatment groups; both were biological replicates. The same mutations as in **d** are plotted.

strong competitive disadvantage compared with the wild-type allele (WT) (Fig. 3e). However, when cultured in the presence of PARPi, cells with the F12A mutation were sensitive to both drugs, niraparib and olaparib (Fig. 3f,g). These findings are consistent with previous work that demonstrated amino acids within position 10–12 are indispensable for *BRCA2*'s ability to interact with PALB2, as only *BRCA2* polypeptide spanning amino acids 10–250, but not a polypeptide of 13–250 can immunoprecipitate PALB2 (ref. 43). Of note, when we performed the same alanine scanning experiment in diploid cells, it failed to reveal the deleterious effect of P9X and M1A during the 3 week window (Extended Data Fig. 3f) and PARPi sensitivity of F12A (Extended Data Fig. 3g,h).

Next, we benchmarked the sensitivity of our loHAP cells to reveal *BRCA2* defects by testing three well-characterized hypomorphic mutations G2508S in exon 15, K2729N in exon 18, and Y3035S in exon 23 in loHAPs (Fig. 4a). These mutations have been shown to impose a moderate cancer risk that reflects their relatively mild defect in facilitating HR and in vivo mESC complementation assays<sup>44</sup>. Engineering these in-frame mutations into *BRCA2* loHAPs was highly efficient, so each mutation was introduced with a frequency of >24.5%. In contrast to frame-shift mutations, the frequency of all three mutations

did not significantly depreciate over time compared with wild-type cells (Fig. 4b). Interestingly, in-frame deletions that were generated proximal to K2729 and Y3038 were depreciated, indicating that these regions of *BRCA2* encode critical functions (Extended Data Fig. 4a–c). More importantly, when treated with niraparib and olaparib, all three mutations revealed a significant and dose-dependent sensitivity to PARP inhibition (Fig. 4c,d) that recapitulated the relative HR competence previously reported for these mutations<sup>44</sup>. Next, we evaluated cell-type-specific differences of VUS to PARP inhibition. We generated 11 VUS in *BRCA2* exon 2 in loHAP hPSCs using HDR-mediated mutagenesis and then differentiating the mutant hPSCs pool to fibroblasts. Exposing these cells to PARPi did not reveal any sensitivity in hPSCs (Fig. 4e and Extended Data Fig. 4d). However, when exposing the differentiated isogenic fibroblasts, we identified a cell-type-specific sensitivity of the SSP *BRCA2* mutations to niraparib and olaparib treatment (Fig. 4f and Extended Data Fig. 4e). Taken together, these results show that our assay has a sensitivity that is on par with the gold-standard assays for annotating *BRCA2* mutations<sup>30,44,45</sup>. Moreover, these data demonstrate that, through the differentiation of edited loHAPs, our assay can reveal cell-type-specific drug responses.

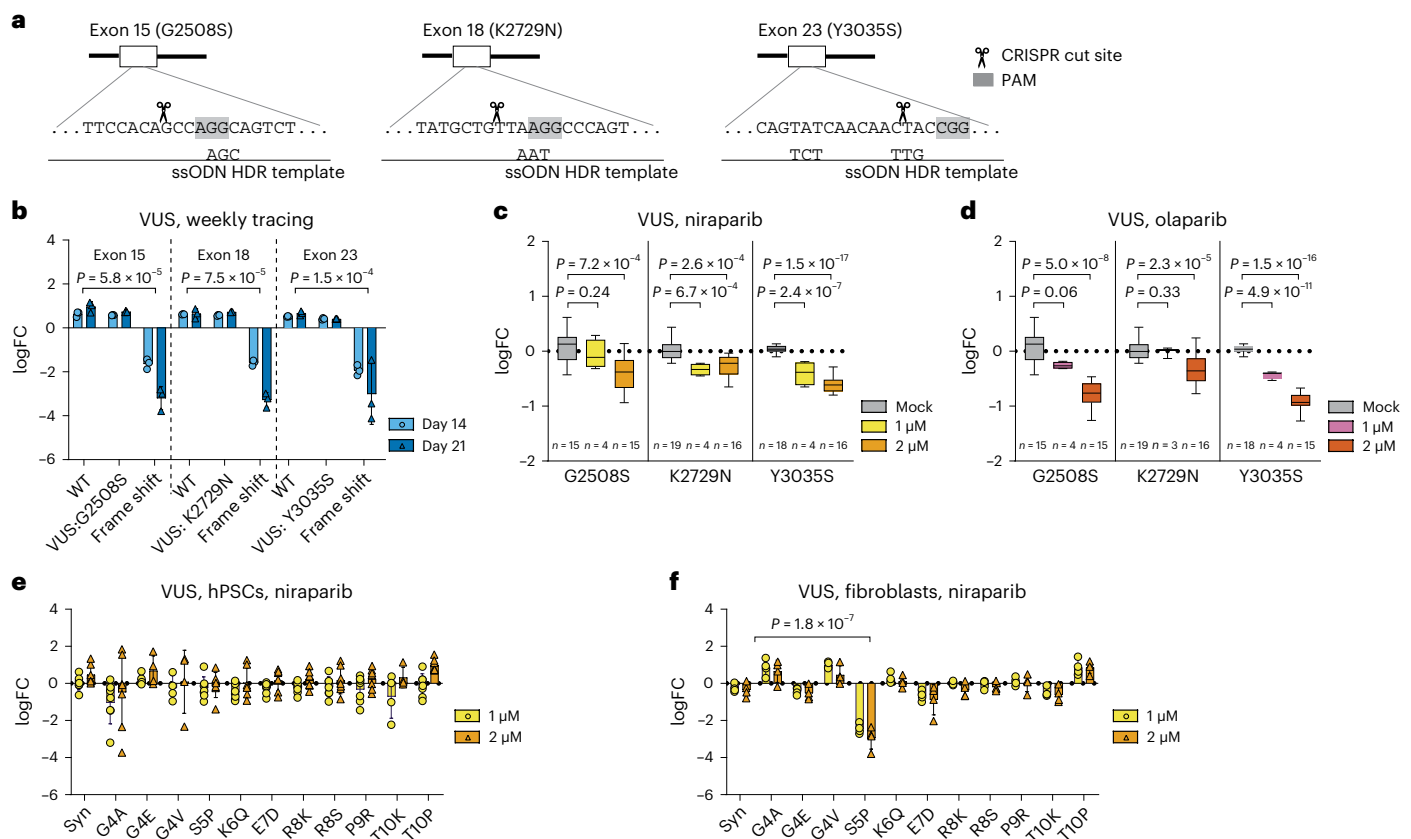


**Fig. 3 | Using HDR-mediated genome engineering in *BRCA2* loHAPs to assess functionality of designer mutations.** **a**, HDR-based editing strategy to introduce designer mutations in exon 11 of *BRCA2*. Shown is the targeted sequence with the PAM shaded in grey. Indicated below are the two HDR single-stranded repair oligos specifying the mutations that result in either synonymous or non-sense mutations. **b**, Quantification of categorized allele frequencies (unedited, in-frame indels, HDR-synonymous mutation and frame-shift mutations) in exon 11 assayed in diploid and *BRCA2* loHAP hiPSCs at days 7 and 14 after editing. Bars showing mean of three biological replicates and error bars showing s.e.m. **c**, Analysis as in **b**, using the HDR template introducing a non-sense mutation. **d**, HDR-based editing strategy to perform alanine scanning in *BRCA2*. Shown is the targeted sequence within exon 2 with the PAM shaded in

grey. Indicated below are examples for the HDR single-stranded repair oligos, specifying the mutations that result in alanine replacements. The start codon is indicated in green. **e**, Allele frequency changes over time of mutations generated in *BRCA2* exon 2 alanine scanning in *BRCA2* loHAP hiPSCs, represented as  $\log_2$  (day 14/day 7), circle, and  $\log_2$  (day 21/day 7), triangle.  $P$ , FDR-corrected two-tailed Student's  $t$ -test versus WT. Bars showing mean and error bars showing s.e.m. Data points represent biological replicates. FC, fold change. **f**, Normalized allele frequency changes upon treatment of PARP inhibitor, niraparib, of *BRCA2* exon 2 alanine mutations in *BRCA2* loHAP hiPSCs, represented as  $\log_2$  (drug/mock), 1  $\mu$ M, circle; 2  $\mu$ M, triangle.  $P$ , FDR-corrected two-tailed Student's  $t$ -test versus synonymous mutations (Syn). Allele frequencies were normalized against corresponding WT. **g**, Same as in **f**, for olaparib treatment.

Thus far, our data show that loHAPs provide a highly sensitive genetic background for the functional annotation of *BRCA2*. Next, we evaluated the scale and resolution that loHAPs can be employed to annotate *BRCA2* function by CRISPR-mediated mutation tiling<sup>46</sup>. To this end, we mutagenized *BRCA2* using every possible sgRNA (as determined by the spCas9 NGG protospacer adjacent motif) in exon 2 (Fig. 5a,b). We tested two approaches: (1) delivery of the editing components as RNPs and (2) infection of cells with a lentiviral library expressing that sgRNA followed by Cas9-protein transfection (Extended Data Fig. 5). While both approaches were highly effective, we focused on the RNP approach as it resulted in overall higher editing efficiencies than the lentiviral approach (~34% compared with ~12%). As before, we followed changes in the mutant allele spectra of three independent triplicates over 3 weeks. Deep sequencing, filtering and alignment of the resulting alleles revealed a highly complex mutation

spectrum, including an almost complete tiling of in-frame deletions that generate single, two or three amino acid deletions ( $\Delta 3$ ,  $\Delta 6$ ,  $\Delta 9$  nucleotides) as well as tiling of frame-shift mutations at almost each amino acid position close to CRISPR cut sites (Fig. 5c). Our analysis of exon 2 identified several classes of mutations that are strongly selected against: (1) frame-shift mutations, (2) in-frame deletions that delete the ATG encoding the translational start site, (3) deletions that alter the core splice-donor site at the end of exon 2, or (4) deletions that are within the *BRCA2* reading frame ( $\Delta 3$ ,  $\Delta 6$ ,  $\Delta 9$  nucleotides, and so on) that generate de novo stop codons at the deletion site (Fig. 5d). Given this validation, we next generated a statistical model to analyse the high complexity of in-frame deletions to evaluate the functional contribution of individual amino acids in exon 2 of *BRCA2* (Fig. 5e). In brief, the method uses maximum likelihood to estimate the functional effects of mutations in each site, combining information from mutations



**Fig. 4 | Using HDR-mediated genome engineering in *BRCA2* loHAPs to evaluate *BRCA2* VUS.** **a**, HDR-based editing strategy to generate G2508S, K2729N and Y3035S. Shown are the targeted sequences with the PAM shaded in grey. Indicated below are the HDR single-stranded repair oligos used. **b**, Normalized allele frequency changes over time of G2508S, K2729N and Y3035S in *BRCA2* loHAP hPSCs, with WT and one frame-shift mutation at each locus as controls, represented as  $\log_2(\text{day 14}/\text{day 7})$ , circle, and  $\log_2(\text{day 21}/\text{day 7})$ , triangle.  $P$ , FDR-corrected two-tailed Student's  $t$ -test versus WT. Bars showing mean of three biological replicates and error bars showing s.e.m. FC, fold change. **c**, Normalized allele frequency changes upon treatment of PARP inhibitor, niraparib, of G2508S, K2729N and Y3035S loHAP cell pools, represented as  $\log_2(\text{drug}/\text{mock mean})$ .  $P$ , FDR-corrected two-tailed Student's  $t$ -test versus

mock. Allele frequencies of G2508S and K2729N were normalized against corresponding WT; Y3035S were normalized against synonymous mutation. Boxes showing quartiles and whiskers showing the 10th and 90th percentile.  $n$  indicates biological replicates. **d**, Same as in **c**, for olaparib treatment. **e**, Normalized allele frequency changes upon treatment of PARP inhibitor, niraparib, of *BRCA2* exon 2 VUS in *BRCA2* loHAP hPSCs, represented as  $\log_2(\text{drug}/\text{mock})$ . 1  $\mu\text{M}$ , circle; 2  $\mu\text{M}$ , triangle. Allele frequencies were normalized against WT. Alleles with frequencies  $>0.5\%$  in both hPSCs and fibroblasts are plotted. Bars showing mean and error bars showing s.e.m. Data points represent biological replicates. **f**, Same as **e**, in mutated *BRCA2* loHAP hPSC-derived fibroblasts.  $P$ , FDR-corrected two-tailed Student's  $t$ -test versus synonymous mutations (Syn).

covering multiple sites and taking amino acid similarity into account. A full description of the model is given in Supplementary Methods. In agreement with our previous result, this model also identified the phenylalanine at position 12 as a critical residue in *BRCA2*. Simultaneous deletion of both F11 and F12 showed a significantly greater impact than single amino acid deletion in both the initial mutation tiling experiments (Fig. 5f) and in a targeted follow-up validation experiment (Fig. 5g). Thus, we demonstrate through alanine substitution and in-frame deletions that amino acid F12 is important for *BRCA2* function, probably by mediation of interactions with PALB2 (ref. 43). Noteworthy, the same tiling deletion experiment, when performed in diploid cells, failed to reveal the deleterious effect of the F11 and F12 double deletion (Extended Data Fig. 6a). From these experiments, we conclude that loHAPs can be used to map critical protein–protein interactions with single amino acid resolution.

Next, we tested if the mutation tiling approach can be used to map the specific amino acid position in exon 27 at which C-terminal truncating mutations render cells sensitive to PARPi. Mutagenizing loHAPs with 22 sgRNA cutting in exon 27 around the RAD51 binding domain (Fig. 6a,b) resulted again in a highly complex allele spectrum, with truncating and in-frame mutations tiling almost the entire

region (Fig. 6c). While mutations at every position in exon 27 persisted until week 3 to some extent, only frame-shift mutations before R3302 showed a significant reduction in fitness (Fig. 6d). Subsequent exposure of these cells to 2 Gy irradiation only changed the allele spectrum mildly (Extended Data Fig. 6b). In contrast, treatment with both niraparib and olaparib independently revealed that frame-shift mutations before and at amino acid position 3302 are sensitive to PARPi. (Fig. 6e,f and Extended Data Fig. 6c). These results were validated in an independent deletion tiling screen (Extended Data Fig. 6d–f) in loHAPs and diploid cells. In diploid cells, the effects of frame-shift mutations binned into groups with and without the amino acid 3,300–3,302 PPR motif were much smaller compared with loHAPs (Extended Data Fig. 6d). Analysing individual mutations in loHAPs identified 86.0% of frame-shift mutations that truncate the amino acids 3,300–3,302 PPR motif as less competitive compared with wild type, while in diploid cells, only 30.0% showed a significant deleterious effect when correcting for a 10% false discovery rate (FDR) (Extended Data Fig. 6e). Similarly, the response to PARPi inhibition of variants without 3,300–3,302 PPR motif was inconsistent between niraparib and olaparib treatment, while both drugs had a significant effect in loHAPs (Extended Data Fig. 6f). Thus, predicting

the performance of individual mutations is more reliable in loHAPs than in diploid cells.

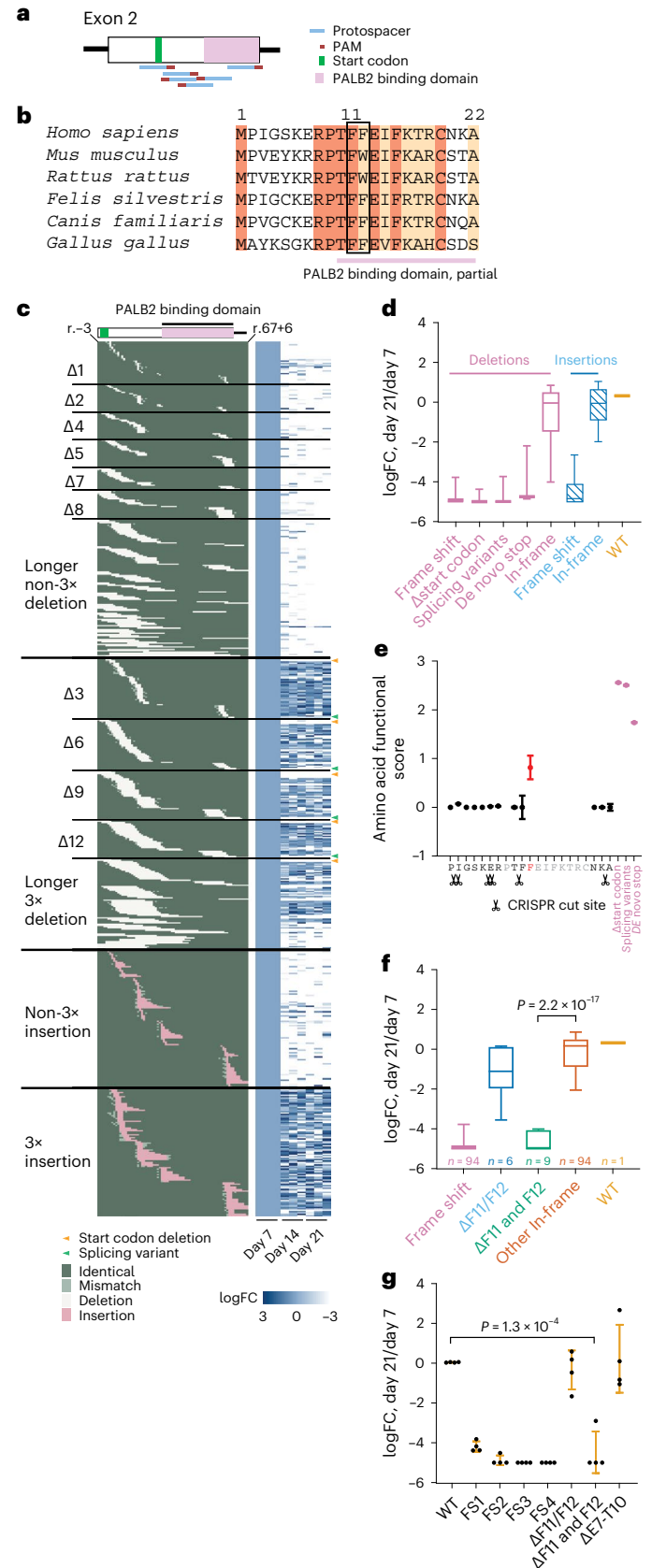
Our conclusion that amino acids 3301 and 3302 are critical amino acids in BRCA2 is again highly consistent with biochemical data that mapped the amino acids 3,280 to 3,305 in BRCA2 to be required for the recruitment of RAD51 (ref. 25) and with mouse data that map this interaction to the very same conserved region<sup>47</sup> (P3301 *Homo sapiens* corresponds to P3224 *Mus musculus* (Fig. 6b)). To further corroborate our conclusion, we deployed our statistical model to evaluate the effect of in-frame mutations within the RAD51 binding domain (Fig. 6b) on BRCA2 function. Using the aforementioned statistical method for quantifying the effect of mutations in individual sites confirmed the importance of amino acids 3,293–3,302 and revealed overlap between evolutionarily conserved amino acids in this region (Extended Data Fig. 7a). Importantly, in-frame mutations that removed the 3,300–3,302 PPR motif and de novo stop mutations, S3303X, C3304X and Y3308X also rendered cells sensitive to PARPi (Extended Data Fig. 7b–e), which is consistent with the clinVar annotation and previous reports<sup>28–31</sup>. Thus, our data correlate the disruption of this well-mapped interaction domain as the most C-terminal frame-shift mutations that result in PARPi sensitivity.

### Discussion

We have established an efficient method to generate loHAPs in which large genomic regions of interest are excised from a single allele in hPSCs. Because only one copy of a gene of interest is present in loHAPs, allele frequencies in a cell population directly report on the fitness of mutations in response to an environmental or genetic perturbation.

**Fig. 5 | Using BRCA2 loHAPs to identify critical amino acids by CRISPR-mediated mutation tiling in BRCA2 exon 2.** a, Schematic of editing strategy in the mutation tiling experiments on BRCA2 exon 2. Shown are protospacers (blue), PAMs (red), start codon (green) and the part of PALB2 binding domain in exon 2 (pink). b, Sequence alignment of amino acids encoded by BRCA2 exon 2 across multiple species. Identical amino acids are marked dark brown, and conserved amino acids are marked light brown. The pink bar indicates the part of PALB2 binding domain in exon 2. The black box highlights the identified critical amino acid F11, F12. c, The aligned indel profile and heatmaps of allele frequency change over time of each mutation at the nucleotide level in each of the triplicate experiments. Indel category is indicated to the left (3x, multiples of 3) and sample labels at the bottom. The identity of each position of each allele is categorized and colour coded as identical to WT (dark green), mismatch (light green), deletion (white) and insertion (pink). Specific mutations shown are start codon deletion (orange arrowheads) and splicing variant (green arrowheads). Allele frequencies of each mutation in each replicate were normalized to the corresponding day 7 frequencies. A total of 679 alleles that appeared in all independent day 7 samples and had at least 20 reads in at least one sample were analysed. FC, fold change. d, Categorized allele frequency changes of in-frame and frame-shift alleles in exon 2 at day 21 compared with day 7, presented as log<sub>2</sub>(day 21/day 7). Boxes showing quartiles and whiskers showing the 10th and 90th percentile. e, Functional score of each amino acid in BRCA2 exon 2 calculated by a statistical model, with CRISPR cut site labelled. Only scores of amino acids covered by one amino acid deletion are plotted. The F12 with outstanding functional score is marked in red. Known detrimental mutation groups are shown in purple as positive controls. Dots indicate mean and error bars indicate standard deviation. f, Categorized allele frequency changes of F11, F12 deletions at day 21 compared with day 7, presented as log<sub>2</sub>(day 21/day 7). P value reports the significance between F11 and F12 double deletion (n = 9) versus other in-frame mutations (n = 94) in two-tailed Student's t-test. Boxes showing quartiles and whiskers showing the 10th and 90th percentile. g, Normalized allele frequency changes of amino acid F11, F12 deletions at day 21 compared with day 7, presented as log<sub>2</sub>(day 21/day 7), in a targeted validation experiment, with allele category labels at the bottom. FS1–4, four examples for different frame-shift mutations and ΔE7–T10 for unaffected in-frame deletions. P value reports the significance between F11 and F12 double deletion versus WT in two-tailed Student's t-test. n = 4 biological replicates. Error bar presents standard deviation.

The pooled approach we use to assess mutations has several advantages: it allows direct comparison of mutant allele frequencies to the unedited WT and simultaneous testing of highly complex mutant pools with drugs classifying mutations into those responsive or resistant to the treatment.



Using the *BRCA2* gene, we provide proof of concept for functional annotation of mutations introduced on the remaining gene copy in loHAPs. We find mutations that disrupt the reading frame, delete the translational start site, introduce a de novo stop codon or impair proper splicing are annotated effectively in loHAPs. Moreover, we establish protocols that evaluate mutations introduced by HDR templates to perform alanine scanning experiments and investigate clinical VUS. When implemented in loHAPs, these protocols allow for more efficient evaluation of mutations compared with diploid counterparts. Moreover, we demonstrate that our pipeline can effectively annotate hypomorphic alleles that are functional in unchallenged cells but are sensitive to treatment with PARPi. For each of these experiments, we demonstrate that loHAPs outperform their diploid counterpart.

We perform these experiments in hESCs and iPSCs as well as differentiated cells, extending the range from the traditional cancer cell lines in which similar structure–function analyses are typically performed towards more diverse cell types with diverse genetic backgrounds. By extending the physiological context and genetic background of VUS analysis, our method can reveal cell-type-specific therapeutic responses of genetic variants. Thus, using loHAPs and their differentiated cell types can elucidate fundamental questions about why mutations with specific driver mutations, such as *BRAC2*, are associated with specific cancer types<sup>37</sup>. Together, our loHAPs assay to evaluate VUS fulfils the suggestions by the American College of Medical Genetics and Genomics/Association for Molecular Pathology clinical variant interpretation guidelines, which states that VUS evaluation ideally employs multiple orthogonal approaches reflecting the physiology, disease context and patient population<sup>15–17</sup>.

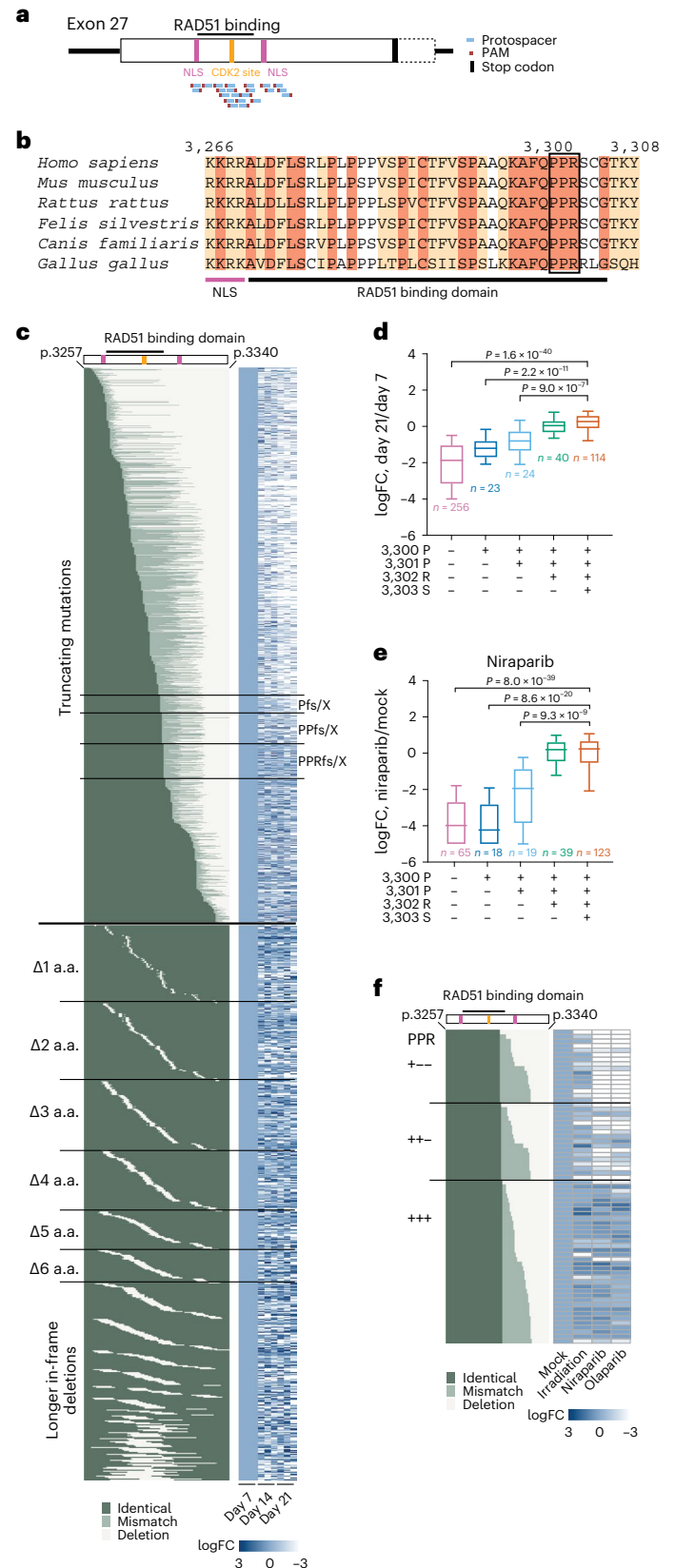
Our benchmarking experiments show that evaluating mutations loHAPs is robust, scalable and sensitive. We generate and test 13 continuous alanine substitutions in exon 2 of *BRCA2*. Moreover, by engineering three clinically relevant and well-characterized hypomorphic

mutations, we demonstrate that our approach has a sensitivity that is similar in detecting defects as established complementation or direct HR activity assays to evaluate VUS in *BRCA2* (refs. 30,44,45).

Beyond the clinical evaluation of VUS, our proof-of-concept experiments highlight how functional annotation of genes in loHAPs can advance biological insights. By comprehensively mutating a region

**Fig. 6 | Using *BRCA2* loHAPs to identify critical amino acids by CRISPR-**

**mediated mutation tiling in *BRCA2* exon 27.** **a**, Schematic of editing strategy in the mutation tiling experiments on *BRCA2* exon 27. Shown are protospacers (blue), PAMs (red), stop codon (black), NLS (purple), CDK2 phosphorylation site (orange) and RAD51 binding domain (black bar on top). **b**, Sequence alignment of amino acids around the RAD51 binding domain (black bar) in exon 27 across multiple species. Identical amino acids are marked dark brown, and conserved amino acids are marked light brown. The first *BRCA2* NLS is indicated with a purple bar. The black box highlights the identified critical amino acids PPR (3,300–3,302). **c**, The aligned full truncating mutation and in-frame deletion profile generated at amino acid (a.a.) level and heatmaps of allele frequency changes over time of each mutation, with mutation category labels to the left, sample labels at the bottom and the labels of truncating mutations (fs/X, frameshift or de novo stop codon) break at the critical P3300, P3301 and R3302 at the right. The identity of each position of each allele is categorized and colour coded as identical to WT (dark green), mismatch (light green) and deletion (white). Allele frequencies of each mutation in each replicate were normalized to the corresponding day 7 allele frequencies. A total of 2,195 alleles that appeared in all independent day 7 samples and had at least 20 reads in at least one sample were analysed. **d**, Categorized allele frequency changes of frame-shift mutations before P3300, at P3300, at P3301, at R3302 or at later positions, at day 21 compared with day 7, presented as  $\log_2(\text{day 21}/\text{day 7})$ . Boxes showing quartiles and whiskers showing the 10th and 90th percentile. *P* values report the significant comparisons versus the last group in two-tailed Student's *t*-test with FDR correction. *n*, number of different alleles in each group. Alleles with at least 100 reads at day 7 were analysed. FC, fold change. **e**, Categorized allele frequency changes on groups as in **d** upon niraparib treatment compared with the mock condition, presented as  $\log_2(\text{niraparib}/\text{mock})$ . Plotting parameters and statistics are shown the same as in **d**. Alleles with at least 100 reads in mock condition were analysed. **f**, The aligned allele profile at amino acid (a.a.) level of truncating mutations at P3300, P3301 or R3302, and heatmaps of allele frequency changes upon the treatment of irradiation or PARPi. Allele frequencies of each mutation were normalized to the corresponding allele frequencies in mock. Alleles with at least 100 reads in mock condition were analysed.





of the gene using every possible PAM, we generate a highly complex mutation spectrum that comprises in-frame deletions that almost completely tile across the region of interest. Using a new statistical model that can combine evidence from multiple mutations, each potentially spanning several sites, we demonstrate that this mutation tiling allows for the resolution of protein function to the single amino acid level. Specifically, we map the amino acids in *BRCA2* exon 2 required for the documented interaction with PALB2, the interacting region in exon 27 of *BRCA2* with RAD51, and the most C-terminal frame-shift mutation that renders cells sensitive to PARP inhibition.

Functionally annotating variants in loHAPs is generally limited to genes that are not fully haploinsufficient. Moreover, we find that the deletion spectrum in our tiling experiments is currently restricted by the availability of SpCas9 PAM sites. To fully cover the entire coding region of a gene might require deploying orthogonal Cas approaches<sup>48,49</sup> that can cut at non-NGG protospacer adjacent motifs.

## Outlook

Locally haploid human pluripotent stem cells provide an approach for the evaluation of VUS in *BRCA2* and to test their sensitivity to PARPi in a cell-type-specific manner. Scaling this approach to the more than 5,000 reported VUS in *BRCA2* could provide a unique dataset to guide strategies for therapeutic interventions (Supplementary Fig. 4). Moreover, we expect that mutation scanning in loHAPs will prove as a general strategy to elucidate how mutations differentially trigger cell-type-specific responses. Finally, future experiments with loHAPs could be extended to analyse phenotypes more complex than cell survival and drug resistance.

## Methods

### Pluripotent stem cell culture

Pluripotent stem cell research is approved under 2012-12-024 by the Stem Cell Research Oversight Committee at the University of California, Berkeley. Human pluripotent stem cells (WTC11 (ref. 50) hiPSCs or WIBR3 (ref. 51) hESCs, National Institutes of Health stem cell registry #0079) were cultured on  $4.1 \times 10^5 \text{ cm}^{-2}$  irradiated mouse embryonic fibroblasts (MEFs) in hPSC medium (Dulbecco's modified Eagle medium/Nutrient Mixture F-12 (DMEM/F12), 20% KnockOut Serum Replacement, 1× Non-Essential Amino Acids (NEAA), 1 mM glutamine, 1× penicillin/streptomycin, 0.1 mM  $\beta$ -mercaptoethanol and 4 ng  $\text{ml}^{-1}$  heat-stable basic fibroblast growth factor). The culture medium was changed daily, and cells were passaged with 1 mg  $\text{ml}^{-1}$  collagenase IV every 5–7 days. The day before and after passage, the medium was supplemented with 10  $\mu\text{M}$  Y27632 (CD0141, Chemdea) to increase cell survival.

### Cas9/sgRNA RNP assembly

Chemically modified sgRNAs were purchased from Synthego. Cas9 protein was bought from the QB3 MacroLab, UC Berkeley. To assemble RNP, 300 pmol sgRNA and 80 pmol Cas9 were mixed in nuclease-free water to a final volume 10  $\mu\text{l}$ , then incubated at room temperature (RT) for 5–10 min before nucleofection.

### Nucleofection of hPSCs and human fibroblasts

hPSCs cultured on MEF were detached from feeder cells by treating with 1 mg  $\text{ml}^{-1}$  collagenase IV and 0.5 U  $\text{ml}^{-1}$  dispase for 25–30 min. Detached colonies were washed with DMEM/F12 once and then dissociated to single cells following incubation with 1× acutase (SCRO05, MilliporeSigma) for 5–7 min at 37 °C. Dissociated cells were washed with DMEM/F12 followed by a wash with phosphate-buffered saline (PBS). Next, 0.5 million hPSCs were pelleted then resuspended in 20  $\mu\text{l}$  Lonza P3 primary nucleofection reagent, mixed with pre-assembled Cas9/sgRNA RNP with or without 100 pmol ssODN HDR donor, and nucleofected using program CA137 on Lonza nucleofactor 4D. For nucleofection of hPSCs-derived fibroblasts, cells were

dissociated with 0.25% trypsin–EDTA and then nucleofected with the same protocol.

### Generation of loHAPs

To generate loHAPs, we optimized previously reported genome editing strategies introducing large deletions<sup>52–55</sup>. Guide RNAs with specificity scores<sup>56</sup> >70 were selected from a region 50–200 bp downstream of the annotated transcriptional end site of the gene of interest and the neighbour gene using the CRISPR Targets track<sup>57</sup> in UCSC genome browser (Supplementary Table 3). The distance between the CRISPR cut site and transcriptional end site was empirically determined by weighing three factors: (1) avoiding indels that would disrupt gene functions on the remaining allele, (2) potential off-target effect when using guide RNAs with lower specificity scores and (3) preserving upstream *cis*-regulatory elements enabling future functional annotation of non-coding variants. As described above, a total of 300 pmol sgRNAs targeting the intended deletion region were delivered to hPSCs as RNPs by nucleofection. After nucleofection, cells were directly seeded onto four MEF 96-well plates at seeding densities of 10, 30, 100 and 300 cells per well, to compensate for batch-to-batch cell survival variability. The medium was changed at days 4, 7, 10, 12 and 13, and 10  $\mu\text{M}$  Y27632 was supplemented at day 13. At day 14, cells were washed with PBS once and then treated with 40  $\mu\text{l}$  0.25% trypsin–EDTA for 5 min at 37 °C, and then 60  $\mu\text{l}$  foetal bovine serum (FBS)/hPSC medium (replacing 10% KnockOut Serum Replacement with 10% FBS in regular hPSC medium) supplemented with 10  $\mu\text{M}$  Y27632 was added to each well to inactivate trypsin. Cells were then gently triturated, and 50  $\mu\text{l}$  cell suspension was transferred to a 96-well PCR plate pre-loaded with 50  $\mu\text{l}$  2× lysis buffer (100 mM KCl, 4 mM  $\text{MgCl}_2$ , 0.9% NP-40, 0.9% Tween-20 and 500  $\mu\text{g ml}^{-1}$  proteinase K, in 20 mM Tris pH 8) for DNA extraction. The remaining 50  $\mu\text{l}$  of cells was reseeded to a new MEF 96-well plate pre-loaded with 100  $\mu\text{l}$  10  $\mu\text{M}$  Y27632 supplemented FBS/hPSC medium and cultured for another 7 days with hPSC medium changed daily. Meanwhile, the cell lysate in the 96-well plates was incubated at 50 °C overnight and then heated to 95 °C for 10 min to inactivate the proteinase K. Cell clones were genotyped by detecting PCR amplicons spanning the deletion junctions using primer pairs flanking the deletion region and usually 100 bp away from the CRISPR cut site to capture potentially extended deletion might be caused by local homology. Amplicons were resolved by agarose gel electrophoresis, and Sanger sequenced or sequenced by NGS. Additionally, an SNP located within the deletion region in the hPSC cell line was identified from whole genome sequencing data. Using primers containing NGS barcode attachment sites (GCTCTCCGATCT), the SNP region was amplified from 2  $\mu\text{l}$  cell lysis from each well with Titan DNA polymerase. Amplicons were then purified using an automated SPRI beads purification protocol at the UC Berkeley DNA Sequencing Facility, i5/i7 barcoded, pooled and sequenced on 150PE iSeq or 300PE MiSeq in the NGS core facility at the Innovative Genomics Institute. NGS data were analysed using the CRISPResso2 (ref. 58) software to identify wells containing only one allele at the SNP. Cells in those wells identified to contain loHAPs were then subcloned by low-density seeding, manual picking and re-genotyped to establish clonal loHAP cell lines.

### Pluripotency marker staining

For immunofluorescent staining, cells were fixed with 4% paraformaldehyde in PBS for 20 min, permeabilized with 0.1% Triton X-100 in PBS for 30 min at RT, blocked in 3% bovine serum albumin in PBS for 1 h at RT and then incubated with primary antibody (PCRP-POU5F1-1D2, MC-813-70 (SSEA-4), Developmental Studies Hybridoma Bank) overnight at 4 °C, and secondary antibody (Thermo Fisher A11001) for 2 h at RT. For alkaline phosphatase staining, cells were fixed with cold 4% paraformaldehyde in PBS for 10 min, equilibrated with pH 9.5 100 mM Tris buffer for 10 min at RT, then incubated with NBT/BCIP (SK-5400, Vector Laboratories) at RT for 2 h. Images were acquired on a Zeiss Axio Observer A1 inverted fluorescence microscope.

### Human fibroblast differentiation from hPSCs

hPSC colonies were collected with  $1 \text{ mg ml}^{-1}$  collagenase IV, washed three times with 5% foetal calf serum in DMEM by gravitational settling, then cultured in suspension to form embryonic bodies in KSR medium (hPSC medium without human fibroblastic growth factor) in ultralow-attachment six-well plates. On day 4, the medium was changed to fibroblast medium (15% FBS in DMEM,  $1 \times$  NEAA,  $1 \times$  penicillin/streptomycin and  $1 \times$  glutamine). At day 7, embryonic bodies were collected and seeded onto 0.2% gelatin-coated 10 cm dishes in fibroblast medium and medium was changed weekly until fibroblast morphology was established. Human fibroblasts were then passaged with 0.25% trypsin–EDTA every week with 1:3 splitting ratio.

### CRISPR-mediated mutagenesis

To introduce indels, 300 pmol sgRNA and 80 pmol purified Cas9 protein were delivered as RNP into hPSCs or hPSC-derived fibroblasts by nucleofection as described above. Targeted designer mutations were introduced with a 150–200 nt ssODN HDR template centred at the CRISPR cleavage site. HDR templates were designed with synonymous CRISPR–Cas9-blocking mutations to increase integration efficiency; when synonymous PAM-ablating mutations disrupted annotated regulatory sequences (for example, splicing sites) or were not feasible, sgRNA-blocking mutations were introduced proximal to the cut site. HDR templates contained a  $\geq 2$  nucleotide change relative to the WT sequence to distinguish designer mutations from sequencing errors. When introducing multiple variants (for example, alanine scanning) within a 30–40 nt window around the sgRNA cut site, ssODN templates were purchased from Integrated DNA Technologies (IDT) and nucleofected in the same HDR pool. In either case, 100 pmol of ssODN ultramer or oligo pool (customized oPools, IDT) were co-delivered with pre-assembled Cas9–RNPs (Supplementary Table 3) into 1 million cells. Cells were seeded back to multiple MEF wells as duplicates or triplicates after nucleofection and passaged with 0.25% trypsin–EDTA once a week as indicated (usually at days 7, 14 and 21 post-nucleofection) while a portion of cells was collected for genomic DNA during each passage. This experimental setting yielded  $\sim 25 \times$  coverage for a variant with a frequency 0.05% with a 5% post-nucleofection survival rate. With 5% empirical HDR efficiency, it allowed the introduction of 25–50 designer mutations in one single nucleofection. To determine allele frequencies in each cell pool, an approximately 250 bp region encompassing the CRISPR targeting site was amplified from 50–100 ng genomic DNA or cell lysis equivalent to  $\sim 10,000$ – $20,000$  genomes with primers containing NGS barcode attachment sites using Titan DNA polymerase or PrimeStar GXL DNA polymerase (R050B Takara). Amplicons were purified, barcoded and NGS sequenced, including an unedited control as described above. Alleles were called and quantified by CRISPResso2. For designer mutations that only differ by one nucleotide from WT, the baseline frequency caused by sequencing errors was measured relative to the unedited control.

### Generation of lentiviral sgRNA library

Oligonucleotides to clone lentiviral sgRNA expression vectors targeting exons 2 and 27 were purchased as oligo pools (IDT) cloned into BstXI/Blp1 (R0113S and R0585, NEB) double digested pJR104 (a gift from Jonathan Weissman) using NEBuilder HiFi DNA Assembly kit (E2627, NEB). Library complexity and quality were confirmed using NGS (each sgRNA had a coverage  $> 100 \times$ ). In total, 10  $\mu\text{g}$  sgRNA library plasmids were transfected into  $\sim 8$  million HEK293T cells together with 7.5  $\mu\text{g}$  psPAX2 (#12260, Addgene) and 2.5  $\mu\text{g}$  pMD2.G (#12259, Addgene) using Lipofectamine 2000 (11668019, ThermoFisher). Lentivirus was collected daily, filtered through a 0.45  $\mu\text{m}$  filter, and frozen at  $-80^\circ\text{C}$ .

### Lentivirus-based mutagenesis

Human PSCs were detached from MEF by collagenase IV and dissociated to single cells by accutase as described above. One million

cells were infected in suspension for 6 h with sgRNA library lentivirus at a multiplicity of infection of 0.15. Cells were then seeded onto puromycin-resistant MEF, selected by  $0.8 \mu\text{g ml}^{-1}$  puromycin 48 h post-infection for one day, and expanded. A total of 100 pmol purified Cas9 protein was then delivered into 0.5 million infected cells by nucleofection, and cells were collected on day 7.

### Dose–response curve

For irradiation,  $5 \times 10^4$  hPSCs were irradiated in suspension at 0, 0.2, 0.5, 1, 2 and 5 Gy on Precision X-RAD 320 irradiator, then seeded back on MEF and cultured for 1 week. For PARP inhibitors,  $5 \times 10^4$  hPSCs were seeded on MEF and treated with 0, 0.5, 1, 2 and 5  $\mu\text{M}$  niraparib (HY-10619, MedChem Express) or Olaparib (O-9201, LC Laboratories) for a week. Relative cell numbers in each condition were quantified by confluency using phase contrast images in ImageJ.

### Drug sensitivity screening of cells with BRCA2 mutations

CRISPR-mutagenized cell pools were cultured for 3–4 weeks after mutagenesis to allow for the stabilization of the alleles present in the cell pools. Cells were then collected with 0.25% trypsin–EDTA as single cells. In triplicates,  $\sim 5 \times 10^4$  cells were either irradiated in suspension with the indicated dose of gamma irradiation and seeded back onto MEF or seeded and then exposed to PARP inhibitor (niraparib or olaparib) using the indicated concentrations. Untreated samples with the same number of cells were seeded in parallel as mock controls. Cells were cultured for 1 week before genomic DNA was isolated. The mutagenized genomic regions in *BRCA2* were then amplified by PCR, NGS sequenced and analysed as described above. For drug sensitivity screening in fibroblasts, cells derived from hPSCs on days 30–50 post-differentiation were treated with PARPi for 10 days.

### RNP-based CRISPR-mediated mutation tiling

Each sgRNA was individually delivered into 1 million hPSCs as RNP, as described above. After nucleofection, cells were combined into two pools by the position of sgRNA they received (6 sgRNAs in exon 2, 22 sgRNA in exon 27), then each cell pool was seeded onto three 15 cm MEF dishes as biological triplicates. At least 0.3 million cells were replated for expansion or collected as genomic DNA every 7 days for 3 weeks. Drug sensitivity screening was started on day 21 and performed as described above, except that at least 0.3 million cells were used in each replicate, and cells were collected on day 28. At least 2  $\mu\text{g}$  phenol–chloroform extracted genomic DNA from each sample was used as a template in PCR amplification of the mutagenized region, then samples were passed through the NGS pipeline as described above and sequenced at depth  $> 1$  million reads per sample on PE250 NovaSeq or PE150 NextSeq2000. NGS reads were first processed using CRISPResso2 to trim off adapter sequences, aligned to a provided reference sequence, and filtered by quality. After this, a custom Python script was used for further analysis. First, observed DNA deletions were categorized by proximity to the target cut site to separate out sequencing artefacts from rare gene editing events. If any part of an observed mutation was within two base pairs of a predicted CRISPR/Cas9 cut site, the allele progressed further into the analysis. If multiple mutations around cut sites were observed, the allele was filtered out, as the experiments outlined above do not permit the interpretation of complex editing events. If there were multiple technical replicates, defined here as independent PCRs performed from the same sample of genomic DNA, those read counts are combined into one file. Next, alleles were categorized into five groups. Alleles with mutagenized start codons are classified as ‘start codon deletions’, and alleles mutagenized at any nucleotide within  $\pm 3$  bp around annotated splicing junctions were classified as ‘splicing variants’. The remaining alleles were further classified by their indel sizes. Alleles with indel sizes that were not multiples of 3 were classified as ‘frame-shift’ and alleles with indel size that were multiple of 3 were classified as ‘in-frame’, unless a novel stop

codon formed at the junctions or within insertions, in which case they were classified as 'de novo stop'. Then, alleles were bioinformatically translated on the basis of where the native splice site junctions are. Read counts from alleles with identical translations were combined to simplify downstream analysis. After this, only alleles that appeared in all mock or week 1 samples and had at least 20 reads in at least one sample were used for downstream analysis. We next normalized allele read counts from each file to the same allele in its respective biological replicate in mock or week 1. The in-frame deletions data were used to estimate the fitness of individual amino acids using a statistical model (Supplementary Methods).

### Data analysis and statistics

Hypothesis tests and multiple comparison correction were done as indicated using R 4.0. Bar graphs and box plots were drawn in GraphPad Prism 9. Heatmaps were generated using Morpheus (Broad Institute) or ggplot2 in R 4.0. Error bars indicate the standard error of mean (s.e.m.) unless otherwise specified.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The main data supporting the results in this study are available within the paper and its Supplementary Information. The raw sequencing data of the mutation-tiling experiments are available in the Gene Expression Omnibus repository under accession number [GSE233683](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE233683). Source data are provided with this paper.

### Code availability

The custom Python code used in the mutation tiling experiments is provided as Supplementary Information.

### References

- King, M. C., Marks, J. H. & Mandell, J. B. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. *Science* **302**, 643–646 (2003).
- Bryant, H. E. et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913–917 (2005).
- Lord, C. J. & Ashworth, A. PARP inhibitors: synthetic lethality in the clinic. *Science* **355**, 1152–1158 (2017).
- Urnov, F. D. et al. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).
- Hockemeyer, D. et al. Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27**, 851–857 (2009).
- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
- Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
- Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
- Kleinstiver, B. P. et al. Engineered CRISPR–Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
- Findlay, G. M. et al. Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **562**, 217–222 (2018).
- Hanna, R. E. et al. Massively parallel assessment of human variants with base editor screens. *Cell* **184**, 1064–1080.e20 (2021).
- Erwood, S. et al. Saturation variant interpretation using CRISPR prime editing. *Nat. Biotechnol.* **40**, 885–895 (2022).
- Brnich, S. E. et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. *Genome Med.* **12**, 3 (2019).
- Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
- Gelman, H. et al. Recommendations for the collection and use of multiplexed functional data for clinical variant interpretation. *Genome Med.* **11**, 85 (2019).
- Smith, S. A., Easton, D. F., Evans, D. G. & Ponder, B. A. Allele losses in the region 17q12-21 in familial breast and ovarian cancer involve the wild-type chromosome. *Nat. Genet.* **2**, 128–131 (1992).
- Futreal, P. A. et al. BRCA1 mutations in primary breast and ovarian carcinomas. *Science* **266**, 120–122 (1994).
- Carette, J. E. et al. Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326**, 1231–1235 (2009).
- Sagi, I. et al. Derivation and differentiation of haploid human embryonic stem cells. *Nature* **532**, 107–111 (2016).
- Yilmaz, A., Peretz, M., Aharoni, A., Sagi, I. & Benvenisty, N. Defining essential genes for human pluripotent stem cells by CRISPR–Cas9 screening in haploid cells. *Nat. Cell Biol.* **20**, 610–619 (2018).
- Bartha, I., di Iulio, J., Venter, J. C. & Telenti, A. Human gene essentiality. *Nat. Rev. Genet.* **19**, 51–62 (2018).
- Li, H. et al. Highly efficient generation of isogenic pluripotent stem cell models using prime editing. *eLife* **11**, e79208 (2022).
- Esashi, F. et al. CDK-dependent phosphorylation of BRCA2 as a regulatory mechanism for recombinational repair. *Nature* **434**, 598–604 (2005).
- Roy, R., Chun, J. & Powell, S. N. BRCA1 and BRCA2: different roles in a common pathway of genome protection. *Nat. Rev. Cancer* **12**, 68–78 (2011).
- Yang, H. et al. BRCA2 function in DNA binding and recombination from a BRCA2–DSS1–ssDNA structure. *Science* **297**, 1837–1848 (2002).
- Alsop, K. 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.* **30**, 2654–2663 (2012).
- Hucl, T. et al. A syngeneic variance library for functional annotation of human variation: application to BRCA2. *Cancer Res.* **68**, 5023–5030 (2008).
- Kuznetsov, S. G., Liu, P. & Sharan, S. K. Mouse embryonic stem cell-based functional assay to evaluate mutations in BRCA2. *Nat. Med.* **14**, 875–881 (2008).
- Naseem, H. et al. Inherited association of breast and colorectal cancer: limited role of CHEK2 compared with high-penetrance genes. *Clin. Genet.* **70**, 388–395 (2006).
- Higgs, J. E. et al. The BRCA2 polymorphic stop codon: stuff or nonsense? *J. Med. Genet.* **52**, 642–645 (2015).
- Mazoyer, S. et al. A polymorphic stop codon in BRCA2. *Nat. Genet.* **14**, 253–254 (1996).
- Baughan, S. & Tainsky, M. A. K3326X and other C-terminal BRCA2 variants implicated in hereditary cancer syndromes: a review. *Cancers* **13**, 447 (2021).
- Landrum, M. J. et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* **46**, D1062–d1067 (2018).

36. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
37. Venkitaraman, A. R. Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell* **108**, 171–182 (2002).
38. Rebbeck, T. R. et al. Association of type and location of BRCA1 and BRCA2 mutations with risk of breast and ovarian cancer. *JAMA* **313**, 1347–1361 (2015).
39. Kass, E. M., Lim, P. X., Helgadottir, H. R., Moynahan, M. E. & Jasin, M. Robust homology-directed repair within mouse mammary tissue is not specifically affected by Brca2 mutation. *Nat. Commun.* **7**, 13241 (2016).
40. Morimatsu, M., Donoho, G. & Hasty, P. Cells deleted for Brca2 COOH terminus exhibit hypersensitivity to gamma-radiation and premature senescence. *Cancer Res.* **58**, 3441–3447 (1998).
41. McAllister, K. A. et al. Cancer susceptibility of mice with a homozygous deletion in the COOH-terminal domain of the Brca2 gene. *Cancer Res.* **62**, 990–994 (2002).
42. Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J. A. & Jasin, M. Gene conversion tracts from double-strand break repair in mammalian cells. *Mol. Cell. Biol.* **18**, 93–101 (1998).
43. Xia, B. et al. Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Mol. Cell* **22**, 719–729 (2006).
44. Shimelis, H. et al. BRCA2 hypomorphic missense variants confer moderate risks of breast cancer. *Cancer Res.* **77**, 2789–2799 (2017).
45. Pierce, A. J., Johnson, R. D., Thompson, L. H. & Jasin, M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev.* **13**, 2633–2638 (1999).
46. Shi, J. et al. Discovery of cancer drug targets by CRISPR–Cas9 screening of protein domains. *Nat. Biotechnol.* **33**, 661–667 (2015).
47. Sharan, S. K. et al. Embryonic lethality and radiation hypersensitivity mediated by Rad51 in mice lacking Brca2. *Nature* **386**, 804–810 (1997).
48. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell* **163**, 759–771 (2015).
49. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR–Cas9 variants. *Science* **368**, 290–296 (2020).
50. Kreitzer, F. R. et al. A robust method to derive functional neural crest cells from human pluripotent stem cells. *Am. J. Stem Cells* **2**, 119–131 (2013).
51. Lengner, C. J. et al. Derivation of pre-X inactivation human embryonic stem cells under physiological oxygen concentrations. *Cell* **141**, 872–883 (2010).
52. Canver, M. C. et al. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J. Biol. Chem.* **289**, 21312–21324 (2014).
53. Liu, Z. et al. Efficient CRISPR/Cas9-mediated versatile, predictable, and donor-free gene knockout in human pluripotent stem cells. *Stem Cell Rep.* **7**, 496–507 (2016).
54. Zheng, Q. et al. Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. *BioTechniques* **57**, 115–124 (2014).
55. Zhu, S. et al. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR–Cas9 library. *Nat. Biotechnol.* **34**, 1279–1286 (2016).
56. Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
57. Haeussler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* **17**, 148 (2016).
58. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).

## Acknowledgements

We thank all members of the Hockemeyer lab for advice and critical comments on the paper, F. Urnov for initial discussions on this project and comments on the paper and B. Cheung and A. Murdock for helping with graphic arts. D.H. and H.L. disclose support for the research described in this study from Innovative Genomics Institute, Siebel Stem Cell Institute. D.H. discloses support for the research described in this study from Chen Zuckerberg Biohub, Pew Charitable Trusts, the Alexander and Margaret Stewart Trust, American Cancer Society (133396-RSG-19-029-01-DMC), National Institutes of Health (R01GM138634-02 and R01HL131744-05) and U.S. Department of Defense (W81XWH-19-1-0586).

## Author contributions

H.L. and D.H. conceptualized the project, designed all experiments and wrote the paper. H.L., R.B., A.H., Y.V. and A.R.N. performed the experiments and analysed the data. N.K. and G.R.P. performed NGS. L.Z. and R.N. developed the statistical models used in this study.

## Competing interests

A provisional patent application, filed in the United States by University of California, has D.H., H.L. and R.B. as inventors. The other authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s41551-023-01065-7>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41551-023-01065-7>.

**Correspondence and requests for materials** should be addressed to Dirk Hockemeyer.

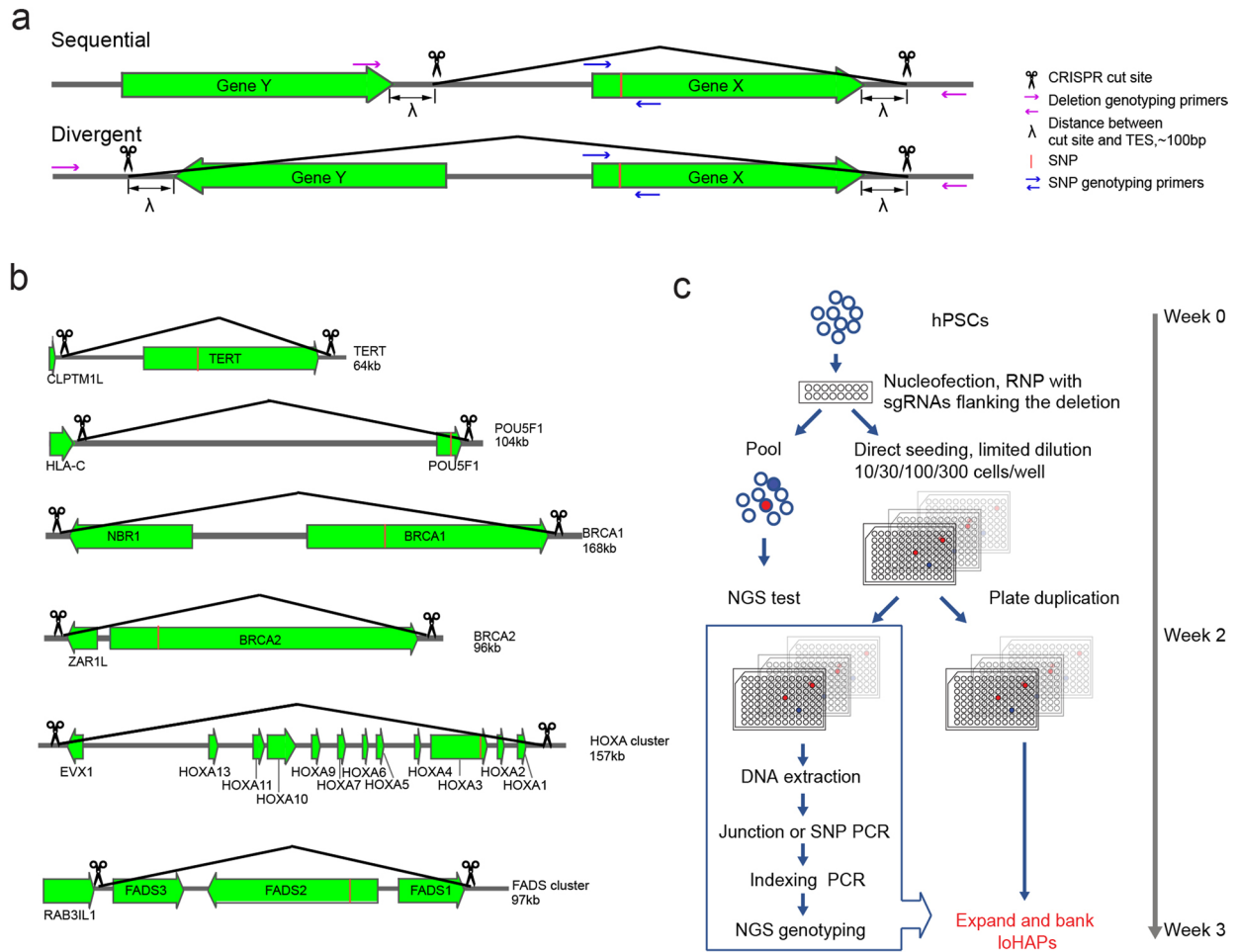
**Peer review information** *Nature Biomedical Engineering* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

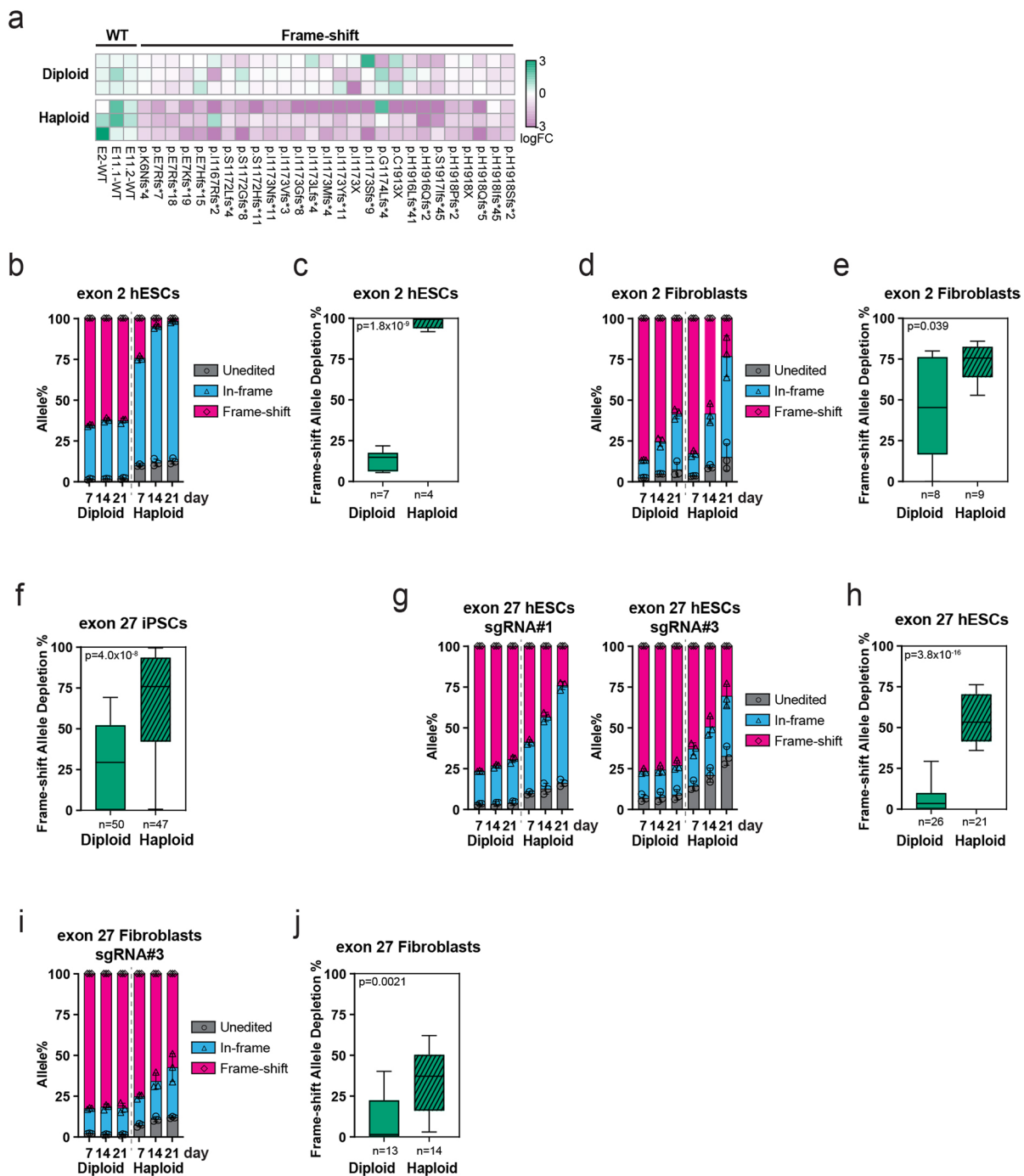
**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

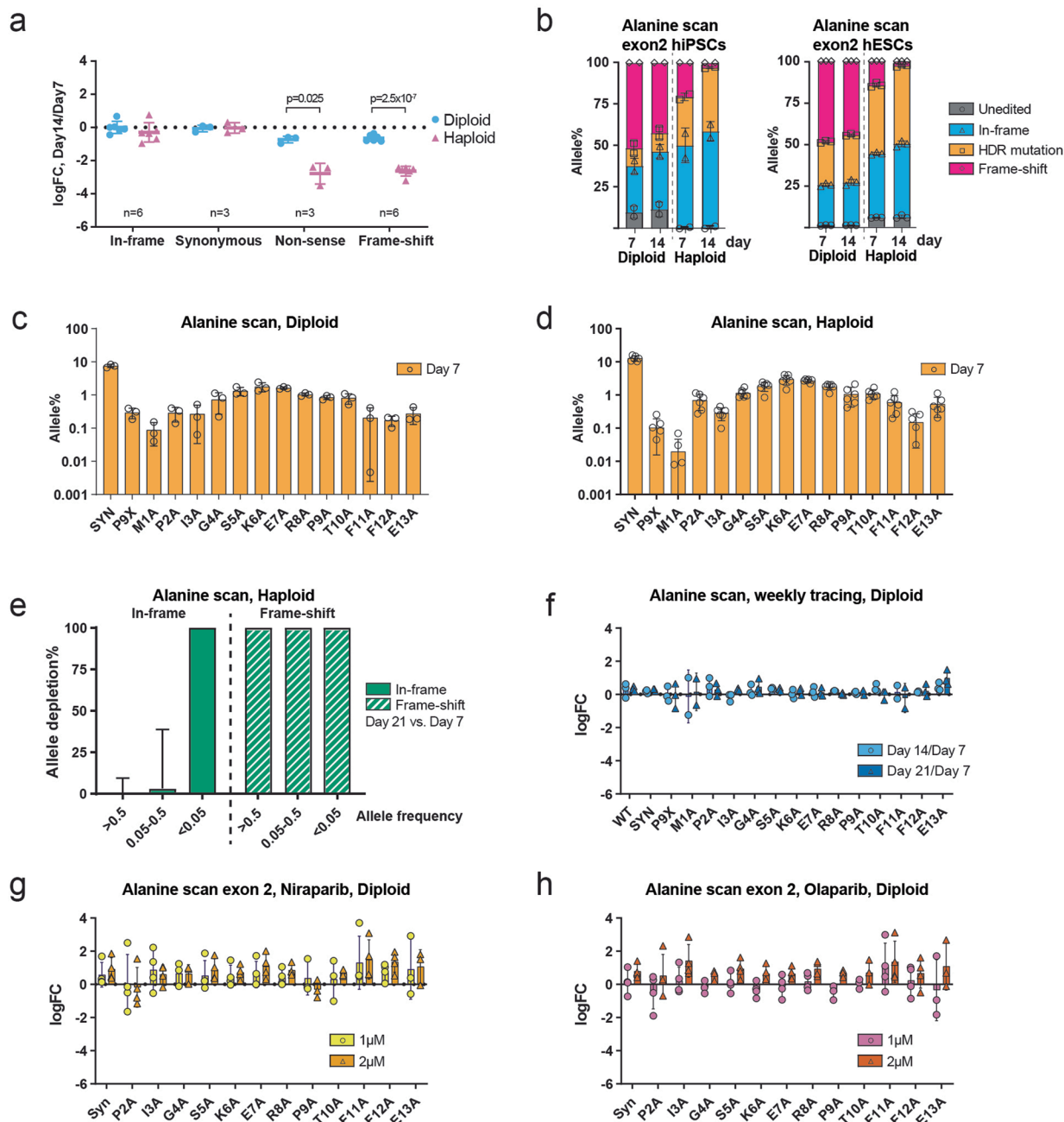


**Extended Data Fig. 1 | Design and gene-editing pipeline of loHAPs. a**, General design of loHAPs, with key parameters. **b**, Specific design of loHAPs for each gene of interest, with deletion size specified. **c**, A scalable high-throughput gene editing pipeline combining limit dilution into 96-well right after nucleofection and multiplex NGS genotyping.



**Extended Data Fig. 2 | Annotating BRCA2 protein domain using CRISPR-mediated mutagenesis in loHAP hiPSCs, hESCs and hPSCs-derived fibroblasts. a**, Heatmap indicating the allele frequency changes,  $\log_2$  (day 14/day 7), of individual frame-shift mutation in 3 biological replicates in hiPSCs. **b**, Quantification of categorized allele frequencies (unedited, in-frame and frame-shift indels) for an sgRNA targeting exon 2 assayed in diploid and *BRCA2* loHAP hESCs at day 7, 14 and 21 after editing. Bars showing mean of 3 biological replicates and error bars showing SEM. **c**, Quantification of the relative depletion of frame-shift mutation in exon 2 comparing samples collected at day

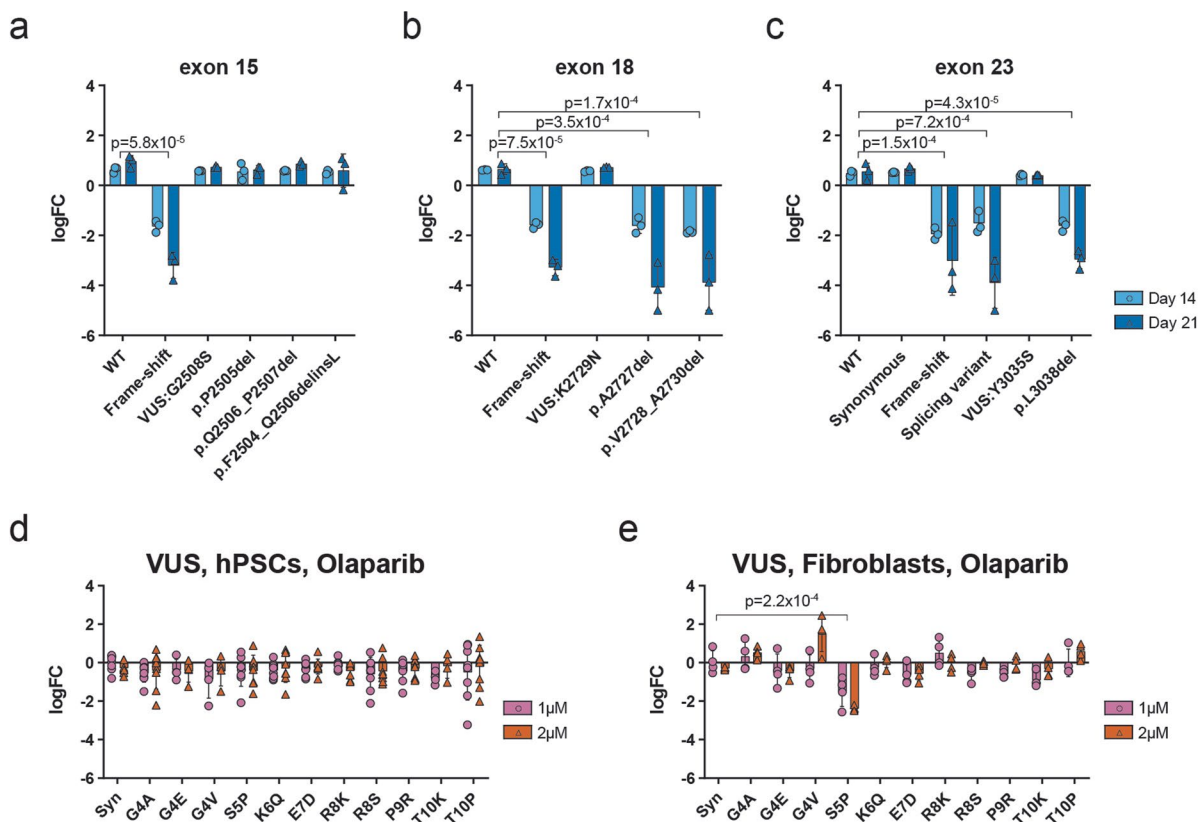
7 and 21 after editing in diploid and *BRCA2* loHAP hESCs. Boxes showing quartiles and whiskers showing the 10th and 90th percentile. Mutations with frequency  $>0.5\%$  are plotted. **p**, two-tailed Student's t-test; **n**, number of alleles. **d**, Same as in **b**, for a sgRNA targeting exon 2 in hPSCs derived fibroblasts. **e**, Same as in **c**, for alleles generated in exon 2 in hPSCs derived fibroblasts. **f**, Same as in **c**, for alleles generated in exon 27 in hiPSCs. **g**, Same as in **b**, for sgRNAs targeting exon 27 in hESCs. **h**, Same as in **c**, for alleles generated in exon 27 in hESCs. **i**, Same as in **b**, for a sgRNA targeting exon 27 in hPSCs derived fibroblasts. **j**, Same as in **c**, for alleles generated in exon 27 in hPSCs derived fibroblasts.



### Extended Data Fig. 3 | Alanine scanning in *BRCA2* loHAP hESCs and hiPSCs.

**a**, Normalized allele frequency changes in categories (in-frame indels, synonymous, non-sense and frame-shift mutations) in exon 11 assayed in diploid and *BRCA2* loHAP hiPSCs, represent as  $\log_2$  (day 21/day 7). P values report the significant comparisons between diploid and *BRCA2* loHAPs in each category, in two-tailed Student's t-test. Centre showing mean and error bars showing SEM. n represents biological replicates. **b**, Quantification of categorized allele frequencies (unedited, in-frame indels, HDR alanine substitution and frame-shift mutations) in exon 2 assayed in diploid and *BRCA2* loHAP hiPSCs and hESCs at day 7 and 14 after editing. Bars showing mean of 2 or 3 biological replicates and error bars showing SEM. **c**, Allele frequencies of designer mutations in *BRCA2* exon 2 alanine scan in diploid hESCs at day 7. Bars showing mean of 3 biological replicates and error bars showing SEM. **d**, Same as in **c**, for *BRCA2* loHAP hESCs. **e**, Quantification of the relative depletion of in-frame and frame-shift mutations

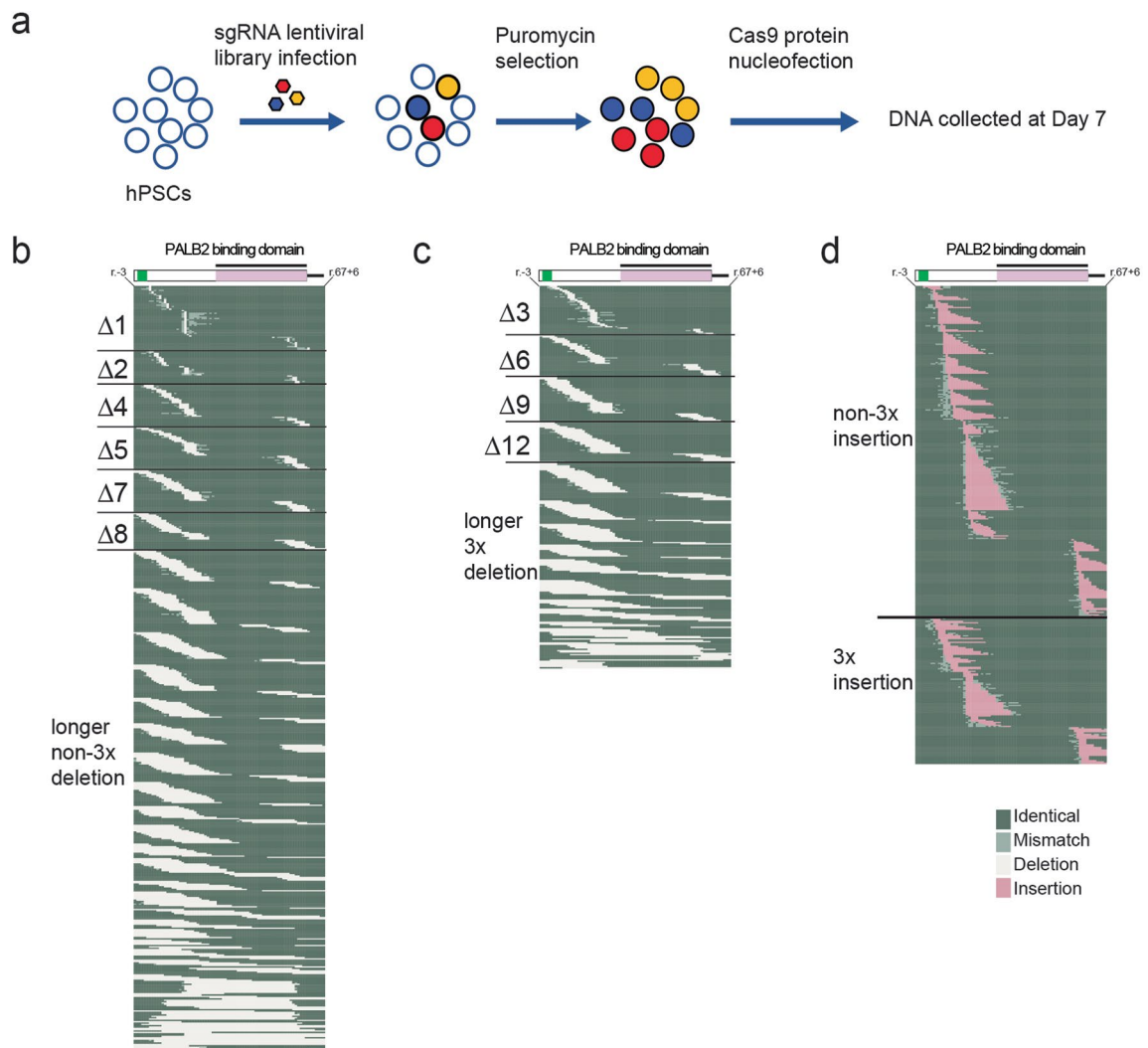
in different frequency groups in *BRCA2* loHAPs. Cells were edited in exon 2 performing alanine scanning as described in Fig. 3d–g and samples collected at day 7 and 21 after editing in *BRCA2* loHAP hESCs. Frame-shift and in-frame alleles were binned into three groups based on their allele frequency (>0.5%, 0.05–0.5% and <0.05%) and the bars showing median and whiskers showing interquartile range. **f**, Allele frequency changes overtime of mutations generated in *BRCA2* exon 2 alanine scan in diploid hESCs, represented as  $\log_2$ (day 14/day 7), circle and  $\log_2$  (day 21/day 7), triangle. Bars showing mean and error bars showing SEM. Data points represent biological replicates. **g**, Normalized allele frequency changes upon treatment of PARP inhibitor, Niraparib, of *BRCA2* exon 2 alanine mutations in diploid hESCs, represented as  $\log_2$ (drug/mock). 1  $\mu$ M, circle; 2  $\mu$ M, triangle. Allele frequencies were normalized against corresponding WT. Bars showing mean and error bars showing SEM. Data points represent biological replicates. **h**, Same as in **g**, for Olaparib treatment.



**Extended Data Fig. 4 | Allele frequency changes overtime of mutations generated around G2508, K2729 and Y3035 and drug sensitivity of VUS in hPSCs and fibroblasts. a,** Allele frequency changes overtime of mutations generated around G2508 in *BRCA2* loHAP hPSCs. Represented as  $\log_2(\text{day 14}/\text{day 7})$ , circle and  $\log_2(\text{day 21}/\text{day 7})$ , triangle, p, FDR-corrected two-tailed student's t-test vs. WT. Bars showing mean of 3 biological replicates and error bars showing SEM. **b,** Same as **a**, for K2729. **c,** Same as **a**, for Y3035. **d,** Normalized allele

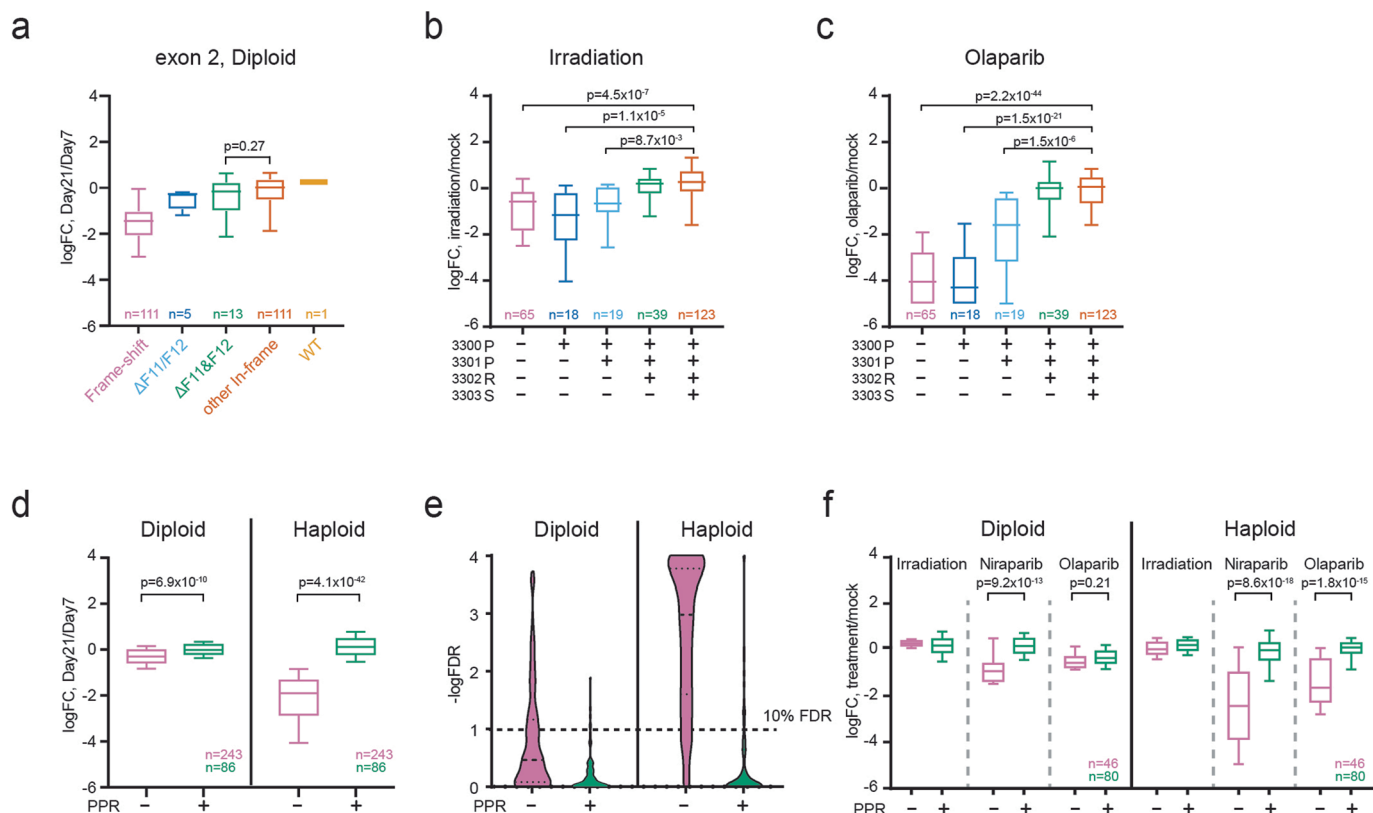
frequency changes upon treatment of PARP inhibitor, Olaparib, of *BRCA2* exon 2 VUS in *BRCA2* loHAP hPSCs, represented as  $\log_2(\text{drug}/\text{mock})$ . 1  $\mu\text{M}$ , circle; 2  $\mu\text{M}$ , triangle. Allele frequencies were normalized against WT. Alleles with frequency  $>0.5\%$  in both hPSCs and fibroblasts are plotted. Bars showing mean and error bars showing SEM. Data points represent biological replicates. **e,** Same as **d**, in mutated *BRCA2* loHAP hPSCs-derived fibroblasts. p, FDR-corrected two-tailed student's t-test vs. synonymous mutations (Syn).





**Extended Data Fig. 5 | A representative indel profile of lentiviral sgRNA library-based mutation tiling in *BRCA2* exon 2. **a**, The workflow of lentivirus-based mutagenesis. **b**, The aligned deletion profile on nucleotide level of alleles with deletion size of non-multiples of 3 nucleotides, with category labels to the**

left. The identity of each position of each allele are categorized and color coded as identical to WT (dark green), mismatch (light green), deletion (white). **c**, Same as **b**, for deletions with deletion size of multiples of 3 nucleotides. **d**, Same as **b**, for insertions with inserted nucleotides in pink.



**Extended Data Fig. 6 | Comparison between diploid and *BRCA2* loHAPs in identifying functionally critical amino acids by CRISPR-mediated mutation tiling.**

**a**, Categorized allele frequency changes of F11, F12 deletions at day 21 comparing to day 7, presented as log<sub>2</sub>(day 21/day 7) in diploid. P value reports the significance between F11&F12 double deletion vs. other in-frame mutations in two-tailed Student's t-test. n shown number of different alleles in each group. Boxes showing quartiles and Whiskers showing the 10th and 90th percentile.

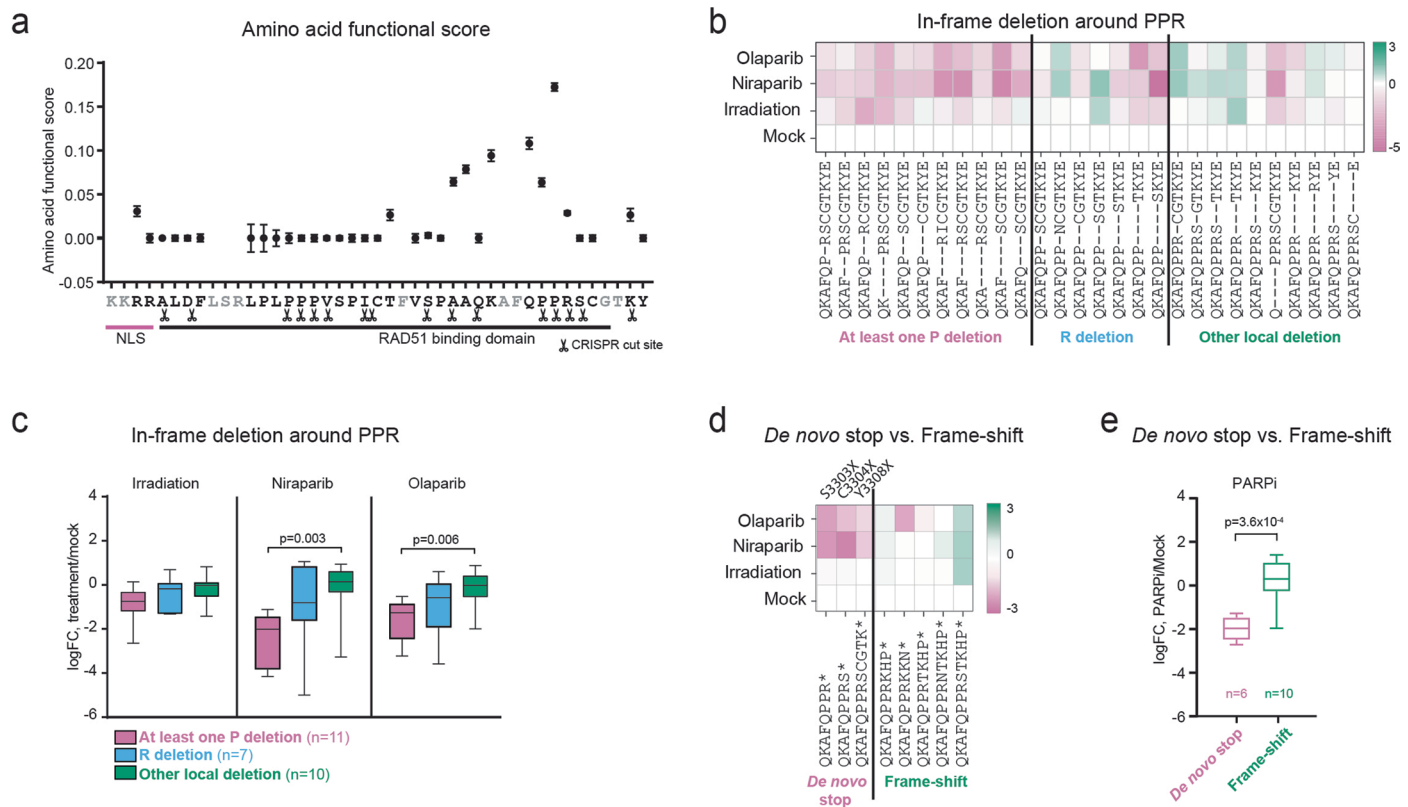
**b**, Categorized allele frequency changes of frame-shift mutations break before P3300, at P3300, at P3301, at R3302 or at later positions, upon irradiation comparing to the mock condition, presented as log<sub>2</sub>(irradiation/mock). Boxes showing quartiles and Whiskers showing the 10th and 90th percentile.

P, FDR-corrected two-tailed student's t-test vs. the PPRS fully preserved group. N, number of different alleles in each group. **c**, Same as **b**, for Olaparib treatment.

**d**, Categorized allele frequency changes of frame-shift mutations with PPR motif preserved (+) or destroyed (-) assayed in diploid and *BRCA2* loHAP hESCs in an independent tiling deletion screen, at day 21 comparing to day 7, presented as log<sub>2</sub>(day 21/day 7). Boxes showing quartiles and Whiskers showing the 10th and 90th percentile. P, two-tailed student's t-test. n, number of different alleles in each group.

**e**, The distribution of FDR-corrected p value of each allele shown in **d**. The p values of alleles that frequencies increased, were assigned to be 1 (-logFDR=0).

**f**, Same as **d**, upon treatment of irradiation, Niraparib or Olaparib comparing to the mock condition, presented as log<sub>2</sub>(treatment/mock).



**Extended Data Fig. 7 | Evaluating the function of critical amino acids by analyzing in-frame deletions generated in CRISPR-mediated mutation tiling.** **a**, Functional score of amino acid in the BRCA2 RAD51 binding domain encoded in exon 27 calculated by a statistical model, with CRISPR cut site labeled. Only scores of amino acids covered by one amino acid deletions are plotted. **b**, The aligned allele profile of  $\leq 4$  amino acids in-frame deletions around PPR (3300–3302) and heatmaps of allele frequency changes upon the treatment of irradiation or PARPi. Alleles were grouped by the amino acid in PPR motif deleted. Allele frequencies of each mutation were normalized to the corresponding allele frequencies in mock. **c**, Categorized allele frequency changes of alleles in **b**. Boxes showing quartiles and Whiskers showing the 10th and 90th percentile. p, FDR-corrected two-tailed student's t-test. n, number of different alleles in each group. **d**, The aligned allele profile of *de novo* stops and frame-shift mutation

that result in stop codons between S3303 and Y3308 and heatmaps of allele frequency changes upon the treatment of irradiation or PARPi as shown in **b**. Note: We identified three *de novo* stop mutations S3303X, C3304X and Y3308X. In each case these *de novo* stops depleted after PARPi suggesting that *de novo* stop codons between 3302 and 3308 have reduced fitness suggesting that the presence of amino acids (even if not generated by the BRCA2 frame) between amino acid 3302 and 3308 contributes to BRCA2 protein function. These findings are consistent with a pathogenic *de novo* stop codon identified at amino acid 3308. **e**, Categorized allele frequency changes of alleles in **d** upon PARPi inhibition. p, two-tailed student's t-test. n, number of measurements in each group. The responses of each allele to Niraparib or Olaparib were considered as separate measurements.

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The main data supporting the results in this study are available within the paper and its Supplementary Information. Source data for the figures are provided with this paper. The raw sequencing data of the mutation-tiling experiments are available in the GEO repository under accession number GSE233683.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

The study did not involve human participants.

Reporting on race, ethnicity, or other socially relevant groupings

—

Population characteristics

—

Recruitment

—

Ethics oversight

—

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

We compared allele frequencies in  $n_1 \geq 6$  null samples to  $n_2 \geq 3$  samples under treatment. The same null samples were used for all treatments. The test employed was aimed at detecting differences in allele frequencies between null and treatment ( $H_0$ : allele frequency in null = allele frequency under treatment). Because the sample sizes (number of reads) were relatively large, we used a t-test as an approximation. Assuming that reads were sampled according to a Poisson distribution, and that the number of reads was much smaller than the number of cells from which DNA was extracted, the variance in the number of reads with a mutation should equal the expected number of reads with that mutation. We used this observation for power calculations. For example, with expected 1,000 reads, the probability of detecting a systematic reduction from an allele frequency of 0.2 to 0.15 of a mutation using a one-sided t-test is 0.999 (calculated using the function `pwr.t2n.test` in R with Cohen's D calculated using the expected Poisson variance). The sample sizes used clearly provide high power to detect allele-frequency differences. In fact, the main reason for this level of replication is for reproducibility and to be able to accommodate increased variance caused by unknown confounders.

Data exclusions

Samples that did not yield PCR amplification and NGS data were excluded.

Replication

Measurements were taken from replicates. Key results were replicated with different sgRNAs and in different cell lines.

Randomization

Cells were equally split into sub-populations of 50,000 cells, which is at 250x coverage of an allele at 0.5% frequency. Those subpopulations were randomly assigned to different treatment groups.

Blinding

The experiments were open-labelled.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials &amp; experimental systems

n/a	Involvement
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

## Methods

n/a	Involvement
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Antibodies

Antibodies used

PCRP-POU5F1-1D2, MC-813-70 (SSEA-4), Developmental Studies Hybridoma Bank; Goat anti-Mouse IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488, A11001 Thermo Fisher.

Validation

PCRP-POU5F1-1D2: this antibody was characterized by the Protein Capture Reagents Program, produced by JHU/CDI, [https://antibodyregistry.org/search.php?q=AB\\_2618968](https://antibodyregistry.org/search.php?q=AB_2618968).

MC-813-70 (SSEA-4): the use of this antibody on immunofluorescence(IF) has been shown in multiple peer reviewed papers, including The EMBO journal 2.12 (1983): 2355-61., Nature 311.5985 (1984 Oct 4-10): 469-72, Scientific reports 6. (2016 Oct 3): 34476., Scientific reports 8.1 (2018 Apr 18): 6168. (information provided by the Developmental Studies Hybridoma Bank)

## Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

WTC11 ( iPSC, Allen institute), WIBR3 (NIH registry #0079)

Authentication

Multiple SNPs matched the published whole-genome sequencing data.

Mycoplasma contamination

The cells were tested for mycoplasma monthly.

Commonly misidentified lines  
(See [ICLAC](#) register)

No commonly misidentified cell lines were used.