ARTICLE

A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution Variants on Protein Function

Lea M. Starita,[1,2,7] Muhtadi M. Islam,[3,7] Tapahsama Banerjee,[3] Aleksandra I. Adamovich,[3] Justin Gullingsrud,[1] Stanley Fields,[1,4,5] Jay Shendure,[1,2,5,*] and Jeffrey D. Parvin[3,6,*]

Loss-of-function pathogenic variants in *BRCA1* confer a predisposition to breast and ovarian cancer. Genetic testing for sequence changes in *BRCA1* frequently reveals a missense variant for which the impact on cancer risk and on the molecular function of BRCA1 is unknown. Functional BRCA1 is required for the homology-directed repair (HDR) of double-strand DNA breaks, a critical activity for maintaining genome integrity and tumor suppression. Here, we describe a multiplex HDR reporter assay for concurrently measuring the effects of hundreds of variants of BRCA1 for their role in DNA repair. Using this assay, we characterized the effects of 1,056 amino acid substitutions in the first 192 residues of BRCA1. Benchmarking these results against variants with known effects on DNA repair function or on cancer predisposition, we demonstrate accurate discrimination of loss-of-function versus benign missense variants. We anticipate that this assay can be used to functionally characterize BRCA1 missense variants at scale, even before the variants are observed in results from genetic testing.

Introduction

Genetic testing for genes associated with hereditary breast and ovarian cancer often reveals a *BRCA1* missense variant whose impact on the molecular function of the encoded protein, and therefore its contribution to cancer risk, is unknown. Such variants are reported as variants of uncertain significance (VUSs). VUSs cause distress for both physicians and individuals with a family history of cancer and can lead to unnecessary surgeries.[1,2] As genetic testing becomes more common in the clinic, reports of missense VUSs are rapidly accumulating in public databases.[3] Additional VUSs arise from the increasingly widespread sequencing of tumor genomes and exomes to guide precision therapy. For rare missense variants, clinically calibrated functional assays are particularly helpful for aiding classification.

The most commonly reported class of VUSs in *BRCA1* is single-nucleotide variants (SNVs), which are predicted to result in missense amino acid substitutions. As of May 2018, 1,794 missense VUSs in *BRCA1* are in the clinical genetics database ClinVar.[3] An additional 231 missense variants have conflicting interpretation reports, suggesting that clinical testing labs apply discordant classification criteria. Across BRCA1's 1,863 amino acids, 12,458 SNVs are predicted to result in a missense substitution that might or might not affect protein function; these are often considered VUSs when identified as either a germline or somatic variant in an individual. Most strategies for variant interpretation fail when missense variants are novel or rare because clinical information for correlation is correspondingly scarce.[4–7] Computational variant-effect prediction algorithms scale without limit, but these are not accurate enough for routine clinical use and furthermore contribute to discordance.[8,9] Functional assays, on the other hand, are considered by the American College of Medical Genetics (ACMG) guidelines as strong evidence for or against the pathogenicity of missense variants.[8] However, performing a post hoc functional assay for each BRCA1 SNV as it is discovered is infeasible at their current rate of accumulation.

BRCA1 is required for the maintenance of genome integrity via the homology-directed repair (HDR) pathway. The effect of variants in BRCA1 on its HDR function can be determined in tissue culture with a GFP-based reporter assay for intact DNA repair function.[10] For BRCA1 variants tested thus far, this assay has stratified the protein function of known benign and pathogenic variants with high sensitivity and specificity.[11–14] Here, we have developed a multiplex version of this HDR assay with the goal of testing all possible protein variants in the BRCA1 N terminus (residues 2–192), which includes the RING domain (residues 7–98). Proper folding of the RING domain is required for the stability and function of the full-length protein.[12,15,16] In addition, all missense substitutions that are known to cause increased cancer risk map to either the RING or BRCT domain.

Using the HDR reporter cell line with integrated *BRCA1* variant libraries, we tested approximately 600 BRCA1 missense variants per experiment. Loss-of-function

[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; [2]Brotman Baty Institute for Precision Medicine, Seattle, WA 98195, USA; [3]Department of Biomedical Informatics, Ohio State University, Columbus, OH 43210, USA; [4]Department of Medicine, University of Washington, Seattle, WA 98195, USA; [5]Howard Hughes Medical Institute, Seattle, WA 98195, USA; [6]Comprehensive Cancer Center, Ohio State University, Columbus, OH 43210, USA
[7]These authors contributed equally to this work
*Correspondence: shendure@u.washington.edu (J.S.), jeffrey.parvin@osumc.edu (J.D.P.)
https://doi.org/10.1016/j.ajhg.2018.07.016.

missense variants were identified by their relative depletion from the subset of GFP^+ cells. BRCA1's first 98 amino acids, which encode the RING domain, were considerably more sensitive to missense substitutions than the subsequent 94 residues. The results of the multiplex assay correlate well with the results from singleton HDR reporter assays and other functional assays. Furthermore, known pathogenic variants can be separated from known benign variants by their recurrent depletion from the functional population in replicate experiments. Our results for 63 VUSs or variants with conflicting reports of pathogenicity from ClinVar show that five are nonfunctional for HDR, whereas 57 are functional. We suggest improvements to our current protocol to potentially increase its throughput and accuracy. Finally, we anticipate that this assay can be used to functionally characterize BRCA1 missense variants at scale in order to provide the additional information necessary for more definitive interpretation of VUSs in the clinic.

Material and Methods

All enzymes were purchased from New England BioLabs unless stated otherwise. Primer sequences can be found in Table S1.

Creating the HeLa-DR-FRT Cell Line

A description of the HeLa-DR (direct repeat) cell line can be found in the report by Ransburgh et al.[12] To create a FLP-in version of HeLa-DR, we stably integrated a flippase recognition target (FRT) sequence into the cells by using the pFRT/lacZeo plasmid (Thermo Fisher Scientific). We tested Zeocin-resistant clones that had a single integration site detected by Southern blot for high-activity integration sites by using the mammalian β-galactosidase activity assay (Gal-Screen, Thermo Fisher Scientific). Clonal expansion of the selected colony established the HeLa-DR-FRT cell line.

Site-Saturation Mutagenesis Libraries and Barcoding

The hemagglutinin (HA)-tagged BRCA1 N-terminal HindIII-EcoRI fragment containing amino acids 1–302 was cloned into the pUC18 plasmid. Two site-saturation mutagenesis libraries of BRCA1 were constructed by a previously reported inverse-PCR-based method.[17] Pool 1 included substitutions in amino acids 2–96, and pool 2 included 97–192. For each codon, 30 base mutagenic primers were ordered with machine-mixed NNK bases at the 5′ end of the sense oligonucleotide (N = ACTG; K = GT). The mutagenized HA-BRCA1 2–302 fragments were ligated into the EcoRI and HindIII sites of the BRCA1 cDNA in the pcDNA5/FRT-TO vector that had been modified to lack a second EcoRI site in the gene for hygromycin resistance and include a second multiple cloning site at the MluI and NruI sites for the barcode. A 16 base degenerate barcode encoded on oligonucleotides (pc5_barcode_longer_ W and _C) that had been annealed, extended, and digested with NotI and SbfI was then ligated into the second multiple cloning site. Assessment of colony-forming units after barcode ligation and transformation detected ~25,000 barcoded clones for each of the pool 1 and pool 2 libraries. Metrics for the variant library cloning steps can be found in Table S4.

Assigning Barcodes to Variants by Using PacBio Long Reads

To prepare the circular SMRTbell templates for the pool 1 and 2 variable regions and barcode, we removed the intervening sequence between the barcode and the BRCA1 N-terminal variable region by using NotI and HindIII restriction digest followed by end repair and blunt-end ligation.[18] The ligations were transformed into *E. coli* to remove concatamers. We then cut the plasmids with SbfI and EcoRI to release the barcode and BRCA1 N-terminal variable region. Custom SMRTbell adapters pb_SbfI and pb_EcoRI were sticky-end ligated to the purified fragment. To make a working stock of 20 μM SMRTbell adaptors in 10 mM Tris, 0.1 mM EDTA, and 100 mM NaCl, we heated them to 85°C and snap cooled them on ice. The ligation reaction contained 0.3 pmol of purified fragment, 0.4 μM of each adaptor, 0.25 μL of EcoRI-HF, 0.25 μL of SbfI-HF, 1× ligase buffer, and 0.5 μL of T4 ligase in a 25 μL reaction. The ligation was performed at room temperature for 30 min and then heat inactivated at 65°C for 20 min. To cleave SMRTbells with the remaining plasmid backbone, we added 0.25 μL of each of XhoI and NdeI and incubated them at 37°C for 15 min. And finally, to digest noncircular DNA, we added 0.5 μL of each of ExoIII (Enzymatics) and ExoVII and incubated them at 37°C for 15 min. The final SMRTbell fragments were purified via AmpurePB (Pacific Biosciences) at 1.8× concentration, washed in 70% ethanol, eluted in 15 μL 10 mM Tris (pH 8), and quantified by BioAnalyzer (Agilent).

Each BRCA1 library was sequenced on four SMRT cells on a Pacific Biosciences RS II sequencer. Barcodes and variable regions were identified as previously described[19] as follows. Base call files were converted from the bax format to the bam format with bax2bam (version 0.0.2), and then bam files for each library from separate lanes were concatenated. Consensus sequences for each sequenced molecule in every library were determined by the Circular Consensus Sequencing (CCS) algorithm (version 2.0.0) with default conditions (ccs and bax2bam can found in the PacBio GitHub repository; see Web Resources). Each resulting consensus sequence was then aligned to a BRCA1 reference sequence with the Burrows-Wheeler Aligner.[17] Barcodes and insert sequences were extracted from each alignment with custom scripts that parsed the CIGAR and MD strings. For barcodes sequenced more than once, if barcode-variant sequences differed, the barcode was assigned to the variant that represented more than 50% of the sequences. Barcodes lacking a majority variant sequence were assigned the variant sequence with the highest average quality score as determined by the CCS algorithm. The barcode-variant extraction and barcode unification scripts can be found in the Shendure lab's GitHub repository (see Web Resources). Pool 1 had 19,809 barcodes assigned to variants with zero or one amino acid substitution encoding 1,602 unique protein variants. Pool 2 had 17,635 barcodes assigned to variants with zero or one amino acid substitution encoding 1,695 unique protein variants. Additional metrics regarding the sequence processing for the barcode-variant assignments can be found in Table S4. For both BRCA1 libraries, a barcode-variant map file was created to contain each barcode and its nucleotide sequence.

Integration of Libraries into Cells

For each of the BRCA1 plasmid libraries (pools 1 and 2), 70–80 million HeLa-DR-FRT cells in ten 10 cm tissue-culture plates were transfected with 200 μg of pOG44 to express a modified flippase enzyme and 100 μg of the pcDNA5/FRT BRCA1 variant

library. Plasmids were diluted in 10 mL of Opti-MEM and incubated for 5 min. 300 μL of Lipofectamine 2000 (Thermo Fisher Scientific) was diluted in 10 mL of Opti-MEM for 5 min. The Lipofectamine and plasmid dilutions were then combined and incubated for 20 min. The mixture was then applied directly to cells. After 24 hr, cells were trypsinized and transferred to ten 15 cm tissue-culture dishes. Because pOG44 flippase has reduced activity at 37°C, 4–8 hr after transfer, the cells were moved from a 37°C humidified incubator to a 30°C humidified incubator for 24 hr. The cells were then returned to 37°C for an additional 24 hr. Approximately 72 hr after the initial transfection, the cells were trypsinized and transferred to 20 15 cm plates containing selection media (50% fresh DMEM supplemented with 10% fetal bovine serum, 50% filter-sterilized conditioned media, and 550 μg/mL hygromycin B). Hygromycin-resistant cells were selected at 37°C for 24 hr. The cells were washed with sterile phosphate-buffered saline (PBS), and the selection media were replaced after the first 24 hr and again every 48 hr until cell colonies were visible without a microscope (about 14 days). Colonies were then counted, trypsinized, resuspended in 20 mL of culture media, and mixed thoroughly (colony counts can be found in Table S4). 15 mL of the resuspended cells was frozen in 1 mL aliquots. 5 mL of the resuspended colony mixture was plated onto three 15 cm plates (3, 1.5, and 0.5 mL) and incubated for 24 hr. The plate that was closest to a confluent monolayer of cells was passaged. We passaged the cells for an additional 2 weeks before performing HDR reporter experiments to assure loss of the unintegrated BRCA1 expression plasmid.

HDR Reporter Assays, Sorting, and gDNA Prep

HDR reporter assays and fluorescence-activated cell sorting (FACS) were performed for each of the BRCA1 variant libraries in quadruplicate. A confluent 10 cm plate of the HeLa BRCA1 variant cell line was trypsinized and resuspended in 10 mL of culture media. 65 μL of the suspension was plated in each of 48 wells across two 24-well tissue-culture plates. 24 hr later, each well was transfected with 30 pmol of small interfering RNA (siRNA) and 1.5 μL of Oligofectamine (Thermo Fisher Scientific). Oligofectamine was diluted with 6 μL of Opti-MEM, and siRNA was diluted with 25 μL of Opti-MEM for 5 min. The dilutions were then combined and incubated for an additional 30 min. The transfection mixture was then applied directly to the cells. 24 hr later, the cells were trypsinized and transferred to four 6-well tissue culture plates and then incubated for another 24 hr. Each well was then transfected with 50 pmol of siRNA, 3 μg of pCBASceI (for I-SceI expression), and 3 μL of Lipofectamine 2000. Plasmid and siRNA were diluted in 125 μL of Opti-MEM, and Lipofectamine was diluted in 125 μL of Opti-MEM and then incubated for 5 min. The dilutions were combined and incubated for an additional 20 min. The transfection mixture was then applied directly to cells. 4–6 hr later, the culture media were replaced with fresh media. siRNA of the BRCA1 3′ UTR, BRCA1 coding sequence, and control were used in both rounds of transfection. After 24 hr, we analyzed one well of cells treated with each condition for GFP expression by using flow cytometry to confirm transfection efficiency. If cells treated with control siRNA and 3′ UTR siRNA fell within 7%–9% and 4%–7% GFP^+ cells, respectively, and cells treated with siRNA of the BRCA1 coding sequence were 1%–2% GFP^+, the experiment would proceed. 72 hr after transfection, the cells were pooled according to treatment and sorted by FACS with an Aria IIu instrument. Cells were resuspended and pooled in filter-sterilized sorting buffer containing 1× PBS (Ca^{2+}- and Mg^{2+}-free), 5 mM EDTA, 25 mM HEPES (pH 7.0), and 1% heat-inactivated fetal bovine serum dialyzed against Ca^{2+} and Mg^{2+} PBS. A minimum of 500,000 GFP^+ cells and a maximum of two million GFP^- cells were collected per pool (Table S5). Genomic DNA (gDNA) was extracted from GFP^+ and GFP^- cells with a DNeasy Blood & Tissue Kit according to the manufacturer's (QIAGEN) instructions. DNAs were eluted in 200 μL of Buffer EB (QIAGEN).

Barcode Amplification and Sequencing

To amplify the barcode from gDNA, we spread the GFP^+ and GFP^- populations over 8–16 reactions each containing 250 μg of gDNA (220,000–440,000 genome equivalents by weight given ~9 pg per triploid HeLa genome; see Table S5). Reactions also contained Kapa2G Robust Polymerase (Kapa Biosystems), a primer that annealed to the SV40 promoter 5′ to the integrated plasmid and a primer adjacent to the barcode (SV40_F and newpc5bc_nexteraR). PCR was performed under the following conditions: 95°C for 5 min; ~28 cycles of 95°C for 40 s, 65°C for 30 s, and 72°C for 3 min; and 72°C for 10 min. The reactions produced a ~3,700 base amplicon specifically from integrated plasmids. The reactions for each sample were combined, and the amplicons were purified by 0.5× AMPure and eluted with 10 mM Tris at 10% of the original reaction volume. 10% of the eluted PCR volume was re-amplified with primers containing sample indexes and Illumina cluster-generating sequences (pc5bc_p5_F, nextIndex). Reactions also contained Kapa2G Robust Polymerase and 0.5× SYBR green II (Thermo Fisher Scientific). PCRs were monitored on a Mini-opticon qPCR machine (Bio-Rad) and removed during exponential amplification under the following conditions: 95°C for 3 min and five to ten cycles of 95°C for 20 s, 65°C for 30 s, and 72°C for 20 s. The PCR produced a 350 base amplicon that was purified according to a double AMPure protocol. First, the large DNA fragments were removed by precipitation by the addition of 0.6× volume AMPure beads to the reactions. Second, the 350 base amplicons were purified from the supernatant with 0.9× volume AMPure beads. The samples were multiplexed, and the barcodes and sample indexes were sequenced (single read, pcDNA5_barcodeSeq_F) on a Nextseq 500 High Output 75 base kit (reads per experiment can be found in Table S5). For the pilot 16-plex experiment without barcodes, the BRCA1 region containing the variants was amplified and directly sequenced.

Variant Scoring, Classifications, and Depletion Score

FASTQ files containing barcodes for each sample and the barcode map for each library were used as input for the software package Enrich2.[20] Enrich2 was used to count the barcodes, associate each barcode with a nucleotide variant, and then translate and count both the unique nucleotide and unique amino acid variants. Barcodes assigned to variants containing insertions or deletions were removed from analysis. We converted the counts for each protein variant to frequencies by dividing them by the total number of variant counts for each sample. The ratio of the frequency of each variant in the GFP^+ population over its frequency in the GFP^- population was calculated. That ratio for each variant was then normalized to the GFP^+/GFP^- ratio for the wild-type (WT) sequence as $\ln(GFP^+/GFP^-) = 0$. Counts and scores for each variant can be found in Table S6. Variants with multiple amino acid substitutions were removed from further analysis.

To construct the binary functional/nonfunctional classifier for variants, we first determined that the standard deviation of the

score decays according to read count in a way that can be modeled by a $\log_{10}$-$\log_{10}$ curve. We then modeled the decay of the standard deviation of scores from the control siRNA experiment. Next, we applied that model to the standard deviation of scores from the BRCA1 siRNA experiment and calculated a p value and a false-discovery-rate-adjusted q value for each variant to determine whether it was similar to or significantly different from that of the control siRNA experiment ($q < 0.055$). Finally, we determined where the false positives in the control experiments occurred along the continuum of read counts to assign a read-count threshold for the BRCA1 siRNA condition. A read-count threshold is usually part of the heuristic applied to remove noise due to stochastic dropout from deep mutational scanning data.[16] We removed variants with a read count below that threshold from further analyses. These analyses were performed with R studio; an R markdown file containing all data manipulations is supplied as Data S1.

The depletion score represents the number of replicates in which a variant was found to be depleted from the BRCA1 siRNA GFP^+ population. Only variants that were present above the read-count threshold in at least three replicates had depletion scores. Depletion scores for those 1,056 variants can be found in Table S7.

Results

A Multiplex Assay for Measuring the Effect of Protein Variants on HDR

The multiplex HDR reporter assay described here is based on an assay developed by the Jasin laboratory.[10] In the reporter assay for single variants of BRCA1, a cell line has two inactive GFP-encoding alleles integrated into its genome. A site-specific double-strand DNA break is induced in one of the inactive GFP alleles by transfection of a plasmid for the expression of the I-SceI endonuclease, and if this break is repaired via homologous recombination with the second inactive GFP allele as the template, then the GFP sequence is corrected, resulting in conversion of a GFP^- cell into a GFP^+ cell (Figure 1A). Loss of BRCA1 activity (e.g., by depletion with siRNA) results in cells that cannot repair the DNA break by HDR, and such cells remain GFP^-. We previously used this assay[11–14] to test the competence of 158 individual BRCA1 missense variants to rescue the loss of endogenous BRCA1. BRCA1 HDR function with these individual assays demonstrates 100% specificity and 93% sensitivity for predicting the cancer risk associated with variants with established interpretations ($n = 48$; Figure S1 and Table S2). The one known pathogenic variant misidentified as benign, c.330A>G (p.Arg71Gly), affects splicing rather than protein function[21] and therefore could not have been classified correctly with an assay that expresses variants within a cDNA copy of *BRCA1* (MIM: 113705). Given this functional assay's established specificity and sensitivity for predicting clinical pathogenicity, we sought to convert it to a multiplex format in order to generate high-confidence predictions for hundreds of BRCA1 variants within a single experiment.

As a pilot version of a multiplex HDR reporter assay (Figure 1B and Table S3), we prepared a pool of plasmids containing the WT *BRCA1* and 15 *BRCA1* variants that had been tested individually (Figure 1C). We integrated these plasmids *en masse* into the HeLa-derived HDR reporter cell line, which we had engineered to contain a single recombinase-based landing pad (HeLa-DR-FRT; Material and Methods). Although it has low efficiency, we used a recombinase-based system to ensure that only a single variant would be integrated into a common site in each cell, thus ensuring consistent expression levels and avoiding complications associated with lentiviruses, e.g., variable integration sites and template switching.[22]

After selecting for cells containing an integrated *BRCA1* variant, we performed the three steps of the HDR reporter assay. First, we depleted endogenous BRCA1 with a siRNA targeting the 3′ UTR of the endogenous *BRCA1* mRNA. Second, we expressed I-SceI to generate a double-strand break in the broken GFP reporter. Third, after allowing sufficient time for double-strand DNA break repair, we sorted cells into GFP^+ and GFP^- populations. We PCR amplified and sequenced the *BRCA1* variants in each sorted population and counted DNA reads for each variant. We calculated a score for each variant by taking the ratio of its frequency in GFP^+ cells to its frequency in GFP^- cells and normalized to the equivalently calculated ratio for the WT transgene. Scores from this pilot assay were largely consistent with published results for these 15 variants (Figures 1C and 1D); the sole exception was the BRCA1 p.His41Arg variant, which demonstrated an intermediate level of activity in the multiplex assay but was defective in the singleton assay.[12] Of note, independent data based on saturation genome editing (see Findlay et al.;[23] discussed further below) clearly support the classification of p.His41Arg as conferring loss of function, i.e., the result of the singleton assay.

Because this was a single experiment, we interpreted this false negative (i.e., the inappropriate classification of p.His41Arg as intermediate rather than loss of function) as indicative of the need to perform replicate assays. We settled on four replicates as a good balance between confidence in results and the expense and difficulty of the procedure. We proceeded to scale up the assay to analyze hundreds of variants per experiment by focusing on BRCA1 residues 2–192.

We created two individual pools of barcoded site-saturation mutagenesis libraries.[13,24] Pool 1 contained variants in residues 2–96, and pool 2 contained variants in residues 97–192 (Table S4). We integrated the plasmids in each pool into the HeLa-DR-FRT cell line. We then performed four replicates of the multiplex HDR reporter assay on each pool by using either a siRNA against endogenous *BRCA1* mRNA or a control siRNA (Table S5). The barcodes were amplified and sequenced from genomes of the GFP^+ and GFP^- cell populations. The score for each variant was calculated as described above for the small-scale assay (Table S6). Variant scores between replicates were well correlated for pool 1 but less so for pool 2, for which most variants scored close to ln(0), indicating little or no depletion (Figure S2)

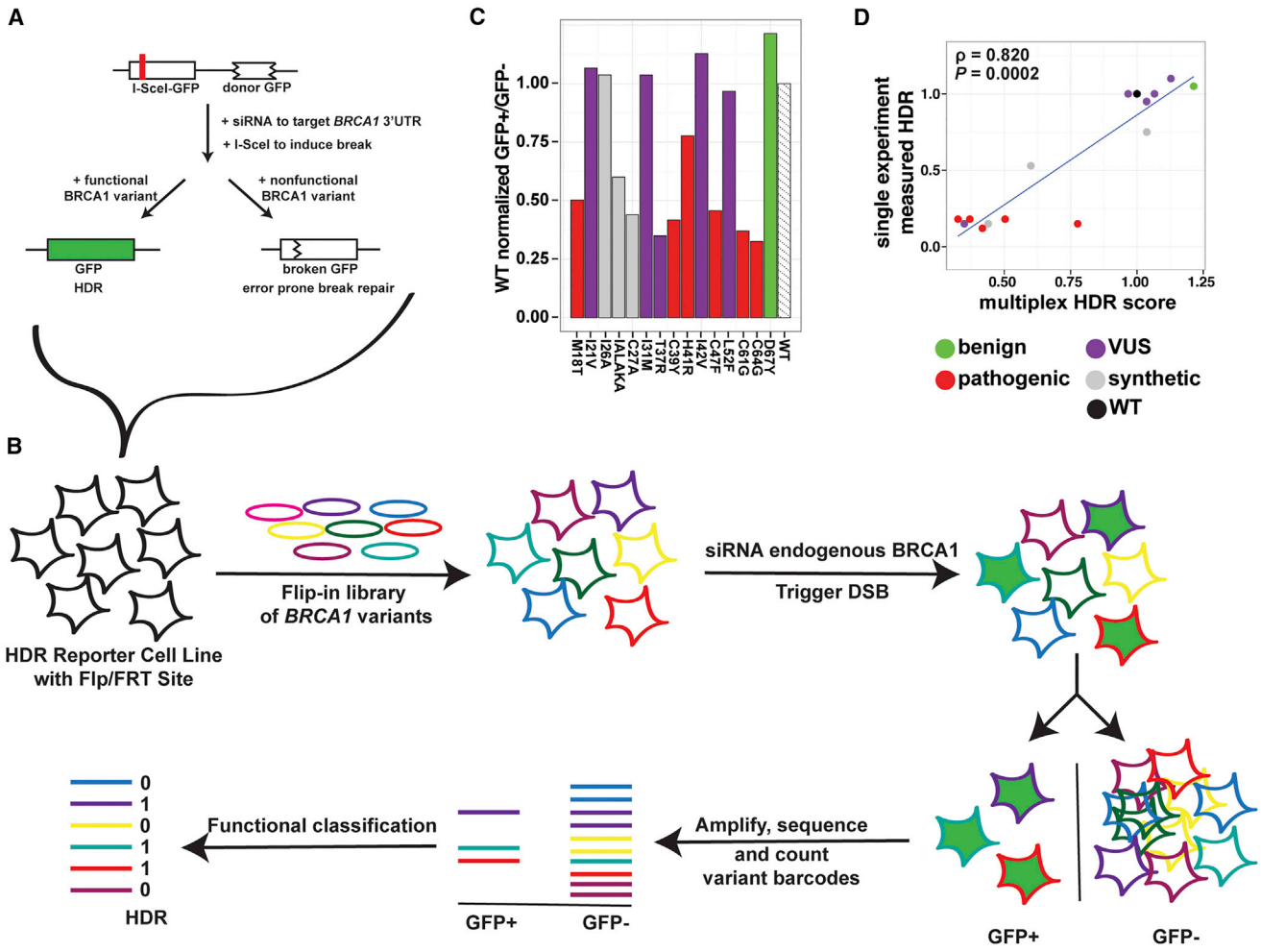

Figure 1. Overview of the Multiplex HDR Reporter Assay

(A) Schematic of the HDR reporter assay;[10] two inactive coding sequences for GFP are integrated in the genome of HeLa cells, and transfection of an expression vector for I-SceI results in a double-strand break in one inactive GFP coding sequence. HDR using the second inactive GFP coding sequence as a template results in the expression of functional GFP in a given cell. Depletion of *BRCA1* by siRNA transfection results in a decrease in GFP^+ cells, and expression of variant *BRCA1* cDNAs rescues GFP^+ status if the BRCA1 variant is functional.[12]

(B) Schematic of workflow for multiplex HDR reporter assay. Individual BRCA1 variants are integrated into the HeLa HDR reporter cell line with a flip-recombinase system. Depletion of the endogenous *BRCA1* by siRNA transfection makes these cells dependent on the integrated *BRCA1* variant. A double-strand break is induced in the HDR reporter by expression of I-SceI. After DNA repair, GFP^+ and GFP^- populations are sorted by FACS. Barcodes marking the variants from each population are amplified and sequenced. Depletion of *BRCA1* variants from the GFP^+ (HDR-functional) population is quantified.

(C) Results from the pilot 16-plex HDR assay testing WT and 15 variants previously analyzed[13] in a multiplex format. WT-normalized, GFP^+/GFP^- ratios are plotted on the y axis, and variant identifications are shown on the x axis.

(D) The correlation between scores from the 16-plex experiment and scores from individual HDR reporter assays.[13] The number of times the WT sequence was counted in the GFP^+ cells was set to 1.0, and the counts of the indicated variants were scored in relation to the WT. Spearman rho and p values are reported. Bars and points are colored according to ClinVar variant interpretation.

Functional Classification of Variants

A characteristic of all sequencing count data, including from multiplex assays for variant effects, is that the variance of scores increases at low read counts as a result of error from Poisson-distributed shot noise and stochastic dropout of variants.[20] This effect is exacerbated in cases such as the HDR reporter assay, which, despite being the gold-standard assay for *BRCA1* missense variants, has a bottleneck. Specifically, at most, only ~10% of the cells become GFP^+; this bottleneck limits the dynamic range of functional scores (Table S5). Because the score depends on read counts, we cannot directly compare the scores of any two variants. Therefore, we chose to construct a binary classifier ("depleted" versus "not depleted") by modeling the relationship between the read counts and scores (Material and Methods). With this model, we compared the significance of each variant's depletion in the *BRCA1* siRNA experiments with that of variants in the control siRNA experiments at the same read count in the same experiment (Figure 2A, Figure S3, Material and Methods, and Table S6). To remove as many variants that could be substantially affected by stochastic dropout as possible, we applied a

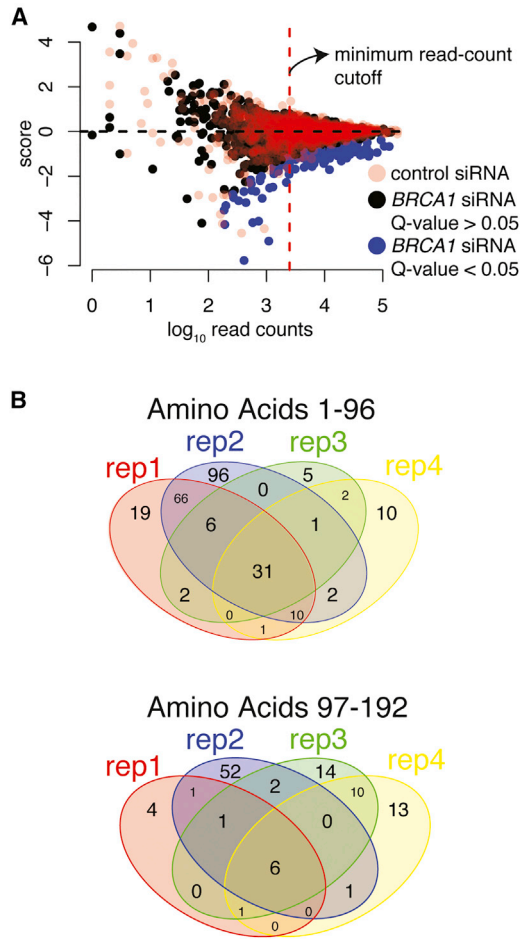

Figure 2. Identifying BRCA1 Variants Depleted from the GFP$^+$ Population

(A) The log of the WT-normalized, GFP$^+$/GFP$^-$ ratios are on the y axis, and $\log_{10}$ read counts are on the x axis for a single replicate of the HDR reporter assay for codons 2–96. Variants from the control (pink) or BRCA1 (black) siRNA conditions are indicated, and variants significantly depleted from the GFP$^+$ population in the *BRCA1* siRNA condition ($q < 0.05$) are colored blue. The dashed line represents the read-count threshold.

(B) Venn diagrams of the number of variants found depleted in multiple or single replicates.

stringent read-count filter to the GFP$^-$ population and removed variants below this threshold from further analysis (Figure S3 and Material and Methods).

For each experimental replicate, the number of variants above the read-count threshold in the GFP$^-$ population varied, as did the number of variants classified as depleted (i.e., nonfunctional variants; Table 1). The Venn diagrams shown in Figure 2B indicate the number of variants found to be depleted among the four replicate experiments for each pool of variants. In each set of replicate experiments, some variants were scored as depleted only once. Most of these singletons were depleted as a result of stochastic dropout, but some were nonsense variants that passed the read-count filter only in one replicate (Table S6). The variants found to be depleted in three or four replicates were enriched at residues with known pathogenic variants ($20\times$ enrichment, $p < 2.0 \times 10^{-16}$, Fisher's exact test). Pool 1 (amino acids 2–96) had the highest proportion of nonfunctional variants, which is expected because the structured RING domain is encoded almost entirely by the positions mutated in pool 1. Pool 2 (amino acids 97–192) had many fewer nonfunctional variants; these included missense variants at the two remaining positions of the RING domain (97 and 98), which were repeatedly depleted from the functional GFP$^+$ population. Many more variants were not depleted in the GFP$^+$ population in both pools, and those variants were also shared between replicates (Figure S4).

Variant Effects Measured in This Multiplex HDR Assay Are Strongly Concordant with Singleton HDR Assays and ClinVar Interpretations

In total, we measured the functional effects of 1,056 BRCA1 missense variants and four nonsense variants that were above the read-count threshold in three or more replicates. The great majority of BRCA1 missense variants were functional for DNA repair—only 59 variants (5.6%) were depleted from the population in three or four replicates and therefore likely to be nonfunctional for DNA repair (Figure 3A). Among the variants successfully tested, results from singleton HDR reporter assays were available for 16 variants.[11,13,14] All ten that had been scored as functional in HDR were depleted in zero replicates in the multiplex assay, and all six that were nonfunctional in singleton assays were depleted in three or four replicates of the multiplex assay (Figure 3B).

The number of times that variants were depleted from the functional population in the multiplex HDR reporter assay was strongly correlated with the results of other multiplex functional assays. For example, we previously used *in vitro* ubiquitin ligase and BARD1-binding yeast two-hybrid scores to predict HDR function for variants within the first 102 amino acids of BRCA1.[13] The 234 variants (boxplots) that overlap in the two assays, including the 25 variants (points) that are also in ClinVar, are highly concordant (Figure 3C). Variants that were never depleted in the multiplex HDR reporter assay correspond to those in the previous study with higher HDR predictions, whereas variants that were repeatedly depleted correspond to those in the previous study with lower HDR predictions. However, the phage- and yeast-based functional assays misidentified the known pathogenic variant p.Leu22Ser as functional, whereas it was correctly found to be defective in both single[13] and multiplex HDR experiments in human cells.

The number of replicates in which variants were depleted from the multiplex HDR reporter assay was also strongly consistent with functional scores from a multiplex BRCA1 functional assay that relies on the survival of a haploid cell line after saturation genome editing (SGE) of exons in *BRCA1*[23] (Figure 3D). To an extent, the exceptions are predictable. The SGE-based functional assay identifies variants that reduce splicing because the variants are edited into their native context in the genome, whereas

Table 1. The Number of Variants in Each Replicate

Pool	Replicate	Total Variants	Variants above Read Threshold	Variants Depleted	Percentage Depleted
Amino acids 1–96	1	1,228	736	138	19%
	2	1,210	795	226	28%
	3	1,341	254	49	19%
	4	1,340	254	61	24%
Amino acids 97–192	1	1,665	784	14	2%
	2	1,657	663	69	10%
	3	1,663	1,091	35	3%
	4	1,667	1,291	32	2%

such variants are false negatives in the HDR reporter assay in which the BRCA1 ORF is not spliced (Figure 3D; red points represent variants with reduced RNA).

Like the multiplex HDR reporter assay, the SGE assay is highly accurate for identifying variants that are damaging for protein function. However, some variants (p.Val14Gly, p.Cys44Gly, p.Glu85Gly, p.Ala92Gly, and p.Ala92Thr) are found to be damaging to protein function by the SGE assay but not the multiplex HDR reporter assay. None of these five variants have data or interpretations of cancer risk in ClinVar. In previous work, we individually compared the effects of amino acid substitutions on BRCA1 function in HDR versus repair by single-strand annealing (SSA).[15] Although HDR and SSA are both related mechanisms for repairing DNA double-strand breaks dependent on BRCA1, we found different tolerances for 7 of the 35 variants tested when we compared the two assays in that study. Although all seven of these were functional in HDR, three were nonfunctional in SSA and four had significantly depleted partial function in SSA.[15] When comparing the two multiplex assays for HDR and SGE, finding five differences among 118 comparisons is a low percentage (4%) and could reflect appropriate biology of the two assays. This is especially true given that the biochemical mechanism affecting cell growth in the SGE assay is unknown. Of the five differences in the comparison of high-throughput HDR and SGE assays, p.Cys44Gly could be a false negative of the multiplex HDR reporter assay given that all amino acid substitutions in the cysteine and histidine residues that bind zinc ions in the BRCA1 RING domain and have been tested to date in singleton functional assays are damaging,[11–13] although this specific p.Cys44Gly variant has not before been tested for HDR function. Alternatively, p.Cys44Gly and the other discordant variants could be slightly destabilized and rescued by the expression level in the HDR reporter assay, or these residues could be necessary for another BRCA1 function critical for cell survival in the SGE assay.

Clinical interpretations for some of the BRCA1 variants that we functionally score here are reported in ClinVar (accessed May 2018).[20] Of the 1,056 variants scored here, five were classified as benign or likely benign in ClinVar. All five were either never depleted or depleted in only one replicate in the multiplex HDR assay. In contrast, we tested eight BRCA1 variants that are established pathogenic variants in ClinVar. Of these, seven were depleted in three or four replicates in the multiplex HDR assay. The eighth pathogenic variant, p.Arg71Gly, had WT-like DNA repair activity in our data, a result consistent with previous HDR reporter assays.[15] p.Arg71Gly causes a defect in RNA splicing,[19] which as discussed above we do not expect to be detectable in our assay (Figure 3E).

In summary, we observed strong concordance between variants depleted in zero or one replicates in our assay and benign status in ClinVar or WT-like function in other assays, as well as between variants depleted in three or four replicates in our assay and pathogenic status in ClinVar or loss of function in other assays. We therefore conclude that the number of replicates in which a variant is found depleted is a reasonable proxy for its functionality and term these depletion scores (range 0–4; Table S7).

Depletion Scores Identify Damaging BRCA1 Variants

Depletion scores for all variants passing the read-count threshold in at least three replicates are shown in the form of a sequence-function map in Figure 4A. 100% of the damaging amino acid substitutions (dark red) were observed in the first 98 amino acids, which comprise the RING domain, whereas residues 99–192 strongly tended to tolerate amino acid substitutions. Nonsense variants, indicated by the asterisk on the bottom row, were nonfunctional. The distribution of variants indicates that the degenerate positions in the mutagenic oligonucleotides used to make the site-saturation mutagenesis libraries contained a much higher fraction of guanines than the 1:1:1:1 ratio of the four nucleotides specified for the synthesis (see Material and Methods). Thus, codons containing guanine in the first or second position (particularly glycine, arginine, alanine, and valine) are the most highly represented. Given that the highest representation of substitutions was to glycine, we mapped the number of times each substitution to glycine was found depleted in

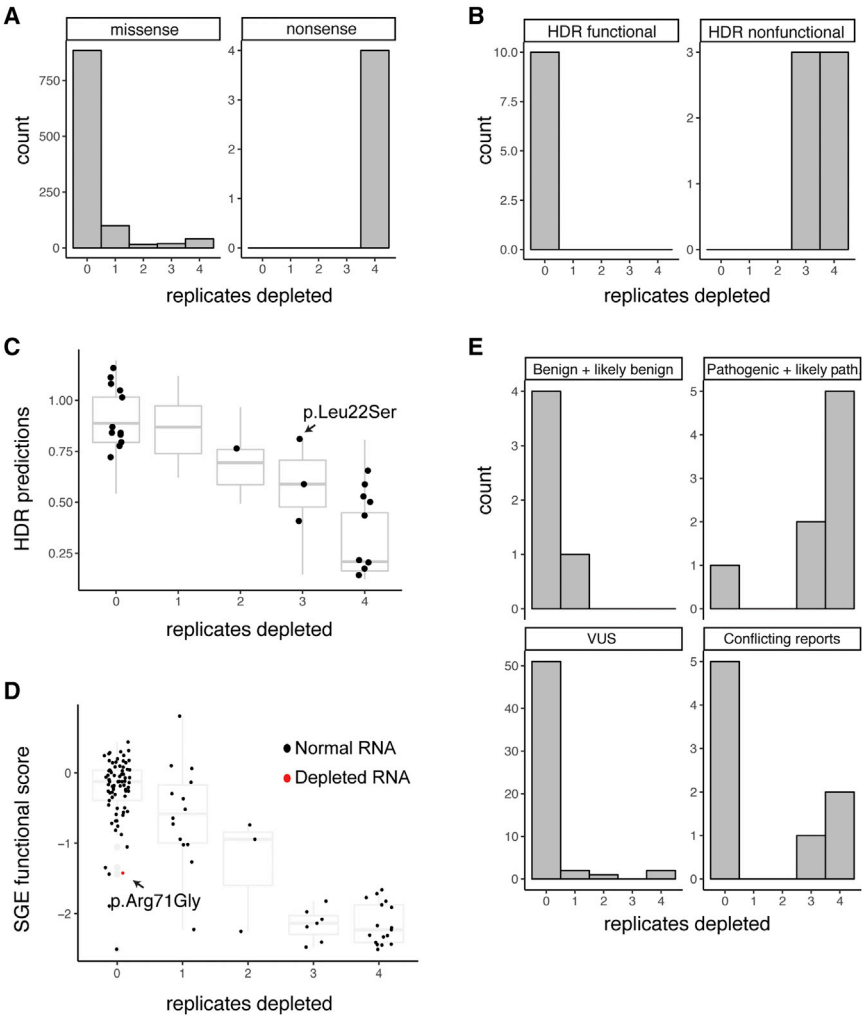

Figure 3. Comparison of Depletion Scores and Scores from Other Functional Assays and ClinVar Classifications

(A) Histogram of depletion scores for all missense or nonsense variants above the read-count threshold in at least three replicates.

(B) Histograms of variant depletion scores that were functional or nonfunctional as measured by individual HDR assays.

(C) Boxplots comparing 243 variant depletion scores (x axis) against HDR predictions (y axis) from Starita et al.[13] The 25 ClinVar variants are indicated by overlaid points, and BRCA1 p.Leu22Ser is indicated.

(D) Boxplots and points comparing variant depletion scores (x axis) against SGE functional scores (y axis[23]). Points marking variants with >80% RNA depletion in the SGE assay are colored red, and BRCA1 p.Arg71Gly is indicated.

(E) Histograms of variant depletion scores for each ClinVar classification. ClinVar variant IDs and classifications can be found in Table S7.

replicate experiments to the solution structure of the BRCA1 and BARD1 RING domain dimer[25] (Figure 4B). Amino acids positions that were the most intolerant to substitution were buried in either the interior of the four-helix bundle, which acts as the BARD1 interface, or in the loops that coordinate zinc ions.

We examined the depletion scores of 63 BRCA1 variants that are ambiguously classified in ClinVar (VUS or conflicting reports of pathogenicity) to predict their functional impact. Of 56 variants tested in the multiplex HDR reporter assay and classified as VUSs in ClinVar, two (p.Val11Gly and p.Ile68Arg) had a depletion score of 3 or 4 and are thus likely to be nonfunctional for HDR. Of eight variants that have conflicting interpretations of pathogenicity in ClinVar, three (p.Thr37Arg, p.His41Leu, and p.Cys47Gly) were nonfunctional in the DNA repair assay (Figure 3E). These seven variants, all ambiguous in ClinVar and nonfunctional in our assay, are at positions that coordinate zinc ions (p.Cys47Gly and p.His41Leu) or positions within the zinc-finger loops (p.Thr37Arg and p.Ile68Arg) or have side chains in the interior of the four-helix bundle (p.Val11Gly). Most known pathogenic missense variants map to these same structural features.

The remaining 58 variants that are ambiguously classified in ClinVar (VUS or conflicting reports of pathogenicity) have a low depletion score in our assay. Of these, 47 lie outside of the RING domain, where we did not identify any damaging amino acid substitutions. The 11 variants found within the structured RING include p.Gln81Glu, p.Ile90Val, and p.Ile90Thr, whose side chain points out of the four-helix bundle where amino acid changes are tolerated (Figure 4B). p.Thr77Met and p.Thr97Arg are substitutions to the amino acids that abut the helices and might not affect BARD1 binding; however, other changes at Thr97 are not tolerated, and VUS p.Thr97Ala has an ambiguous depletion score of two (Figure 4A). p.Ile42Leu and p.Ile68Val are in RING loops but are conservative changes and therefore might be tolerated; p.Glu29Gly, p.Glu33Ala, p.Gly57Arg, and p.Pro58Ala are also in the loops, but their side chains point away from the interior of the structure, which could be why they are tolerated. In studies of ubiquitin ligase activity, substitutions at Glu29 are damaging to ubiquitin ligase function[13] but not to HDR. This finding adds to the growing evidence that ligase activity is not required for BRCA1's HDR function.[26,27]

Discussion

We developed a multiplex reporter assay to measure the effect of hundreds of amino acid substitutions in BRCA1 on HDR, the molecular function most closely associated with

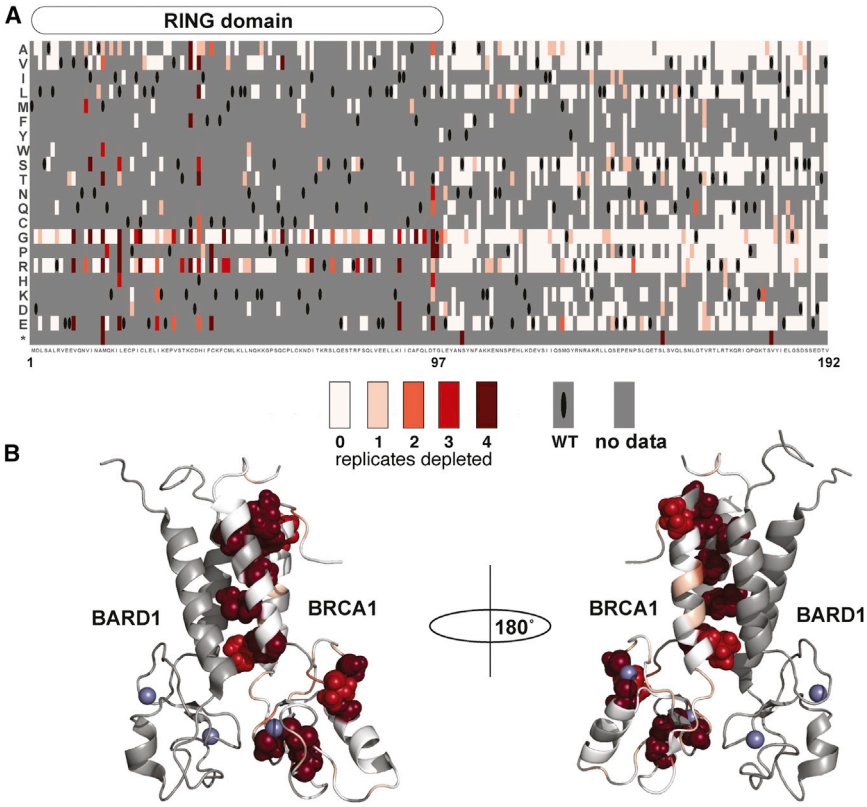

Figure 4. The Effect of Amino Acid Substitutions on the DNA Repair Function of BRCA1 Residues 2–192

(A) Sequence-function map of the effect of amino acid substitutions in BRCA1 residues 2–192. The depletion score for each variant is the count of experimental replicates in which it was found depleted from the functional population. Each position in BRCA1 (2–192) is arranged along the x axis, and the position of the RING domain is diagrammed above. The amino acid substitutions, grouped by side-chain properties, are on the y axis, and nonsense codons are indicated by asterisks. The depletion scores range from never depleted or likely functional (white) to likely nonfunctional (dark red). Black ovals demarcate the WT residue, and gray indicates missing data.

(B) The depletion score for amino acid substitutions to glycine are mapped to the solution structure of the BRCA1 1–102 and BARD1 26–125 dimer (PDB: 1JM7).[25] The color scale is the same as in (A), and spheres are shown for side chains at amino acid positions with a depletion score of 3 or 4.

its role in tumor suppression. We reproducibly measured 1,056 of the possible 3,820 amino acid or nonsense substitutions (191 codons × 20 amino acid or stop codon changes). Although these 1,056 constitute only 28% of all possible substitutions, to date, our assay is the highest-throughput assay that specifically analyzes the DNA repair function of BRCA1. Of the 12 variants known to be benign or pathogenic and assayed here, we classified them with 87.5% sensitivity and 100% specificity.[3] The assay was only 87.5% sensitive because it misclassified p.Arg71Gly, a cancer-associated variant that is pathogenic because of its impact on splicing, a feature not tested in the DNA repair assay. Coverage of more variants of known effect in future experiments will be required for more thoroughly validating this assay.

In this multiplex HDR reporter assay, we used a cell line (HeLa) that was not derived from breast or ovarian tissue. The requirement for a high transfection efficiency restricted us to only a few choices of human cell lines. The function of BRCA1 in HDR is considered ubiquitous for all human cell types. Although the use of HeLa cells for HDR reporter assays produced results that accurately predict hereditary breast and ovarian cancer risk, this assay does not address the breast and ovarian specificity of loss of BRCA1 activity.

The nonuniform coverage of amino acid substitutions in our variant libraries was a technical challenge. Two features of the protocol employed here to make variant libraries[24] contributed to this nonuniformity. First, we performed inverse PCRs individually for each amino acid position by using individually synthesized oligonucleotides. Therefore, failed reactions and uneven mixing of the PCR products caused loss of all or most substitutions at some positions (Figure 4A). Second, a strong guanine bias at each degenerate nucleotide position during oligonucleotide synthesis led to a bias in the encoded amino acids (Figure 4A). We anticipate that if the distribution of variants in the library were more uniform, we would be able to query more variants per experiment (up to a theoretical maximum of ~4,000 variants per experiment under ideal conditions) at our current number of sorted cells. Alternative approaches using array-derived oligonucleotides could create more uniformly distributed variant libraries.[28,29] It might also be useful to limit the number of amino acid changes to those accessible by single-nucleotide changes because these are the missense variants relevant to human disease. However, this limitation would also reduce the information content of the multiplex experiments.

In addition to improving library uniformity, restricting the number of barcoded variants in each multiplex HDR reporter experiment should result in more variants that pass the read-count filter. For example, a higher percentage of variants passed the read-count threshold in pool 2 (50% on average) than in pool 1 (29%). The variant libraries for pools 1 and 2 contained similar numbers of barcoded variants. However, more variants were damaging to BRCA1 HDR function in pool 1, and thus only ~7% of the

cells in this pool converted to being GFP^+ after the double-strand break, further limiting the dynamic range of pool 1 experiments.

In summary, we describe the development of a multiplex assay for measuring the effects of amino acid substitutions on a protein's function in double-strand DNA break repair in human cells. We analyzed 1,056 amino acid substitutions of BRCA1 residues 2–192 and found that this approach yielded results comparable to those of low-throughput HDR analysis of single variants and other functional assays. More importantly, our results are concordant with BRCA1 variants known to predispose to cancer and with BRCA1 variants previously identified to damage protein function. This assay can be repurposed to measure the effect of variants in other HDR pathway proteins that are also hereditary breast and ovarian cancer tumor suppressors, such as BRCA2[30–32] and BARD1.[33,34] As genetic testing for cancer risk becomes more common and additional genes are added to testing panels, the number of rare missense variants that inevitably become VUSs will continue to increase. We anticipate that multiplex functional assays can be used to functionally characterize such variants at scale, even before the variants are observed in results from genetic testing.

Accession Numbers

The accession number for the raw data reported in this article is GEO: GSE116427.

Supplemental Data

Supplemental Data include four figures, seven tables, and a test for variant depletion (provided in two file formats) and can be found with this article online at https://doi.org/10.1016/j.ajhg.2018.07.016.

Acknowledgments

We thank Jason Underwood and Katy Munson of the University of Washington PacBio Sequencing Services for their assistance with long-read sequencing, Ronald Hause and Alan Rubin for their helpful suggestions regarding statistical methods, Martin Kircher for his help with the analysis of long read sequences, Ethan Ahler for his assistance with figures, and the Ohio State University Comprehensive Cancer Center Analytical Cytometry Shared Resource for the sorting of GFP^+ cells. This work was supported by National Institutes of Health grants to S.F. (Biomedical Technology Research Resource project P41GM103533) and J.S. (Director's Pioneer Award DP1HG007811-05) and by the Basser Center for BRCA (Basser BRCA1 Innovation Award to J.D.P.). S.F. and J.S. are investigators of the Howard Hughes Medical Institute. M.M.I. was supported by a Pelotonia cancer training fellowship. This study was funded in part by the Brotman Baty Institute for Precision Medicine.

Declaration of Interests

The authors declare no competing interests.

Received: March 30, 2018
Accepted: July 19, 2018
Published: September 12, 2018

Web Resources

Burrows-Wheeler Aligner, http://bio-bwa.sourceforge.net/
Gene Expression Omnibus, https://www.ncbi.nlm.nih.gov/geo/
OMIM, http://omim.org/
pbccs: Generate Accurate Consensus Sequences from a Single SMRTbell, https://github.com/PacificBiosciences/unanimity/blob/master/doc/PBCCS.md
RSCB Protein Data Bank, https://www.rcsb.org/pdb/home/home.do
Scripts to assign a barcode to a variable region from PacBio SMRT reads, https://github.com/shendurelab/AssemblyByPacBio/

References

1. Murray, M.L., Cerrato, F., Bennett, R.L., and Jarvik, G.P. (2011). Follow-up of carriers of BRCA1 and BRCA2 variants of unknown significance: Variant reclassification and surgical decisions. Genet. Med. *13*, 998–1005.
2. Welsh, J.L., Hoskin, T.L., Day, C.N., Thomas, A.S., Cogswell, J.A., Couch, F.J., and Boughey, J.C. (2017). Clinical decision-making in patients with variant of uncertain significance in BRCA1 or BRCA2 genes. Ann. Surg. Oncol. *24*, 3067–3072.
3. Landrum, M.J., Lee, J.M., Riley, G.R., Jang, W., Rubinstein, W.S., Church, D.M., and Maglott, D.R. (2014). ClinVar: Public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. *42*, D980–D985.
4. Chenevix-Trench, G., Healey, S., Lakhani, S., Waring, P., Cummings, M., Brinkworth, R., Deffenbaugh, A.M., Burbidge, L.A., Pruss, D., Judkins, T., et al.; kConFab Investigators (2006). Genetic and histopathologic evaluation of BRCA1 and BRCA2 DNA sequence variants of unknown clinical significance. Cancer Res. *66*, 2019–2027.
5. Sweet, K., Senter, L., Pilarski, R., Wei, L., and Toland, A.E. (2010). Characterization of BRCA1 ring finger variants of uncertain significance. Breast Cancer Res. Treat. *119*, 737–743.
6. Osorio, A., de la Hoya, M., Rodríguez-López, R., Martínez-Ramírez, A., Cazorla, A., Granizo, J.J., Esteller, M., Rivas, C., Caldés, T., and Benítez, J. (2002). Loss of heterozygosity analysis at the BRCA loci in tumor samples from patients with familial breast cancer. Int. J. Cancer *99*, 305–309.
7. Easton, D.F., Deffenbaugh, A.M., Pruss, D., Frye, C., Wenstrup, R.J., Allen-Brady, K., Tavtigian, S.V., Monteiro, A.N.A., Iversen, E.S., Couch, F.J., and Goldgar, D.E. (2007). A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. Am. J. Hum. Genet. *81*, 873–883.
8. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al.; ACMG Laboratory Quality Assurance Committee (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet. Med. *17*, 405–424.
9. Harrison, S.M., Dolinsky, J.S., Knight Johnson, A.E., Pesaran, T., Azzariti, D.R., Bale, S., Chao, E.C., Das, S., Vincent, L., and Rehm, H.L. (2017). Clinical laboratories collaborate to resolve differences in variant interpretations submitted to ClinVar. Genet. Med. *19*, 1096–1104.

10. Pierce, A.J., Johnson, R.D., Thompson, L.H., and Jasin, M. (1999). XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. Genes Dev. *13*, 2633–2638.
11. Lu, C., Xie, M., Wendl, M.C., Wang, J., McLellan, M.D., Leiserson, M.D.M., Huang, K.L., Wyczalkowski, M.A., Jayasinghe, R., Banerjee, T., et al. (2015). Patterns and functional implications of rare germline variants across 12 cancer types. Nat. Commun. *6*, 10086.
12. Ransburgh, D.J.R., Chiba, N., Ishioka, C., Toland, A.E., and Parvin, J.D. (2010). Identification of breast tumor mutations in BRCA1 that abolish its function in homologous DNA recombination. Cancer Res. *70*, 988–995.
13. Starita, L.M., Young, D.L., Islam, M., Kitzman, J.O., Gullingsrud, J., Hause, R.J., Fowler, D.M., Parvin, J.D., Shendure, J., and Fields, S. (2015). Massively parallel functional analysis of BRCA1 RING domain variants. Genetics *200*, 413–422.
14. Towler, W.I., Zhang, J., Ransburgh, D.J.R., Toland, A.E., Ishioka, C., Chiba, N., and Parvin, J.D. (2013). Analysis of BRCA1 variants in double-strand break repair by homologous recombination and single-strand annealing. Hum. Mutat. *34*, 439–445.
15. Drost, R., Bouwman, P., Rottenberg, S., Boon, U., Schut, E., Klarenbeek, S., Klijn, C., van der Heijden, I., van der Gulden, H., Wientjens, E., et al. (2011). BRCA1 RING function is essential for tumor suppression but dispensable for therapy resistance. Cancer Cell *20*, 797–809.
16. Wu, W., Sato, K., Koike, A., Nishikawa, H., Koizumi, H., Venkitaraman, A.R., and Ohta, T. (2010). HERC2 is an E3 ligase that targets BRCA1 for degradation. Cancer Res. *70*, 6384–6392.
17. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.
18. Travers, K.J., Chin, C.-S., Rank, D.R., Eid, J.S., and Turner, S.W. (2010). A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. *38*, e159–e159.
19. Matreyek, K.A., Starita, L.M., Stephany, J.J., Martin, B., Chiasson, M.A., Gray, V.E., Kircher, M., Khechaduri, A., Dines, J.N., Hause, R.J., et al. (2018). Multiplex assessment of protein variant abundance by massively parallel sequencing. Nat. Genet. *50*, 874–882.
20. Rubin, A.F., Gelman, H., Lucas, N., Bajjalieh, S.M., Papenfuss, A.T., Speed, T.P., and Fowler, D.M. (2017). A statistical framework for analyzing deep mutational scanning data. Genome Biol. *18*, 150.
21. Vega, A., Campos, B., Bressac-De-Paillerets, B., Bond, P.M., Janin, N., Douglas, F.S., Domènech, M., Baena, M., Pericay, C., Alonso, C., et al. (2001). The R71G BRCA1 is a founder Spanish mutation and leads to aberrant splicing of the transcript. Hum. Mutat. *17*, 520–521.
22. Sack, L.M., Davoli, T., Xu, Q., Li, M.Z., and Elledge, S.J. (2016). Sources of error in mammalian genetic screens. G3 (Bethesda) *6*, 2781–2790.
23. Findlay, G.M., Daza, R.M., Martin, B., Zhang, M.D., Leith, A.P., Gasperini, M., Janizek, J.D., Huang, X., Starita, L.M., and Shendure, J. (2018). Accurate classification of *BRCA1* variants with saturation genome editing. Nature *562*. Published online September 12, 2018. https://doi.org/0.1038/s41586-018-0461-z.
24. Jain, P.C., and Varadarajan, R. (2014). A rapid, efficient, and economical inverse polymerase chain reaction-based method for generating a site saturation mutant library. Anal. Biochem. *449*, 90–98.
25. Brzovic, P.S., Rajagopal, P., Hoyt, D.W., King, M.C., and Klevit, R.E. (2001). Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. Nat. Struct. Biol. *8*, 833–837.
26. Shakya, R., Reid, L.J., Reczek, C.R., Cole, F., Egli, D., Lin, C.-S., deRooij, D.G., Hirsch, S., Ravi, K., Hicks, J.B., et al. (2011). BRCA1 tumor suppression depends on BRCT phosphoprotein binding, but not its E3 ligase activity. Science *334*, 525–528.
27. Reid, L.J., Shakya, R., Modi, A.P., Lokshin, M., Cheng, J.-T., Jasin, M., Baer, R., and Ludwig, T. (2008). E3 ligase activity of BRCA1 is not essential for mammalian cell viability or homology-directed repair of double-strand DNA breaks. Proc. Natl. Acad. Sci. USA *105*, 20876–20881.
28. Majithia, A.R., Tsuda, B., Agostini, M., Gnanapradeepan, K., Rice, R., Peloso, G., Patel, K.A., Zhang, X., Broekema, M.F., Patterson, N., et al.; UK Monogenic Diabetes Consortium; Myocardial Infarction Genetics Consortium; and UK Congenital Lipodystrophy Consortium (2016). Prospective functional classification of all possible missense variants in PPARG. Nat. Genet. *48*, 1570–1575.
29. Kitzman, J.O., Starita, L.M., Lo, R.S., Fields, S., and Shendure, J. (2015). Massively parallel single-amino-acid mutagenesis. Nat. Methods *12*, 203–206, 4, 206.
30. Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., and Micklem, G. (1995). Identification of the breast cancer susceptibility gene BRCA2. Nature *378*, 789–792.
31. Moynahan, M.E., Pierce, A.J., and Jasin, M. (2001). BRCA2 is required for homology-directed repair of chromosomal breaks. Mol. Cell *7*, 263–272.
32. Guidugli, L., Pankratz, V.S., Singh, N., Thompson, J., Erding, C.A., Engel, C., Schmutzler, R., Domchek, S., Nathanson, K., Radice, P., et al. (2013). A classification model for BRCA2 DNA binding domain missense variants based on homology-directed repair activity. Cancer Res. *73*, 265–275.
33. Couch, F.J., Shimelis, H., Hu, C., Hart, S.N., Polley, E.C., Na, J., Hallberg, E., Moore, R., Thomas, A., Lilyquist, J., et al. (2017). Associations between cancer predisposition testing panel genes and breast cancer. JAMA Oncol. *3*, 1190–1196.
34. Lee, C., Banerjee, T., Gillespie, J., Ceravolo, A., Parvinsmith, M.R., Starita, L.M., Fields, S., Toland, A.E., and Parvin, J.D. (2015). Functional analysis of BARD1 missense variants in homology-directed repair of DNA double strand breaks. Hum. Mutat. *36*, 1205–1214.