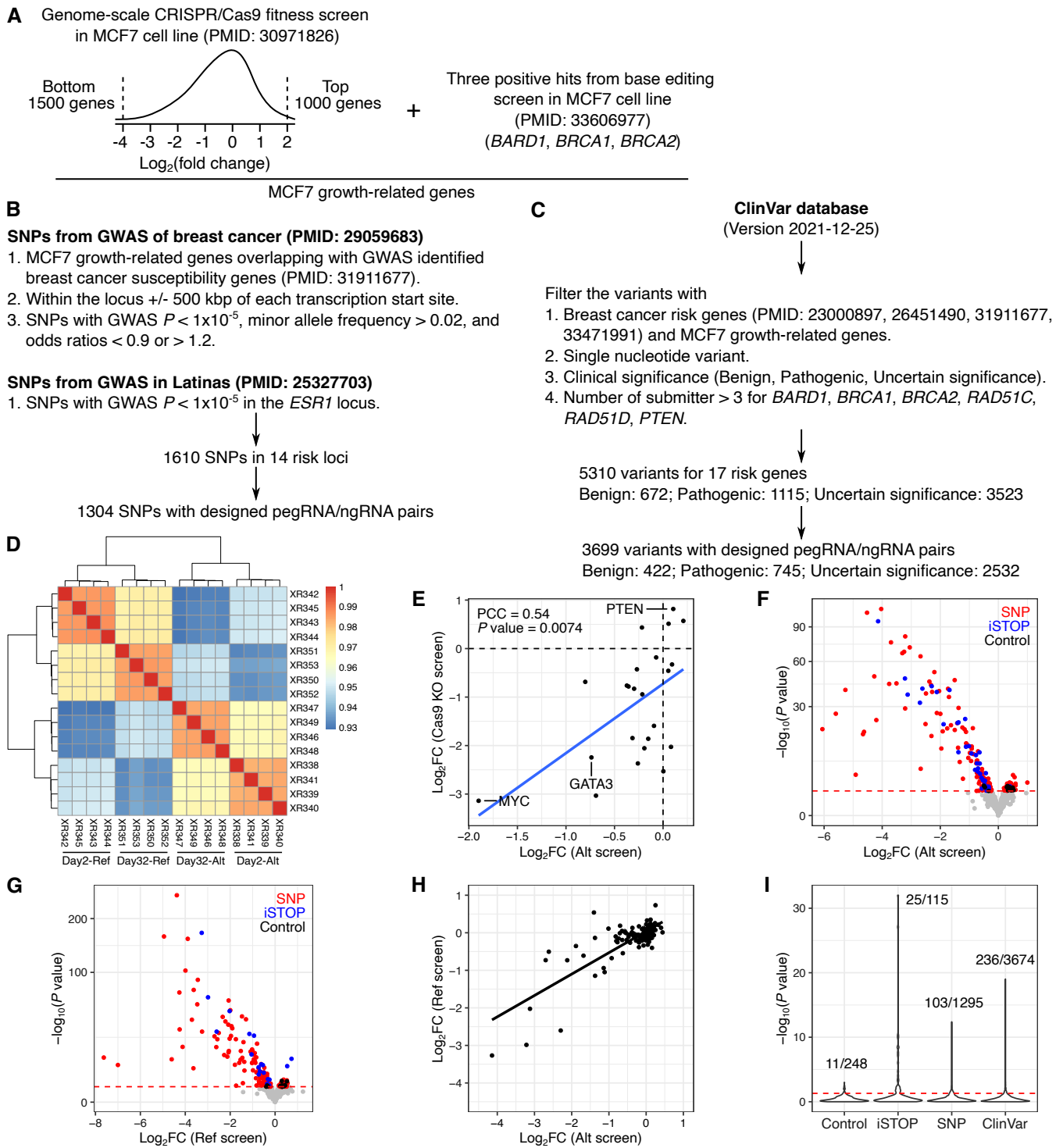**Figure S1. Optimizing PE efficiency in MCF7 cell line, related to Figure 1.**
(A) PE efficiency and indel rate by co-infection of pegRNA, ngRNA and nCas9/RT expressing lentiviruses in MCF7 cells. (B) Immunofluorescent staining showing the localization of nCas9/RT (Red, FLAG tagged) in the nucleus (Blue, DAPI) in MCF7-nCas9/RT cells. Scale bars, 1000 μm. (C) Editing efficiency and indel rate by PE using three different structured RNA motifs to the 3' terminus of pegRNAs at 2-week and 4-week post infection in MCF7-nCas9/RT cells. Error bars represent the s.d.
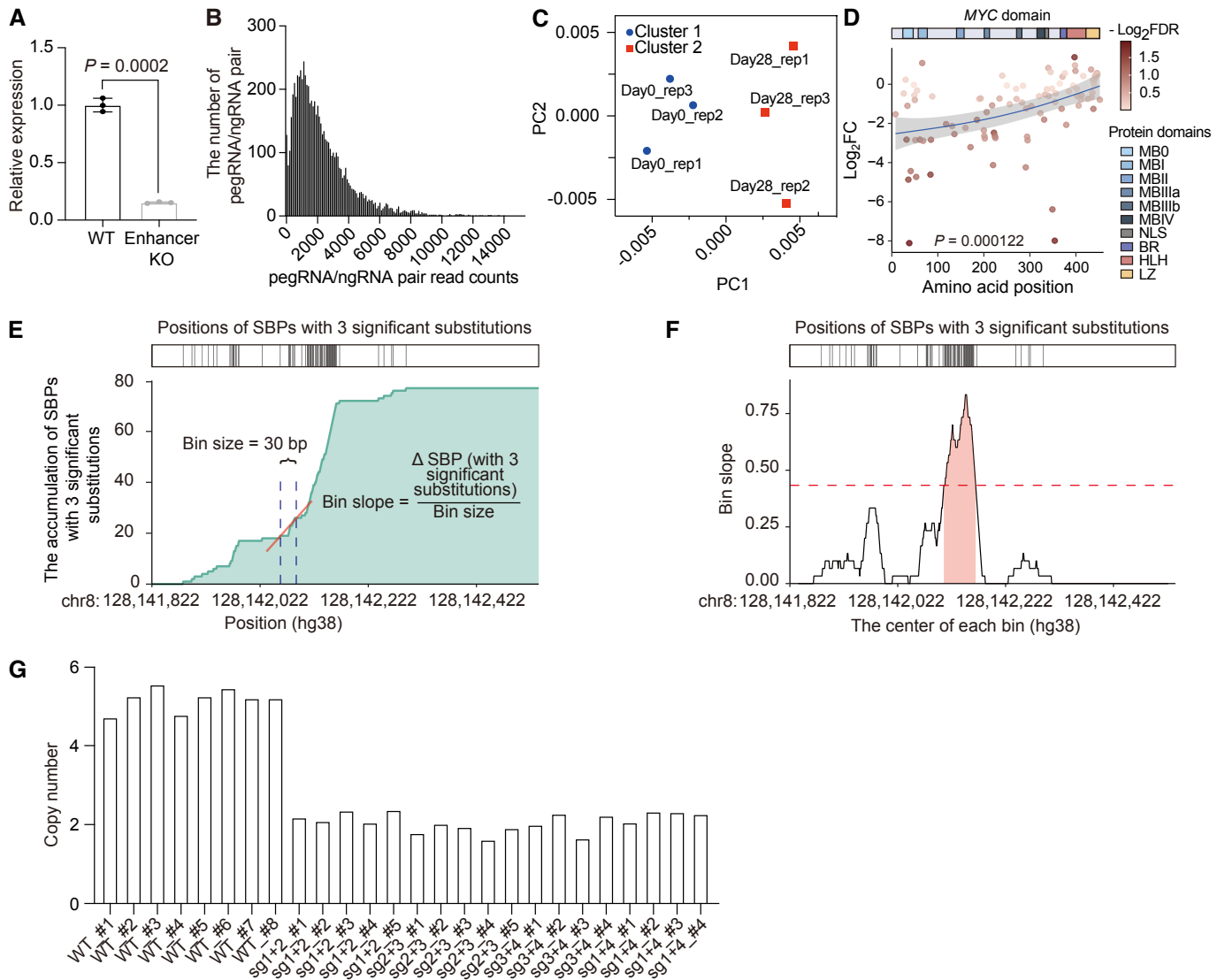
Figure S2. Characterize the *MYC* enhancer sequences in MCF7 cells, related to Figure 2.
(A) CRISPR/Cas9 knockout of the *MYC* enhancer in MCF7 decreased *MYC* expression. *P* values were calculated by two-tailed two-sample t-test. Error bars represent the s.e.m. (B) Distribution of pegRNA/ngRNA pair read counts in the cloned plasmid library. (C) Multidimensional scaling analysis demonstrates the high reproducibility of PRIME between biological replicates. (D) The correlation between locations of PE-introduced stop codons and their effect sizes. The blue line and *P* value were calculated using generalized additive models. The shaded areas indicate 95% confidence intervals. (E) (Top) The position of SBPs with three significant substitutions. (Bottom) Cumulative distribution plot of SBPs with three significant substitutions along the *MYC* enhancer and the formula for calculating the slope of each continuous bin. (F) Line plot of slopes for each continuous bin along the *MYC* enhancer. The red dashed line is the cutoff for a significant slope, which is based on a slope with a Z score-derived *P* value equal to 0.05. The red region is the core enhancer region, derived from the bins' slopes greater than the cutoff (slope > 0.43). (G) The copy numbers of the *MYC* enhancer in wild type clones and edited clones were determined by qPCR and normalized to *RPP30*.

A Genome-scale CRISPR/Cas9 fitness screen
in MCF7 cell line (PMID: 30971826)

Bottom
1500 genes

Top
1000 genes

-4 -3 -2 -1 0 1 2
Log$_2$(fold change)

+ Three positive hits from base editing
screen in MCF7 cell line
(PMID: 33606977)
(*BARD1*, *BRCA1*, *BRCA2*)

MCF7 growth-related genes

B

SNPs from GWAS of breast cancer (PMID: 29059683)
1. MCF7 growth-related genes overlapping with GWAS identified breast cancer susceptibility genes (PMID: 31911677).
2. Within the locus +/- 500 kbp of each transcription start site.
3. SNPs with GWAS $P < 1\times10^{-5}$, minor allele frequency > 0.02, and odds ratios < 0.9 or > 1.2.

SNPs from GWAS in Latinas (PMID: 25327703)
1. SNPs with GWAS $P < 1\times10^{-5}$ in the *ESR1* locus.

↓

1610 SNPs in 14 risk loci

↓

1304 SNPs with designed pegRNA/ngRNA pairs

C

ClinVar database
(Version 2021-12-25)

↓

Filter the variants with
1. Breast cancer risk genes (PMID: 23000897, 26451490, 31911677, 33471991) and MCF7 growth-related genes.
2. Single nucleotide variant.
3. Clinical significance (Benign, Pathogenic, Uncertain significance).
4. Number of submitter > 3 for *BARD1*, *BRCA1*, *BRCA2*, *RAD51C*, *RAD51D*, *PTEN*.

↓

5310 variants for 17 risk genes
Benign: 672; Pathogenic: 1115; Uncertain significance: 3523

↓

3699 variants with designed pegRNA/ngRNA pairs
Benign: 422; Pathogenic: 745; Uncertain significance: 2532

**D**

E

F

G

H

I

Figure S3. Strategies for prioritizing GWAS-identified breast cancer-associated SNPs and clinical variants, and quality control and primary analysis of PRIME results, related to Figure 3.
(A) The MCF7 growth-related genes were selected from the CRISPR/Cas9 knockout screen and base editing screen in MCF7 cells. (B) The strategy used for selecting breast cancer-associated SNPs. (C) The strategy used for selecting clinical variants. (D) Heatmap with pairwise correlations and hierarchical clustering of read counts from PRIME. (E) Pearson correlations between the log$_2$(fold change) of iSTOPs in the Alt library screen and the log$_2$(fold change) of gRNAs in the CRISPR/Cas9 knockout screen for each target gene. (F) Volcano plot of the results from the Alt library screen. (G) Volcano plot of the results from the Ref library screen. (H) The log$_2$(fold change) for each iSTOP from the Alt and Ref library screens. (I) Violin plot showing the 5% FDR cutoff used for the relative effect analysis comparing the Alt and Ref libraries. Numbers above peaks indicate the significant data points versus the total data points in each category when using 5% FDR. We used the 5% percentile of *P* values from negative controls as the empirical significance threshold to achieve an FDR of 5% indicated by the red dashed line in F, G and I.
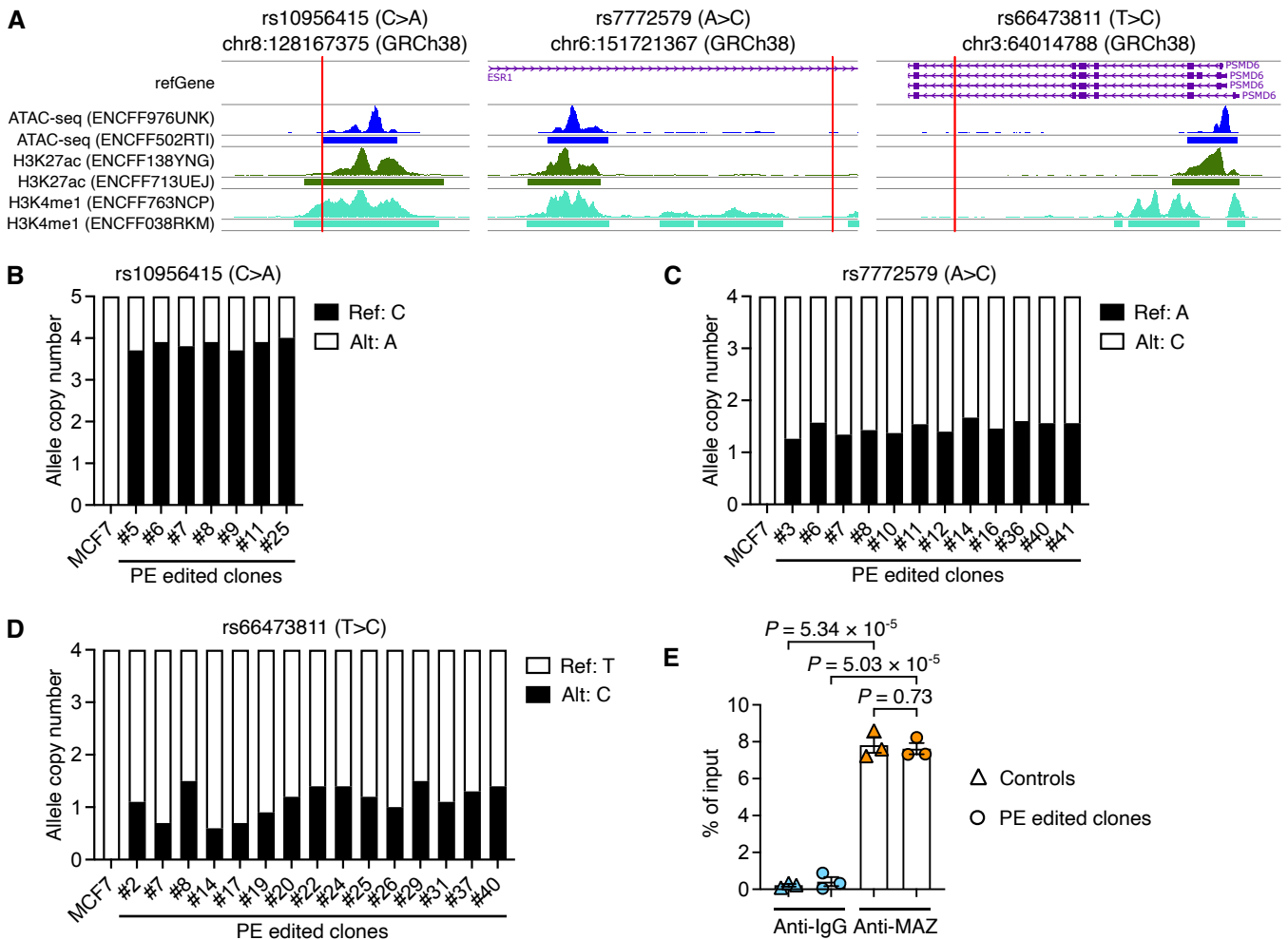
Figure S4. Functional validation of PRIME identified SNPs, related to Figure 4.
(A) Snapshot of epigenomic profiles at genomic loci with three validated GWAS SNPs. The red lines indicate SNP positions. ATAC-seq, H3K27ac, H3K4me1 datasets in MCF7 are from the ENCODE project with corresponding access numbers indicated. (B-D) Copy numbers of alternative and reference alleles in MCF7 cells and PE edited clones for each SNP. (E) MAZ binding at rs66473811 locus assessed by ChIP-qPCR. Data are displayed in mean with s.e.m. P values were calculated by two-tailed two-sample t-test. Dots show individual replicate values.
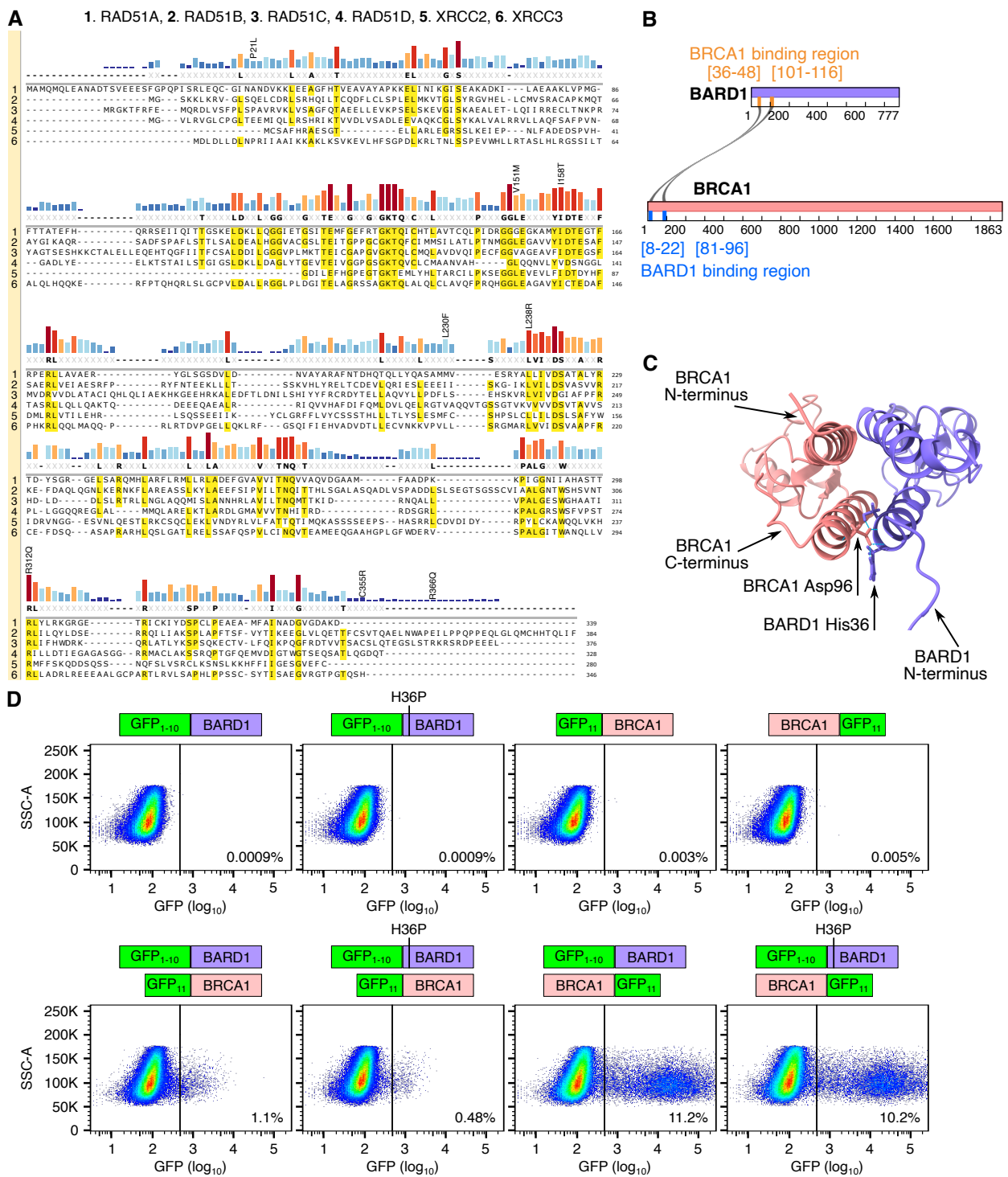
Figure S5. Sequence conservation of RAD51 family proteins, and evaluating the BARD1-BRCA1 interaction efficiency in MCF7 using the split GFP system, related to Figure 5.

(A) Sequence conservation of RAD51 family proteins. Alignment of RAD51 family proteins using MUSCLE. Functional VUS identified by PRIME in RAD51C are labeled. (B) Graphic showing the binding regions between BARD1 and BRCA1. (C) The Alphafold predicted protein structure showing the spatial distance between termini of the wild type BARD1 and BRCA1 complex. The distance between the BARD1 N-terminus and the BRCA1 C-terminus is closer than the distance between the BARD1 N-terminus and the BRCA1 N-terminus. (D) Flow cytometry plots showing GFP signals from MCF7 transfected with different split GFP vectors. No GFP signal was observed from individual vectors (upper panel). The strong GFP signals were observed from cells transfected with GFP$_{1-10}$-BARD1 or GFP$_{1-10}$-BARD1^{H36P} + BRCA1-GFP$_{11}$, not from cells transfected with GFP$_{1-10}$-BARD1 or GFP$_{1-10}$-BARD1^{H36P} + GFP$_{11}$-BRCA1 (bottom panel), consistent with the spatial distance between BARD1 and BRCA1 terminuses in B and highlighting the high sensitivity of the split GFP system in detecting protein-protein interaction.