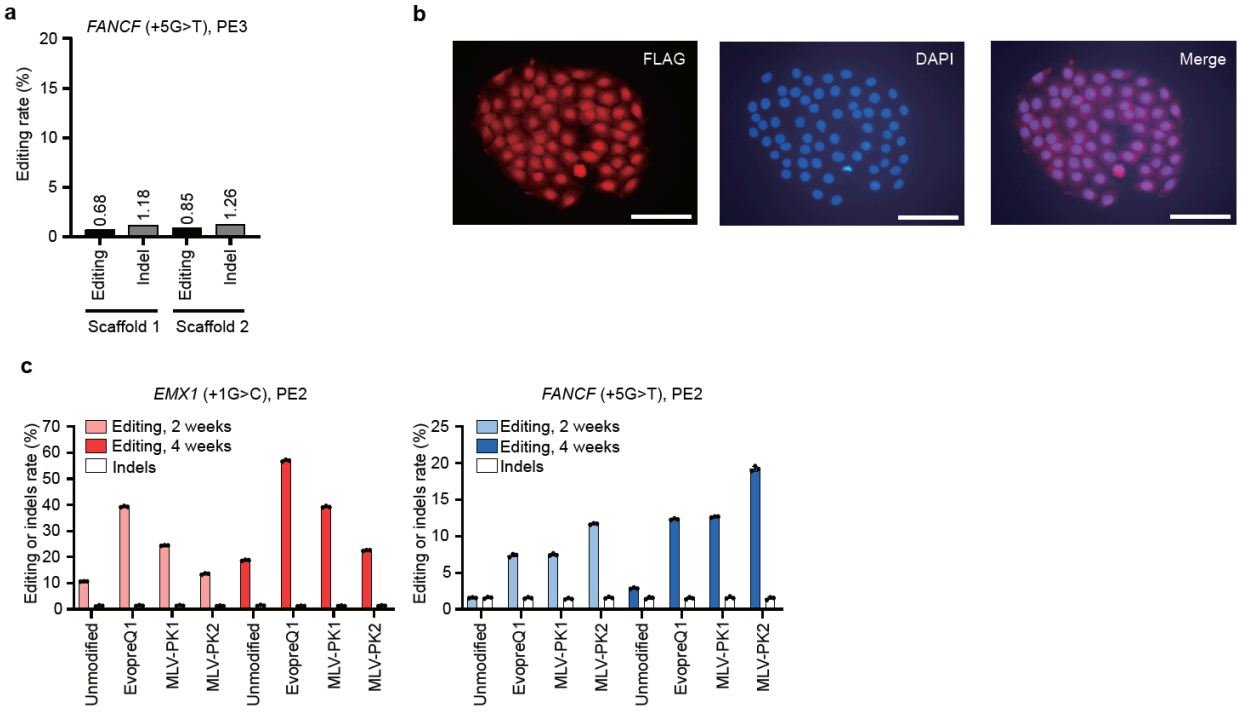
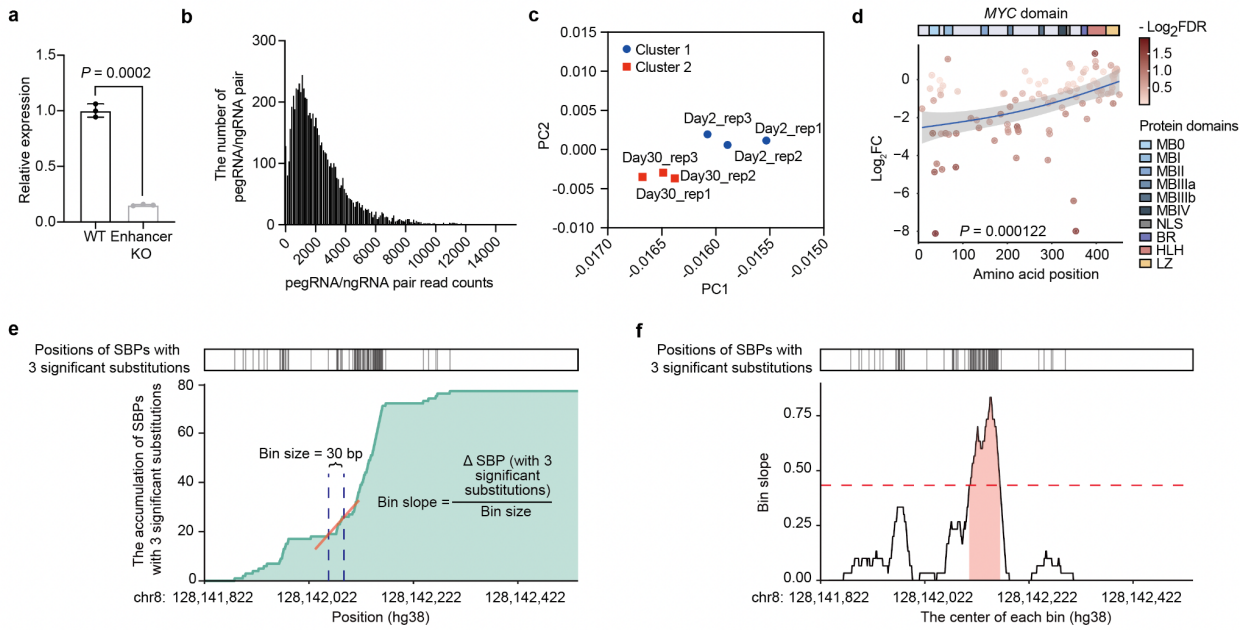
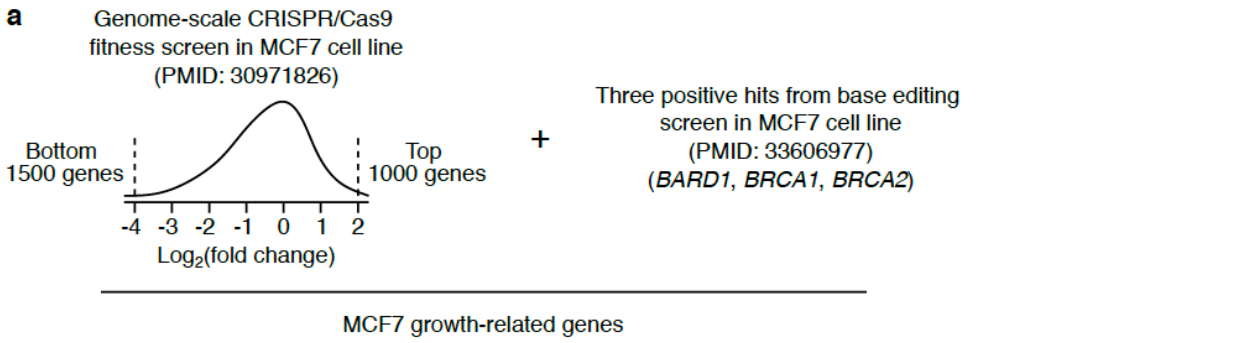


Supplementary Figure 1. Optimizing PE efficiency in MCF7 cell line.

(a) Prime editing 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 and 4 weeks post infection in MCF7-nCas9/RT cells. Error bars represent the s.d.



Supplementary Figure 2. Characterize enhancer function and results of PRIME in MCF7 cells. (a) CRISPR/Cas9 knockout of the *MYC* enhancer in MCF7 decreased *MYC* expression. *P* values were calculated using a 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) PCA analysis demonstrates the high reproducibility of PRIME between biological replicates. (d) The correlation between locations of PE-induced 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).



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 $P < 1x10^{-5}$, minor allele frequency < 0.02, and odds ratios < 0.9 or > 1.2

SNPs from GWAS in Latinas (PMID: 25327703)

1. SNPs with GWAS $P < 1x10^{-5}$ 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*

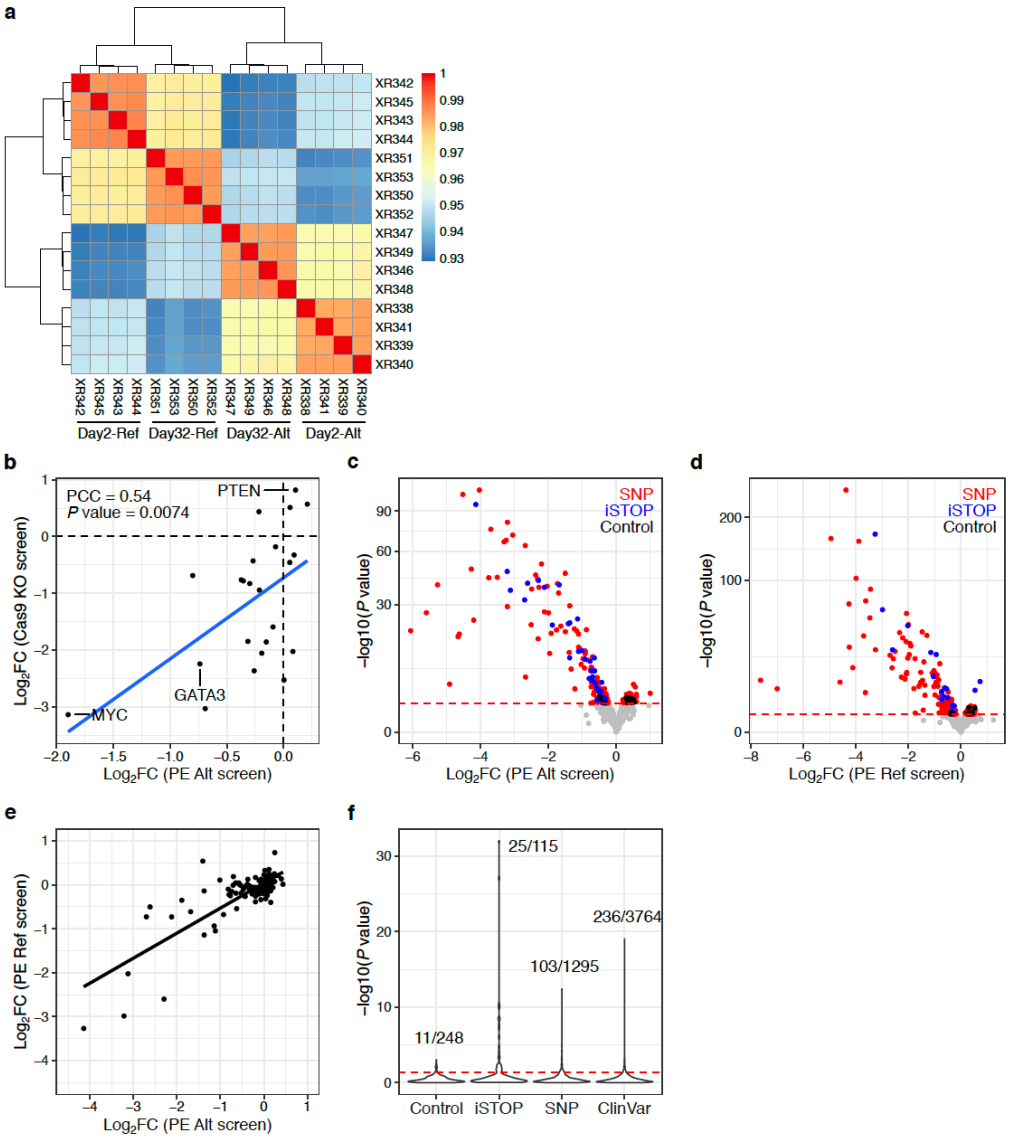
↓

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

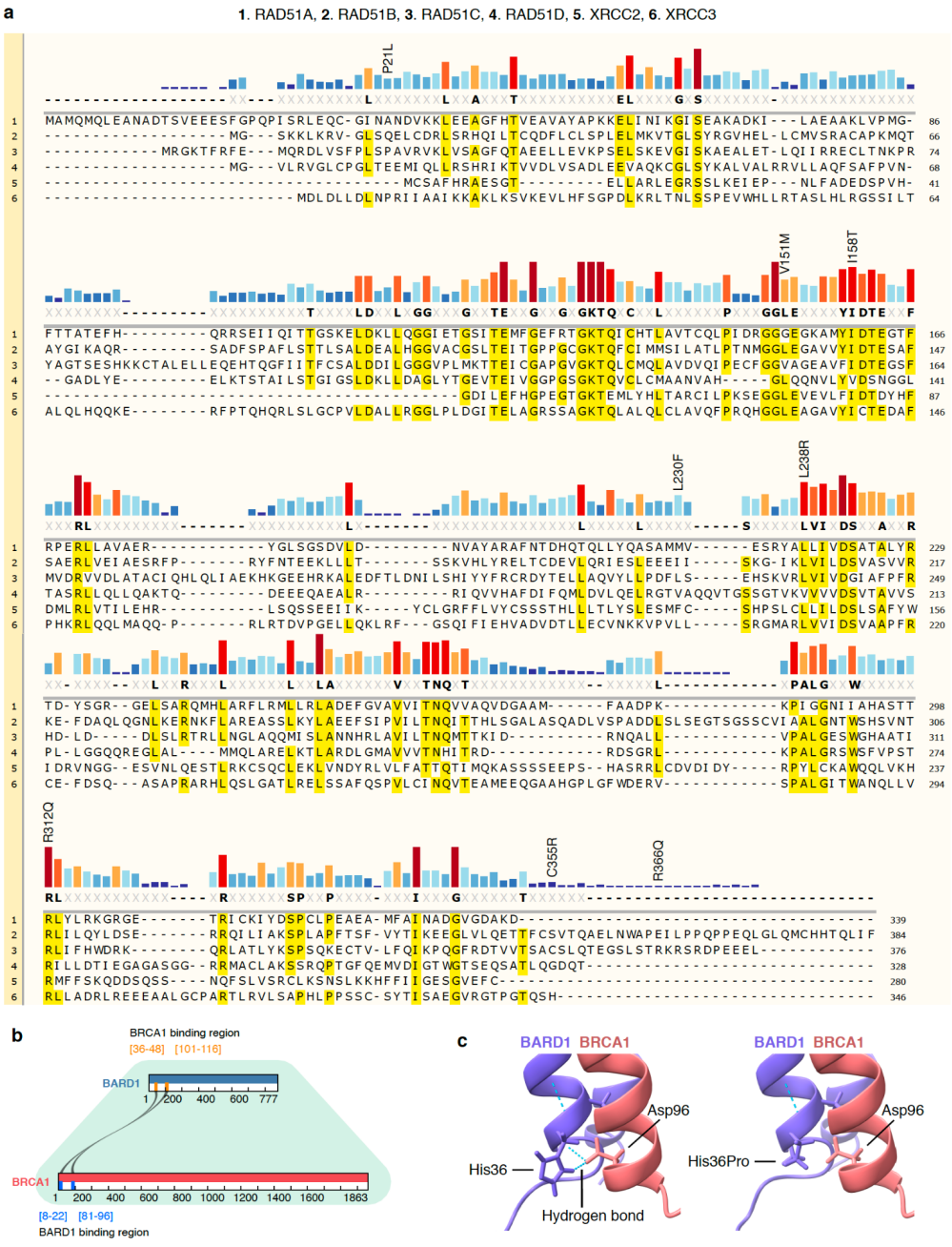
↓

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

Supplementary Figure 3. Strategies for prioritizing genomic loci and clinical variants. (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-related SNPs. (c) The strategy used for selecting clinical variants.



Supplementary Figure 4. Quality control and primary analysis of disease variants. (a) Heatmap with pairwise correlations and hierarchical clustering of read counts from PRIME. (b) 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. (c) Volcano plot of the results from the Alt library screen. (d) Volcano plot of the results from the Ref library screen. (e) The $\log_2$(fold change) for each iSTOP from the Alt and Ref library screens. (f) 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 a false discovery rate (FDR) of 5% indicated by the red dashed line in d-f.



Supplementary Figure 5. Examples of functional VUS with their potential consequences. (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 of the BARD1 and BRCA1 complex. Two hydrogen bonds were identified between wild type His36 in BARD1 and Asp96 in BRCA1, but lost following the BARD1 His36Pro mutation.