Supplemental Information for

Genome-scale targeted mutagenesis in *Brassica napus* using a pooled CRISPR library

Jianjie He, Kai Zhang, Shuxiang Yan, Mi Tang, Weixian Zhou, Yongtai Yin, Kang Chen, Chunyu Zhang, Maoteng Li

Corresponding authors: Chunyu Zhang and Maoteng Li
Email: zhchy@mail.hzau.edu.cn; limaoteng426@hust.edu.cn

This PDF file includes the following:

Supplemental Figures S1 to S11
Summaries for Supplemental Tables S1 to S9
Summaries for Supplemental Code S1 to S13
Summaries for Supplemental Datasets S1 to S8

Other supplemental materials for this manuscript include the following:

Supplemental Tables S1 to S9
Supplemental Code S1 to S13
Supplemental Datasets S1 to S8

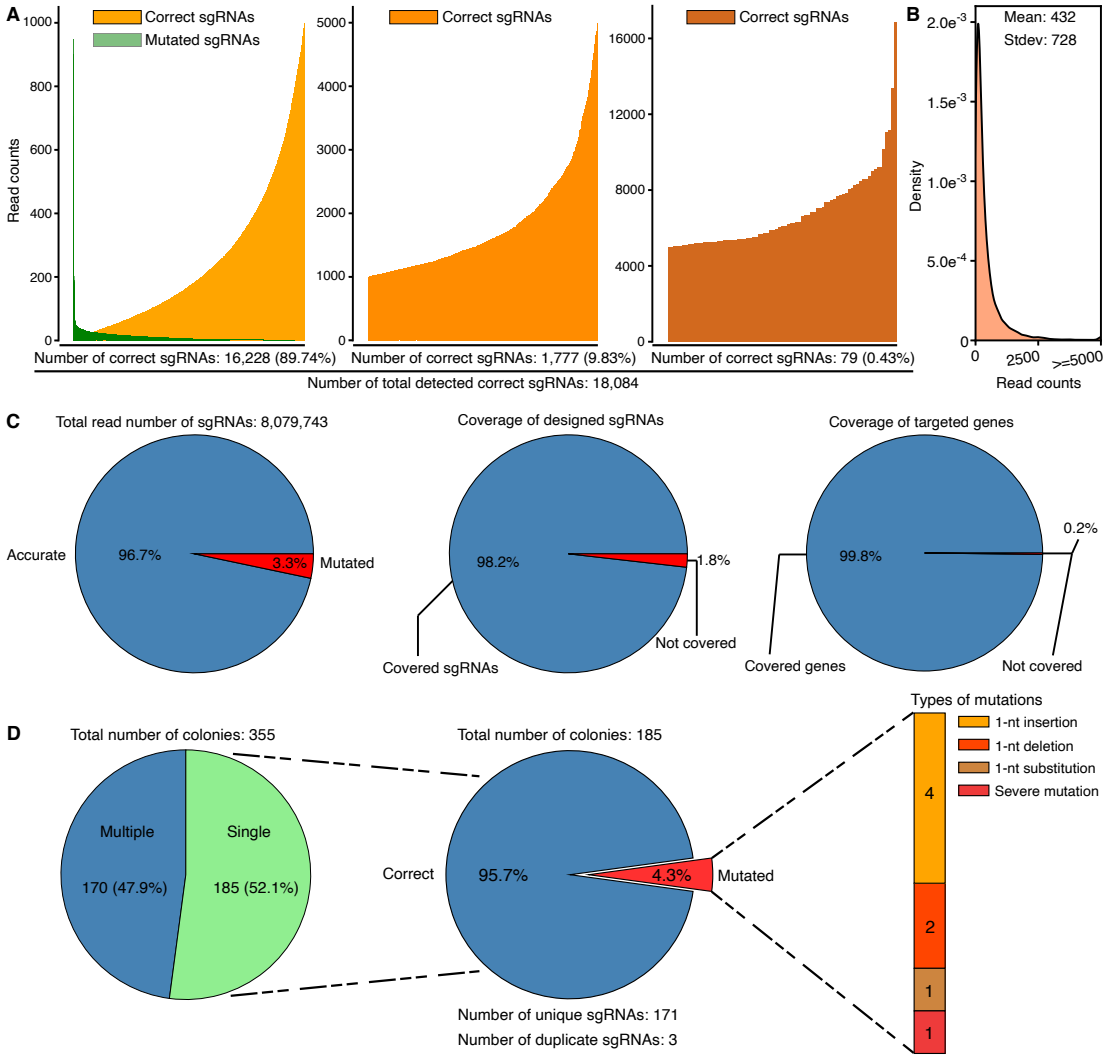
Supplemental Figures

A

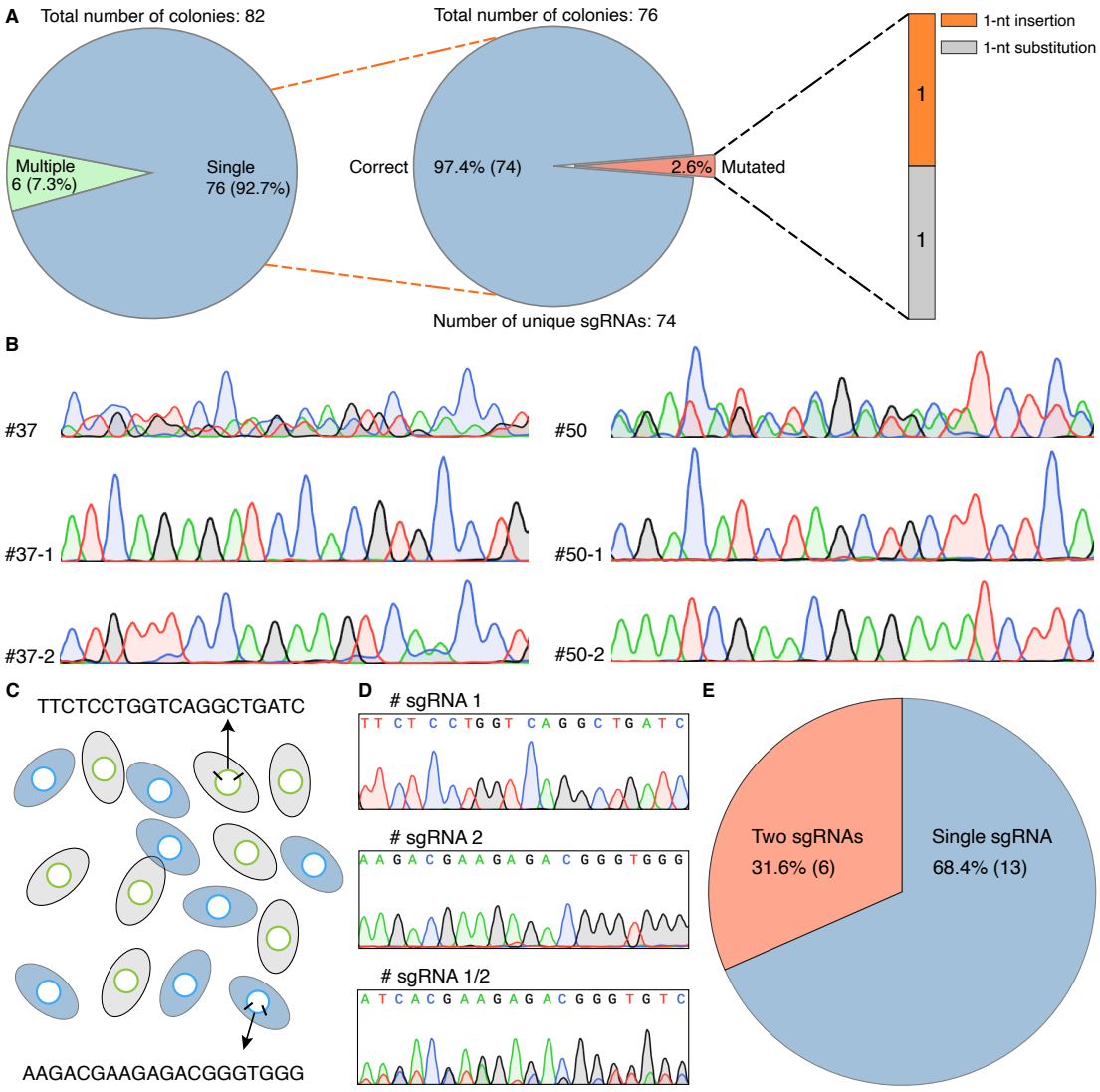
Term	Gene number	Term	Gene number	Term	Gene number	Term	Gene number
Flower	520	Ovule	69	Endosperm	59	Acyltransferase	173
Pollen	488	Ovary	2	Fatty acid	133	Filling	1
Anther	69	Floral	128	Oleosin	16	Desiccation	27
Stamen	35	Pollination	4	Lipid droplet	8	Dormant	1
Pistil	13	Pod	11	Aleurone	2	Dormancy	29
Stigma	19	Seed	738	Filament	72	GA4	1
Carpel	113	Seed coat	39	Embryo	723	Transcriptome	391
Total genes of interest: 3,054 (after removing duplicates)							

B

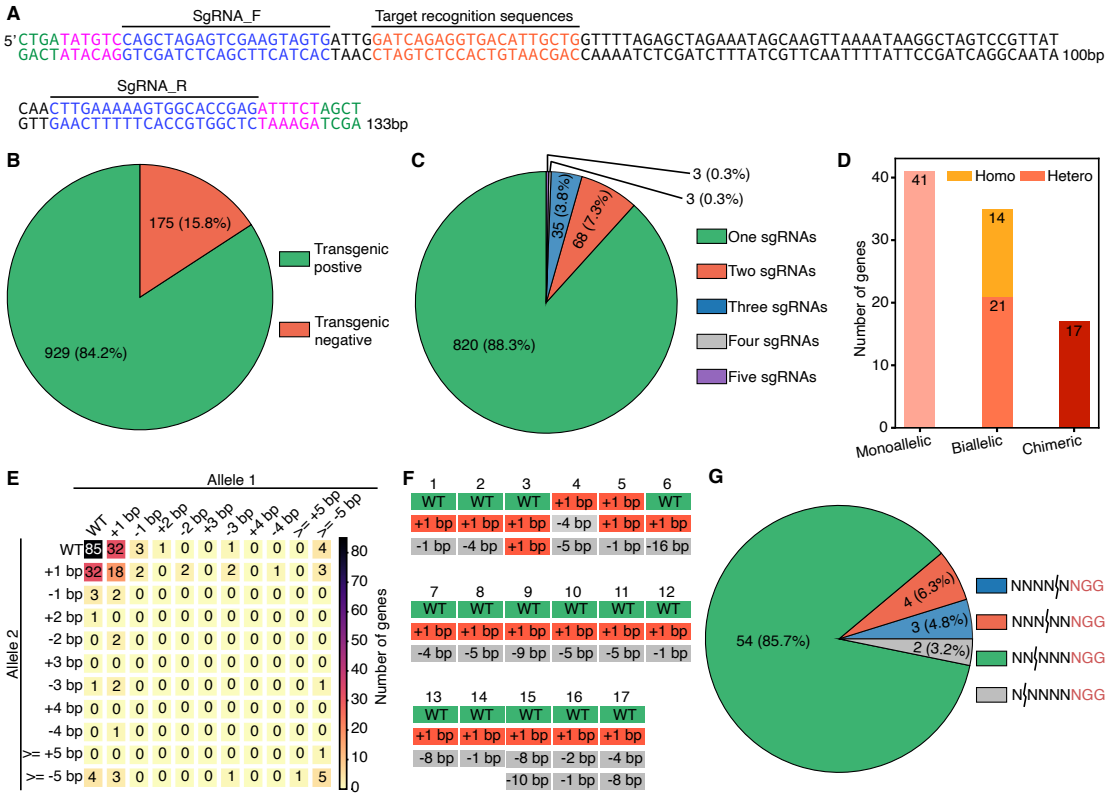
Supplemental Fig. S1. Screening for *A. thaliana* genes and vector backbone of the transfer DNA (T-DNA). **(A)** Screening for *A. thaliana* genes. Among the total genes, 2,662 were obtained by searching within the TAIR website using 26 queries (light green, except for GA4) related to reproductive organs. The gene *GA4*, of great interest, which involves the gibberellic acid biosynthetic pathway, was also picked up. The last 391 genes were obtained following the process below: First, 808 highly expressed rapeseed genes (with average FPKM > 125) were identified in our transcriptomes of seeds under six developing stages (10 DAF, 17 DAF, 24 DAF, 30 DAF, 36 DAF, and 42 DAF); Next, 391 *A. thaliana* genes homologous to the 808 characterized *B. napus* ones were obtained. **(B)** Primary functional elements in the T-DNA backbone of the CRISPR-Cas9 expression vector DP-05. Hyg represents the hygromycin-resistant gene.



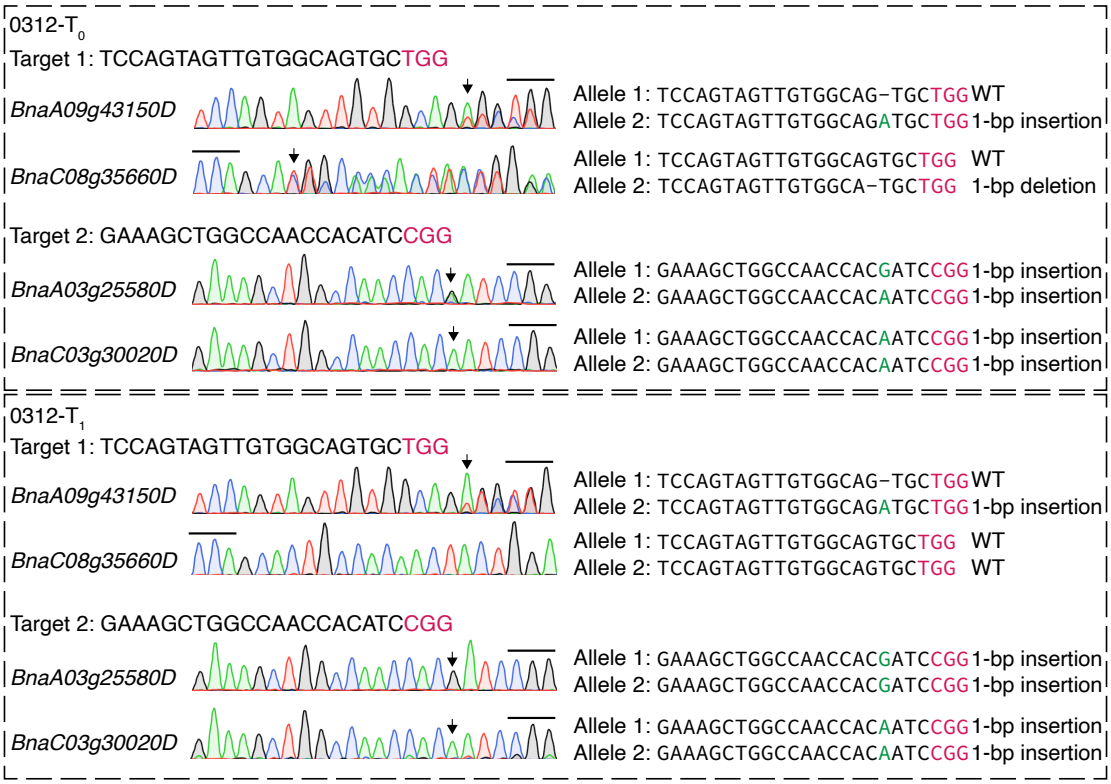
Supplemental Fig. S2. Quality evaluation of the pooled plasmids and agrobacteria. **(A)** SgRNA read counts from the constructed pooled plasmids detected by high-throughput sequencing (NGS). **(B)** Density plot of the correct sgRNAs with different read counts derived from the NGS data. The average number and the standard deviation are shown above. **(C)** Quality evaluation of the plasmid library derived from the sequencing data. The left panel is the ratio of the accurate sgRNAs to the mutated ones; the middle panel is the proportion of sgRNAs detected in the NGS data to the total selected; the right panel is the proportion of target genes derived from the NGS data to the whole designed. **(D)** Accuracy and representation analysis of the sgRNAs utilizing 355 randomly chosen *Agrobacterium* colonies.



Supplemental Fig. S3. Evaluation of single sgRNA and characterization of positive transgenic lines co-transformed with two distinct sgRNAs. **(A)** Single sgRNA analysis with 82 randomly selected *Agrobacterium* colonies from those transformed with the three times diluted pooled plasmids. **(B)** Sequencing results of 20-nt target recognition sequences in two *Agrobacterium* colonies of the first generation and four colonies of the subsequent generation. Two different sgRNAs were found in the first generation; however, these sgRNAs were separated into distinct single colonies in the following descent. **(C)** The schematic diagram shows the pooled agrobacteria solution harboring two distinct sgRNAs in equal amounts. The arrows point to the 20-nt target recognition sequences. **(D)** Representative Sanger sequencing chromatograms of the 20-nt target recognition sequences in the 19 positive transgenic plants transformed with the mixed agrobacteria described in **(C)**. **(E)** The ratio of the transgenics containing two distinct sgRNAs to those harboring a single one.



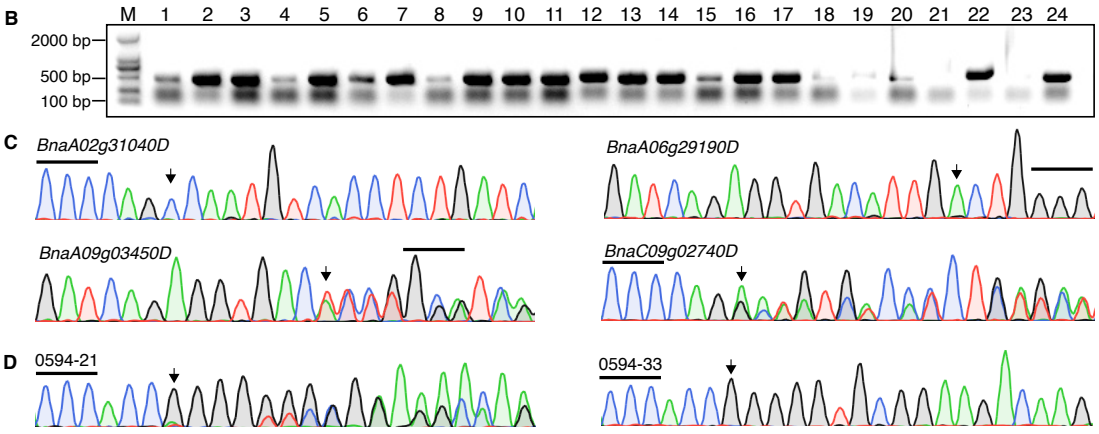
Supplemental Fig. S4. Identifying 20-nt target recognition sequences in the transgenic rapeseed plantlets and characterizing the genotyping results in the T_0 generation. **(A)** An example of the sgRNA amplicon for high-throughput sequencing. The four base pairs marked in green are the protected nucleotides, and nearby are the 6-bp barcodes. **(B)** The ratio of the positive transgenic plants to the negative ones. **(C)** Percentages of the transgenics harboring different numbers of distinct sgRNAs. **(D)** Summary of the genotyping results of the edited genes. Homo, homozygote; Hetero, heterozygote. **(E)** Overview of the genes with two alleles in the editing identification of the target sites. **(F)** Detailed information for the 17 genes with at least three identified alleles in the editing interrogation of the target sites. **(G)** Percentage of cleavage sites cut by the Cas9 protein derived from the edited alleles with 1-bp insertions.



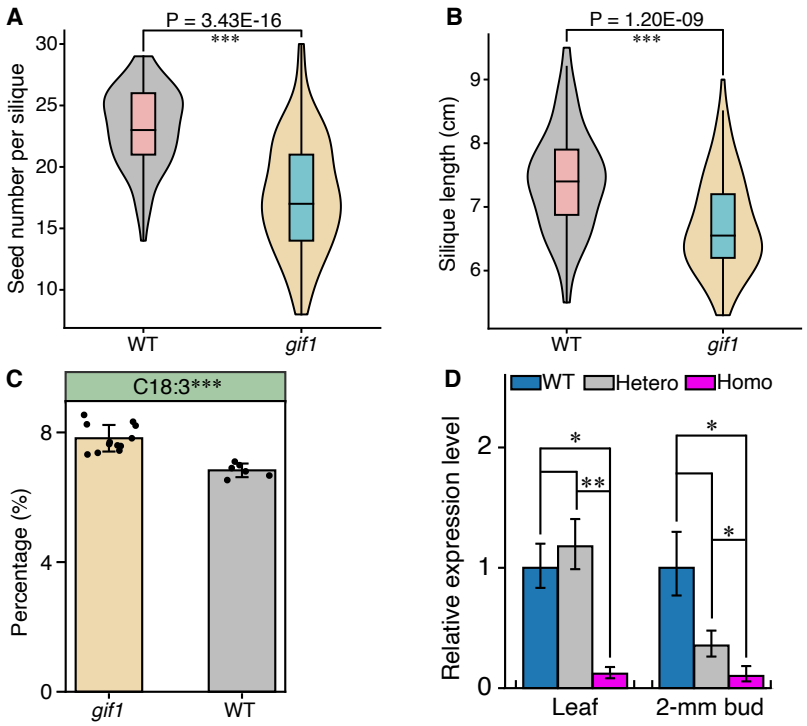
Supplemental Fig. S5. Representative sequencing results of the four target sites in line 0312 harboring two distinct sgRNAs in the T_0 and T_1 generation. PAM sequences are marked in red and indicated by the black bars above the sequencing chromatograms. Inserted nucleotides in the target sites are marked in green and pointed by solid arrows.

A

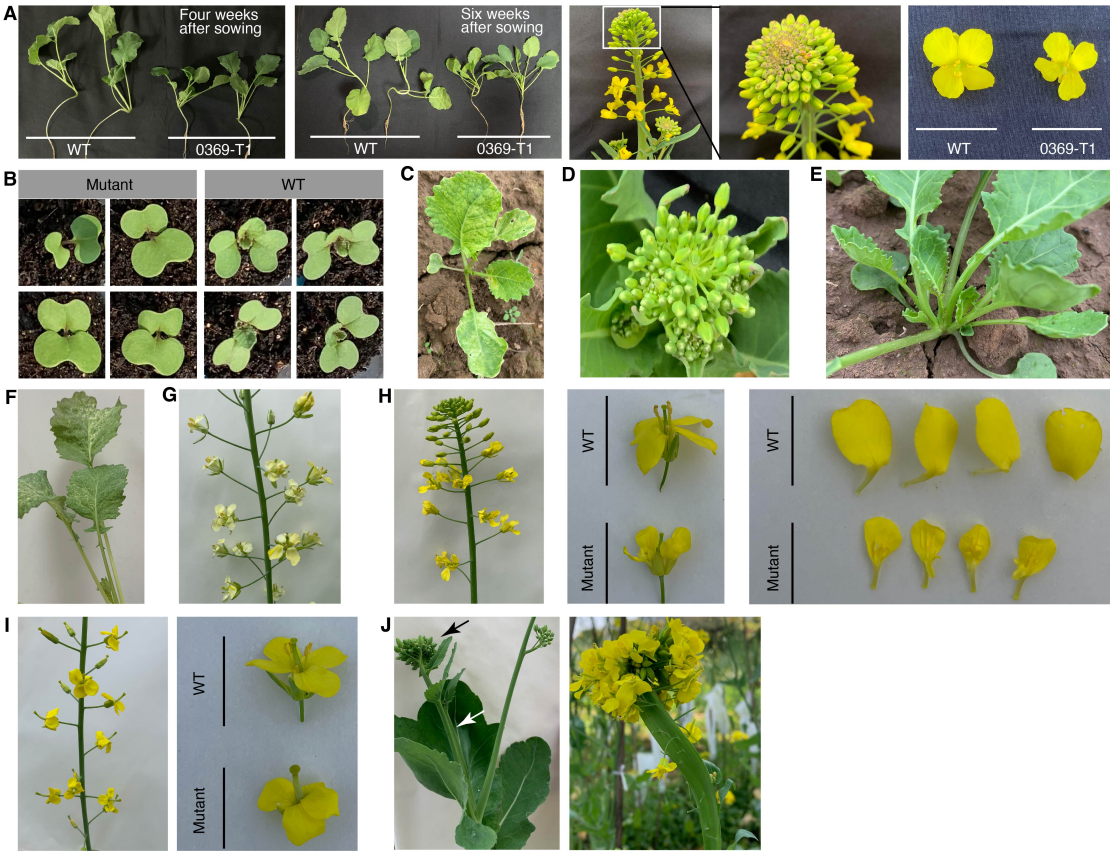
Line	Target	Marker free	Genes	Mutations
0369-1	GATCAGAGGTGACATTGCTG	Yes	*BnaA02g31040D*	+1 bp/+1 bp (homozygote)
			BnaA06g29190D	+1 bp/+1 bp (homozygote)
			BnaA09g03450D	-3 bp/-3 bp (homozygote)
			BnaC09g02740D	+1 bp/+1 bp (homozygote)
0369-4	GATCAGAGGTGACATTGCTG	Yes	*BnaA02g31040D*	+1 bp/+1 bp (homozygote)
			BnaA06g29190D	+1 bp/+1 bp (homozygote)
			BnaA09g03450D	-3 bp/-3 bp (homozygote)
			BnaC09g02740D	-1 bp/-1 bp (homozygote)
0369-19	GATCAGAGGTGACATTGCTG	Yes	*BnaA02g31040D*	+1 bp/+1 bp (homozygote)
			BnaA06g29190D	+1 bp/+1 bp (homozygote)
			BnaA09g03450D	-3 bp/-34 bp (heterozygote)
			BnaC09g02740D	+1 bp/-1 bp (heterozygote)
0369-21	GATCAGAGGTGACATTGCTG	Yes	*BnaA02g31040D*	+1 bp/+1 bp (heterozygote)
			BnaA06g29190D	+1 bp/+1 bp (homozygote)
			BnaA09g03450D	-3 bp/-3 bp (homozygote)
			BnaC09g02740D	+1 bp/+1 bp (homozygote)
0594-21	CTTGTCTTCCGCACCCCGGT	No	*BnaA09g28510D*	+1 bp/WT (heterozygote)
0594-33	CTTGTCTTCCGCACCCCGGT	No	*BnaA09g28510D*	+1 bp/+1 bp (homozygote)



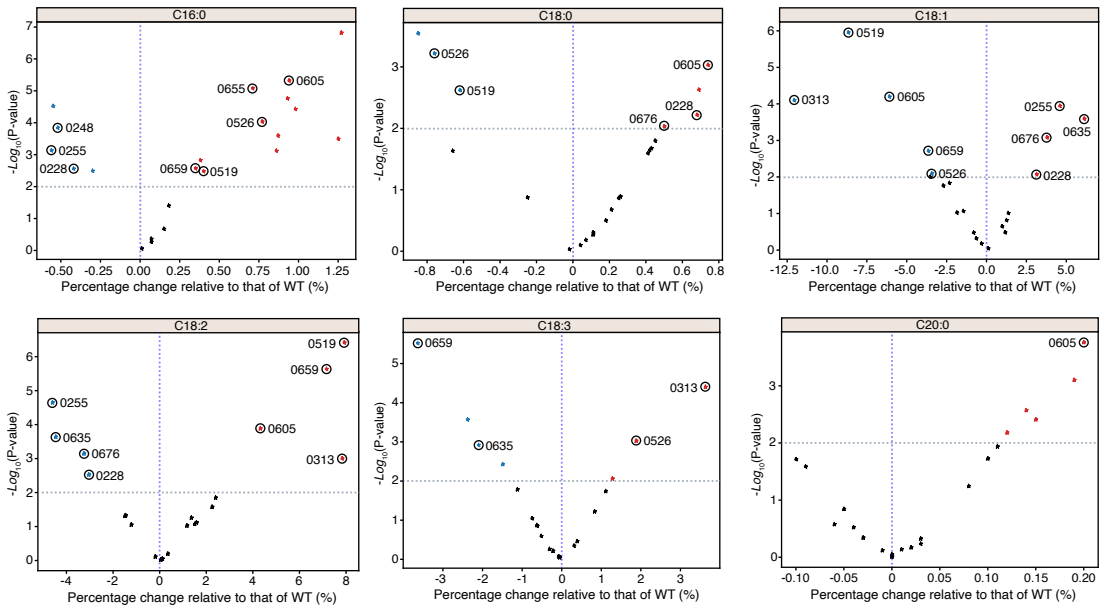
Supplemental Fig. S6. Mutation examination and identification of marker-free plants in the *gif1* and *kcs6.A09* mutants of the T_1 generation. **(A)** Detailed editing information for the *gif1* (T_0 ID: 0369) and *kcs6.A09* (T_0 ID: 0594) mutants used for reverse genetics. **(B)** Detection of marker-free plants by agarose gel electrophoresis. The above numbers indicate the plant IDs of *gif1* mutants in the T_1 generation. **(C)** Sanger sequencing chromatograms of the four target sites in *gif1* mutant (0369-19). **(D)** Sanger sequencing chromatograms of the target site in the two *kcs6.A09* mutants. In **(C)** and **(D)**, PAM sequences are marked by solid black lines, and the arrows indicate the editing sites.



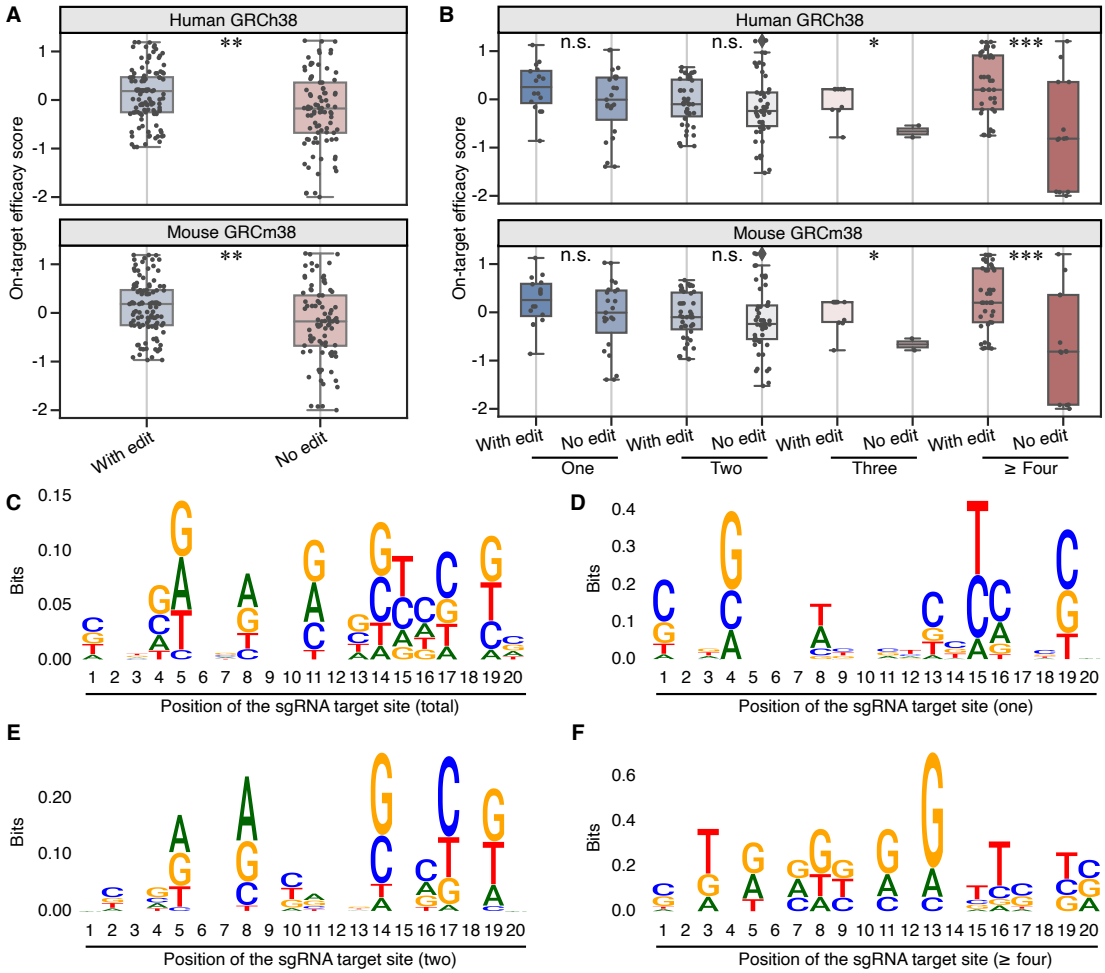
Supplemental Fig. S7. Capability of reverse genetic screening with the post-genotyped positive transgenic plants of T_1 generation. **(A)** Decrease of seed number per silique in the *gif1* mutant (n=100) compared to wild type (n=70). **(B)** Decrease of silique length in the *gif1* mutant relative to the wild type (n=108). **(C)** Increase of linolenic acid (C18:3) percentage in the *gif1* mutant (n=12) relative to that of the wild type (n=6). **(D)** Relative expression level of *BnKCS6.A09* gene in leaf and 2-mm bud tissues. WT, wild type; Hetero, heterozygote; Homo, homozygote. In **(A)** and **(B)**, the widest part of the violin plot indicates the highest point density and the top and bottom are the maximum and minimum data, respectively. The middle part of the violin plot is a box plot, the center line represents the median, and the top and bottom lines indicate the 25th and 75th percentiles. In **(A)**-**(D)**, Student's *t*-test (**P*<0.05, ***P*<0.01, and ****P*<0.001) was used for the statistical analysis. Error bars in **(C)**-**(D)** are standard deviations.



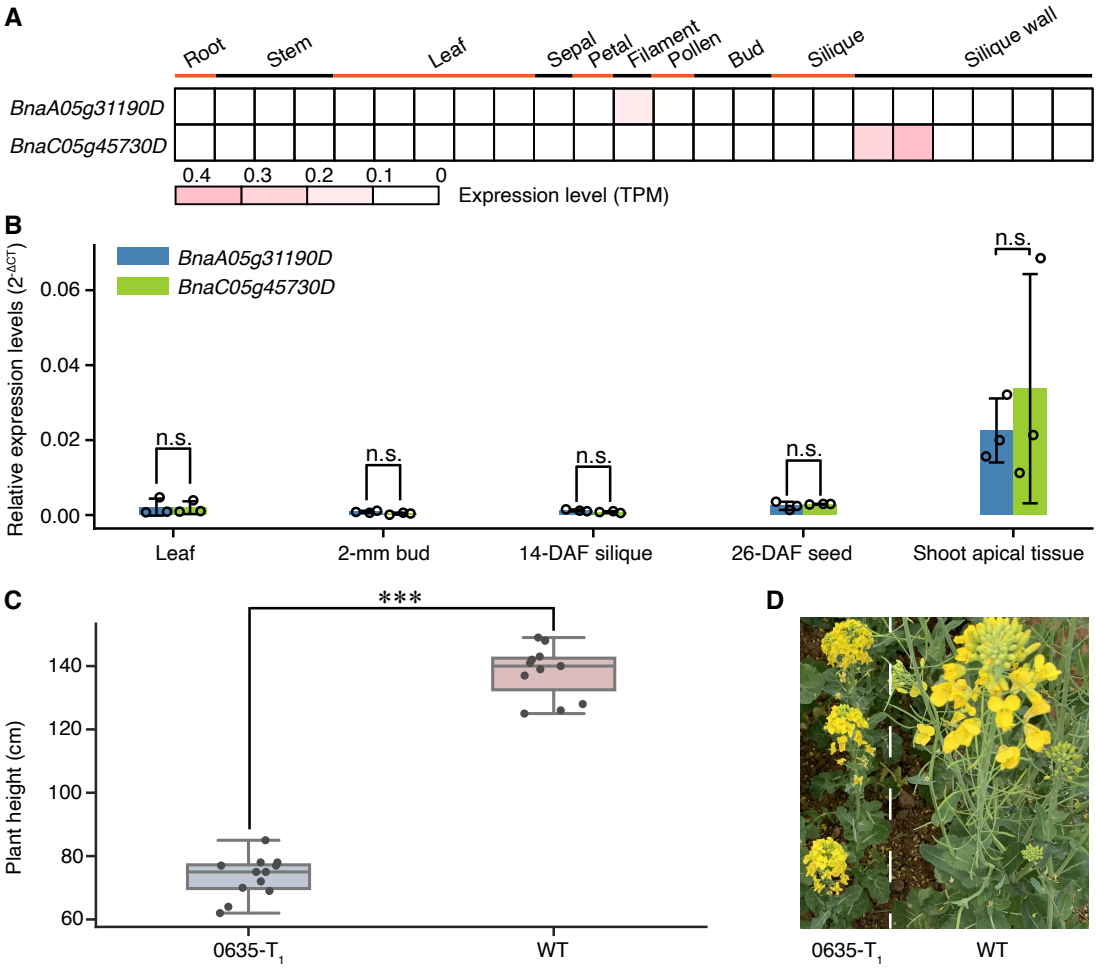
Supplemental Fig. S8. Other observed abnormal phenotypes in the constructed mutant collection. **(A)** Representative photos show the abnormal phenotypes of the *gif1* mutant in the T_1 generation relative to WT's. Perturbed primary inflorescences, leafier traits, and smaller flowers with narrow petals than WT were seen in the mutants. **(B)** Mutants with delayed emergence of true leaves in the T_1 generation. **(C)** The mutant with wrinkled yellow leaves in the T_1 generation. **(D)** Abnormal primary inflorescence was found in the T_0 generation. **(E)** The transgenic seedling displays altered architecture in the T_1 generation. **(F)** The mutant with wrinkled white leaves in the T_1 generation. **(G)** The mutant shows white flowers in the T_0 generation. **(H)** Abnormal petals were found in the T_0 generation. **(I)** Mutant displays shorter filaments found in the T_0 generation. **(J)** Fasciculate flowers in the T_0 generation.



Supplemental Fig. S9. Fatty acid compositions of the self-pollinated seeds from the 23 randomly selected transgenic lines ahead of editing interrogation of the target sites. These seeds were from the plants of the T_0 generation. Each dot represents a transgenic line, and IDs of the plants used for genotyping and showing significant changes in fatty acid compositions are marked alongside. Student's *t*-test was used for the statistical analysis.



Supplemental Fig. S10. On-target efficacy scores and sequence compositions of the examined sgRNAs. **(A)** On-target efficacy scores of total interrogated non-mutagenic and mutagenic sgRNAs. Mann-Whitney U test (***P*<0.01) was used for the statistical analysis. **(B)** The on-target efficacy scores of sgRNAs targeting different numbers of homoeologs. Student's *t*-test (n.s.–not significant, **P*<0.05, and ****P*<0.001) was used for the statistical analysis. Numbers of the homoeologs are indicated below the solid black lines. In (**A**) and (**B**), the scores were predicted by GPP sgRNA Designer with the human and mouse reference genomes. In the box plots, the center line represents the median, and the top and bottom lines indicate the 25th and 75th percentiles. Each dot represents a data point. (**C-F**) Sequence compositions of total mutagenic sgRNAs (**C**), mutagenic sgRNAs targeting only one homoeolog (**D**), two homoeologs (**E**), and ≥ four homoeologs (**F**). The numbers 1 to 20 indicate the nucleotide positions of the sgRNAs.



Supplemental Fig. S11. Expression features of the target genes and dwarf phenotype of line 0635 in the T_1 generation. (**A**) The expression profile of the target genes of line 0635 in ten indicated tissues of the ZS11 cultivar. Genes with TPM ≥ 1 were identified as expressed. (**B**) Relative expression levels of the two target homoeologs in the five tissues of J9709 WT. DAF, days after flowering. The statistical analysis was performed by Student's *t*-test (n.s.–not significant). Error bars indicate the standard deviation. (**C**) Comparison of the plant height between the mutants of line 0635 in the T_1 generation and the J9709 WT. Each dot represents a data point of an individual plant. Student's *t*-test (***P<0.001) was utilized for the statistical analysis. In the box plots, the center line represents the median, and the top and bottom lines represent the 25th and 75th percentiles. (**D**) The representative photo shows the dwarf phenotype of the line 0635 mutants in the T_1 generation compared to J9709 WT.

Summaries for Supplemental Tables

Supplemental Table S1. Details of selected *A. thaliana* genes and their homologies in *B. napus*.

Supplemental Table S2. Details of the 18,414 selected 20-bp target sites and the corresponding target genes in *B. napus*.

Supplemental Table S3. Sequencing results of the 20-nt target recognition sequences within the sgRNAs in each regenerated *B. napus* line.

Supplemental Table S4. Detailed information for editing evaluation of the target sites in randomly selected transgenic lines.

Supplemental Table S5. Details of sequence variations between the reference and the recipient target sites.

Supplemental Table S6. Detailed information for off-target effect examination in selected Mutants.

Supplemental Table S7. Detailed information for mutation interrogation of (off-) target sites in selected transgenic lines that show abnormal phenotypes.

Supplemental Table S8. Primers used for amplifying the sgRNAs.

Supplemental Table S9. Primers used for RT-qPCR experiments.

Summaries for Supplemental Code

Supplemental Code S1. Source data and code used for generating Fig. 1.

Supplemental Code S2. Source data and code used for generating Fig. 2.

Supplemental Code S3. Source data and code used for generating Fig. 3.

Supplemental Code S4. Source data and code used for generating Fig. 4.

Supplemental Code S5. Source data and code used for generating Fig. 5.

Supplemental Code S6. Source data and code used for generating Supplemental Fig. S2.

Supplemental Code S7. Source data and code used for generating Supplemental Fig. S3.

Supplemental Code S8. Source data and code used for generating Supplemental Fig. S4.

Supplemental Code S9. Source data and code used for generating Supplemental Fig. S7.

Supplemental Code S10. Source data and code used for generating Supplemental Fig. S9.

Supplemental Code S11. Source data and code used for generating Supplemental Fig. S10.

Supplemental Code S12. Source data and code used for generating Supplemental Fig. S11.

Supplemental Code S13. Python script used for extracting the 20-nt target recognition sequences of the sgRNAs. In the "combination.txt" file, the 6-nt barcodes in the forward and reverse primers are structured respectively in the first and second columns; the third column shows the well coordinate of each 96-well plate. The first column of the generated file indicates the read counts of the corresponding sgRNAs, followed by the well coordinate in the second column. The sgRNA sequences are given in the third column. Even though noise is presented in the generated results, each plant's exact sgRNA sequence(s) can be easily determined according to the read counts.

Summaries for Supplemental Datasets

Supplemental Dataset S1. Sanger sequencing results of the sgRNAs amplified from the 355 randomly selected *Agrobacterium* single colonies. Forward primers were used for the sequencing.

Supplemental Dataset S2. Sanger sequencing results of the sgRNAs amplified from the 82 randomly selected *Agrobacterium* single colonies. Reverse primers were used for the sequencing.

Supplemental Dataset S3. SgRNA sequences of the 28 selected positive transgenic plants obtained from the NGS were verified by Sanger sequencing. Forward or reverse primers were used for the sequencing.

Supplemental Dataset S4. Sanger sequencing data for the editing interrogation of the target sites.

Supplemental Dataset S5. Sanger sequencing data for the off-target effect evaluation.

Supplemental Dataset S6. Sanger sequencing data for the editing interrogation of the target sites in T_1 plants of lines 0369 and 0594.

Supplemental Dataset S7. Sanger sequencing data for the editing interrogation of the target sites in the plants used for the forward genetic screening.

Supplemental Dataset S8. SgRNA design data for selecting the target sites following the three principles.