

Genome editing

Insertion of a GFP sequence into the *TRAC* locus using the long insertion workflow in TrueDesign Genome Editor

Highlights

- With TrueDesign Genome Editor, users can now design complex knock-in experiments such as inserting a CAR transgene into the *TRAC* locus. Insert sequences can range in size up to 10 kb with varying lengths of homology arms of up to 1 kb. Here, we demonstrate a use case for the insertion of a green fluorescent protein (GFP) sequence into the *TRAC* locus.
- Several factors, including GC content, repetitive DNA, and homopolymer content, are checked in the designed donor DNA to ensure manufacturing feasibility.
- TrueDesign Genome Editor scores and ranks gRNAs based on state-of-the-art on-target and off-target algorithms.
- TrueDesign Genome Editor also helps design various primer pairs for verification of knockout and knock-in.
- Users can visualize SNPs from 8 different populations to assist with variant-aware design of genome editing experiments.

Introduction

Genome editing technologies have given scientists the power to unlock a variety of new applications and experimental approaches. The [Invitrogen™ TrueDesign™ Genome Editor](#) is free design software that enables scientists to easily design and order reagents for successful genome editing experiments, including our flagship [Invitrogen™ TrueCut™ Cas9 proteins](#) and [TrueGuide™ Synthetic gRNAs](#). Recent advancements and future expectations for applications of gene editing technologies have inspired the latest upgrades to TrueDesign Genome Editor software, including:

- **Large donor insertions**—Design and order donor DNA up to 12 kb (including varying lengths of homology arms). This feature will enable users to design large knock-in or tagging experiments, such as inserting a chimeric antigen receptor (CAR) transgene into the *TRAC* locus as described in this application note.

- **Targeting and editing in non-protein-coding regions**—Easily design and order experimental reagents to target noncoding genes with the addition of 40,000 new genes, such as long noncoding RNAs and miRNAs. Regulatory regions such as promoter and enhancer regions can also be targeted.
- **Inclusion of genetic variation data**—Conveniently visualize and utilize SNP information of 8 different populations from the Genome Aggregation Database (gnomAD) [1] during the design process. It is critical to consider genetic diversity when designing a genome editing experiment, in order to enable efficient designs and specificity in the target population.
- **Updated on-/off-target scoring algorithms**—Utilize an updated scoring algorithm for precise gRNA designs. TrueDesign Genome Editor software now includes a Cutting Frequency Determination (CFD) score [2] for enhanced gRNA specificity, along with the addition of a new and updated on-target scoring algorithm, Rule Set 3 [3], to increase efficacy.

This application note describes how the TrueDesign Genome Editor can now be used to design large knock-in experiments with insertion sequences ranging anywhere from 31 bp to 10 kb along with varying lengths of homology arms. Large donor sequences designed using TrueDesign software will be synthesized by [Invitrogen™ GeneArt™ Gene Synthesis Services](#). To demonstrate this feature, we will describe the insertion of a green fluorescent protein (GFP) sequence into the *TRAC* locus.

Use case: Insertion of a GFP sequence in the *TRAC* locus

Chimeric antigen receptors (CARs) are synthetic receptors used to reprogram T cells to identify specific antigens on cancer cells and kill them. With the advent of genome editing technologies such as the CRISPR-Cas9 system, directing a large transgene such as a CAR sequence to a desired location is now feasible and effective.

The T cell receptor (TCR) consists of alpha and beta chains expressed as heterodimers. The knockout of one of these chains is sufficient to disrupt TCR surface expression. The T cell receptor alpha constant (*TRAC*) locus is an excellent target for knockout of TCR surface expression and simultaneous knock-in of the CAR transgene. Using the new features of the TrueDesign Genome Editor, we can design highly efficient and specific CRISPR-Cas9 gRNA or TALEN™ mRNA pair sequences to knock in the CAR transgene to the *TRAC* locus and order all required reagents to support the workflow (Figure 1).

To help ensure efficient insertion of the CAR transgene, we will first test this system with a GFP sequence as described in this application note. The endogenous promoter of the TCR enables transcription of the insert sequence encoding a self-cleaving peptide and GFP. This knock-in will result in the expression of cytosolic GFP and knockout of TCR expression, which can be easily detected via flow cytometry.

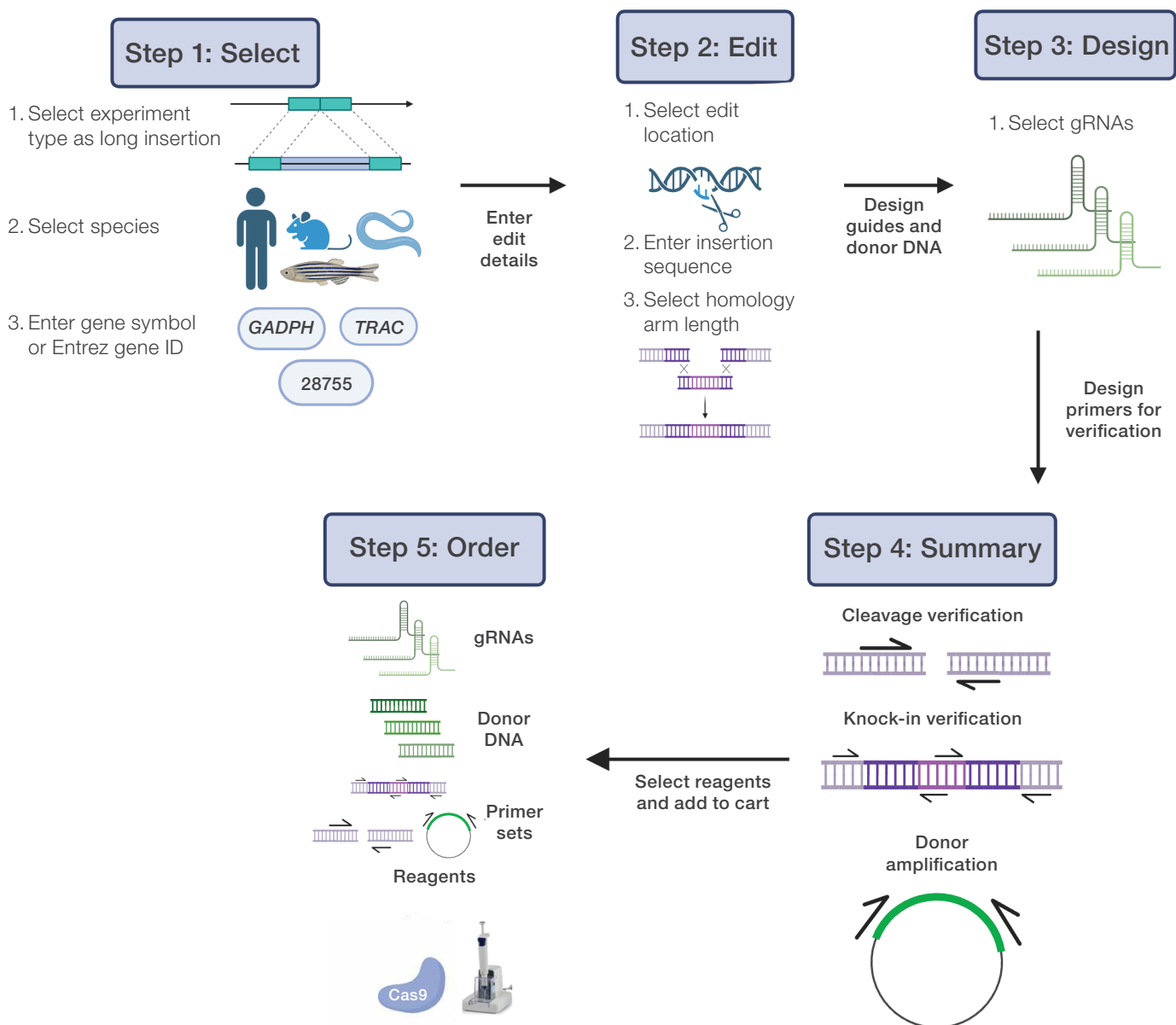


Figure 1. Overview of steps to insert a transgene into the *TRAC* locus. Created with [BioRender.com](https://www.biorender.com).

Step 1: Select

1. Select “Long Insertion (31 to 10k bases)” as your experiment type (Figure 2) and click “Next”.

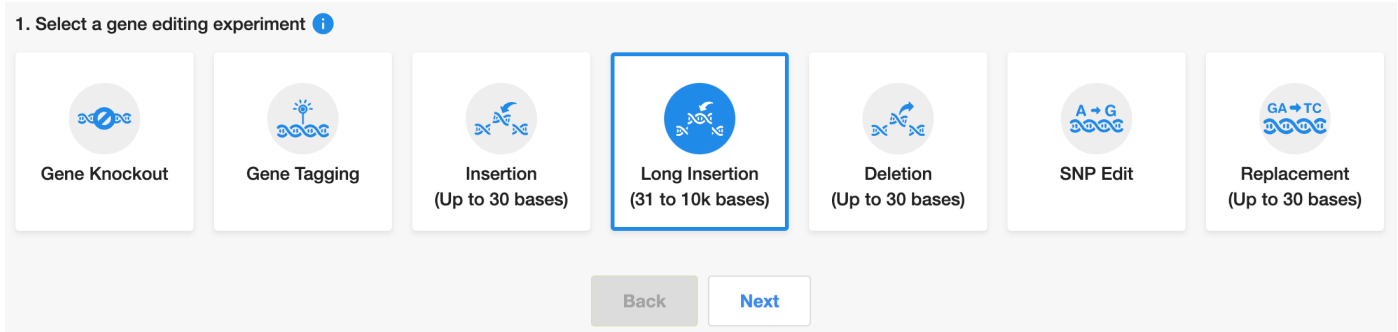


Figure 2. Selection of “Long Insertion” experiment type.

2. Select “Human” for “Species” and click “Next”. Then, select “Gene Symbol or Entrez ID” for “Input Type” and click “Next”. In the input box presented on the next page, type “TRAC” or its Entrez ID of “28755” and click “Search Gene”.
3. The transcript table presented on the following page displays a single transcript for *TRAC*. As NCBI’s RefSeq database does not have a defined transcript for *TRAC*, TrueDesign Genome Editor displays the corresponding Ensembl™ Transcript ID (ENST00000611116.2). Select the transcript and click “Next”.

Step 2: Edit

4. Navigate to the start of exon 1 by using the slider on the Transcript Topology graphic, and click on sequence location 22,547,531 (edit location) to insert the GFP sequence (Figure 3).



Figure 3. Navigation to start of exon 1 (upper red arrow) and selection of insertion location (lower red arrow).

- Paste the insert sequence (Table 1) in the text box titled “Enter Sequence to Insert (31 to 10k bases)” (Figure 4). Once you enter the sequence, the length of the sequence is displayed; ensure that the length of the sequence is correct.

Table 1. Insert sequence used in this application note.

Sequence name	Sequence
Insert sequence	GCTACTAACTTCAGCCTGCTGAAGCAGGCTGGAGACGTGGAGGAGAACCCTGGACCTATGGTGAGCAAGGGCGAGGAGCT GTTACACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGC GAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCACCC TCGTGACCACCTTCACCTACGGCGTGCACTGCTTCGCCCGCTACCCCGACCACATGAAGCAGCAGGACTTCTCAAGTCC GCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGA AGTTCGAGGGCGACACCCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCA CAAGCTGGAGTACAACACTACAACAGCCACAAGGTCTATATACCCGCCGACAAGCAGAAGAACGGCATCAAGGTGAACCTCAA GACCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGGCGACGGCCC CGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATG GTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA

Figure 4. Text box for entering the insert sequence.

- Once you click “OK”, TrueDesign Genome Editor will check the entered GFP sequence for manufacturing feasibility against the criteria that GeneArt Gene Synthesis Services uses to evaluate sequences.
 - Sequence passes diagnostic check:** A success message will be displayed (Figure 5), and a green arrow indicating the size and location of the insertion sequence will be displayed on the Sequence Editor.

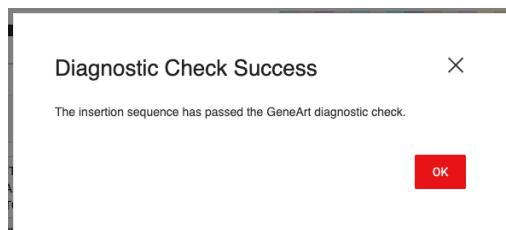


Figure 5. Diagnostic check success message.

- Once your sequence passes the diagnostic check, the preferred length for homology arms can be chosen from the provided options, or a custom length (between 35 and 1,000 bp) can be entered. For optimal performance of the GFP knock-in, we recommend entering 300 bp using the Custom Length option for optimal performance (Figure 9). Click “Next” to design the gRNAs and homology arms.

ATCACTGGCATCTGGACTCCAGCCTGGGTTGGGCAAGAGGGAAATGAGATCATGTCTCAACCTGATCCCTCTGTCCACAGATATCCAGAACCCTGACCCCTGCGGTGACACAGCTGAGAGACTAAATCCAGTGACAAGTCTGTCCCTATCCAGGATTTGATTCTCAACAATGTGTACAAAGTAAGGATAGTGAACCGTAGACCTGAGGTCGACCAACCCCTTCTCCCTTACTCTAGTACAGATTGGGACTAGGAGAAGGGTGTCTATAGTCTTGGGACTGGGACCGGCACATGGTGCAGCTCTGAGATTTAGGTCAGCTGTTGAGACAGACGGATAGGCTAAACTAAGAGTTTGTACACAGTGTTCATTCCTA

4777-bp insertion at 22547531

5' Flank Exon 1

22547440 22547460 22547480 22547500 22547520 22547540 22547560 22547580 22547600 22547620

Enter Sequence to Insert (31 to 10k bases)

TACCAGCAGAACCACCCATCGGCGACGGCCCGTGTCTGCCCCGACAACCACTACCTGAGCACCAGTCCGCGCTGAGCAAGACCCCAACGAGAAAGCGGATCACATGTCTGCTGAGTTCTGACCCGCCCGGGATCACTCTGGCATGGACGAGCTGTACAAGTAA

Please note that insertion sequence will undergo GeneArt diagnostic check.

Length (bp) : 777

OK

Select Homology Arm Length

35bp 100bp 500bp Custom Length (bp): 300

Range: 35bp to 1000bp

Back Next

Edit Type	Reference Locus	Sequence	Actions
Long Insertion	22547531	Custom Sequence	

Figure 9. Edit page showing insertion site, length, and selection of homology arm length.

Step 3: Design

- TrueDesign Genome Editor will display gRNAs and donor DNA preferentially designed to target closest to the intended edit location, up to 40 bp away. The further the gRNA is from the edit location, the lower the knock-in efficiency. All designed donor DNA sequences (now including the homology arms) will be checked for GeneArt Gene Synthesis manufacturing feasibility. If the donor DNA fails the check, you will not be able to select that gRNA and the row will be grayed out. In such cases, please select different gRNAs or TALEN pairs, if available, or change your insertion location or homology arm length.

For each CRISPR-Cas9 gRNA and TALEN pair, the donor DNA sequence can be viewed by clicking on the eye icon in the Donor DNA column. To view and annotate the donor DNA sequence in your chosen cloning software, download the FASTA file by clicking on the download icon in the Donor DNA column. Ensure that the donor DNA is in frame to achieve optimal expression of GFP.

Note: When designing donor DNA, silent mutations or appropriate modifications may be introduced to prevent it being recut. To achieve this, TrueDesign Genome Editor will either disrupt the PAM sequence or the seed sequence (8 PAM-proximal nucleotides) automatically using the following rules:

- Disruption of PAM sequence:
 - If the PAM sequence is in the coding region of a gene, the wobble position of the codon is modified.
 - If the PAM sequence is in a noncoding region, the 5' G of the PAM is modified to a C.
- Modification of seed sequence:
 - If the seed is in the coding region of a gene, the first seed nucleotide in a wobble position, starting from the most PAM-proximal nucleotide, is modified.

- The top three gRNAs (indicated by the green check marks, Figure 10) are recommended based on the updated scoring algorithm (combination of Rule Set 3 on-target score and an aggregate Cutting Frequency Determination off-target score).

Points to note:

- Scores colored green are predicted to have high on-target activity and low off-target activity.
- The top-ranked gRNA overlaps with a low-frequency variant (SNP) as indicated in the CRISPR gRNA Targets table (Figure 10). The allele frequency of this variant in the overall population is 8%, indicating that this gRNA will be most efficient in ~92% of the population. Selecting gRNAs or TALEN pairs that overlap with SNPs with higher allele frequencies might result in less-effective targeting. A workaround might be to select gRNAs or TALEN pairs that do not overlap with SNPs.

CRISPR gRNA Targets		TALEN Targets		Score (%)	Off-Targets in Reference Genome	Off-Targets in Variant Genome	Edit site distance	Donor DNA
<input type="checkbox"/>	CRISPR							
<input type="checkbox"/>	TCAGGG			61.87	32		27 bp	
<input type="checkbox"/>	TGGATTAGAGTCTCTCAGC	TGG		51.84	37	Click to perform off-target search including variants <input type="button" value="Search"/>	8 bp	
<input type="checkbox"/>	GCTGGTACACGGCAGGGTCA	GGG		50.84	54		10 bp	
<input type="checkbox"/>	AGAGTCTCTCAGCTGGTACA	CGG		49.06	52		1 bp	

Figure 10. CRISPR gRNA Targets table on the design page.

- The off-target sites (TrueDesign Genome Editor reports all off-target sites with NGG, NGA, and NAG PAM sites with up to 3 mismatches to the gRNA) for each gRNA can be viewed by clicking on the number in the “Off-Targets in Reference Genome” column. This displays a pop-up table (Figure 11) with the:

- Off-target recognition sites
- Highlighted mismatched nucleotides with respect to the gRNA
- Locations of the mismatched nucleotides
- Genomic locations of the off-target sites
- Annotations of whether the off-target site is in an exon, intron, or intergenic region and if the off-target site overlaps with an oncogene or a tumor suppressor gene (as annotated by Vogelstein et al. [4])
- CFD scores, which signify the probability of cutting at the off-target site (i.e., if the CFD score is 0.317, the probability of cutting at that site is 31.7%).

Off-target recognition site	PAM	Mismatches Count (Position)	Genomic Locations	Annotations	CFD Score
TCAGAGTTCTGGATGCTCTGA	AGG	3(5,15,20)	chr5[29633302]	intergenic	0.317
TTAAGGTTCTGAATATCTGT	TAG	3(2,4,12)	chrX[82257482]	intergenic	0.198
TCAGAGTTCTGAATTTCTGT	AGG	3(5,12,15)	chr1[186351303]	intron: TPR	0.162
TCAGGTTCCAGGATATCTGT	CAG	3(6,8,10)	chr11[72679292]	intergenic	0.148
TCATGGTTGTGCATATCTGT	GGG	3(4,9,12)	chr1[176112459]	intron: COP1	0.119
TGAGGCTTGTGGATATCTGT	CAG	3(2,6,9)	chr3[102283671]	intergenic	0.076

Figure 11. Off-target sites in the reference genome for a gRNA.

- To evaluate the effects of the variants in the off-target space for the gRNAs, click on the Search button in the Off-Targets in Variant Genome column (Figure 10). Please note that this search takes on average ~2 minutes and adds additional time to your design process. Once this search is completed, the Off-Targets in Variant Genome column will be populated (Figure 12). To view the off-target sites for a gRNA in the variant genome, click on the number in the Off-Targets in Variant Genome column.

		CRISPR gRNA Targets		TALEN Targets			
<input type="checkbox"/>	CRISPR Target Sequence ⓘ	PAM Site	Score (%) ⓘ	Off-Targets in Reference Genome ⓘ	Off-Targets in Variant Genome ⓘ	Edit site distance	Donor DNA ⓘ
<input type="checkbox"/>	TCAGGGTCTGGATATCTGTG.....	GGG ...	61.87	32	37	27 bp	
<input type="checkbox"/>	TGGATTTAGAGTCTCTCAGC	TGG ...	51.84	37	42	8 bp	
<input type="checkbox"/>	GCTGGTACACGGCAGGGTCA	GGG ..G	50.84	54	53	10 bp	
<input type="checkbox"/>	AGAGTCTCTCAGCTGGTACA	CGG	49.06	52	54	1 bp	

Figure 12. Off-target sites in the variant genome.

Points to note:

- The number of off-target sites in the variant genome can be higher than the number of off-target sites in the reference genome if new valid PAM sites (NGG, NAG, or NGA) are generated because of SNPs or if the number of mismatches is reduced to 3.
- The number of off-target sites in the variant genome can be lower than the number of off-target sites in the reference genome if the number of mismatches is greater than 3 for some sites or if PAM sites become invalid.
- The number of off-target sites in the variant genome can remain the same as the number of off-target sites in the reference genome. In this case, attention should be paid to the sites for which an orange warning symbol is displayed (Figure 13). This indicates that the off-target site has become more potent because of an SNP.

Off Targets [TCTCTCAGCTGGTACACGGC - AGG] ⓘ							
	Off-target recognition site	PAM	Mismatches Count (Position)	Genomic Locations ⓘ	Annotations ⓘ	CFD Score ⓘ	Allele Frequency [Overall] Position in sgRNA: Freq
	Reference: TCAATCAGCTGGTGCACGGC Variant : TCAATCAGCTGGTGCACGGC	TGG TGG	3(3,4,14) 3(3,4,14)	chr1[236924556]	Intergenic	0.281 0.281	- -
	Reference: TCTCACAGCTGGAACACAGC Variant : TCTCACAGCTGGAACACAGC	AGG AGG	3(5,13,18) 3(5,13,18)	chr8[10769744]	Intron: PINX1	0.240 0.240	- -
	Reference: TCTCATAGCTGGTACATGGC Variant : TCTCATAGCTGGTACATGGC	GGG GGG	3(5,6,17) 3(5,6,17)	chr15[100865585]	Intron: GCAWKR	0.217 0.217	- -
	Reference: TTTCTCAGCTGGTACATGGA Variant : TTTCTCAGCTGGTACATGGA	GGG GGG	3(2,17,20) 3(2,17,20)	chr1[247923626]	Intergenic	0.212 0.212	- -
	Reference: TCCATCAGCTGCTGCACGGC Variant : TCCATCAGCTGCTGCACGGC	AGG AGG	4(3,4,12,14) 3(4,12,14)	chr3[78639793]	exon: ROBO1	0.089 0.208	- 3: 0.034
	Reference: TCTCAGAGCTGGTACACAGC	AGG	3(5,6,18)	0.173	-

* - Off-target site has higher predicted off-target activity when variants are considered.

Figure 13. Off-target sites in the variant genome for a gRNA.

The Allele Frequency column indicates the position of the SNP, with 1 being the most PAM-distal nucleotide of the off-target site, and the allele frequency of the SNP in the population selected for the design.

12. To achieve higher insertion efficiency, select a gRNA that is closer to the edit location. Although the score of the gRNA shown here is color-coded orange, the edit site distance is 3 bp and the number of off-target sites in both the reference and the variant genome is not worrisome. Select this gRNA (it would appear as the 5th CRISPR Target Sequence in Figures 10 and 12) and click “Next”.

<input type="checkbox"/>	TCTCTCAGCTGGTACACGGC	AGG	48.68	35	36	3 bp		
--------------------------	----------------------	-----	-------	----	----	------	--	--

Step 4: Summary

13. The summary page first displays a warning message regarding the manufacturing feasibility of the donor DNA (Figure 14). Although TrueDesign Genome Editor checks the insertion sequence and the donor DNA against the GeneArt services criteria, the final price and the feasibility may change once the order is received. Please note that the donor DNA will be delivered to you as sequence-verified plasmid DNA in the standard GeneArt production vector with kanamycin resistance. If a different vector is desired, please order the donor DNA through the GeneArt Services Dashboard.

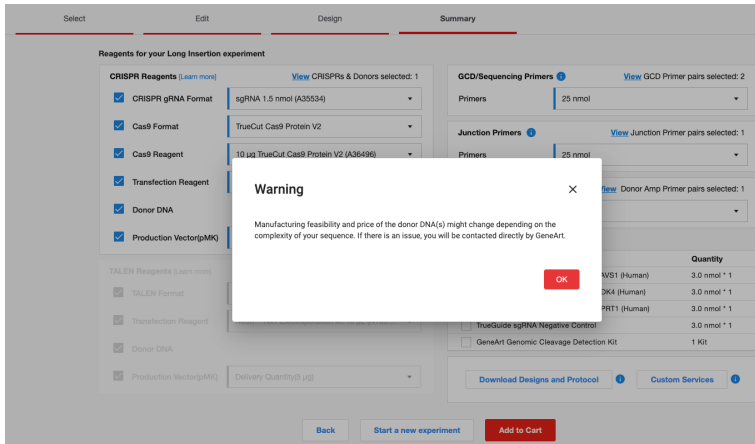


Figure 14. Warning message on the summary page.

14. Three types of primers will be designed and displayed for the selected gRNA and donor DNA (Figure 15). Primers are selected by default to be added to the cart.

- **GCD/sequencing primers:** These primer sets can be used to amplify the nuclease cut site to assess editing efficiency. They are compatible with Sanger sequencing and the Invitrogen™ GeneArt™ Genomic Cleavage Detection Kit.
- **Junction primers:** These primers can be used to help verify precise insertion of the donor DNA at the site of interest. They consist of two sets of primers for each donor DNA, one to verify the junction of the left homology arm and the insertion sequence and the other to verify the junction of the right homology arm and the insertion sequence.
- **Donor amplification primers:** These primers can be used to amplify the donor DNA from the plasmid vector sent to you by GeneArt Gene Synthesis Services. In cases where the design of donor amplification primers fails, please use the restriction sites (Sfil) provided in the vector to extract your donor DNA.

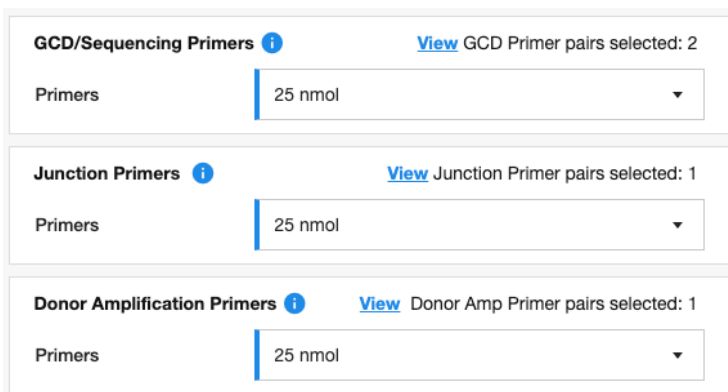


Figure 15. Three types of primers on the summary page.

15. Positive and negative controls can be selected from the Other Products You May Need section.
16. Before adding all the products to your cart, click on “Download Designs and Protocol”. A Microsoft™ Excel™ file with multiple tabs including your design details will be downloaded. This file contains all the relevant information regarding gRNA or TALEN sequences, off-target information, donor DNA sequences, primer sequences, and protocols to perform your experiment.
17. Finally, click on “Add to Cart” to add all the selected products to the cart and order. If “Add to Cart” is not enabled in your region, you can use the downloaded Microsoft Excel report and place your order through the local sales representative or customer service.

Conclusions

In this application note, we have introduced the new features of the TrueDesign Genome Editor and demonstrated how the user can easily achieve GFP expression from the *TRAC* locus. This workflow can also be adapted for the expression of a CAR sequence from the *TRAC* locus. We also provide general recommendations for the long insertion workflow, including designing a donor sequence up to 10 kb that can be easily manufactured and ordering reagents required for your genome editing experiment.

References

1. Chen S, Francioli LC, Goodrich JK et al. (2024) A genomic mutational constraint map using variation in 76,156 human genomes. *Nature* 625(7993):92–100.
2. Doench JG, Fusi N, Sullender M et al. (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* 34(2):184–191.
3. DeWeirdt PC, McGee AV, Zheng F et al. (2022) Accounting for small variations in the tracrRNA sequence improves sgRNA activity predictions for CRISPR screening. *Nat Commun* 13(1):5255.
4. Vogelstein B, Papadopoulos N, Velculescu VE et al. (2013) Cancer genome landscapes. *Science* 339(6127):1546–1558.

Ordering information

Product	Quantity	Cat. No.
TrueDesign Genome Editor	–	thermofisher.com/truedesign
TrueCut Cas9 Protein v2	10 µg	A36496
TrueCut HiFi Cas9 Protein	10 µg	A50574
TrueGuide Modified Synthetic gRNA	–	A35534
TrueGuide sgRNA Negative Control, non-targeting 1	3 nmol	A35526
Neon NxT Electroporation System 10 µL kit	96 x 2 reactions	N1096
Lipofectamine CRISPRMAX Transfection Reagent	0.1 mL	CMA000001
GeneArt Genomic Cleavage Detection Kit	20 reactions	A24372

 Get started today at thermofisher.com/truedesign

invitrogen