

TP53 C-terminal mCherry Knock-in -- Design Report

Endogenous tagging in HEK293T | SpCas9 + HDR | GRCh38 / NM_000546.6

AI Bioinformatician by Labterna | 29 June 2026 | GRCh38/hg38

Design complete and cross-validated -- one biology caveat to read first



A primary + backup SpCas9 guide at the TP53 stop codon (validated across three independent tools -- Cas-OFFinder + CRISPOR + FlashFry; **no pseudogene / p53-family off-targets**), an 800/800 bp HDR donor with codon-optimized Bsal/BsmBI-free mCherry and built-in re-cut protection, and a full genotyping set are delivered. **Caveat:** HEK293T expresses SV40 large T antigen, which binds and functionally inactivates p53 -- p53 stress dynamics in this line are non-physiological (Section 7).

1. Objective

Design an endogenous C-terminal mCherry knock-in at TP53 (NM_000546.6, GRCh38) in HEK293T for live-cell imaging of p53 under genotoxic stress, using SpCas9 + HDR. Requested: (i) guide(s) at the TP53 stop codon cross-validated across ≥ 2 tools (a prior design failed on an undetected pseudogene off-target); (ii) an HDR donor with ~800 bp arms, a GS linker before mCherry, and PAM-blocking silent mutations; (iii) 5' and 3' junction genotyping primers anchored outside the arms plus an internal Sanger pair; (iv) a full off-target report. mCherry: human codon-optimized, no internal Bsal/BsmBI.

2. Target biology and tagging strategy

TP53 lies on chr17p13.1 (**minus strand**); NM_000546.6 encodes the canonical 393-aa p53. The C-terminus (...EGPDS₃₉₃) is in the last exon, and the entire 1,188-nt 3'UTR is intronless and contiguous with the stop codon, so both homology arms come cleanly from genomic sequence. mCherry is inserted **in-frame immediately before the native TGA stop** (chr17:7669609-7669611), reusing the endogenous stop and 3'UTR. Fusion ORF: p53(1-393)-(GGGGS) $\times 3$ -mCherry-stop.

Why C-terminal (and its caveat). An N-terminal tag would block the MDM2-binding site (~residues 17-29) and artificially stabilize p53, so the C-terminus is conventional. It abuts the **C-terminal regulatory domain (CTD, ~363-393)** -- the site of MDM2 ubiquitination (Lys370-386) and many PTMs (Kruse & Gu 2009). The tag is placed after residue 393 (CTD lysines preserved) with a flexible (GGGGS) $\times 3$ linker, but tagged-p53 function must be validated (Section 7).

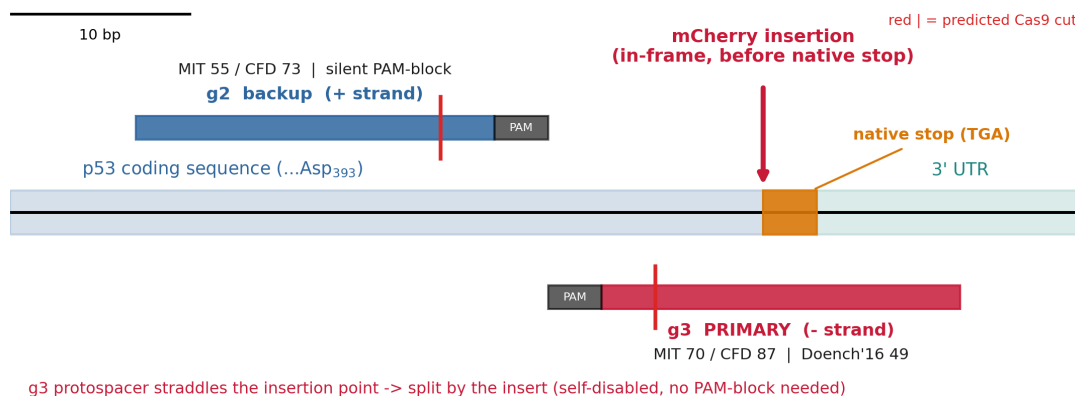
Endogenous fluorescent tagging of p53 for live single-cell dynamics is well precedented (the p53-Mdm2 pulsing literature, Lahav et al. 2004 [9], and later endogenous reporter lines), which motivates tagging at the native locus over overexpression. This is a de novo guide+donor design for that purpose; a parameter-by-parameter comparison against a specific published TP53 knock-in line is a possible extension.

TP53 C-terminal mCherry Knock-in -- Design Report

Endogenous tagging in HEK293T | SpCas9 + HDR | GRCh38 / NM_000546.6

AI Bioinformatician by Labterna | 29 June 2026 | GRCh38/hg38

A. SpCas9 guides at the TP53 stop codon (C-terminal insertion site)



B. C-terminal mCherry fusion relative to p53 domain structure

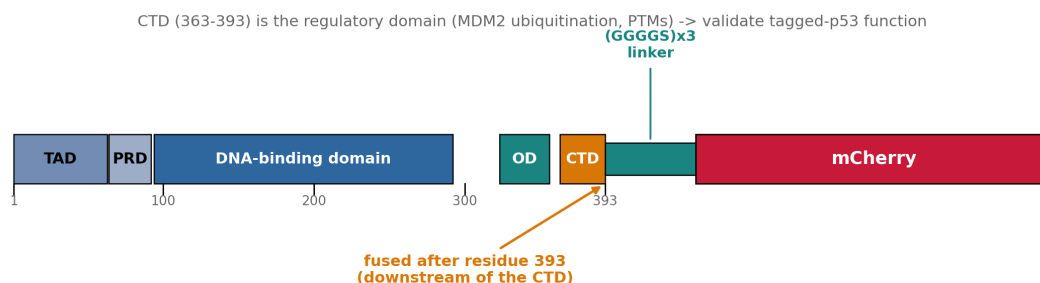


Figure 1. SpCas9 guides at the TP53 stop codon (top) and the C-terminal mCherry fusion relative to p53 domain structure (bottom). The insertion sits just downstream of the regulatory CTD.

3. Guide RNA design and selection

All NGG protospacers cutting within 35 bp of the insertion point were enumerated on both strands (7 candidates).

On-target/specificity scores are from CRISPOR; genome-wide off-targets independently from Cas-OFFinder (Section 4).

Two guides are recommended:

Guide (role)	Spacer 5'->3'	PAM	Str.	Cut vs ins.	GC%	MIT	CFD	Doench16
TP53ct_g3 (PRIMARY)	GGAGAATGTCAGTCTGAGTC	AGG	-	-6 bp	50	70	87	49
TP53ct_g2 (backup)	AACTCATGTTCAAGACAGAA	GGG	+	-18 bp	35	55	73	63

g3 (primary) cuts 6 bp from the insertion (efficient HDR), is 50% GC, starts with G (U6-ready), and has the highest MIT (70) and CFD (87) specificity of any guide at the stop codon. Its protospacer straddles the insertion point, so the insert splits the target -- the edited allele is no longer a Cas9 substrate and **no PAM-block is needed**. On-target efficiency is moderate (Doench'16 49); g3 is chosen for specificity, cut-to-insert proximity and intrinsic re-cut immunity rather than peak predicted cutting.

On the backup. For a single-point knock-in both guides must sit at the stop codon, so the standard "backup ≥ 100 bp away" rule does not apply -- g3 and g2 overlap (cuts ~12 bp apart). g2's value is its **different seed, strand and PAM** (immune to a SNP under g3's seed or to g3-specific inefficiency) and its higher predicted efficiency (Doench'16 63) -- the natural fallback if g3 cuts weakly. It does **not** cover a locus-level deletion/SV spanning the stop codon (a residual shared-failure mode inherent to point knock-ins). g2's target lies in the 5' arm, so a silent PAM-block is built into the donor (Section 5). g4-g7 scored lower; the GC-rich g5-g7 are promiscuous (622-908 sites) and not recommended.

4. Off-target / specificity analysis (three independent tools)

Because the prior design failed on a pseudogene off-target missed by a single tool, specificity was assessed with **three independent genome-wide methods**: **Cas-OFFinder** (local exhaustive enumeration vs hg38; $NGG \leq 4$ mm + $NAG \leq 3$ mm, every hit annotated to RefSeq genes), **CRISPOR** (web, BWA-based, CFD/MIT scoring + annotation), and **FlashFry** (local, against a purpose-built hg38 SpCas9-NGG database; same community metrics). All three agree closely.

Tool (hg38)	g3 off-targets	g3 ≤ 1 mm	g3 top-OT CFD	g2 off-targets	g2 ≤ 1 mm	p53-family / pseudogene
Cas-OFFinder (local)	202	0	n/a	299	0	0
CRISPOR (web)	184	0	0.612	289	0	0
FlashFry (local)	176	0	0.612	262	0	0*

* FlashFry confirms the count and mismatch distribution on a third algorithm; gene/pseudogene annotation is from Cas-OFFinder + CRISPOR (both 0). Cas-OFFinder enumerates exhaustively and does not compute CFD; the g3/g2 top-off-target CFD (0.61 / 0.65) is reported identically by CRISPOR and FlashFry. FlashFry's tabulated count is genomic sites (otCount); Figure 2B bins unique off-target sequences (slightly fewer, as some map to several loci).

Pseudogene check (the prior-failure mode): cleared. For g3, all three tools report **0 off-targets at ≤ 1 mismatch**, and the two annotating tools find **0 hits in any TP53 pseudogene or p53-family gene (TP63/TP73)**. All exonic off-targets occur only at 4 mismatches (max CFD 0.17, predicted inactive); the same loci (UBR1, FGA, NAALADL2) are recovered independently. CRISPOR and FlashFry flag the identical single highest-risk off-target (CFD 0.61).

Binding-site variation. A common SNP in a spacer seed or PAM would cause allele-specific cutting failure -- relevant because the edit is in HEK293T's own genome. Screening Ensembl Variation (GRCh38) across the g3 and g2 protospacer+PAM windows returned 69 cataloged variants, **none common (MAF $\geq 1\%$)** under either guide -- no allele-dropout risk (g3 PASS, g2 PASS). The locus is densely catalogued with rare single-submission variants that do not threaten the assay.

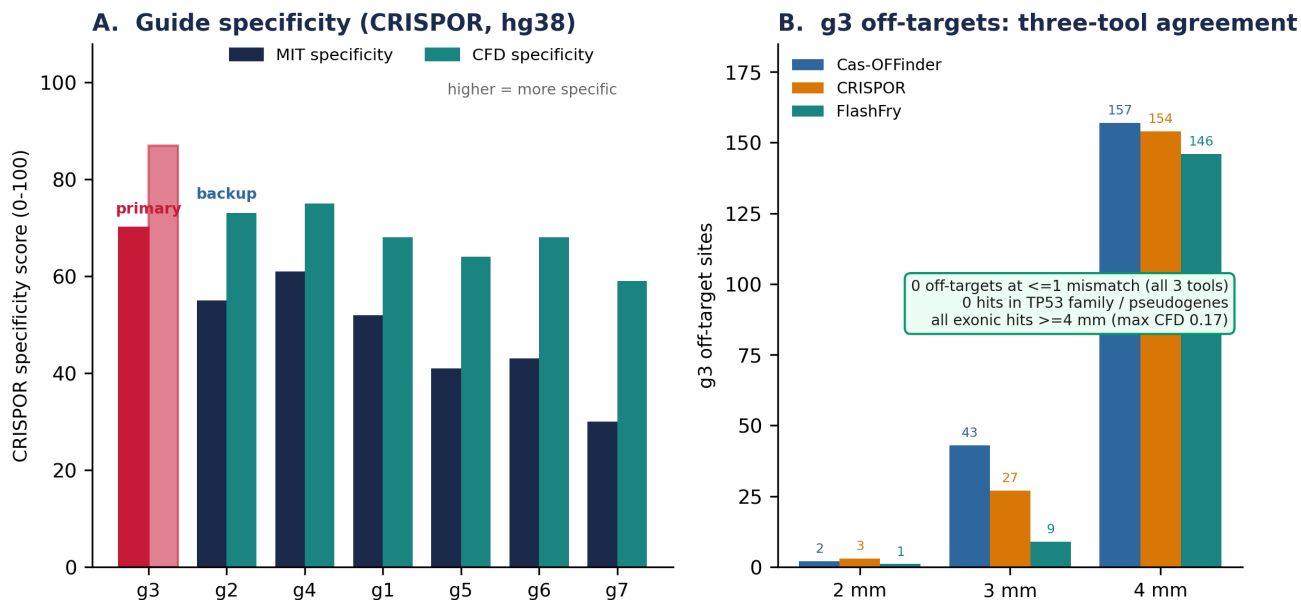


Figure 2. (A) CRISPOR specificity across the seven candidates -- g3 is the most specific guide at the stop codon. (B) g3 off-target counts by mismatch agree across Cas-OFFinder, CRISPOR and FlashFry; no hit below 2 mismatches in any tool and none in p53-family/pseudogenes.

5. HDR donor design

The donor is a **2353 bp dsDNA** (plasmid or synthetic long dsDNA): 5' arm (800 bp) - (GGGG) $\times 3$ linker - mCherry - 3' arm (800 bp).

TP53 C-terminal mCherry Knock-in -- Design Report

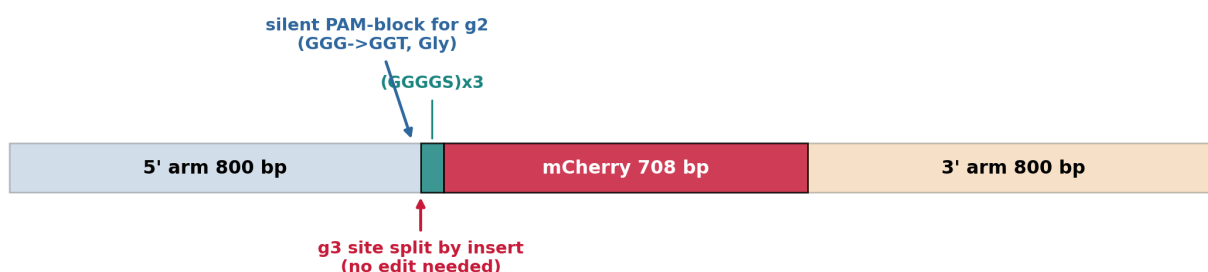
Endogenous tagging in HEK293T | SpCas9 + HDR | GRCh38 / NM_000546.6

AI Bioinformatician by Labterna | 29 June 2026 | GRCh38/hg38

Component	Length	Location / context	Notes
5' homology arm	800 bp	chr17:7669612-7670411	genomic; ends at Asp393; carries g2 silent PAM-block
(GGGS)x3 linker	45 bp	insert	flexible, codon-diversified (no tandem repeat)
mCherry CDS (no stop)	708 bp	insert	human codon-opt, GC 61.9%, 0 BsaI / 0 BsmBI
3' homology arm	800 bp	chr17:7668812-7669611	genomic; native TGA stop + 3'UTR
Full donor	2353 bp	5'arm + insert + 3'arm	insert BsaI/BsmBI = 0; arms carry 2 BsaI + 1 BsmBI (native genomic)

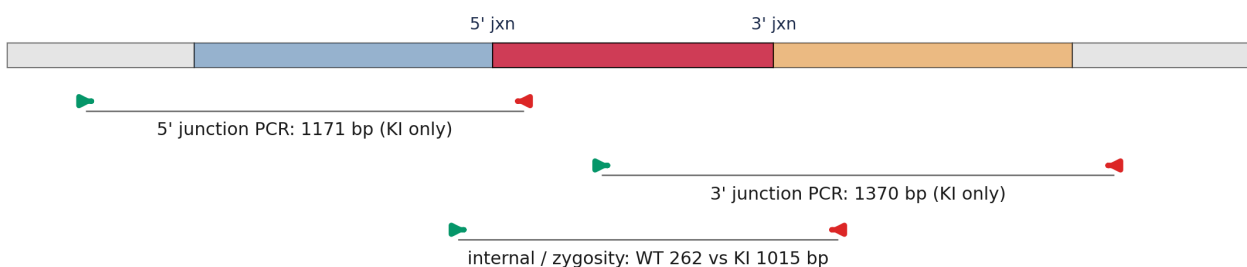
Re-cut protection. g3 needs none (the insertion splits its target). For g2, a single **silent** change in the 5' arm (codon 389 GGG→GGT, both Glycine) destroys the NGG PAM; the CDS translation is unchanged and g2 can no longer cut the donor or the edited allele -- one donor serves either guide. **mCherry** (Shaner 2004; GenBank AY678264, verified byte-identical) is human codon-optimized to 708 nt with 0 internal BsaI/BsmBI as requested.

A. HDR donor (2,353 bp dsDNA)



Insert (linker+mCherry): 0 BsaI / 0 BsmBI sites (re-clonable). Homology arms carry 2 BsaI + 1 BsmBI (native genomic).

B. Genotyping & Sanger primers (junctions anchored OUTSIDE the arms)



Green = forward, red = reverse. Junction PCRs give NO product on WT or donor plasmid -> band only on correct genomic integration.

Figure 3. (A) HDR donor architecture with built-in re-cut protection. (B) Genotyping/Sanger primers; junction PCRs give a product only on correct genomic integration (not on WT or donor plasmid).

6. Genotyping and validation strategy

Assay	Forward / Reverse (Tm)	Product
5' junction PCR	F CGCTTCGAGATGTTCCGAGA (59.9C, outside 5' arm) R TGATGGCCATGTTGTCTCCTCC (60.0C, in insert)	1171 bp (KI only; none on WT/donor)
3' junction PCR	F ATCCCCGACTACCTGAAGCT (60.0C, in insert) R GAACCTGTGGTCCCAGCTAC (60.0C, outside 3' arm)	1370 bp (KI only; none on WT/donor)
Internal / zygosity	F TGTCTCCTACAGCCACCTGA (59.9C, 5' arm) R GCAGGCCAACTTGTTTCAGTG (60.0C, 3' arm)	WT 262 bp vs KI 1015 bp

TP53 C-terminal mCherry Knock-in -- Design Report

Endogenous tagging in HEK293T | SpCas9 + HDR | GRCh38 / NM_000546.6

AI Bioinformatician by Labterna | 29 June 2026 | GRCh38/hg38

All genomic primers are unique on hg38 (isPcr); junction-PCR logic was confirmed by in-silico PCR on WT, donor and knock-in templates. Junction primers are anchored outside the arms so they cannot prime on residual donor -- a band proves genuine genomic integration. The internal pair distinguishes wild-type / heterozygous / homozygous clones and provides a Sanger amplicon across the whole insert -- but its primers lie within the arms, so it also amplifies residual donor (same KI-size band); use the junction PCRs as the definitive integration proof and account for donor carryover when scoring zygosity.

7. Critical caveats and recommendations

(a) HEK293T expresses SV40 large T antigen. Large T binds p53 directly, **stabilizes** it (raising basal levels) while **inhibiting transactivation** and blocking p53-dependent MDM2 induction (Lilyestrom 2006). p53 stress dynamics in 293T are therefore not physiological. For physiological dynamics, a p53-wild-type line without a viral oncoprotein (MCF7, RPE1-hTERT, A549, U2OS, HCT116) is preferable; if 293T is required, interpret with this caveat. **This is the most important point in this report.**

(b) C-terminal tag vs p53 regulation. Validate that tagged p53 stabilizes after genotoxic stress (doxorubicin, nutlin-3a), localizes to the nucleus, and induces canonical targets (CDKN1A/p21, MDM2), against an untagged control.

(c) HEK293 ploidy. 293/293T is hypotriploid (~3 TP53 copies); expect multiple alleles to edit. The internal/zygosity PCR reveals allele dosage; all-allele tagging may need more clones.

(d) Delivery & HDR. Deliver the donor as plasmid or long dsDNA (800 bp arms are too long for ssODN); co-deliver SpCas9 + g3 (RNP or plasmid). Validate by junction PCR → Sanger across the insert → mCherry fluorescence and tagged-p53 function. Guide cloning oligos are provided for pX330/pSpCas9(BB)-family BbsI vectors.

Design quality-control summary

Check	Status	Detail
Guide cross-validation across 3 tools (Cas-OFFinder + CRISPOR + FlashFry)	PASS	All rank g3 most specific; counts + top-OT CFD agree
No off-target at <=1 mismatch (primary guide g3)	PASS	0 hits in all three tools
No off-target in a TP53 pseudogene / p53 family gene	PASS	0 hits (prior-failure mode); annotating tools Cas-OFFinder + CRISPOR
mCherry human codon-optimized, no internal BsaI/BsmBI	PASS	708 nt, GC 61.9%, 0 sites
Donor re-cut protection	PASS	g3 split by insert; g2 silent PAM-block (Gly)
In-frame C-terminal fusion, native stop + 3'UTR reused	PASS	single stop; ...MDELYK-stop verified
Junction primers KI-specific (no WT / no donor product)	PASS	isPcr unique; in-silico verified
No common SNP under guide binding sites (Ensembl, MAF>=1%)	PASS	69 cataloged variants in window, all rare
Cell line supports physiological p53 dynamics	WARN	HEK293T SV40 large T antigen perturbs p53
C-terminal tag and p53 CTD-dependent regulation	WARN	Validate tagged-p53 function

8. Methods, databases and auditability

TP53 NM_000546.6 from NCBI; genomic sequence from local GRCh38/hg38 2bit; stop codon anchored by exact sequence match (build-independent). On-target/specificity from CRISPOR v5.2 (hg38; batch retained); exhaustive off-targets from Cas-OFFinder v2.4.1 (hg38, CPU) and independently from FlashFry (local hg38 SpCas9-NGG database, cached for reuse). mCherry codon optimization with DNACHisel (h_sapiens) enforcing BsaI/BsmBI absence; primers with primer3; specificity by isPcr vs hg38; binding-site variant screen via Ensembl Variation (GRCh38). Data retrieved 2026-06-29/30. The numbered scripts/ pipeline ships as the reproducible record.

9. Output files

- **TP53_mCherry_Knockin_Report.pdf** -- this report
- **TP53_mCherry_Knockin_Design.xlsx** -- guides, off-targets, donor components, primers, cloning oligos

- **sequences.fasta** -- guides, mCherry / insert, donor + arms, primers, BbsI cloning oligos
- **figures/** -- knock-in strategy; specificity evidence; donor + genotyping

*The **scripts/** folder ships as the reproducible numbered pipeline.*

10. References

- [1] O'Leary NA, et al. (2016). Reference sequence (RefSeq) database at NCBI. *Nucleic Acids Res* 44:D733-D745. TP53 transcript NM_000546.6.
- [2] Shaner NC, Campbell RE, Steinbach PA, Giepmans BNG, Palmer AE, Tsien RY (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nat Biotechnol* 22:1567-1572. (mCherry; GenBank AY678264)
- [3] Concordet JP, Haeussler M (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res* 46:W242-W245.
- [4] Bae S, Park J, Kim JS (2014). Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* 30:1473-1475.
- [5] Doench JG, et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* 34:184-191. (CFD / Rule Set 2)
- [6] Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8:2281-2308.
- [7] Liliestrom W, Klein MG, Zhang R, Joachimiak A, Chen XS (2006). Crystal structure of SV40 large T-antigen bound to p53: interplay between a viral oncoprotein and a cellular tumor suppressor. *Genes Dev* 20:2373-2382.
- [8] Kruse JP, Gu W (2009). Modes of p53 regulation. *Cell* 137:609-622. (C-terminal domain: ubiquitination, acetylation, PTMs)
- [9] Lahav G, Rosenfeld N, Sigal A, Geva-Zatorsky N, Levine AJ, Elowitz MB, Alon U (2004). Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat Genet* 36:147-150.