Journal of High School Science

From convergence to an iterative Optimization-Circumvention-Collapse framework in CRISPR bioengineering

Filippone-Thaulero F

Submitted: June 15, 2026, Revised: version 1, July 13, 2026
Accepted: July 15, 2026

Abstract
CRISPR-Cas9 gene editing is constrained by unintended off-target effects (OTEs) and inefficient homology-directed repair (HDR). Although modern CRISPR engineering increasingly co-evaluates activity, specificity, repair outcome, delivery, and genomic stability, OTEs and HDR are often analyzed through distinct intervention frameworks. This review formalizes their interdependence as three levels of convergence that define a Pareto-like optimization frontier. Operationally, finite cell numbers, *ex vivo* editing constraints, and delivery limitations prevent sequential optimization. Mechanistically, Cas9 exposure time influences both off-target accumulation and synchronization with HDR-competent cell-cycle windows. Structurally, on-target activity, Cas9 specificity, and HDR efficiency form a mutually constraining trade-off in which no single point maximizes all three objectives. Convergence therefore extends optimization into an iterative design framework with two additional engineering responses. If the architecture of DSB-based CRISPR systems prevents the optimization frontier from intersecting the clinical acceptance region, circumvention through base- or prime-editor architectures may be preferable. Conversely, when an architecture is clinically viable but constrained by strong mechanistic coupling, collapse weakens those couplings and reshapes the frontier toward greater clinical utility. This review therefore develops an iterative optimization-circumvention-collapse framework in which convergence generates the optimization frontier, identifies mechanistic couplings that may be collapsed, and indicates when architectural dependencies should be circumvented.

Keywords
CRISPR-Cas9, Gene editing, Genome engineering, Off-target effects, Homology-directed repair, Multi-objective optimization, Pareto frontier, Base editing, Prime editing, Bioengineering

Federico Filippone-Thaulero, Marymount International School Rome, Via di Villa Lauchli 180, Rome, 00191 RM, Italy. fedefth@yahoo.com

https://doi.org/10.64336/001c.165176

1. Introduction

CRISPR-Cas9 gene editing technology is actively being used in clinical trials to treat genetic diseases and cancer (1). Nonetheless, despite its apparent precision, CRISPR-Cas9 editing remains limited by unintended off-target effects and eukaryotic predominance of non-homologous end joining repair mechanisms (2, 3). CRISPR-Cas9 systems use a single-guide RNA (sgRNA) and Cas9 protein complex to cleave DNA. More specifically, Cas9 is a nuclease designed to cut DNA at specific locations via a double-strand break (DSB). sgRNA is a synthetic RNA sequence designed to bind to a specific DNA sequence (4). Importantly, the Cas9-sgRNA complex will only bind to the target gene if the protospacer adjacent motif (PAM) is present, critical to initiate R-loop formation for DNA cleavage (Figure 1) (5). Once cut, the cell's own machinery determines the repair mechanism.

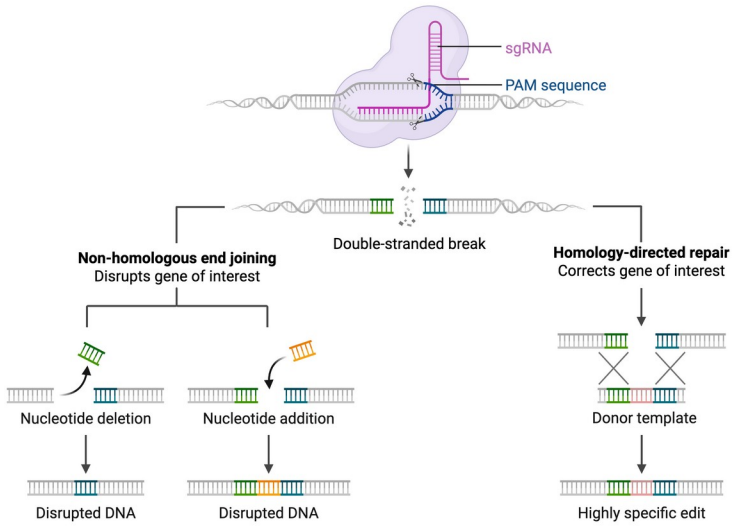

Figure 1. CRISPR-Cas9 target recognition and the two double-strand break repair outcomes. The Cas9–sgRNA complex binds the target sequence adjacent to the PAM and generates a double-strand break (DSB), resolved either by homology-directed repair (HDR), which uses a donor template for a highly specific edit, or by non-homologous end joining (NHEJ), which produces error-prone insertions or deletions. Created in BioRender. Thaulero, F. (2026) https://BioRender.com/ona45ik.

Off-target effects (OTEs) are unintended interactions in which Cas9 binds to the incorrect DNA sequence, thus creating a DSB at the wrong location. This can lead to major, unpredictable changes. Unintended edits can activate oncogenes or knock out tumor suppressors (6), cause genomic instability through large insertions and deletions (6), and have been linked to immune responses against edited products (7). Another limitation of CRISPR-Cas9 technologies is the lack of control over post-editing DSB repair pathways. Non-homologous end joining (NHEJ) is characterized by the rapid ligation of broken DNA ends without a template. NHEJ is ideal for gene-knockout disrupting genes by causing random, error-prone insertions or deletions (indels) (8). Quantitative analyzes in human cells confirm that NHEJ is often more active than

homology-directed repair (HDR), though ratios are highly context-dependent and can be shifted via bioengineering interventions. As mentioned, HDR is the second major repair pathway. Contrary to NHEJ, it is a highly precise mechanism able to insert specific mutations or insertions as it uses a homologous sequence (a donor template) to repair the DSB. HDR is significantly less efficient due to its high precision (9).

Although OTE reduction and HDR promotion are increasingly co-evaluated in modern CRISPR engineering, they are often analyzed through distinct intervention logics. This review formalizes the interdependence of specificity, activity, and repair-pathway control through three levels of convergence: operational convergence, where clinical delivery constraints prevent sequential optimization; mechanistic convergence, in which shared variables such as Cas9 exposure time improve both constraints; and structural convergence, where activity, specificity, and repair (intended as HDR efficiency) form a Pareto-like trade-off.

However, this convergence does not merely show that CRISPR bioengineering is an optimization problem; it also provides a framework to ascertain when direct optimization may be the sub-optimal engineering strategy. Namely, as long as a frontier remains clinically reachable, *optimization* is appropriate. If convergence reveals that the architecture's required dependencies prevent the frontier from reaching clinical viability, *circumvention* may be required. If the architecture has high clinical potential but remains limited by severely coupled convergent variables, *collapse* may reshape the very frontier such that it intersects the clinical acceptance region. This review therefore investigates how convergence and optimization may be intrinsic to gene editing through the examination of DSB-based CRISPR-Cas9, and then how such logic extends to an iterative optimization-circumvention-collapse framework for CRISPR bioengineering.

2. Convergence as the basis for optimization

2.1 Framework for convergence

OTE reduction and HDR enhancement are often analyzed through distinct bioengineering frameworks, even though modern CRISPR engineering increasingly evaluates activity, specificity, repair outcome, viability, delivery, and genomic stability together. This review formalizes this interdependence as an optimization frontier generated by three different levels of convergence. First, clinical contexts in which CRISPR-Cas9 is actually deployed force simultaneous optimization of both. Second, certain interventions share success in both domains as they both target a common control variable. Third, interventions designed for one domain non-trivially influence the other. These three different observations represent three distinct levels of convergence, namely operational, mechanistic, and structural. This section will therefore analyze how HDR-maximization and OTE-reduction strategies converge, where the convergence becomes visible, and why such convergence is the necessary first step before optimization, circumvention, or collapse can be rationally considered.

2.2 Operational convergence

Bioengineering strategies for specificity and repair control are typically constructed in idealised systems with *in vitro* cell lines and

single-variable experiments. Clinical interventions operate under fundamentally different parameters. Sequential optimization becomes operationally unfeasible due to limited time and resources, and the question becomes whether they can be deployed together.

In clinical contexts, the same physical resources both impact HDR efficiency and OTE control. There are several key HDR constraints. For example, cell number is fixed by primary tissue yields and cannot be expanded. Consequently, iterative trial-and-error bioengineering is not a viable strategy, as each unsuccessful edit permanently depletes the finite pool of available cells. The manipulation window is also no longer indefinite, now fixed by cell viability *ex vivo* which prevents sequential optimization and favors simultaneous intervention. And lastly, delivery vehicle dose simultaneously determines HDR efficiency (donor template concentration) and OTE risk (extended Cas9 exposure) (10-12). In effect, clinical application forces a dose-response convergence that prevents either objective from being maximized in isolation.

This convergence is demonstrated by CRISPR-Cas9 with adeno-associated virus serotype 6 (AAV6) editing in primary therapeutic immune cells, limited by several finite variables such as cell recovery, electroporation stress, AAV6 timing, and AAV6 titer rather than HDR efficiency alone. Higher AAV6 titer resulted in 40-60% HDR increase at $\geq 4 \times 10^4$ vg/cell yet plateaued beyond 4×10^4 vg/cell, indicating that titer increase does not automatically increase HDR (13). Cas9 delivery dose is also a fidelity constraint: reduced Cas9 exposure effectively reduces off-target edits, while increased lifetime increases OTEs (14). Therefore, clinical editing creates an operational trade-off: the same finite intervention window must deliver enough donor template and nuclease activity to enhance HDR, while limiting reagent dose, cell stress, and off-target cleavage. Therefore, clinical applications establish why HDR and specificity must be evaluated as a single constrained system within clinical gene editing.

This convergence is not limited to *ex vivo* systems, as *in vivo* lipid nanoparticle (LNP) delivery actually amplifies it. *In vivo* systems introduce even stricter coupling due to added constraints such as systemic dose ceilings (NTLA-2001 systems), delivery format dependency on tissue tropism, and immune response risk (15, 16). Thus, *in vivo* systems also induce finite-resource convergence, only with different and significantly harsher limitations, which may render such DSB-based systems disqualifying for some demanding *in vivo* correction applications.

Ultimately, operational coupling establishes that HDR and OTEs must be optimized simultaneously within an integrated clinical framework. Nonetheless, this clinical convergence implies that HDR enhancement and OTE control share mechanistic similarities which allow for single interventions to impact both variables simultaneously.

2.3 Mechanistic convergence

Both systems rely on a shared regulatory axis. A specific class of interventions that manipulate the temporal profile of Cas9 modulate both HDR and OTEs via mechanistic similarity rather than chance.

Specifically, time dictates *when* (HDR) and for *how long* (OTEs) Cas9 is active.

Cas9 binding is inherently stochastic, with the probability of cleavage being highly time-dependent, $P(t) = A \times (1 - e^{-k_{obs}t})$ (eq.1), where P(t) is the probability of cleavage at time t; A is maximum efficiency, and k_{obs} is the observed rate (17). Reducing Cas9 exposure by an order of magnitude therefore disproportionately reduces cumulative Cas9 activity at partially homologous (off-target) sites. Since mismatched duplexes bind weakly and commit to catalytic conformation slower than true on-target matches, they will be the first to fail cleavage once the activity window is reduced.

Analogously, HDR efficiency is also directly dependent on time: the very possibility of HDR pathway selection is contingent upon cell cycle phase. Cas9 activity that falls outside of the S/G2 range contributes negligibly to HDR maximization regardless of cleavage efficiency. In this way, HDR depends on a slightly different facet of temporal control: the timing of Cas9 delivery relative to cell cycle window rather than its total duration. Both outcomes, though, are modulated by the same bioengineering variable, namely Cas9's activity profile.

Ribonucleoprotein (RNP) delivery is a key bioengineering intervention that shows temporal convergence. Recombinant Cas9 is mixed with an artificial sgRNA, to form a pre-assembled complex before cellular delivery (18, 19). Pre-assembled RNPs have a cellular half-life of hours, whereas plasmid-delivered Cas9 is expressed for days or weeks (20). Beyond rather large 10 to 100-fold reductions in OTEs due to restricted activity time, RNP delivery also concentrates Cas9 into HDR-competent windows (20). Therefore, the same temporal control that suppresses OTE accumulation also exploits synchronization with the cell cycle window that elevates HDR.

Geminin-Cas9 fusion demonstrates convergence via a different temporal engineering: gating rather than compression. To be exact, Cas9 is fused to the N-terminal degron of Geminin, which APC/C-Cdh1 recognizes and ubiquitinates in G1 (21). In S/G2, however, CDKs phosphorylate Cdh1 at multiple sites effectively silencing APC/C-Cdh1. Geminin, in turn, accumulates during these phases. Originally engineered as an HDR solution, Geminin-Cas9 effectively halves the cumulative active time of Cas9, experimentally demonstrating synergistic enhancement of HDR and reduction of OTEs (22). The unintended OTE benefit confirms the mechanistic convergence of HDR and OTEs, both affected by different interventions of temporal control.

S1mplex systems embody this functional convergence, deliberately combining three advantage points into a single packet. At the molecular level, the system is composed of S1m-sgRNA; streptavidin, which connects it to biotinylated single-stranded oligodeoxynucleotide (ssODN, where biotin-ssODN is the donor template); and lastly Cas9 forming the RNP complex with the S1m-sgRNA (Figure 2) (23). S1mplex was designed at the convergence of RNP delivery and donor proximity. While the former benefits both OTE and HDR, spatial control via donor tethering specifically targets HDR, demonstrating that even though the main convergence lies on the temporal axis, more specialized interventions can also be

integrated into this bioengineering framework to further optimize the HDR-OTE problem.

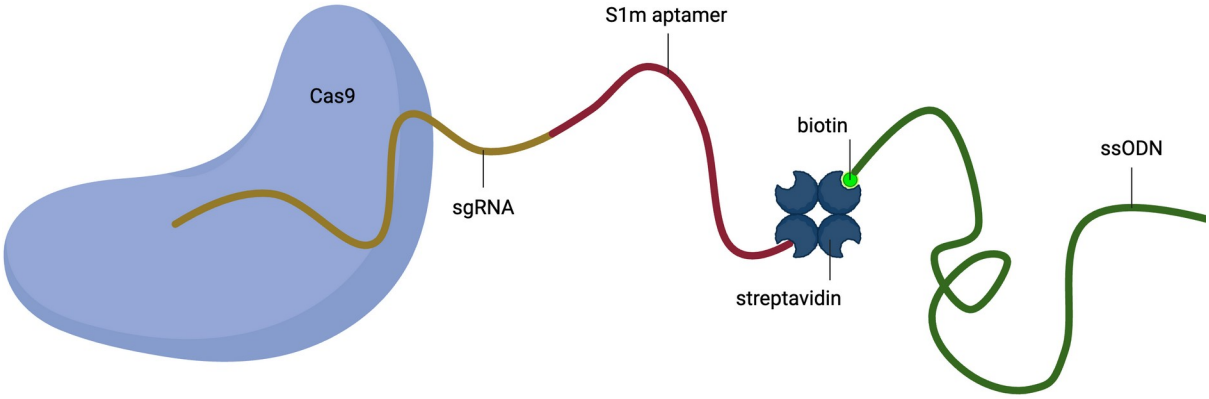

Figure 2. Architecture of the S1mplex system. The Cas9–sgRNA ribonucleoprotein is linked through an S1m aptamer and streptavidin to a biotinylated ssODN donor, tethering the repair template directly to the editing complex. The convergence lies in the combination of transient RNP delivery, which limits Cas9 exposure and OTE accumulation, with donor tethering, which increases local donor availability for HDR. Based on the S1mplex architecture described in (23). Created in BioRender. Thaulero, F. (2026) https://BioRender.com/q1ekoah.

HDR primarily occurs in the S and G2 phases of interphase as it requires a repair template, which in this case corresponds to the sister chromatids. Even if an artificial donor is supplied, and therefore the presence of sister chromatids is irrelevant, the end-resection machinery that initiates HDR is only active during S/G2 (24, 25). Outside this window, NHEJ dominates. Therefore, in non-dividing cells such as post-mitotic neurons, temporal control of Cas9 activity does not help HDR because there is no S/G2 phase to synchronize with. Convergence is present but not universal. Furthermore, temporal control is the most empirically evident shared axis, explaining the reason behind some interventions succeeding in both domains. Thus, mechanistic convergence demonstrates that progress is not necessarily achieved simply by maximizing each variable: identifying shared control points reshapes the optimization frontier itself.

2.4 Structural convergence

A third level of convergence emerges even without shared upstream variables. On-target Cas9 activity, off-target reduction, and HDR efficiency construct a triangle of mutually influencing elements where change in one directly affects the others, complicating the clinical usefulness of isolated maximization. This is structural coupling: the strategies share a substrate, the DSB itself.

High-fidelity Cas9 variants illustrate the first variable of this triangle. Each variant (SpCas9-HF1, HypaCas9, eSpCas9(1.1), evoCas9) reduces OTEs by raising the energetic or conformational threshold for cleavage (26-29). The threshold for cleavage, however, also applies to on-target sites, where SpCas9-HF1 retains 85% efficiency and HypaCas9 70%: on-target reductions persist across variants (26, 27). The trade-off was so robust that the field actively targeted it via Sniper2L, an

evolved version of SniperCas9. This is a recent partial exception, retaining wild-type comparable activity by enhancing unwinding failure in the presence of even a single mismatch (30). However, no engineered variant, including Sniper2L, has been accepted for direct clinical use. An integrated bioengineering framework considering the activity-specificity trade-off still has to be truly optimized.

The engineering strategy that increases specificity in the dominant class of variants, raising binding or cleavage thresholds, extends to on-target sites. Because HDR recruitment requires a DSB, any reduction in activity propagates directly to HDR. The impact is dramatic since HDR efficiency is already low in primary cells. Only at the end of a chain of conditional events, on-target cleavage, donor access, repair-pathway choice, and edited-cell survival does a usable and high-fidelity HDR edit emerge (9). In primary cells, baseline HDR efficiencies are typically 4 to 10-fold lower than in immortalized cell lines, and in long-term engrafting HSCs without selection, HDR allele prevalence often falls below 2.5% (31, 32). Therefore, small reductions in nuclease activity can have drastic downstream effects on HDR efficiency, as its baseline values are already remarkably low. Even though Sniper2L is an exception, it must however be noted that original characterization reports indel rates, and performance in clinical contexts has not yet been tested. The activity-HDR vertex therefore holds across every variant currently used clinically, and remains uncharacterized even for the most prominent variant that decouples activity from specificity *in vitro*.

Table 1. Summary of how the principal interventions are distributed across the three axes

Intervention	Control axis	Activity	Specificity (OTE↓)	HDR	Reported quantitative effect
SpCas9-HF1 / HypaCas9	Energetic / conformational threshold	↓ (85% / >70% WT)	↑↑	↓ (via activity)	OTEs near-undetectable by GUIDE-seq (26, 27, 33)
Sniper2L	Mismatch-sensitised unwinding	≈	↑↑	≈ (in vitro)	Partial exception; not clinically validated (30)
RNP delivery	Temporal compression (hrs vs days)	≈	↑↑	↑	10–100× OTE reduction; concentrates Cas9 in S/G2 (20)
Geminin-Cas9 fusion	Temporal gating to S/G2	≈ (slight ↓)	↑	↑	~Halves cumulative active time; synergistic HDR↑/OTE↓ (21, 22)
S1mplex	RNP + donor tethering (spatial)	≈	↑↑	↑↑	Combines temporal + proximity control (23)
AAV6 donor dose	Donor concentration	n/a	↓ (↑ exposure)	↑	40–60% HDR gain at ≥4×10^4 vg/cell; plateaus (13)

Legend: ↑ increase; ↑↑ strong increase; ↓ decrease; ≈ approximately unchanged vs wild-type; n/a not applicable to this axis.

The activity-specificity-HDR trade-off is thus best reinterpreted as a bioengineering Pareto-like frontier: a constrained solution within a multi-objective framework where no single

point maximizes all three metrics (Figure 3) (34). Optimization becomes context-dependent, with different clinical interventions prioritizing different vertices of the triangle. For example, *ex vivo* correction of monogenic disorders via HDR, such as the trial using Cas9/AAV6 delivery for β-thalassemia, prioritizes HDR efficiency because outcomes that only generate indels cannot replace the HBA1 gene with an entire HBB gene (35). Conversely, *in vivo* somatic editing, exemplified by NTLA-2001 for transthyretin amyloidosis, prioritizes specificity because it is an irreversible systemic CRISPR intervention that needs a high therapeutic index (15). Therefore, structural convergence becomes the point at which optimization is inevitable. Once activity, specificity, and HDR become mutually constraining, the engineering challenge shifts from maximizing individual objectives to evaluating whether the resulting frontier is clinically acceptable or whether it must be escaped through circumvention or reshaped through collapse.

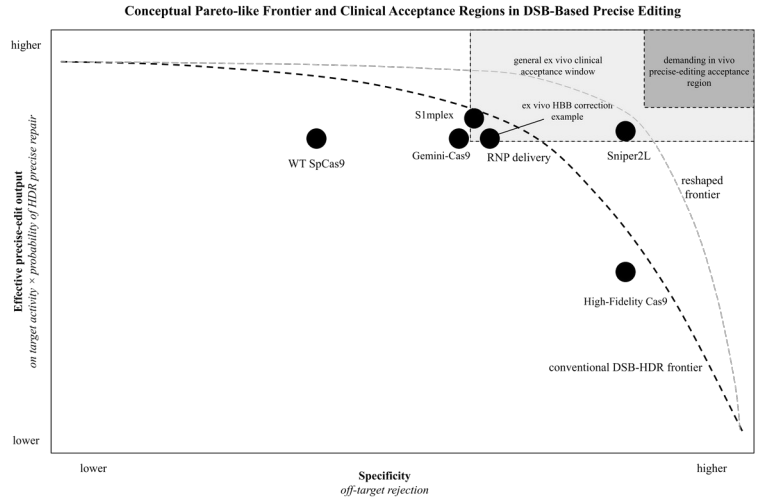

Figure 3. Conceptual Pareto-like frontier and clinical acceptance regions in DSB-based precise editing. The dashed curve represents the Pareto-like frontier of the theoretically optimized conventional DSB-HDR trade-offs. A one-dimensional schematic is used because this review treats HDR as partly downstream of activity in DSB-based precise editing, rather than as an independently plotted third axis. Positions are schematic and indicate the relative direction of intervention effects, not absolute percentages of performance. RNP delivery, Geminin-Cas9, and S1mplex are placed near the frontier to show optimization within the DSB-based system, while Sniper2L is placed beyond the conventional frontier because it represents possible partial collapse of the activity-specificity coupling. Distances between points are not quantitative; the RNP, Geminin-Cas9, and S1mplex cluster is visually expanded to distinguish their relative effects, even though these interventions remain broadly close to WT SpCas9 in on-target activity. The *ex vivo* HBB correction example is shown as a reachable precise-editing context, whereas the demanding in vivo precise-editing region represents a higher clinical threshold that DSB-HDR systems may fail to intersect. These regions are deliberately broad and schematic; real clinical applications would refine them according to disease severity, target tissue, delivery route, acceptable risk, required editing fraction, and repair-outcome tolerance. It must be noted that the model is purely theoretical and not empirically derived: it does not solve OTEs, HDR inefficiency, or genomic instability, but provides a benchmarking framework for comparing interventions against application-specific clinical requirements.

2.5 Synthesis

The prior sections examined three levels of convergence. Operational convergence reveals how clinical editing cannot optimize OTE reduction and HDR enhancement sequentially because finite cell number, delivery dose, and manipulation windows force simultaneous intervention. Mechanistic convergence shows that upstream control variables, mainly Cas9's exposure time, can be manipulated to improve both on-target editing and HDR-competent repair. Structural convergence then demonstrates that activity, specificity, and HDR efficiency form a mutually constraining frontier within DSB-based CRISPR-Cas9 systems.

Convergence thus becomes significant not simply because it demonstrates that inter-variable interaction can occur, but because it extends these interactions into a frontier that optimization can act upon. Once such convergent variables are identified, clinical success depends on efficient navigation of the optimization frontier. Yet, a frontier may be navigable, but also clinically unacceptable if the architecture rests on a clinically disqualifying mechanism, or potentially clinically viable but heavily restricted by strong mechanistic coupling between variables. This transition from optimization within a frontier to evaluation of the frontier-generating architecture motivates the next section. In this way, convergence becomes the motor of an iterative optimization-circumvention-collapse framework.

3. Beyond optimization: circumvention and collapse

3.1 The limits of optimization

Because convergence generates a Pareto-like frontier within DSB-based CRISPR-Cas9, the first engineering impulse may be optimization: to select, engineer, or experimentally identify the best balance among convergent variables within the current editing architecture to meet specific clinical requirements. Nonetheless, optimization alone represents only one possible response to convergence. A Pareto frontier describes the best compromise available within a definite editing framework, but does not determine whether such architecture should be preserved, circumvented, or fundamentally reshaped. If the current editing frontier overlaps with the clinical acceptance region, optimization may be sufficient. If the frontier fails to intersect that region because one of the architecture's required mechanistic dependencies (e.g., DSB-based systems require DSBs) imposes a level of control, efficiency, predictability, or safety that cannot be achieved at acceptable cost or risk, circumvention may be necessary. If the frontier is clinically promising but compressed by a strong mechanistic coupling between variables, collapse becomes the more powerful strategy because it does not abandon the architecture, but instead reshapes the frontier itself.

This paper therefore extends convergence into an iterative framework of optimization, circumvention, and collapse. Optimization refers to the navigation of the existing Pareto frontier within a given architecture. Circumvention refers to the invention or adoption of a new editing paradigm whose required mechanistic dependencies differ from those of the original architecture. This may occur either because the current frontier is clinically unviable, or because an alternative

architecture may possess a more favorable optimization frontier. Collapse refers to the dissociation of conventionally coupled variables, such that improvement in one variable no longer necessarily worsens another. For example, a collapse of the activity-specificity coupling would mean that increased specificity no longer produces a major loss of editing activity. Unlike circumvention, collapse does not escape the architecture: it alters the internal structure of the trade-off. It must also be noted that collapse remains largely theoretical within the CRISPR field.

The following sections discuss these two strategies sequentially. Circumvention is first considered as an architectural replacement, either a response to a clinically-limiting constraint or pursued in search of a more favorable editing frontier. Then, collapse is interpreted as mechanistic decoupling between convergent variables within an architecture possessing highly unfavorable editing frontiers (due to severe coupling between variables). Only following a thorough analysis of these strategies can the full iterative framework be established.

3.2 Circumvention

Circumvention operates at the architectural level. Unlike optimization, which determines the best point on the frontier to meet an objective, circumvention questions the architecture that generates the frontier. Circumvention thus becomes significant when the current editing architecture is clinically insufficient or when alternative architectures may generate a frontier that is more amenable to optimization.

Architecture-intrinsic constraints emerge from a required mechanistic dependency integrated in the editing logic of the architecture. Within DSB-based precise editing, the architecture-intrinsic constraint is not HDR alone, but rather the dependency of precise editing on double-strand breaks followed by endogenous repair pathway choice producing the desired outcome with sufficient efficiency, predictability, and safety. It was already established that HDR efficiency is coupled to the activity and specificity of Cas9. Nonetheless, activity and specificity of Cas9 do not impact the architecture's deeper contingency upon endogenous DSB repair. This DSB-based editing architecture thus requires DSBs to be repaired using a therapeutically precise pathway while avoiding NHEJ, MMEJ, large deletions, indels, rearrangements, p53 activation, and translocations. Such risks are therefore not intrinsic to the endogenous HDR repair pathway itself, but to the broader DSB-dependent editing architecture.

An architecture-intrinsic constraint is considered disqualifying for a given clinical objective when the control burden imposed by that required dependency prevents the intersection of the architecture's Pareto frontier with the clinical acceptance region. Because clinical objectives are context-dependent, disqualification is not absolute for any given editing architecture. For instance, even though DSB-based systems with HDR are used in some *ex vivo* contexts, they may be unacceptable in harder *in vivo* applications where delivery, cell-cycle state, repair heterogeneity, and genomic risk-tolerance become significantly harder to control. In other words, even a Pareto-optimal point may be disqualifying if it possesses unacceptable

risks of oncogenic mutations, translocations, deletions, or chromosomal rearrangements. In this context, the complexity of optimizing all features of DSB-dependent HDR into the clinical acceptance region may outweigh the benefit of preserving the architecture.

The engineering response therefore depends on the kind of constraint identified. In the presence of a disqualifying architecture-intrinsic constraint, circumvention has priority. This form may be called mandatory circumvention, because the current architecture cannot reach the clinical acceptance region without replacing one of its required dependencies. More precisely, even if a coupling constraint within that architecture was resolved through collapse, the architecture would remain clinically unviable. For example, even if the activity-specificity coupling of Cas9 were fully collapsed and independently optimized, the DSB-based architecture remains clinically insufficient since it depends on endogenous repair of the DSB towards the desired HDR result. In such cases, circumventing the current architecture's limiting mechanism is mandatory to generate a new system, such as base editing or prime editing, that no longer depends on the same DSB mechanism.

Base and prime editors both represent empirical examples of circumvention (36, 37). DSB-based systems' potentially disqualifying (depending on the clinical objective) architecture-intrinsic constraint is the need to generate a double-strand break and then rely on a naturally less favored precise repair outcome. Base editors circumvent this dependency by avoiding double-strand breaks altogether, using instead Cas-guided chemical base conversion (36). The new architecture-intrinsic constraint is therefore not endogenous DSB repair, but the requirement that the desired edit be achievable through a chemically permitted base conversion within the editor's accessible activity window (36). Similarly, prime editors bypass DSBs by replacing them with pegRNA-guided reverse transcription and incorporation of the intended edit (37). The fundamental causal node of this architecture is therefore the need for the pegRNA-encoded edit to be reverse-transcribed and successfully incorporated into the target locus. Such empirical instances support the proposition that new architectures do not merely eliminate architecture-intrinsic constraints; on the contrary, such constraints are replaced by new mechanistic dependencies that may be non-disqualifying according to the clinical or engineering objective. It should be noted that broader optimization variables such as product purity, efficiency, delivery, and bystander editing still emerge, yet these are downstream frontier variables rather than the core architectural constraint (38, 39).

Yet, circumvention does not guarantee that the architecture replacement will be better than the previous one in terms of clinical viability. Indeed, the new editing architecture's dependencies and convergent variables should be analyzed to ascertain whether this new system simply replaced a former disqualifying architecture-intrinsic constraint with a new one. In such a case, circumvention has not solved the problem for the intended application, and new editing paradigms with non-disqualifying architecture-intrinsic constraints should be sought.

Circumvention may also be pursued strategically (as opposed to *mandatory* circumvention). Once a non-disqualifying

editing architecture has been found, the focus shifts onto the severity of coupling constraints. If constraints are strongly negatively coupled, collapse and circumvention may both be pursued. While collapse is attempted within the original architecture, circumvention may be pursued strategically to explore alternative architectures whose architecture-intrinsic constraints are non-disqualifying and whose convergent variables are less severely coupled. Therefore, strategic circumvention does not replace collapse; rather, it protects against the uncertainty of collapse.

In both mandatory and strategic circumvention, the design cycle restarts. Simply because a new editing system lacks the previous disqualifying architecture-intrinsic constraint does not guarantee that no equally disqualifying constraint has emerged. Hence, its required dependencies and convergent variables must be identified before optimization or collapse can be rationally pursued.

3.3 Collapse

Collapse is sought because of the extreme return on investment: a successful collapse weakens or eliminates one mechanistic coupling, thereby making the old trade-off much less restrictive. Whereas optimization navigates the frontier, collapse alters its very shape. Therefore, collapse becomes most effective within an architecture that lacks disqualifying constraints, but whose convergent variables are so strongly coupled that optimization alone is insufficient to reach the clinically acceptable region.

Given that collapse remains largely conceptual within the CRISPR domain, the closest relevant example is nucleoside-modified mRNA. In early mRNA therapeutics, there was a severe negative coupling between protein expression and innate immune activation, such that increasing protein expression increased antigen production alongside unwanted effects such as increased innate immune sensing, inflammation, and translational suppression (40). The incorporation of modified nucleosides, namely pseudouridine and later N1-methylpseudouridine, weakened the coupling by maintaining high translation with low innate immune activation (40, 41). A similar example of collapse is Sniper2L, which may weaken the activity-specificity coupling by retaining high on-target editing while increasing specificity (30). However, because it operates within DSB-based architectures, this partial collapse would not remove the DSB-break dependency and may remain insufficient for demanding *in vivo* precise-editing clinical applications. Moreover, such technology has not been clinically applied as nucleoside-modified mRNA has, which remains one of the most eminent examples of how mechanistic decoupling can reshape the optimization frontier.

Coupling constraints, generally motivating a collapse-oriented solution, arise when two or more convergent variables mutually limit each other. For instance, in the activity-specificity-HDR triangle, activity and specificity represent a coupling constraint, as they are negatively coupled: increasing one decreases the other. Such coupling constraints are a direct consequence of gene editing's convergence property: variables do not merely coexist, but rather become mutually restrictive within the same clinical objective.

The distinction between strategic circumvention and collapse is extremely significant when encountering a strong negative coupling constraint. Given the extremely complex nature of collapse, which requires weakening the causal mechanism that links competing objectives without destroying the function that mechanism originally enabled, such a strategy should be treated as a high-investment, high-return research program. For instance, if an architecture possesses no disqualifying architecture-intrinsic constraint but suffers from an extreme activity-product purity coupling, collapse would aim to weaken that coupling directly. However, because such decoupling is highly unlikely, it is reasonable to investigate another architecture whose frontier may be easier to optimize.

In this sense, collapse and optimization are not mutually exclusive: researchers may pursue collapse-oriented engineering within one architecture while optimizing another system for more immediate clinical applications. If collapse succeeds, optimization does not become irrelevant. The new reshaped frontier must be optimized according to the clinical or engineering goal; thus, collapse simply changes the frontier upon which optimization acts.

For example, suppose that activity and specificity in the activity-specificity-HDR triangle were uncoupled. Consequently, an increase in activity increases HDR without a major increase in OTEs. Still, this does not mean that optimization is no longer applicable. Several other variables that have not been considered in the triangle must still be optimized, namely HDR efficiency, donor-template design, delivery format, editor exposure time, cell type, toxicity, and acceptable repair outcome. Therefore, collapse simply removes or weakens the coupling between two vertices, but it does not follow that all other convergent variables are automatically eliminated.

3.4 The iterative Optimization-Circumvention-Collapse framework

Having established optimization, circumvention, and collapse as the primary engineering pathways, the full iterative framework can now be assembled. This model does not treat these strategies as mutually exclusive checkpoints. Instead, optimization is a response to convergent variables, collapse is motivated by severely coupled variables, and circumvention follows from architectural dependencies whose clinical consequences are measured through the changes in those variables.

Convergence analysis therefore diagnoses the Pareto frontier that optimization subsequently navigates, although optimization-oriented experiments may later reveal additional limiting variables. This enables optimization of trade-offs within the existing editing architecture by selecting the best achievable balance among convergent variables according to a pre-defined clinical or engineering goal. However, the identification of convergence enables the diagnosis of the constraints that determine whether optimization, circumvention, or collapse should be pursued.

This model focuses mainly on architecture-intrinsic and coupling constraints because these are often the key determinants of architecture preservation, replacement, or reshaping. Other constraints certainly exist,

such as delivery format, dosage, guide design, expression duration, donor concentration, and experimental timing. Yet, these are generally more directly tunable and can be addressed through optimization unless they become incisive enough to interfere with the architecture's clinical viability.

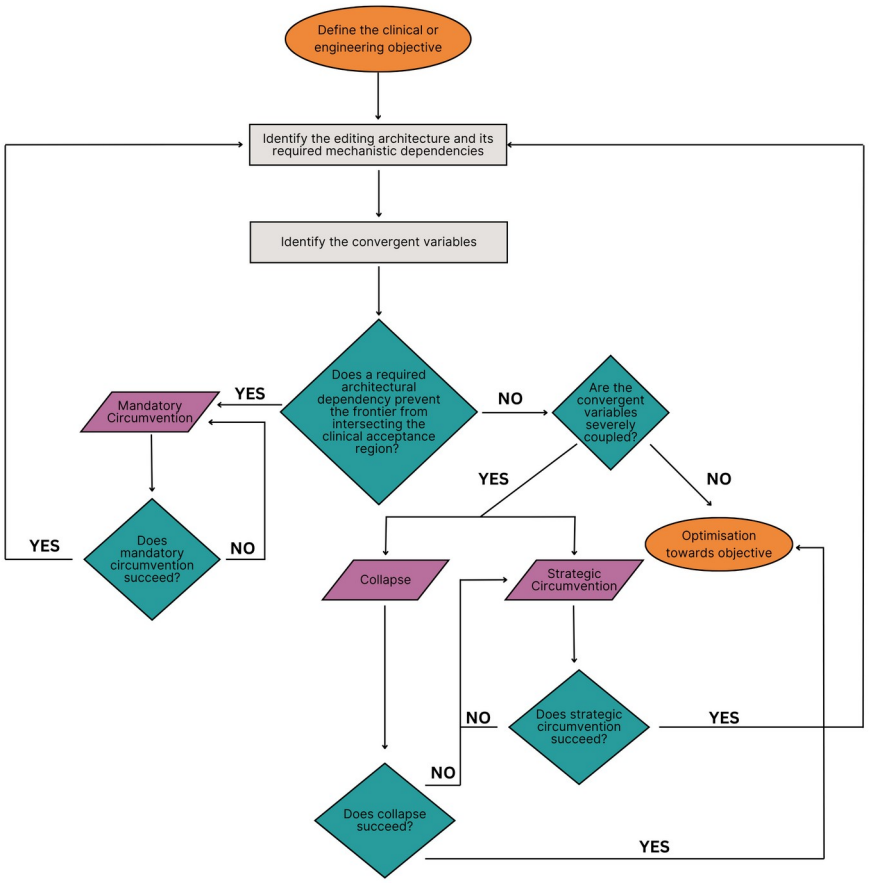

Figure 4. The Optimization-Circumvention-Collapse model. The flowchart begins with a defined clinical or engineering objective. The relevant editing architecture and its required mechanistic dependencies are then identified, followed by the convergent variables that determine the system's optimization frontier. If a required architectural dependency prevents the frontier from intersecting the clinical acceptance region, mandatory circumvention is pursued, and the new architecture is evaluated from the beginning. If no architecture-intrinsic constraint is disqualifying, the severity of coupling constraints determines the next response. If coupling is moderate, optimization within the current frontier may be sufficient. If coupling is severe, collapse may be pursued within the current architecture while strategic circumvention explores alternative architectures in parallel. In both cases, optimization remains necessary because every preserved, circumvented, or collapsed system must eventually be tuned toward the clinical objective. Orange nodes indicate the starting objective and final operational optimization step, purple parallelograms indicate the other engineering responses, teal diamonds indicate decision points, and grey rectangles indicate analytical identification steps.

Convergence remains the central motor of the iterative framework. Convergence identifies the coupling constraints, which may motivate collapse; convergent variables are the performance metrics used to measure architecture-intrinsic constraints that drive circumvention; and convergence of variables into a trade-off enables optimization. While circumvention explores architecture alternatives and collapse reshapes the intra-architecture frontier, optimization ultimately remains the final operational step because every architecture, whether preserved, circumvented, or collapsed, must still be optimized toward the clinical or engineering objective.

4. Conclusion and perspective

This review has formalized how OTE reduction and HDR maximization, though increasingly co-evaluated in modern CRISPR engineering, converge at three levels within a Pareto-like design framework. Operationally, finite cell numbers, fixed *ex vivo* systems, limited intervention time, and shared delivery doses of editing technology force the two to be optimized together. At the mechanistic level, the temporal dimension of Cas9 controls both off-target accumulation and synchronization with HDR-competent windows, hence single interventions such as RNP delivery or Geminin-Cas9 fusion simultaneously improve both. Lastly, structural convergence is observed, as on-target activity, specificity, and HDR efficiency shape a mutually limiting triangle best interpreted as a Pareto-like frontier: no single point maximizes all three variables. Hence, improvement in DSB-based systems shifts from a search for the "best" possible system to a context-dependent analysis of this frontier.

However, the broader significance of convergence is diagnostic. That is, convergence does not simply generate a frontier to be optimized; it also helps identify the appropriate engineering response. Optimization–to select, engineer, or experimentally identify the best balance among convergent variables–is the best response if the frontier already intersects the clinical acceptance region. On the contrary, if a frontier is clinically promising but compressed by severe coupling, collapsing the mechanistic relation between them redefines the trade-off, and, by extension, the optimization frontier. If, however, convergence reveals that the architecture's required dependency prevents the frontier from reaching its clinical goal, circumvention may be necessary. Concretely, in DSB-based CRISPR systems, the architecture-intrinsic constraint is the reliance on double-strand break formation followed by endogenous repair-pathway choice producing the desired precise outcome. In demanding clinical applications, this causal node may impose an excessive burden on safety, efficiency, and predictability.

The optimization-circumvention-collapse framework therefore relocates optimization as the final operational step. Circumvention of DSB-based systems via base and prime editors does not nullify architecture-intrinsic constraints; it simply replaces the frontier with a potentially more favorable one. Similarly, collapse reshapes the frontier upon which optimization acts. Additionally, convergence remains the central motor because it generates the frontier to be optimized, identifies couplings that may be collapsed, and reveals when architectural dependencies should be circumvented.

Future research should therefore reconstruct context-specific frontiers rather than a singular, universal one. Such frontiers will vary with cell type, locus, guide RNA, donor design, delivery system, and clinical context. Furthermore, these convergent variables should all be measured within the same unified experimental system, not inferred across studies, to effectively build an architecture-coherent frontier. Then, frontier mapping would guide clinical and engineering decisions by exposing which variables should be tuned to maximize intersection with the clinical acceptance region, and therefore whether the next iteration should optimize delivery or dosage, target a coupling mechanism, or test an alternative architecture. Another priority is identifying which mechanistic couplings within promising architectures could be collapsed. Given the complexity of redesigning a causal mechanism deeply ingrained in molecular biology, claims of collapse should be substantiated with proof that the trade-off itself has been weakened or eliminated, not merely fine-tuned towards better performance. Ultimately, progress in gene editing depends on choosing, reshaping, and optimizing the frontier most compatible with a given clinical objective.

5. References

1. Farasati Far, B., Akbari, M., Habibi, M. A., Katavand, M., & Nasseri, S. (2025). *CRISPR technology in disease management: An updated review of clinical translation and therapeutic potential*. *Cell Proliferation, 58*(11), e70099. https://doi.org/10.1111/cpr.70099

2. Guo, C., Ma, X., Gao, F., & Guo, Y. (2023). *Off-target effects in CRISPR/Cas9 gene editing*. *Frontiers in Bioengineering and Biotechnology, 11*, 1143157. https://doi.org/10.3389/fbioe.2023.1143157

3. Vogt, A., & He, Y. (2023). *Structure and mechanism in non-homologous end joining*. *DNA Repair, 130*, 103547. https://doi.org/10.1016/j.dnarep.2023.103547

4. Doudna, J. A., & Charpentier, E. (2014). *The new frontier of genome engineering with CRISPR-Cas9*. *Science, 346*(6213), 1258096. https://doi.org/10.1126/science.1258096

5. Acharya, S., Ansari, A. H., Das, P. K., Hirano, S., Aich, M., Rauthan, R., et al. (2024). *PAM-flexible engineered FnCas9 variants for robust and ultra-precise genome editing and diagnostics*. *Nature Communications, 15*, 5471. https://doi.org/10.1038/s41467-024-49233-w

6. Kalter, N., Fuster-García, C., Silva, A., Ronco-Díaz, V., Roncelli, S., Turchiano, G., et al. (2025). *Off-target effects in CRISPR-Cas genome editing for human therapeutics: Progress and challenges*. *Molecular Therapy: Nucleic Acids*, 102636. https://doi.org/10.1016/j.omtn.2025.102636

7. Ewaisha, R., & Anderson, K. S. (2023). *Immunogenicity of CRISPR therapeutics—critical considerations for clinical translation*. *Frontiers in Bioengineering and Biotechnology, 11*, 1138596. https://doi.org/10.3389/fbioe.2023.1138596

8. Guo, T., Feng, Y.-L., Xiao, J.-J., Liu, Q., Sun, X.-N., Xiang, J.-F., et al. (2018). *Harnessing accurate non-homologous end joining for efficient precise deletion in CRISPR/Cas9-mediated genome editing*. *Genome Biology, 19*(1), 170. https://doi.org/10.1186/s13059-018-1518-x

9. Liao, H., Wu, J., VanDusen, N. J., Li, Y., & Zheng, Y. (2024). *CRISPR/Cas9-mediated homology-directed repair for precise gene editing*. *Molecular Therapy: Nucleic Acids, 35*, 102344. https://doi.org/10.1016/j.omtn.2024.102344

10. Roth, T. L., Puig-Saus, C., Yu, R., Shifrut, E., Carnevale, J., Li, P. J., et al. (2018). *Reprogramming human T cell function and specificity with non-viral genome targeting*. *Nature, 559*(7714), 405–409. https://doi.org/10.1038/s41586-018-0326-5

11. Schiroli, G., Conti, A., Ferrari, S., della Volpe, L., Jacob, A., Albano, L., et al. (2019). *Precise gene editing preserves hematopoietic stem cell function following transient p53-mediated DNA damage response*. *Cell Stem Cell, 24*(4), 551–565.e8. https://doi.org/10.1016/j.stem.2019.02.019

12. Frangoul, H., Altshuler, D., Cappellini, M. D., Chen, Y.-S., Domm, J., Eustace, B. K., et al. (2021). *CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia*. *New England Journal of Medicine, 384*(3), 252–260. https://doi.org/10.1056/NEJMoa2031054

13. Lam, A. J., Lin, D. T. S., Gillies, J. K., Uday, P., Pesenacker, A. M., Kobor, M. S., et al. (2021). *Optimized CRISPR-mediated gene knockin reveals FOXP3-independent maintenance of human Treg identity*. *Cell Reports, 36*(5), 109494. https://doi.org/10.1016/j.celrep.2021.109494

14. Hsu, P. D., Lander, E. S., & Zhang, F. (2014). *Development and applications of CRISPR-Cas9 for genome engineering*. *Cell, 157*(6), 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010

15. Gillmore, J. D., Gane, E., Taubel, J., Kao, J., Fontana, M., Maitland, M. L., et al. (2021). *CRISPR-Cas9 in vivo gene editing for transthyretin amyloidosis*. *New England Journal of Medicine, 385*(6), 493–502. https://doi.org/10.1056/NEJMoa2107454

16. Walther, J., Wilbie, D., Tissingh, V. S. J., Öktem, M., van der Veen, H., Lou, B., et al. (2022). *Impact of formulation conditions on lipid nanoparticle characteristics and functional*

delivery of CRISPR RNP for gene knock-out and correction. *Pharmaceutics, 14*(1), 213. https://doi.org/10.3390/pharmaceutics14010213

17. Liu, M.-S., Gong, S., Yu, H.-H., Jung, K., Johnson, K. A., & Taylor, D. W. (2020). *Engineered CRISPR/Cas9 enzymes improve discrimination by slowing DNA cleavage to allow release of off-target DNA*. *Nature Communications, 11*, 3576. https://doi.org/10.1038/s41467-020-17411-1

18. Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., et al. (2014). *Structures of Cas9 endonucleases reveal RNA-mediated conformational activation*. *Science, 343*(6176), 1247997. https://doi.org/10.1126/science.1247997

19. Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., et al. (2014). *Crystal structure of Cas9 in complex with guide RNA and target DNA*. *Cell, 156*(5), 935–949. https://doi.org/10.1016/j.cell.2014.02.001

20. Kim, S., Kim, D., Cho, S. W., Kim, J., & Kim, J.-S. (2014). *Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins*. *Genome Research, 24*(6), 1012–1019. https://doi.org/10.1101/gr.171322.113

21. Gutschner, T., Haemmerle, M., Genovese, G., Draetta, G. F., & Chin, L. (2016). *Post-translational regulation of Cas9 during G1 enhances homology-directed repair*. *Cell Reports, 14*(6), 1555–1566. https://doi.org/10.1016/j.celrep.2016.01.019

22. Matsumoto, D., Kishi, K., Matsugi, E., Inoue, Y., Nigorikawa, K., & Nomura, W. (2023). *Cas9-Geminin and Cdt1-fused anti-CRISPR protein synergistically increase editing accuracy*. *FEBS Letters, 597*(7), 985–994. https://doi.org/10.1002/1873-3468.14608

23. Carlson-Stevermer, J., Abdeen, A. A., Kohlenberg, L., Goedland, M., Molugu, K., Lou, M., et al. (2017). *Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing*. *Nature Communications, 8*, 1711. https://doi.org/10.1038/s41467-017-01875-9

24. Anand, R., Jasrotia, A., Bundschuh, D., Howard, S. M., Ranjha, L., Stucki, M., et al. (2019). *NBS1 promotes the endonuclease activity of the MRE11-RAD50 complex by sensing CtIP phosphorylation*. *The EMBO Journal, 38*(7), e101005. https://doi.org/10.15252/embj.2018101005

25. Zhao, F., Kim, W., Kloeber, J. A., & Lou, Z. (2020). *DNA end resection and its role in DNA replication and DSB repair choice in mammalian cells*. *Experimental & Molecular Medicine, 52*(10), 1705–1714. https://doi.org/10.1038/s12276-020-00519-1

26. Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., et al. (2016). *High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects*. *Nature, 529*(7587), 490–495. https://doi.org/10.1038/nature16526

27. Chen, J. S., Dagdas, Y. S., Kleinstiver, B. P., Welch, M. M., Sousa, A. A., Harrington, L. B., et al. (2017). *Enhanced proofreading governs CRISPR-Cas9 targeting accuracy*. *Nature, 550*(7676), 407–410. https://doi.org/10.1038/nature24268

28. Slaymaker, I. M., Gao, L., Zetsche, B., Scott, D. A., Yan, W. X., & Zhang, F. (2016). *Rationally engineered Cas9 nucleases with improved specificity*. *Science, 351*(6268), 84–88. https://doi.org/10.1126/science.aad5227

29. Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., et al. (2018). *A highly specific SpCas9 variant is identified by in vivo screening in yeast*. *Nature Biotechnology, 36*(3), 265–271. https://doi.org/10.1038/nbt.4066

30. Kim, Y.-H., Kim, N., Okafor, I., Choi, S., Min, S., Lee, J., et al. (2023). *Sniper2L is a high-fidelity Cas9 variant with high activity*. *Nature Chemical Biology, 19*(8), 972–980. https://doi.org/10.1038/s41589-023-01279-5

31. Xu, X., Gao, D., Wang, P., Chen, J., Ruan, J., Xu, J., et al. (2018). *Efficient homology-directed gene editing by CRISPR/Cas9 in human stem and primary cells using tube electroporation*. *Scientific Reports, 8*(1), Article 11649. https://doi.org/10.1038/s41598-018-30227-w

32. Shin, J., Schröder, M., Caiado, F., Wyman, S. K., Bray, N., Bordi, M., et al. (2020). *Controlled cycling and quiescence enables efficient HDR in engraftment-enriched adult hematopoietic stem and progenitor cells*. *Cell Reports, 32*(9), 108093. https://doi.org/10.1016/j.celrep.2020.108093

33. Tsai, S. Q., Zheng, Z., Nguyen, N. T., Liebers, M., Topkar, V. V., Thapar, V., et al. (2015). *GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases*. *Nature Biotechnology, 33*(2), 187–197. https://doi.org/10.1038/nbt.3117

34. He, L., Friedman, A. M., & Bailey-Kellogg, C. (2012). *A divide-and-conquer approach to determine the Pareto frontier for optimization of protein engineering experiments*. *Proteins: Structure, Function, and Bioinformatics, 80*(3), 790–806. https://doi.org/10.1002/prot.23237

35. Cromer, M. K., Camarena, J., Martin, R. M., Lesch, B. J., Vakulskas, C. A., Bode, N. M., et al. (2021). *Gene replacement of α-globin with β-globin restores hemoglobin balance in β-*

thalassemia-derived hematopoietic stem and progenitor cells. *Nature Medicine, 27*(4), 677–687. https://doi.org/10.1038/s41591-021-01284-y

36. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., & Liu, D. R. (2016). *Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage*. *Nature, 533*(7603), 420–424. https://doi.org/10.1038/nature17946

37. Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al. (2019). *Search-and-replace genome editing without double-strand breaks or donor DNA*. *Nature, 576*(7785), 149–157. https://doi.org/10.1038/s41586-019-1711-4

38. Chen, P. J., & Liu, D. R. (2023). *Prime editing for precise and highly versatile genome manipulation*. *Nature Reviews Genetics, 24*(3), 161–177. https://doi.org/10.1038/s41576-022-00541-1

39. Arbab, M., Shen, M. W., Mok, B., Wilson, C., Matuszek, Ż., Cassa, C. A., et al. (2020). *Determinants of base editing outcomes from target library analysis and machine learning*. *Cell, 182*(2), 463–480.e30. https://doi.org/10.1016/j.cell.2020.05.037

40. Karikó, K., Buckstein, M., Ni, H., & Weissman, D. (2005). *Suppression of RNA recognition by Toll-like receptors: The impact of nucleoside modification and the evolutionary origin of RNA*. *Immunity, 23*(2), 165–175. https://doi.org/10.1016/j.immuni.2005.06.008

41. Andries, O., Mc Cafferty, S., De Smedt, S. C., Weiss, R., Sanders, N. N., & Kitada, T. (2015). *N1-methylpseudouridine-incorporated mRNA outperforms pseudouridine-incorporated mRNA by providing enhanced protein expression and reduced immunogenicity in mammalian cell lines and mice*. *Journal of Controlled Release, 217*, 337–344. https://doi.org/10.1016/j.jconrel.2015.08.051