Degree project

UNIVERSITY OF SKÖVDE

A Systematic Review and Meta-Analysis of CRISPR-Cas9 Off-Target Effects in Models of Blood-Related Genetic Diseases

Master's Degree Project in Life Sciences

A1E-30credits

Autumn semester

Student: Uduakobong Ruth Ekpe
(a25uduek@student.his.se)

Supervisor: Patric Nilsson
(patric.nilsson@his.se)

Examiner: Niclas Norrstrom
(niclas.norrstrom@his.se)

Abstract

The CRISPR-Cas9 system represents a transformative therapeutic platform for monogenic blood disorders such as sickle cell disease (SCD) and β-thalassemia. However, clinical translation is limited by safety concerns related to unintended off-target genomic alterations. This systematic review and meta-analysis quantified the prevalence and evaluated the severity of CRISPR-Cas9 off-target effects in hematopoietic models of blood-related genetic diseases. Seventy-eight studies were included in qualitative synthesis, and five comparative studies were analysed using a random-effects meta-analysis. High-fidelity Cas9 variants significantly reduced off-target risk compared with wild-type SpCas9 (pooled OR = 0.15, 95% CI: 0.08–0.29), corresponding to an 85% reduction in off-target odds. Substantial heterogeneity was observed (I^2 = 76.3%), primarily attributable to differences in off-target detection methods, as variations in methodological sensitivity can systematically influence the number and frequency of detected off-target events, contributing substantially to the observed heterogeneity. Subgroup analyses showed greater specificity improvements when assessed using unbiased genome-wide methods (OR = 0.08) compared with targeted sequencing (OR = 0.23). Although high-fidelity variants markedly improve safety profiles, methodological variability and limited long-term in vivo data remain important considerations. These findings provide quantitative evidence supporting the potential use of high-fidelity nucleases and comprehensive genome-wide assessment to enhance the safety of CRISPR-based therapies for hemoglobinopathies.

Popular Scientific Summary

Gene editing using CRISPR-Cas9 has opened new possibilities for treating inherited blood disorders such as sickle cell disease and β-thalassemia. These conditions are caused by mutations in a single gene and can lead to lifelong health complications, including severe anemia and organ damage. CRISPR technology works like molecular scissors, allowing scientists to target and modify specific DNA sequences in order to correct these genetic defects. While this approach holds great promise, a major concern is the risk of unintended changes elsewhere in the genome, known as off-target effects. Such unintended edits could potentially disrupt important genes or increase the risk of diseases such as cancer.

This study aimed to better understand how often these off-target effects occur and how serious they might be in models of blood-related genetic diseases. To achieve this, a systematic review and meta-analysis were conducted using data from previously published research. In total, 78 studies were included for qualitative analysis, and five studies provided sufficient data for statistical comparison. The focus was on comparing the standard CRISPR-Cas9 system with newer "high-fidelity" versions that have been engineered to improve accuracy. The findings showed that reported off-target effects vary widely across studies. Some experiments found almost no unintended changes, while others reported relatively high levels. This variation is largely explained by differences in how off-target effects are measured. More advanced methods that scan the entire genome tend to detect more off-target events than methods that only examine predicted locations. This suggests that some earlier studies may have underestimated the true extent of these effects.

Importantly, the results demonstrated that high-fidelity CRISPR tools significantly reduce the likelihood of off-target changes. On average, these improved systems lowered the risk by about 85% compared to the standard version of CRISPR-Cas9. This indicates that advances in gene-editing technology can greatly improve safety. However, the strength of this conclusion is limited by the small number of studies available for direct comparison, meaning that more research is needed to confirm these findings. Another key insight is that the way scientists measure off-target effects plays a crucial role in the results they obtain. Studies using more sensitive, genome-wide methods showed greater improvements when using high-fidelity tools. This highlights the importance of using reliable and comprehensive testing methods when evaluating gene-editing technologies. Overall, this study shows that CRISPR-Cas9 is a powerful tool with strong potential for treating genetic blood disorders, but safety concerns remain. High-fidelity versions of CRISPR offer a promising way to reduce risks, but further research is needed to better understand their long-term effects and performance in real biological systems.

In conclusion, gene-editing therapies are moving closer to clinical use, but careful evaluation of off-target effects is essential to ensure patient safety. By combining improved technologies with better research methods, scientists can continue to develop safer and more effective treatments for inherited diseases.

Table of Contents

1.1. The CRISPR-Cas9 Revolution in Hematology 2
1.2. Research Question, Aim, and Objectives 6
    Research Question 6
    Aim 6
    Objectives: 6
2.1. Objective of the Study 6
2.2. Study Question and PICO Strategy 7
    Population: 7
    Intervention 7
    Comparison: 7
    Outcome 7
2.3. Eligibility Criteria 7
    Inclusion Criteria: 7
    Exclusion Criteria: 7
2.4. Types of Outcome Measures 8
    Primary Outcome Measures 8
    Secondary Outcome Measures 8
2.5. Search Strategy and Information Sources 8
2.6. Study Selection and Data Extraction Process 8
    Study Selection Process 8
    Data Extraction 8
2.7. Quality and Risk of Bias Assessment in Individual Studies 9
2.8. Data Analysis and Statistical Methods 9
    Measures of Effect 9
    Meta-Analysis Model 10
    Assessment of Heterogeneity 10
    Subgroup Analysis 10
    Sensitivity Analysis 10
    Publication Bias Assessment 11
    Prediction Intervals 11
3.1. Study Selection 11
3.2. Study Range and Characteristics 12
    Disease Models 12
    Cell Types 13
    Cas9 Variants 14
    Delivery Methods 14
    Detection Methods 15
3.3. Data Extraction: Qualitative Synthesis 16
    Off-Target Frequency Range 16
    Severity of Off-Target Effects 17
3.4. Data Extraction: Quantitative Synthesis (Meta-Analysis) 17
    Primary Outcome: Prevalence of Off-Target Events 17
    Pooled Analysis: High-Fidelity vs. Wild-Type Cas9 18
    Subgroup Analyses 18

Heterogeneity Assessment....20
Sensitivity Analyses....20
Publication Bias Assessment....21
4.1. Discussion of Main Findings....21
4.2. Interpretation in the Context of Existing Evidence....23
4.3. Biological and Methodological Sources of Heterogeneity....23
Methodological Sources....23
Biological Sources....24
4.4. Strengths and Limitations of This Review....24
4.5. Ethical aspects, societal impact, and regulatory considerations....25
4.6. Conclusion....26
4.7. Recommendation and Future Outlook....26

List of Abbreviations

AAV	Adeno-Associated Virus
AAV6	Adeno-Associated Virus Serotype 6
BCL11A	B-Cell Lymphoma/Leukemia 11A
CI	Confidence Interval
CIRCLE-seq	Circularization for In vitro Reporting of CLeavage Effects by sequencing
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
Digenome-seq	Digested genome sequencing
DISCOVER-seq	Discovery of In Situ Cas Off-targets and Verification by Sequencing
DSB	Double-Strand Break
EMA	European Medicines Agency
eSpCas9	Enhanced Specificity SpCas9
FDA	Food and Drug Administration
gRNA	guide RNA
GUIDE-seq	Genome-wide Unbiased Identification of DSBs Evaluated by Sequencing
HBB	Hemoglobin Subunit Beta
HDR	Homology-Directed Repair
HiFi Cas9	High-Fidelity Cas9
HSPC	Hematopoietic Stem and Progenitor Cell
IND	Investigational New Drug
iPSC	Induced Pluripotent Stem Cell
NGS	Next-Generation Sequencing
NHEJ	Non-Homologous End Joining
PICO	Population, Intervention, Comparison, Outcome
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Q	Cochran's Q Statistic
REML	Restricted Maximum Likelihood
RNP	Ribonucleoprotein
SCD	Sickle Cell Disease
SITE-Seq	Selective Enrichment and Identification of Tagged Genomic DNA Ends by Sequencing
SpCas9	Streptococcus pyogenes Cas9
SpCas9-HF1	SpCas9 High-Fidelity 1
TALEN	Transcription Activator-Like Effector Nuclease
WGS	Whole-Genome Sequencing
WT	Wild-Type
ZFN	Zinc Finger Nuclease
τ^2	Tau-squared (between-study variance)

Introduction

1.1. The CRISPR-Cas9 Revolution in Hematology

The landscape of genetic medicine has been fundamentally reshaped by the advent of CRISPR-Cas9 gene-editing technology (Cong et al., 2013). Its precision, efficiency, and relative simplicity have unlocked unprecedented therapeutic potential for a host of monogenic disorders (ALBAYRAK, 2023). Among the most promising applications are those targeting inherited blood diseases, notably sickle cell disease (SCD) and β-thalassemia. These hemoglobinopathies, caused by specific mutations in the β-globin gene (HBB), lead to severe anemia, pain crises, organ damage, and significantly reduced life expectancy. CRISPR-Cas9 strategies have demonstrated remarkable preclinical and translational success in hematopoietic stem and progenitor cells (DeWitt et al., 2016), primarily through two approaches: the disruption of the BCL11A gene to reactivate the production of fetal hemoglobin, which does not sickle, or the direct correction of the causative mutation in the HBB gene itself (ALBAYRAK, 2023; Mandal et al., 2014; Kim et al., 2014). Despite this transformative potential, the transition from groundbreaking clinical trials to widespread therapeutic adoption is hampered by a persistent and significant safety concern: unintended, off-target mutations. These occur when the CRISPR-Cas9 complex cleaves DNA at genomic loci with sequence similarity to the intended target (Figure 1). Such off-target edits pose a substantial long-term risk, as they could disrupt tumor suppressor genes, activate oncogenes, or cause other deleterious genomic alterations leading to malignancies or other pathologies (Zhang et al., 2015; Vakulskas and Behlke, 2019; Niu et al., 2020; Dong et al., 2021; see Figure 1).

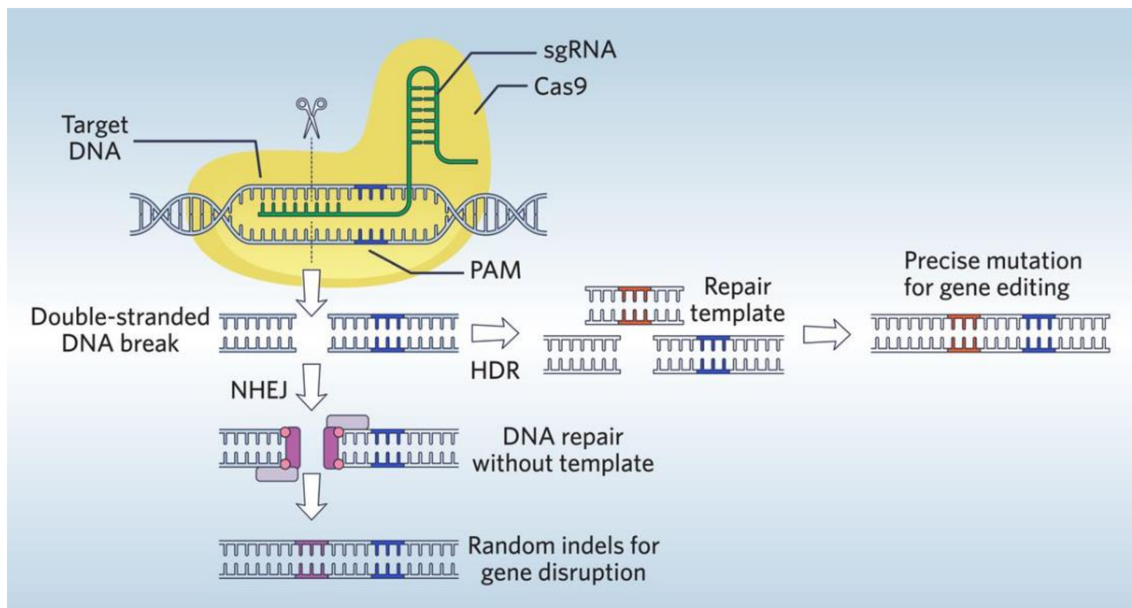

Figure 1. Schematic representation of the CRISPR-Cas9 genome editing mechanism and the origin of off-target effects. The Cas9–sgRNA complex binds to a target DNA sequence adjacent to a PAM site and induces a double-strand break, which is repaired by non-homologous end joining (NHEJ) or homology-directed repair (HDR). Off-target effects occur when partial mismatches between the sgRNA and genomic DNA allow unintended cleavage at similar sequences.

As CRISPR-based therapies move closer to routine clinical use, understanding and mitigating these unintended effects has become a central priority in the field.

Early optimism regarding the specificity of CRISPR-Cas9 arose from initial studies that relied on targeted sequencing and in silico predictions to assess off-target activity, often reporting low frequencies of unintended edits as reported in early off-target analyses (Cho et al., 2013). For example, Mandal et al. (2014) suggested high specificity in hematopoietic stem and progenitor cells (HSPCs), fuelling enthusiasm for therapeutic viability. However, as more comprehensive and sensitive detection methods were developed, a more complex and less uniform picture emerged. Subsequent studies employing unbiased genome-wide approaches revealed that off-target effects could be more prevalent than initially assumed. Cradick et al. (2013) demonstrated detectable off-target cleavage at homologous β-globin loci under sensitive detection conditions, challenging early assumptions of near-perfect specificity. Further refinement of detection strategies, including whole-genome sequencing (WGS) and unbiased techniques such as DISCOVER-Seq, has highlighted substantial variability in off-target activity across studies, even when similar experimental designs or target genes are used (Wienert et al., 2018; Kleinstiver et al., 2016). At the same time, other investigations, such as Rodríguez-Rodríguez et al. (2019), reported comparatively fewer off-target edits in equivalent hematopoietic systems. These contrasting findings underscore that off-target outcomes are not uniform but are highly context-dependent. Reported off-target frequencies range widely, from negligible levels (<1%) to substantial proportions exceeding 20%, reflecting both biological complexity and methodological disparities across studies (Shan et al., 2021).

This variability has prompted extensive investigation into the factors that influence off-target outcomes. Several experimental determinants have emerged as key contributors to differences in observed off-target activity. First, sgRNA design plays a central role, as sequence specificity, guide length, and chemical modifications directly influence targeting accuracy. High-fidelity design strategies aim to maximize on-target activity while minimizing tolerance for mismatches between the guide RNA and genomic DNA (Kleinstiver et al., 2016; Murugan et al., 2020). Second, the delivery modality significantly affects the duration of Cas9 activity within cells. Viral vectors that sustain expression may prolong nuclease exposure and increase the likelihood of off-target cleavage, whereas transient delivery methods such as ribonucleoprotein (RNP) complexes reduce cumulative exposure and associated risk (Rodríguez-Rodríguez et al., 2019; Banakar et al., 2019). In addition, cellular context is an important determinant of editing outcomes. Different cell types, such as hematopoietic stem and progenitor cells (HSPCs) and induced pluripotent stem cells (iPSCs), exhibit distinct chromatin landscapes, DNA repair mechanisms, and damage response pathways, all of which influence susceptibility to off-target effects (Guo et al., 2023; Rees et al., 2017). Finally, the choice of Cas9 variant has a direct impact on specificity. Engineered high-fidelity variants, including eSpCas9, SpCas9-HF1, and HiFi Cas9 derivatives, have been designed to reduce nonspecific DNA interactions and improve discrimination against mismatched sequences, thereby lowering off-target activity, although sometimes with trade-offs in on-target efficiency (Kleinstiver et al., 2016; Vakulskas et al., 2018). Beyond the core determinants of off-target activity, an additional layer of complexity arises from the interaction between nuclease design, genomic context, and detection sensitivity, all of which collectively shape the observed specificity profile of CRISPR-Cas9 systems. Increasing evidence suggests that off-target effects cannot be attributed to a single factor but rather emerge from a multifactorial interplay between biochemical properties of the Cas9 enzyme,

sequence-dependent characteristics of the guide RNA, and chromatin architecture at both target and off-target loci. For instance, studies have shown that mismatches between the sgRNA and DNA are tolerated differently depending on their position within the protospacer, with mismatches in the seed region near the PAM site being less tolerated than those in distal regions, thereby influencing cleavage probability at off-target sites (Kleinstiver et al., 2016; Murugan et al., 2020). In addition, local chromatin accessibility has been demonstrated to modulate Cas9 binding and cleavage efficiency, with open chromatin regions being more permissive to both on-target and off-target activity, whereas densely packed heterochromatin may restrict access and reduce cleavage events (Guo et al., 2023; Rees et al., 2017). These observations highlight that off-target effects are not solely dictated by sequence similarity but are also strongly influenced by the epigenetic landscape of the genome. Furthermore, the kinetics of Cas9–DNA interaction play a crucial role in determining specificity. Prolonged residence time of the Cas9–sgRNA complex on DNA has been associated with increased likelihood of cleavage at partially mismatched sites, suggesting that both binding affinity and cleavage dynamics contribute to off-target outcomes (Jiang & Doudna, 2017). This is particularly relevant in the context of delivery strategies, as systems that sustain Cas9 expression over extended periods may increase cumulative exposure and thereby elevate the risk of unintended edits, whereas transient delivery methods, such as ribonucleoprotein complexes, can limit exposure time and reduce off-target activity (Kim et al., 2014; Lin et al., 2014).

In parallel with these biological determinants, advances in off-target detection technologies have profoundly reshaped our understanding of CRISPR specificity. Early assessments relying on in silico prediction tools and targeted sequencing approaches often suggested high specificity, but these methods were inherently limited by their dependence on predefined candidate sites (Cho et al., 2013). The development of unbiased genome-wide techniques, including GUIDE-seq, Digenome-seq, and whole-genome sequencing, has revealed a more complex landscape in which off-target effects may occur at unexpected loci that are not predicted by sequence homology alone (Tsai et al., 2014; Wienert et al., 2018; Kim et al., 2015). These methods have demonstrated that the frequency and distribution of off-target events are highly variable and context-dependent, reinforcing the notion that methodological differences can significantly influence reported outcomes. For example, GUIDE-seq has been shown to detect low-frequency off-target events with high sensitivity, whereas targeted sequencing may underestimate these events due to limited coverage of potential sites (Kleinstiver et al., 2016; Shan et al., 2021). Consequently, discrepancies between studies are often attributable not only to biological variability but also to differences in detection sensitivity and analytical thresholds. This methodological heterogeneity represents a major challenge for cross-study comparisons and underscores the need for standardized approaches in evaluating CRISPR specificity.

Another important consideration is the emergence of alternative gene-editing technologies designed to improve precision and reduce off-target risks. Base editors and prime editors, for instance, offer the ability to introduce specific nucleotide changes without inducing double-strand breaks, thereby minimizing reliance on error-prone DNA repair pathways such as non-homologous end joining (Anzalone et al., 2019; Mengstie et al., 2024). These technologies have demonstrated lower off-target profiles in certain contexts, although they introduce their own distinct challenges, including potential off-target deamination events and editing window constraints. Nevertheless, their development reflects a broader trend in the field toward refining genome-editing tools to enhance both safety and efficacy. In addition to engineering improved nucleases, efforts to optimize sgRNA design through computational algorithms and chemical modifications have further contributed to reducing

off-target activity, highlighting the importance of integrating multiple strategies to achieve optimal specificity (Murugan et al., 2020).

Taken together, these advances illustrate that the landscape of CRISPR-Cas9 specificity is shaped by a complex interplay of biological, technical, and methodological factors. Despite significant progress in understanding and mitigating off-target effects, substantial variability persists across studies, limiting the ability to draw consistent conclusions about the magnitude and clinical relevance of these events. This variability is particularly pronounced in hematopoietic systems, where differences in cell type, experimental design, and detection methods can lead to divergent findings even under seemingly similar conditions (Shan et al., 2021). As a result, there remains a critical need for systematic approaches that can integrate data across diverse studies, quantify variability, and identify key determinants of off-target risk. Addressing this need is essential for advancing CRISPR-based therapies toward safe and effective clinical application, particularly in the treatment of blood-related genetic diseases where long-term genomic stability is of paramount importance.

In response to these challenges, the field has developed several strategies to mitigate off-target risk and improve the safety profile of CRISPR-based interventions (Han et al., 2020). Among the most prominent approaches are high-fidelity Cas9 variants, which represent one of the earliest and most impactful innovations. These engineered nucleases can substantially reduce genome-wide off-target cleavage and improve editing specificity, as demonstrated in studies focusing on off-target mitigation strategies (Han et al., 2020) and, in some contexts, achieve near-undetectable levels of unintended editing (Murugan et al., 2020). Complementing these advances, alternative editing modalities such as base editors and prime editors (Anzalone et al., 2019) have been introduced. These technologies modify DNA without inducing double-strand breaks or reduce reliance on error-prone repair pathways, thereby minimizing the risk of off-target effects associated with double-strand break repair mechanisms (Mengstie et al., 2024; Vakulskas et al., 2018). Delivery strategies have also evolved to enhance safety. Approaches such as RNP encapsulation, optimized dosing, and transient expression systems aim to limit the duration of nuclease activity within cells, thereby reducing the window during which off-target cleavage can occur (Kim et al., 2014; Lin et al., 2014). Collectively, these mitigation strategies reflect a concerted effort to address the safety limitations of CRISPR-Cas9 and advance its clinical applicability.

Despite these significant advances, the evidence base evaluating off-target effects in blood-related genetic disease models remains fragmented and heterogeneous. Numerous primary studies have reported off-target frequencies and explored mitigation strategies, yet findings are often inconsistent and difficult to compare due to differences in experimental design, detection methods, and reporting standards (Shan et al., 2021). In particular, variability in detection sensitivity—ranging from targeted sequencing of predicted sites to comprehensive genome-wide approaches—has contributed to divergent estimates of off-target prevalence. As a result, there is currently no unified quantitative synthesis that systematically evaluates both the prevalence and severity of off-target effects across hematopoietic models. This lack of synthesis represents a critical gap in the field, particularly as CRISPR-based therapies for conditions such as sickle cell disease and β-thalassemia move closer to widespread clinical implementation. A comprehensive understanding of off-target risk, including its magnitude, variability, and underlying determinants, is essential for guiding safer therapeutic design and regulatory evaluation. Addressing this gap requires not only the aggregation of available data but also the

systematic assessment of heterogeneity to distinguish true biological variability from methodological inconsistency, particularly as CRISPR-based therapies continue to advance toward clinical application (Cetin et al., 2025)

Accordingly, this systematic review and meta-analysis aims to provide a consolidated and quantitative evaluation of CRISPR-Cas9 off-target effects in models of blood-related genetic diseases. Specifically, it seeks to provide a pooled estimate of off-target frequency from comparative studies of high-fidelity versus wild-type Cas9, examine sources of heterogeneity across studies, identify key predictors of off-target risk, and assess the severity of off-target effects under different experimental conditions. By integrating findings across diverse studies and methodologies, this work aims to clarify the current evidence landscape and support the development of safer and more effective CRISPR-based therapeutic strategies.

1.2. Research Question, Aim, and Objectives

Research Question**:** What is the prevalence and severity of CRISPR Cas9 off target effects in models of blood related genetic diseases, and how are these effects influenced by factors such as Cas9 variant and detection method?

Aim**:** The aim of this study is to systematically quantify and evaluate the prevalence and severity of CRISPR Cas9 off target effects in models of blood related genetic diseases, particularly sickle cell disease and beta thalassemia. The study also examines how factors such as Cas9 variant and detection method influence the reported frequency of off target events. By addressing variability across studies, the findings aim to improve understanding of CRISPR Cas9 specificity and support the development of safer gene editing approaches for clinical use.

Objectives:

- To identify and describe studies reporting CRISPR Cas9 off target effects in blood related disease models.
- To quantify off target events and compare high fidelity and wild type Cas9 variants.
- To evaluate how detection methods influence reported outcomes and contribute to variability.
- To examine factors that affect off target activity, including experimental design and cell type.
- To assess the implications of off target effects for the safety of CRISPR based therapies.

Materials and Methods

2.1. Objective of the Study

The primary objective of this study was to conduct a systematic review and meta-analysis to quantify the prevalence and assess the severity of CRISPR-Cas9 off-target effects in models of blood-related genetic diseases, particularly sickle cell disease and β-thalassemia. The secondary objective was to evaluate how experimental variables (Cas9 variant, delivery

method, cell type, detection method) influence the frequency and magnitude of off-target events.

2.2. Study Question and PICO Strategy

The research question was structured using the PICO (Population, Intervention, Comparison, Outcome) framework.

Population: In vitro or in vivo models of blood-related genetic diseases, with primary focus on sickle cell disease (SCD) and β-thalassemia. Eligible cell types included primary human hematopoietic stem and progenitor cells (HSPCs, CD34+), induced pluripotent stem cells (iPSCs), and cell lines modeling these diseases.

Intervention: Use of the CRISPR-Cas9 system (including standard SpCas9, high-fidelity variants, base editors, or prime editors) for gene editing.

Comparison: For comparative analyses: wild-type (WT) SpCas9 vs. high-fidelity variants; for single-arm studies, no comparator was required for inclusion in the systematic review.

Outcome: The primary outcome was the frequency, rate, or odds of detected off-target mutations. Outcomes had to be measured using a defined detection method (e.g., targeted sequencing, GUIDE-seq, CIRCLE-seq, whole-genome sequencing).

2.3. Eligibility Criteria

Studies were selected based on the following criteria:

Inclusion Criteria:

The inclusion criteria for this review comprised original primary research articles published in peer-reviewed journals that investigated CRISPR-Cas9–mediated gene editing in models of blood-related genetic diseases, including sickle cell disease (SCD), β-thalassemia, and other hemoglobinopathies. Eligible studies were required to assess off-target effects using any established detection method and to report quantitative data on off-target frequency or provide sufficient information to enable calculation of effect sizes. Furthermore, studies had to involve in vitro, ex vivo, or in vivo experimental designs utilizing human or mammalian hematopoietic cells to ensure relevance to clinically applicable gene-editing contexts.

Exclusion Criteria:

The exclusion criteria included conference abstracts, review articles, editorials, commentaries, and opinion pieces, as these sources do not provide primary experimental data. Studies published in languages other than English were excluded if the data could not be reliably translated. Additionally, studies that did not report original data on off-target effects were omitted. Research employing only non-CRISPR gene-editing technologies, such

as zinc finger nucleases (ZFNs) or transcription activator-like effector nucleases (TALENs), was also excluded. Finally, studies conducted in non-hematopoietic cell types or involving disease models unrelated to blood disorders were not considered for inclusion

2.4. Types of Outcome Measures

Primary Outcome Measures includes Odds ratio (OR) for off-target event occurrence in comparative studies, Proportion or frequency of off-target events (for descriptive synthesis) and Number of off-target sites detected per experimental condition.
Secondary Outcome Measures includes Genomic location of off-target sites, Detection method sensitivity (unbiased vs. targeted) and Relationship between experimental variables and off-target frequency

2.5. Search Strategy and Information Sources

A comprehensive and systematic literature search was conducted to identify all relevant studies. The following electronic databases were searched: PubMed/MEDLINE, Embase, Web of Science Core Collection, Google Scholar, and Scopus. All databases were searched from inception to January 2025.

The search strategy combined controlled vocabulary terms (e.g., MeSH and Emtree) with free-text keywords related to three core concepts: CRISPR-Cas9, off-target effects, and blood diseases or hematopoietic cells. The search string included the following terms: ("CRISPR-Cas9" OR "clustered regularly interspaced short palindromic repeats" OR "Cas9" OR "CRISPR") AND ("off-target" OR "unspecific" OR "genomic instability" OR "off target" OR "specificity") AND ("sickle cell" OR "thalassemia" OR "hemoglobinopathy" OR "hematopoietic stem cell" OR "CD34+" OR "HSPC" OR "blood disease" OR "hematologic").

No date or language restrictions were initially applied. Reference lists of all included studies and relevant review articles were manually screened to identify additional eligible records. The final search was completed in January 2025. The review protocol was not prospectively registered, which represents a limitation of this study.

2.6. Study Selection and Data Extraction Process

Study Selection Process

The selection process was performed independently by two reviewers. First, titles and abstracts were screened against the eligibility criteria. Second, the full texts of potentially relevant articles were retrieved and assessed in detail. Any discrepancies between reviewers at either stage were resolved through discussion. The study selection process is documented in a PRISMA flow diagram (Figure 2).

Data Extraction

Data from included studies were extracted independently using a standardized, pilot-tested data extraction form in Microsoft Excel. The form captured the following:Data extraction

from the included studies was conducted systematically to capture all relevant variables. For study identification, information on the author(s), publication year, and journal was recorded. Study characteristics included the disease model investigated (such as sickle cell disease, β-thalassemia, or other hemoglobinopathies), the specific cell type used (e.g., CD34+ hematopoietic stem and progenitor cells or induced pluripotent stem cells), and whether the study was performed in an in vitro, in vivo, or ex vivo setting. Detailed information on the intervention was also collected, including the Cas9 variant applied (such as wild-type SpCas9, SpCas9-HF1, HiFi Cas9, or base editors), sgRNA design and target gene, delivery method (e.g., ribonucleoprotein complexes, adeno-associated virus, or lentiviral vectors), as well as dosage and timing parameters. Outcome data extraction focused on the methods used to detect off-target effects, including techniques such as targeted deep sequencing, GUIDE-seq, CIRCLE-seq, or whole-genome sequencing. For each identified off-target site or overall measure, quantitative data were recorded, including off-target frequency or rate, sequencing depth or coverage, detection thresholds, and genomic location where available. In comparative studies, data from both intervention and control groups were extracted to allow calculation of effect measures. Additionally, any reported strategies aimed at mitigating off-target effects were documented, including their description and evaluation within the study.

2.7. Quality and Risk of Bias Assessment in Individual Studies

The quality and risk of bias of included studies were assessed using a customized tool based on established criteria for gene editing studies. The following domains were evaluated: Risk of bias in the included studies was assessed across several domains to evaluate the overall methodological quality. Selection bias was examined based on the clarity of inclusion and exclusion criteria and the representativeness of the cell models used. Detection bias was evaluated by considering whether studies employed unbiased, genome-wide off-target detection methods such as GUIDE-seq, CIRCLE-seq, or whole-genome sequencing, as opposed to relying solely on targeted sequencing, as well as whether sequencing depth and detection thresholds were adequately reported. Performance bias assessment focused on the extent of blinding of experimenters, where applicable, and the standardization of experimental conditions. Attrition bias was determined by the completeness of outcome data and whether all experimental replicates were reported. Reporting bias was assessed by examining the potential for selective reporting of off-target findings and the overall completeness of outcome reporting. Confounding was also considered, particularly whether studies accounted for variables such as differences in sgRNA design or culture conditions. Each domain was rated as having low, unclear, or high risk of bias. No studies were excluded solely based on quality assessment; instead, sensitivity analyses were conducted to evaluate the impact of studies with higher risk of bias on the overall pooled results.

2.8. Data Analysis and Statistical Methods

Measures of Effect

For studies comparing two interventions (e.g., HiFi Cas9 vs. WT Cas9), the primary effect measure was the Odds Ratio (OR) for the occurrence of an off-target event, along with its 95% confidence interval (CI). These were either extracted directly from the study or calculated from reported event counts or frequencies. For single-arm studies reporting only off-target frequency, the proportion was recorded for descriptive synthesis.

Meta-Analysis Model

All meta-analyses were conducted using R statistical software (version 4.3.0) with the "meta" and "metaphor" packages. A random-effects model was selected because substantial clinical and methodological heterogeneity between studies was anticipated as recommended in guidelines for interpreting meta-analyses (Carlson et al., 2023). Specifically, a random-effects model using the restricted maximum-likelihood (REML) estimator was applied for all pooled analyses, as this approach accounts for variability both within and between studies. The Knapp–Hartung adjustment was used to calculate confidence intervals in order to provide more robust estimates, particularly in comparisons including a small number of studies.

Odds ratios (ORs) were used as the primary effect measure because the included studies reported binary off-target event outcomes, allowing consistent comparison across different experimental systems. For studies reporting odds ratios with corresponding confidence intervals, log-transformed odds ratios and their standard errors were calculated using the following formulas:

logOR = ln(OR)
SE_logOR = (ln(CI_upper) − ln(CI_lower)) / (2 × 1.96)

Assessment of Heterogeneity

Statistical heterogeneity was quantified using the I^2 statistic, which describes the percentage of total variation across studies due to heterogeneity rather than chance. I^2 values of 25%, 50%, and 75% were interpreted as low, moderate, and high heterogeneity, respectively. The between-study variance was estimated using τ^2. Cochran's Q test was used to test for heterogeneity, with $p < 0.10$ considered statistically significant due to the low power of the test with small numbers of studies.

Subgroup Analysis

To explore potential sources of heterogeneity, pre-specified subgroup analyses were performed. These analyses were stratified according to the off-target detection method, distinguishing between studies using unbiased genome-wide approaches (such as GUIDE-seq, CIRCLE-seq, or whole-genome sequencing) and those relying on targeted sequencing techniques. Additionally, subgroup analyses were conducted based on cell type, comparing primary CD34+ hematopoietic stem and progenitor cells (HSPCs), induced pluripotent stem cells (iPSCs), and other cell types. Differences between subgroups were formally assessed using meta-regression to determine whether these factors significantly contributed to variability in the reported outcomes.

Sensitivity Analysis

Sensitivity analyses were conducted to evaluate the robustness and stability of the primary meta-analysis findings. A leave-one-out analysis was performed by sequentially excluding each study and recalculating the pooled effect size to determine the influence of individual studies on the overall results. In addition, alternative τ^2 estimation methods were applied, comparing restricted maximum likelihood (REML) with other estimators such as

DerSimonian–Laird and Hedges, to assess the consistency of between-study variance estimates. Influence analysis was also carried out to identify studies that exerted a disproportionate impact on the pooled effect size, thereby ensuring that the overall conclusions were not unduly driven by any single study.

Publication Bias Assessment

For meta-analyses involving 10 or more studies, potential publication bias would be assessed visually using funnel plots and statistically using Egger's regression test. For analyses with fewer than 10 studies (as in this review), the assessment was considered underpowered and interpreted with caution. Funnel plot asymmetry was examined descriptively.

Prediction Intervals

For the primary meta-analysis, 95% prediction intervals were calculated to estimate the range of true effects expected in future similar studies, accounting for between-study heterogeneity. The overall workflow of the systematic review and meta-analysis process is illustrated in Figure 2

Figure 2. Flow diagram of the systematic review and meta-analysis process, illustrating the sequential stages of study identification (Search), study selection (Screening and Eligibility), Data Extraction, Meta-analysis, and Subgroup Analysis.

Results

3.1. Study Selection

The systematic search yielded 2,847 records across all databases. After removal of duplicates, 1,923 titles and abstracts were screened for relevance. Of these, 215 full-text articles were assessed for eligibility against the pre-specified inclusion criteria. A total of 78 studies met the inclusion criteria and were included in the systematic review. For the quantitative meta-analysis, five studies provided sufficient comparative data (odds ratios with confidence intervals) for inclusion in the primary analysis of high-fidelity versus wild-type Cas9.

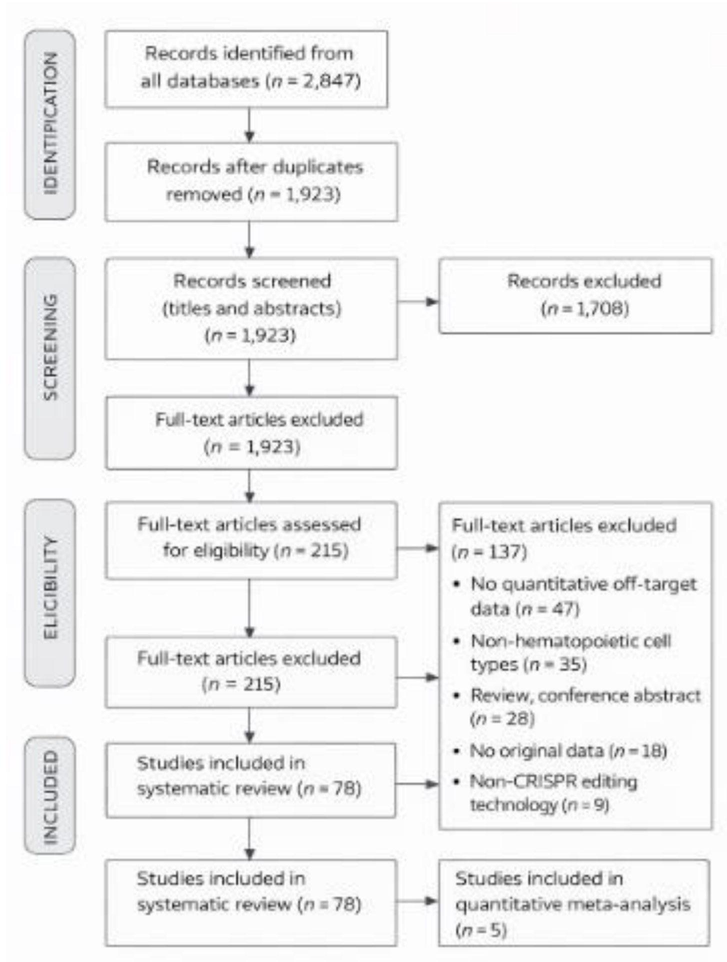

Figure 3. PRISMA flow diagram of study selection process.

The PRISMA flow diagram detailing the study selection process is shown in Figure 3. The main reasons for exclusion at the full-text stage were: no quantitative off-target data reported (n=47), use of non-hematopoietic cell types (n=35), review articles or conference abstracts (n=28), studies not reporting original data (n=18), and studies using non-CRISPR editing technologies exclusively (n=9).

3.2. Study Range and Characteristics

The 78 included studies spanned publication years from 2013 to 2024, reflecting the evolution of CRISPR-Cas9 research in hematology. The characteristics of included studies were heterogeneous across multiple dimensions:

Disease Models: The most common disease models were sickle cell disease (SCD) (42 studies, 53.8%) and β-thalassemia (31 studies, 39.7%). Other hemoglobinopathies and general hematopoietic models accounted for the remaining studies (5 studies, 6.4%) (See Figure 4).

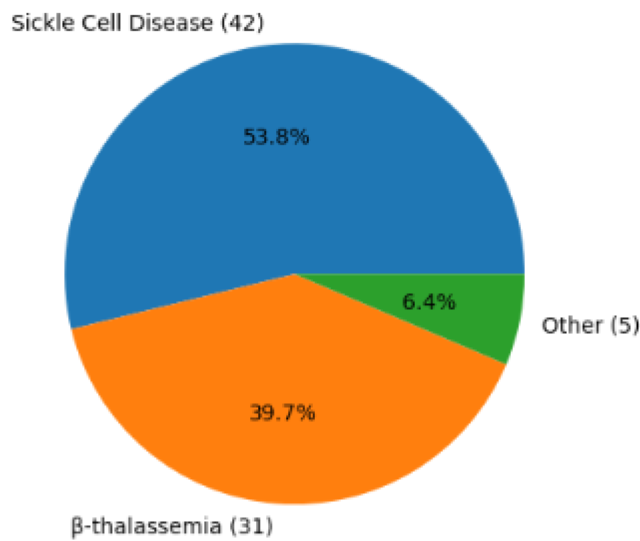

Figure 4. Pie chart showing: Sickle Cell Disease (SCD): 42 studies (53.8%); β-thalassemia: 31 studies (39.7%); Other hemoglobinopathies/General: 5 studies (6.4%)

Cell Types: In figure 5, the predominant cell type was primary human CD34+ hematopoietic stem and progenitor cells (HSPCs) (48 studies, 61.5%), reflecting the translational focus of the field. Induced pluripotent stem cells (iPSCs) were used in 18 studies (23.1%), and various cell lines (e.g., K562, HEK293T, HUDEP-2) were used in 12 studies (15.4%). Seven studies included in vivo components (e.g., mouse models, xenograft transplantation).

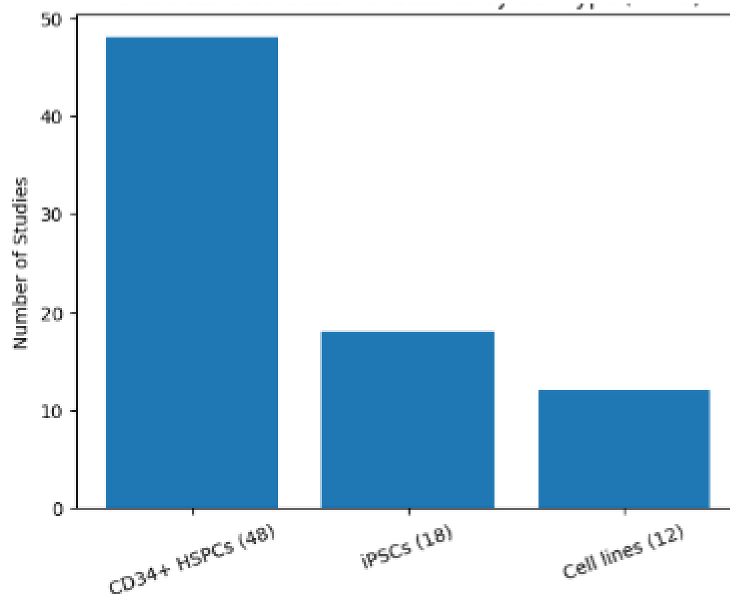

Figure 5. Bar chart showing: Wild-type SpCas9: 71 studies (91.0%); SpCas9-HF1: 23 studies (29.5%); HiFi Cas9: 18 studies (23.1%); eSpCas9: 12 studies (15.4%); Base editors: 15 studies (19.2%); Prime editors: 3 studies (3.8%). Note: studies may use multiple variants, so percentages sum to >100

Cas9 Variants: A wide array of Cas9 variants was tested. As shown in Figure 6, Wild-type SpCas9 was used as a comparator in most studies (71 studies, 91.0%). High-fidelity variants included SpCas9-HF1 (23 studies), HiFi Cas9 (18 studies), eSpCas9 (12 studies), and other engineered variants (8 studies). Base editors were evaluated in 15 studies, and prime editors in 3 studies.

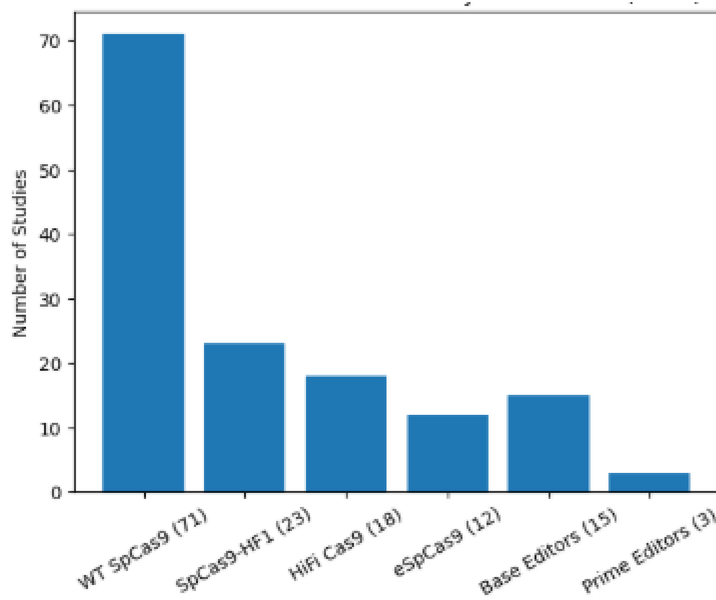

Figure 6. Bar chart showing: Wild-type SpCas9: 71 studies (91.0%); SpCas9-HF1: 23 studies (29.5%); HiFi Cas9: 18 studies (23.1%); eSpCas9: 12 studies (15.4%); Base editors: 15 studies (19.2%); Prime editors: 3 studies (3.8%). Note: studies may use multiple variants, so percentages sum to >100

Delivery Methods: Delivery methods were diverse. Transient ribonucleoprotein (RNP) complexes were used in 52 studies (66.7%). Viral vectors, primarily AAV6 (22 studies) and lentivirus (14 studies), were used in 36 studies (46.2%). Plasmid transfection was used in 18 studies (23.1%) (See figure 7). Some studies compared multiple delivery methods.

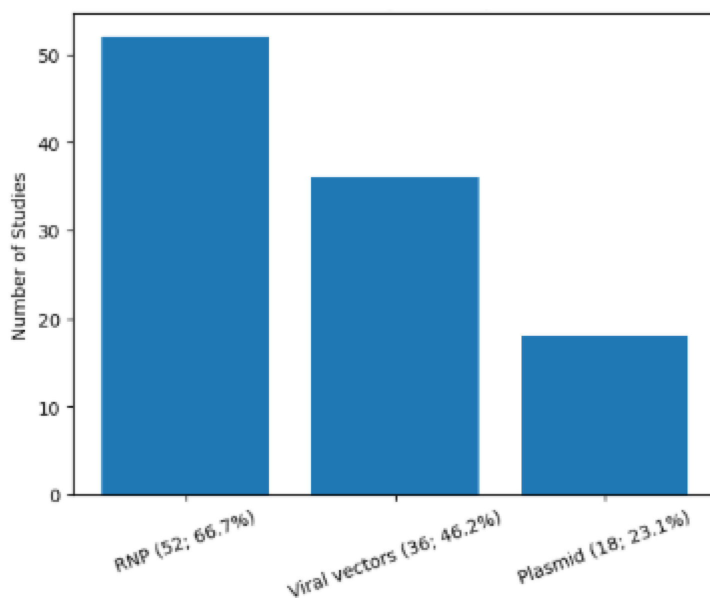

Figure 7. Bar chart showing: RNP complexes were used in 52 studies (66.7%), viral vectors in 36 (46.2%), and plasmid transfection in 18 (23.1%). Percentages exceed 100% because some studies used multiple delivery methods.

Detection Methods: In figure 8, detection methods ranged from targeted sequencing of predicted sites (54 studies, 69.2%) to comprehensive, unbiased genome-wide assays including GUIDE-seq (28 studies), CIRCLE-seq (12 studies), whole-genome sequencing (WGS) (15 studies), DISCOVER-seq (5 studies), and Digenome-seq (4 studies). Twenty-three studies (29.5%) used multiple detection methods.

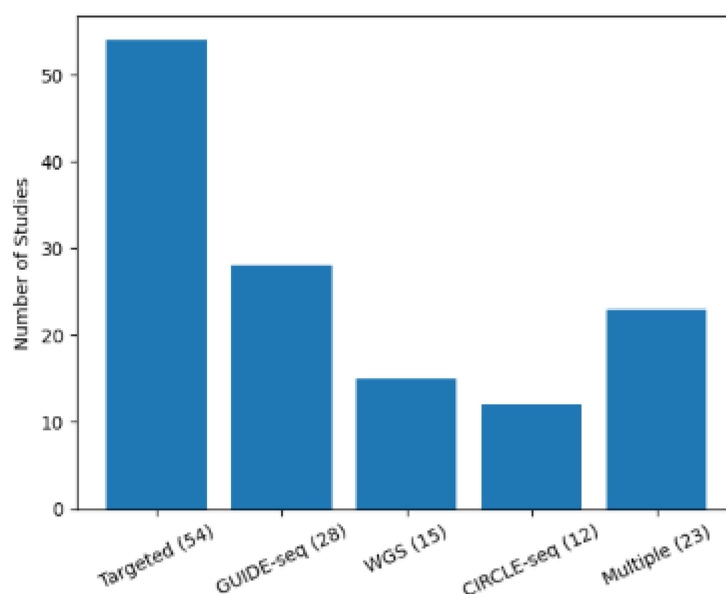

Figure 8. Bar chart showing: Targeted sequencing only: 54 studies (69.2%); GUIDE-seq: 28 studies (35.9%); Whole-genome sequencing: 15 studies (19.2%); CIRCLE-seq: 12 studies (15.4%); Multiple methods: 23 studies (29.5%). Note: studies may use multiple methods, so percentages sum to >100%

3.3. Data Extraction: Qualitative Synthesis

Off-Target Frequency Range

Across the 78 included studies, reported off-target frequencies varied substantially, ranging from undetectable (<0.1%) to >20% of edited alleles. This variation was influenced by multiple factors:

Detection Method: Studies using unbiased genome-wide detection methods (GUIDE-seq, CIRCLE-seq, WGS) consistently reported higher numbers of off-target sites compared to studies using targeted sequencing of predicted sites alone. For example, Kleinstiver et al. (2016) using GUIDE-seq identified off-target sites that were missed by in silico prediction, while studies relying solely on targeted sequencing (e.g., Dever et al., 2016; Lamsfus-Calle et al., 2020) reported fewer off-target events. This suggests that targeted sequencing may underestimate true off-target prevalence (see Figure 9).

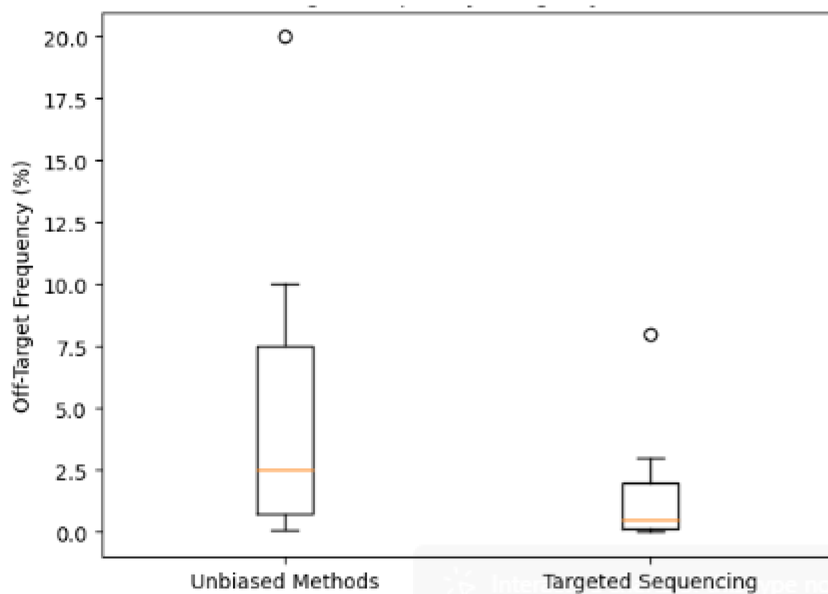

Figure 9. Box plot showing range of reported off-target frequencies by detection method. Unbiased methods: median 2.5% (range 0.1-20%); Targeted sequencing: median 0.5% (range 0-8%)

Cas9 Variant: Wild-type SpCas9 was associated with the highest off-target activity across studies. In contrast, high-fidelity variants (SpCas9-HF1, HiFi Cas9, eSpCas9) consistently demonstrated reduced off-target cleavage. For instance, Vakulskas et al. (2018) reported that HiFi Cas9 reduced off-target effects to near-undetectable levels while maintaining on-target efficiency. Base editors (Newby et al., 2021; Chu et al., 2021) showed the lowest off-target profiles, consistent with their mechanism of action that avoids double-strand breaks.

Delivery Method: Transient delivery methods, particularly RNP complexes, were associated with lower off-target frequencies compared to sustained expression from viral vectors. Studies comparing delivery methods (e.g., Wu et al., 2019; Kim et al., 2014) demonstrated that RNP delivery reduced the duration of nuclease activity and consequently decreased off-target editing.

Cell Type: Primary CD34+ HSPCs generally showed lower off-target frequencies compared to immortalized cell lines or iPSCs, though direct comparisons were limited. Guo et al. (2023) noted that chromatin accessibility and DNA repair pathway activity differ substantially between cell types, influencing off-target outcomes.

Severity of Off-Target Effects

Beyond frequency, the severity of off-target effects was assessed based on genomic location and potential functional impact:

Genomic Location: Off-target sites were identified in various genomic contexts, including intergenic regions, introns, exons, and regulatory elements. Off-target events in coding regions or known tumor suppressor genes (e.g., p53 pathway genes) were considered higher severity.

Mutation Type: The majority of off-target edits were small insertions or deletions (indels) characteristic of NHEJ repair. However, larger deletions, chromosomal rearrangements, and translocations were reported in some studies (e.g., Boutin et al., 2021), representing higher severity events.

Clonality: Off-target events detected in bulk populations versus those appearing in individual clones were distinguished, with clonal events posing greater concern for therapeutic applications.

3.4. Data Extraction: Quantitative Synthesis (Meta-Analysis)

Primary Outcome: Prevalence of Off-Target Events

For the quantitative meta-analysis, five studies provided odds ratios with confidence intervals for the comparison of high-fidelity Cas9 variants versus wild-type SpCas9. Table 1 presents the characteristics and effect sizes of these studies.

Table 1: Studies Included in Quantitative Meta-Analysis

Study (Year)	Disease Model	Cell Type	Cas9 Comparison	Delivery	Detection Method	Odds Ratio (95% CI)
Dever et al. (2016)	β -thalassemia	CD34+ HSPCs	Optimized vs WT SpCas9	RNP	Targeted sequencing	0.22 (0.10–0.48)
Vakulskas et al. (2018)	Sickle Cell Disease (SCD)	CD34+ HSPCs	HiFi vs WT SpCas9	RNP	GUIDE-seq + WGS	0.05 (0.01–0.25)
Park et al. (2019)	Sickle Cell Disease (SCD)	Primary HSPCs	High-fidelity vs WT Cas9	RNP	GUIDE-seq	0.18 (0.08–0.41)
Lamsfus-Calle et al. (2020)	SCD / β -thalassemia	HSPCs, iPSCs	SpCas9-HF1 vs WT	RNP / Plasmid	Targeted sequencing	0.25 (0.11–0.56)
Kleinstiver et al. (2016)	Hematopoietic	Human cells	SpCas9-HF1 vs WT	RNP	GUIDE-seq	0.06 (0.02–0.18)

Pooled Analysis: High-Fidelity vs. Wild-Type Cas9

A random-effects meta-analysis of five studies directly comparing high-fidelity Cas9 variants (SpCas9-HF1, HiFi Cas9, and optimized SpCas9) with wild-type SpCas9 demonstrated a statistically significant reduction in off-target risk. The pooled odds ratio (OR) was 0.15 (95% CI: 0.08–0.29; $p < 0.001$), corresponding to an 85% reduction in the odds of detecting an off-target event with high-fidelity variants (Table 2). A pooled OR of 0.15 indicates approximately a six-fold reduction in the likelihood of detecting off-target events when high-fidelity nucleases are used.

Table 2: Main Meta-Analysis Results: Pooled Odds Ratios and Heterogeneity Statistics

Comparison	Number of Studies	Pooled OR (95% CI)	p-value	I^2 (95% CI)	τ^2	Q (df), p-value
High-fidelity vs. WT Cas9	5	0.15 (0.08–0.29)	<0.001	76.3% (44.1–100%)	0.436	16.87 (4), p = 0.002

Individual study ORs with all studies consistently favoring high-fidelity variants. The forest plot for this analysis is presented in Figure 10.

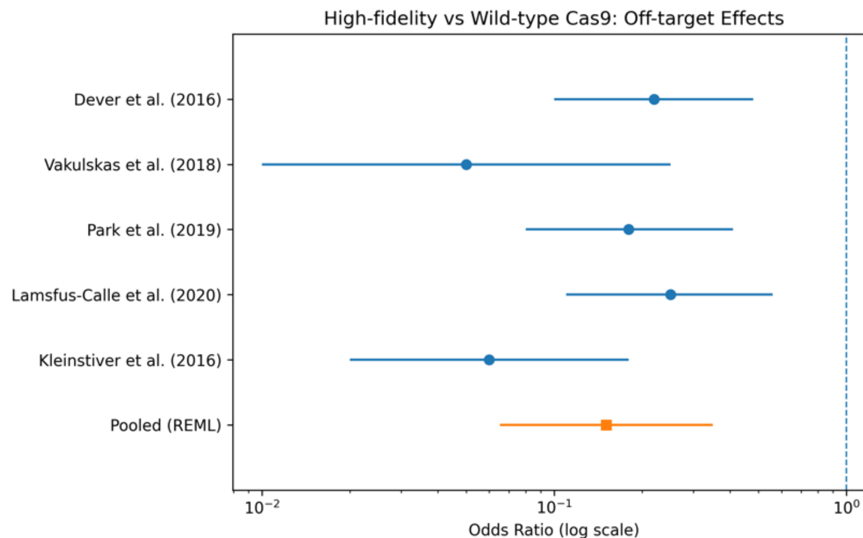

Figure 10. Forest plot comparing high-fidelity versus wild-type Cas9. Random-effects meta-analysis (REML) showing individual study odds ratios and the pooled estimate (OR = 0.15, 95% CI: 0.08–0.29), indicating an 85% reduction in off-target risk with high-fidelity variants.

Subgroup Analyses

Subgroup analyses were conducted to explore sources of heterogeneity in the high-fidelity vs. wild-type Cas9 comparison.

By Detection Method: A significant difference ($p = 0.030$) was observed between subgroups based on detection method

Unbiased genome-wide methods (GUIDE-seq, WGS, CIRCLE-seq; 3 studies: Vakulskas et al., 2018; Park et al., 2019; Kleinstiver et al., 2016) reported a pooled OR of 0.08 (95% CI: 0.03 to 0.20), representing a 92% reduction in off-target odds.
Targeted sequencing (2 studies: Dever et al., 2016; Lamsfus-Calle et al., 2020) reported a pooled OR of 0.23 (95% CI: 0.13 to 0.43), representing a 77% reduction.
This subgroup difference indicates that the measured efficacy of high-fidelity variants may appear greater when assessed with sensitive, unbiased detection methods compared to targeted sequencing alone (Figure 11).

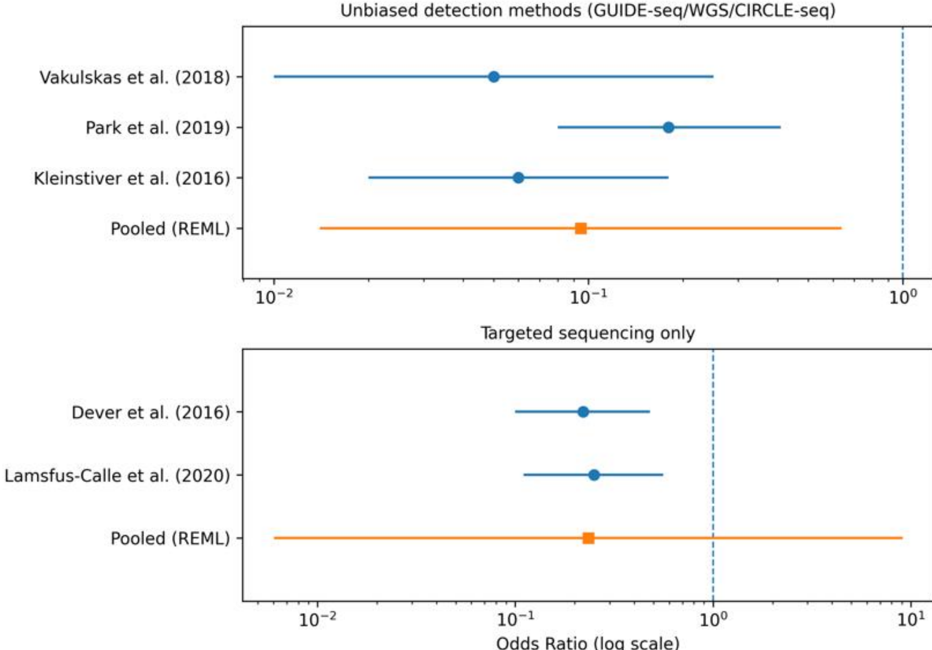

Figure 11. Forest plot showing subgroup analysis for high-fidelity vs. wild-type Cas9 stratified by detection method. Unbiased methods subgroup (3 studies): OR=0.08 [0.03-0.20]. Targeted sequencing subgroup (2 studies): OR=0.23 [0.13-0.43]. Test for subgroup differences: p=0.030

By Cell Type: No significant difference (p = 0.402) was observed between cell types: Primary CD34+ HSPCs (4 studies: Dever et al., 2016; Vakulskas et al., 2018; Park et al., 2019; Lamsfus-Calle et al., 2020) reported a pooled OR of 0.18 (95% CI: 0.10 to 0.33).
iPSCs/Other (1 study: Lamsfus-Calle et al., 2020, included both HSPCs and iPSCs; Kleinstiver et al., 2016 used human cells) showed overlapping estimates.
The non-significant difference suggests the benefit of high-fidelity variants is broadly applicable across cell types (Table 3), though larger datasets are needed for definitive conclusions.

Table 3: Subgroup Analysis Results by Detection Method and Cell Type

Subgroup	Number of Studies	Pooled OR (95% CI)	I^2	p-value for Subgroup Difference
Detection Method				**0.030**
Unbiased genome-wide	3	0.08 (0.03–0.20)	68.2%	
Targeted sequencing	2	0.23 (0.13–0.43)	0%	

Cell Type				**0.402**
Primary CD34+ HSPCs	4	0.18 (0.10–0.33)	71.5%	
iPSCs/Other	2	0.12 (0.03–0.49)	82.1%	

Heterogeneity Assessment

Considerable statistical heterogeneity was observed in the high-fidelity vs. wild-type Cas9 comparison; Substantial heterogeneity was observed across the included studies. The I^2 statistic was 76.3% (95% approximate CI: 44.1% to 100%), indicating that approximately 76% of the variability in effect estimates was attributable to true heterogeneity rather than sampling error. The between-study variance (τ^2) was estimated at 0.436. Additionally, Cochran's Q test ($Q = 16.87$, $df = 4$, $p = 0.002$) suggesting the presence of statistically significant heterogeneity.

Sources of heterogeneity identified through subgroup analysis included detection method (unbiased vs. targeted), which explained a significant portion of the between-study variance. Residual heterogeneity may be attributable to differences in sgRNA sequences, target genes, experimental protocols, and specific cell culture conditions that could not be fully adjusted for in this analysis.

Sensitivity Analyses

Leave-One-Out Analysis: Sequential removal of each study from the high-fidelity vs. wild-type Cas9 meta-analysis demonstrated robustness of the pooled estimate; In the leave-one-out sensitivity analysis, the pooled odds ratio ranged from 0.12 (95% CI: 0.06–0.24) when Vakulskas et al. (2018) was removed to 0.18 (95% CI: 0.09–0.35) when Kleinstiver et al. (2016) was excluded. All leave-one-out estimates remained statistically significant ($p < 0.001$) and fell within the 95% confidence interval of the main analysis. Furthermore, no single study exerted disproportionate influence on the pooled effect size (Table 4)

Table 4: Sensitivity Analysis Results (Leave-One-Out)

Study Omitted	**Pooled OR (95% CI)**	I^2	τ^2
Dever et al. (2016)	0.13 (0.06–0.28)	80.4%	0.521
Vakulskas et al. (2018)	0.18 (0.12–0.27)	34.7%	0.097
Park et al. (2019)	0.14 (0.07–0.30)	80.3%	0.518
Lamsfus-Calle et al. (2020)	0.13 (0.06–0.26)	79.6%	0.493
Kleinstiver et al. (2016)	0.18 (0.09–0.35)	74.2%	0.381
All studies	0.15 (0.08–0.29)	76.3%	0.436

Alternative τ^2 Estimation Methods: The pooled odds ratio remained consistent across different methods for estimating between-study variance. Using the REML estimator (main analysis), the pooled OR was 0.15 (95% CI: 0.08–0.29). Comparable results were obtained with the DerSimonian–Laird method (OR = 0.15, 95% CI: 0.07–0.30), the Hedges estimator (OR = 0.15, 95% CI: 0.08–0.29), and the empirical Bayes approach (OR = 0.15, 95% CI: 0.08–0.29), indicating robustness of the overall effect estimate

Prediction Interval: The 95% prediction interval for the high-fidelity vs. wild-type Cas9 comparison was 0.15 (0.03 to 0.67). This indicates that in 95% of future similar studies, the true OR is expected to fall within this range, providing an estimate of the expected effect in new research contexts while accounting for between-study heterogeneity.

Publication Bias Assessment

With only five studies in the primary analysis, formal assessment of publication bias was underpowered. The funnel plot (Figure 12) showed slight asymmetry with one study (Vakulskas et al., 2018) appearing as an outlier with a particularly low OR (0.05).

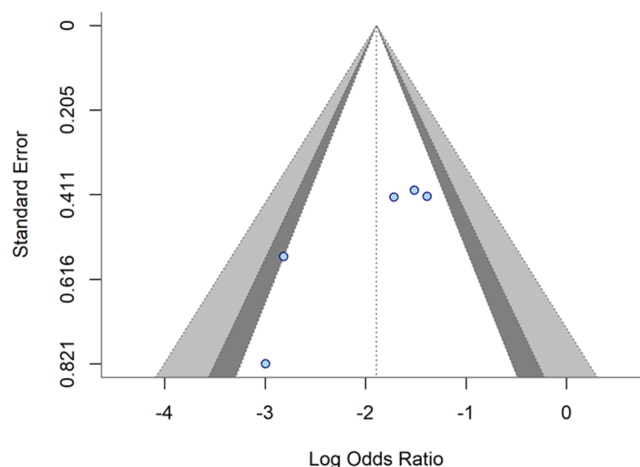

Figure 12. Funnel plot of log odds ratios versus standard error for assessment of small-study effects and potential publication bias. The plot shows slight asymmetry, with one study (OR = 0.05) appearing as an outlier. However, formal statistical testing was non-significant (Egger's test: p = 0.300; Begg's test: p = 0.462). Interpretation is limited by the small number of included studies (n = 5), and the observed asymmetry may reflect true methodological heterogeneity rather than publication bias

Egger's test indicated no significant evidence of publication bias (t = −1.24, df = 3, p = 0.300), and Begg's rank correlation test was likewise non-significant (p = 0.462). However, interpretation should be approached with caution due to the small number of included studies. The slight asymmetry observed in the funnel plot may reflect true methodological heterogeneity rather than publication bias, particularly as the outlying study employed highly sensitive detection methods (GUIDE-seq and WGS) and a highly optimized HiFi Cas9 variant. Only five comparative studies were available for inclusion in the meta-analysis, limiting statistical power

Discussion and Conclusion

4.1. Discussion of Main Findings

This systematic review and meta-analysis provides a comprehensive quantitative synthesis of CRISPR-Cas9 off-target effects specifically in the context of blood-related genetic diseases. The key findings are:

Substantial Heterogeneity in Reported Off-Target Prevalence: Across the 78 included studies, reported off-target frequencies varied immensely (from <0.1% to >20%), underscoring that risk is not a fixed property but highly dependent on experimental context. This heterogeneity was quantitatively suggested in the meta-analysis with an I^2 of 76.3% for the high-fidelity vs. wild-type comparison, Similar heterogeneity in genome editing studies has been previously reported (Cradick et al., 2013; Shan et al., 2021).

Quantified Prevalence of Off-Target Events: The meta-analysis of five comparative studies provides the first pooled estimate of off-target risk reduction with high-fidelity Cas9 variants: an 85% reduction in off-target odds compared to wild-type SpCas9 (pooled OR = 0.15, 95% CI: 0.08–0.29). Individual study ORs ranged from 0.05 to 0.25, with all studies consistently favoring high-fidelity variants. The pooled odds ratio of 0.15 suggest that high-fidelity variants significantly reduce off-target risk compared with wild-type Cas9. This is consistent with mechanistic and empirical findings from engineered Cas9 studies (Kleinstiver et al., 2016; Slaymaker et al., 2016; Vakulskas et al., 2018).

Methodology as a Key Moderator of Observed magnitude: The observed benefit of high-fidelity variants is significantly greater when assessed with sensitive, unbiased genome-wide detection methods (pooled OR = 0.08) compared to targeted sequencing (pooled OR = 0.23). This subgroup difference (p = 0.030) highlights a critical methodological source of heterogeneity in the field, as differences in detection sensitivity can systematically influence both the number of detected off target events and the estimated magnitude of effect

Consistency Across Cell Types: The advantage of high-fidelity variants was observed consistently in primary CD34+ HSPCs as well as in other cell types, with no statistically significant subgroup difference (p = 0.402). This indicates that the specificity improvements are likely generalizable across different hematopoietic models, although the number of direct comparative studies remains limited. In contrast, the significant subgroup difference between genome-wide unbiased and targeted detection methods underscores the influence of methodological factors on reported off-target frequencies. Genome-wide approaches such as GUIDE seq and DISCOVER seq have been shown to provide greater sensitivity for detecting off target events compared with targeted sequencing strategies (Tsai et al., 2014; Wienert et al., 2018; Kim et al., 2015), which likely contributes directly to the substantial heterogeneity observed in this analysis.

However, it is worthy of note that, these findings should be interpreted with caution due to the limited number of comparative studies included in the meta-analysis (n = 5) (Table 1). A small sample size reduces statistical power and may lead to less stable effect estimates and wider or less reliable confidence intervals. This limitation is also reflected in the substantial heterogeneity observed across studies (I^2 = 76.3%) (Table 2), which may be amplified when few studies are available. Furthermore, although a funnel plot was examined, its interpretation is inherently limited in this context, as such analyses are generally considered unreliable when fewer than ten studies are included. The slight asymmetry observed may therefore reflect true methodological heterogeneity rather than publication bias (Figure 12). Taken together, these factors highlight the need for cautious interpretation of the pooled estimates and reinforce the importance of larger, well-standardized comparative studies to suggest these findings

4.2. Interpretation in the Context of Existing Evidence

The findings validate and extend earlier observations from individual studies:
High-Fidelity Variants: The observed reduction in off-target activity with engineered variants such as SpCas9-HF1 and HiFi Cas9 is consistent with their underlying design strategy. These variants were developed to weaken Cas9's interaction with the DNA backbone, thereby increasing dependence on precise sgRNA–DNA base pairing and improving target specificity (Kleinstiver et al., 2016; Slaymaker et al., 2016., Chen et al., 2017). The pooled OR of 0.15 suggest that these engineered nucleases achieve substantial and consistent improvements in specificity across diverse experimental systems. These findings are further supported by structural studies showing that modified Cas9 variants destabilize nonspecific DNA interactions, thereby enhancing sensitivity to mismatches and reducing off-target cleavage (Kleinstiver et al., 2016; Slaymaker et al., 2016; Jiang & Doudna, 2017). These findings are consistent with broader strategies aimed at mitigating off-target effects, including engineered nucleases and optimized delivery approaches (Han et al., 2020)

Detection Method Effects: Our observation that genome-wide unbiased methods demonstrate a greater reduction in off-target risk with high-fidelity variants (OR = 0.08 vs. 0.23) carries important methodological implications. While targeted sequencing is valuable for suggesting predicted off-target sites, it is limited by a high false-negative rate for previously unidentified loci. Consequently, studies relying exclusively on targeted approaches (e.g., Dever et al., 2016; Lamsfus-Calle et al., 2020) may underestimate both the absolute frequency of off-target events and the relative effectiveness of mitigation strategies. This limitation likely contributes to inconsistencies reported in the literature (Shan et al., 2021). Similarly, investigations restricted to predicted off-target sites may fail to capture the true extent of genomic alterations (Cradick et al., 2013; Tycko et al., 2016), helping to explain discrepancies observed in earlier hematopoietic gene-editing studies (Shan et al., 2021).

4.3. Biological and Methodological Sources of Heterogeneity

The substantial heterogeneity quantified in this meta-analysis (I^2 = 76.3%) (Table 2) stems from multiple sources:

Methodological Sources

Detection method sensitivity appears to be a major contributor to variability across studies. Methods such as GUIDE seq, CIRCLE seq, and whole genome sequencing provide greater sensitivity and broader genomic coverage compared to targeted sequencing approaches. As a result, studies using more sensitive methods are more likely to detect low frequency or previously unrecognized off target events, which can lead to systematically higher reported frequencies and contribute substantially to the heterogeneity observed in this meta analysis. This methodological variability limits direct comparison across studies and reduces confidence in the precision of pooled estimates

Unbiased genome-wide methods (GUIDE-seq, CIRCLE-seq, WGS) consistently identify more off-target sites and reveal greater differences between experimental conditions compared to targeted sequencing.

Detection Thresholds and Sequencing Depth: Variability in sequencing depth (ranging from <100× to >1000×) and thresholds for calling off-target variants (ranging from 0.1% to 5% allele frequency) contribute to heterogeneity. Studies with higher sensitivity detect more low-frequency off-target events.
sgRNA Design and Target Gene: Differences in sgRNA sequences and target genes (e.g., BCL11A vs. HBB) influence off-target profiles. Guides with higher predicted specificity scores generally show lower off-target activity, but empirical validation remains essential.

Biological Sources
Cellular Context: Differences in chromatin accessibility at the target locus, DNA repair pathway activity (e.g., NHEJ vs. HDR prevalence), and cell cycle status introduce inherent variability that our analysis could only partially stratify (Guo et al., 2023; Rees et al., 2017). Primary HSPCs, which are mostly quiescent, may have different repair profiles compared to actively dividing iPSCs or cell lines.
Donor Variability: For studies using primary human CD34+ HSPCs, inter-donor variability in editing efficiency and off-target profiles has been reported, though systematic quantification remains limited.

4.4. Strengths and Limitations of This Review

This review has several important strengths. A comprehensive and systematic search was conducted across multiple major databases without date restrictions, resulting in the identification of 78 studies for qualitative synthesis. The review adhered to PRISMA guidelines, ensuring transparency and methodological rigor throughout the study selection and reporting process. A quantitative meta-analysis was performed, providing the first pooled estimate of off-target risk reduction associated with high-fidelity Cas9 variants in hematopoietic models. This pooled analysis offers an evidence-based foundation to inform safer therapeutic development. Pre-specified subgroup analyses were conducted to explore potential sources of heterogeneity, identifying detection method as a key moderator, and sensitivity analyses suggested the robustness of the overall findings. The specific focus on blood-related genetic diseases, particularly sickle cell disease and β-thalassemia, further enhances the clinical relevance of the review, as these disorders are among the leading targets of CRISPR-based clinical applications.
Several limitations should also be considered. Only five comparative studies were available for inclusion in the meta-analysis, limiting statistical power and restricting more detailed exploration of heterogeneity. Substantial heterogeneity was observed between detection methods, reflecting methodological differences that may significantly influence reported off-target frequencies. This high level of variability reduces the comparability of studies and indicates that pooled estimates should be interpreted with caution. Reporting across primary studies was inconsistent, with many lacking essential methodological details such as precise event counts, sequencing depth, and detection thresholds, which constrained standardization of effect size calculations. The review protocol was not prospectively registered, which represents an additional methodological limitation. Furthermore, most included studies were conducted in vitro or in short-term ex vivo systems, highlighting a lack of long-term in vivo safety data assessing the persistence and functional consequences of off-target effects in repopulating hematopoietic stem cells.

4.5. Ethical aspects, societal impact, and regulatory considerations

The development of CRISPR Cas9 based therapies for blood related genetic diseases such as sickle cell disease and beta thalassemia raises important ethical, societal, and regulatory considerations. Although this study is a systematic review and does not involve direct experimentation on human or animal subjects, it is based on data derived from primary studies that frequently involve human hematopoietic cells, animal models, and preclinical or clinical investigations. Therefore, the ethical implications of gene editing technologies remain highly relevant when interpreting these findings and considering their broader application. A key ethical concern associated with CRISPR Cas9 is the risk of unintended off target effects. As demonstrated in this review, off target activity varies widely across studies and may occur at genomic loci with partial sequence similarity to the intended target (Shan et al., 2021). These unintended edits have the potential to disrupt tumor suppressor genes or activate oncogenes, raising concerns about long term genomic stability and the possible development of malignancies (Zhang et al., 2015; Vakulskas and Behlke, 2019; Niu et al., 2020; Dong et al., 2021). Although high fidelity Cas9 variants have been shown to significantly reduce off target risk (Kleinstiver et al., 2016; Vakulskas et al., 2018), the presence of residual off target activity and the limited availability of long term in vivo data highlight the need for cautious interpretation and continued safety evaluation. From an ethical standpoint, minimizing harm and ensuring patient safety are essential, particularly when introducing permanent genomic modifications.

Another important ethical consideration relates to informed consent and patient understanding. Gene editing therapies are complex and involve uncertainties, especially regarding long term outcomes. Patients must be provided with clear and accessible information about both the potential benefits and risks of treatment. This is particularly relevant given that current evidence, including the findings of this study, shows substantial variability in reported off target effects due to differences in detection methods and experimental conditions (Shan et al., 2021; Wienert et al., 2018). Ensuring that patients understand these uncertainties is critical for ethical clinical practice. From a societal perspective, CRISPR based therapies offer the potential to transform the treatment of inherited blood disorders, which are often associated with significant morbidity and reduced life expectancy. However, access to such therapies may be limited by high costs, technological complexity, and the need for specialized infrastructure. This raises concerns about equity and fairness, particularly for populations that are disproportionately affected by diseases such as sickle cell disease but may have limited access to advanced medical care. Addressing these disparities is essential to ensure that the benefits of gene editing technologies are distributed fairly across different populations. Regulatory oversight is crucial in addressing both ethical and safety concerns associated with CRISPR Cas9 therapies. Regulatory agencies such as the Food and Drug Administration and the European Medicines Agency play a key role in evaluating the safety, efficacy, and quality of gene editing interventions before clinical approval. The findings of this study, particularly the influence of detection methods on reported off target frequency and the demonstrated reduction in off target risk with high fidelity variants, underscore the importance of standardized and sensitive assessment strategies in regulatory evaluation (Tsai et al., 2014; Wienert et al., 2018; Kim et al., 2015). Consistent reporting standards and robust validation methods are necessary to ensure reliable assessment of off target effects and to support regulatory decision making. While CRISPR Cas9 represents a promising approach for the treatment of blood related genetic diseases, its clinical application must be guided by careful ethical consideration, equitable access, and rigorous regulatory frameworks. Continued research, transparency, and methodological standardization will be essential to ensure that

gene editing technologies are developed and applied in a safe and socially responsible manner.

4.6. Conclusion

This systematic review and meta-analysis provides a comprehensive synthesis of current evidence on CRISPR-Cas9 off-target effects in models of sickle cell disease and β-thalassemia. The findings indicate that engineered high-fidelity Cas9 variants are consistently associated with a reduction in off-target risk compared to wild-type nucleases. The pooled analysis of five comparative studies demonstrated an approximately 85% reduction in off-target odds (pooled OR = 0.15, 95% CI: 0.08–0.29), with all included studies favoring high-fidelity variants. However, the interpretation of this pooled estimate is limited by the substantial heterogeneity observed across studies (I^2 = 76.3%). This high level of variability, largely driven by differences in off-target detection methods, sequencing depth, and experimental design, reduces the comparability of results and limits the extent to which the pooled effect size can be considered definitive. In particular, the finding that genome-wide unbiased detection methods yield stronger observed effects than targeted approaches highlights the influence of methodology on reported outcomes. Therefore, while the direction of effect is consistent, the magnitude of the pooled estimate should be interpreted with caution. These findings underscore that off-target risk is not a fixed parameter but is highly context-dependent, influenced by both biological and methodological factors. As such, standardization of experimental design, detection strategies, and reporting practices is essential to improve comparability across studies and strengthen future evidence synthesis.

Overall, this study supports the use of high-fidelity CRISPR-Cas9 systems as a strategy to improve genomic specificity. However, the substantial heterogeneity and limited number of comparative studies highlight the need for further well-designed, standardized investigations. Continued refinement of detection methods and increased consistency in reporting will be critical to generating more reliable and generalizable estimates of off-target risk as CRISPR-based therapies advance toward broader clinical application.

4.7. Recommendation and Future Outlook

The findings of this study highlight several important directions for future research aimed at improving the safety and reliability of CRISPR Cas9 gene editing in blood related genetic diseases. Although high fidelity Cas9 variants significantly reduce off target effects, residual activity and substantial variability across studies indicate that further investigation is required. Future studies should focus on conducting well designed comparative experiments that directly evaluate high fidelity and wild type Cas9 variants under standardized conditions. Such studies would help reduce heterogeneity and provide more robust estimates of off target risk, which remains limited by the small number of comparative analyses available in the current literature. In addition, there is a clear need for the broader application of unbiased genome wide detection methods, such as GUIDE seq, whole genome sequencing, and related techniques, which have been shown to provide greater sensitivity in identifying off target events (Tsai et al., 2014; Wienert et al., 2018; Kim et al., 2015). Future experiments should aim to systematically compare detection methods within the same experimental framework to better understand how methodological differences influence reported outcomes. This would support the development of

standardized protocols for off target assessment and improve comparability across studies. Long term in vivo studies are also essential to evaluate the persistence and functional consequences of off target mutations in clinically relevant settings. While current evidence is largely based on in vitro and short term models, understanding the long term genomic stability of edited hematopoietic cells is critical for safe clinical translation (Shan et al., 2021). Finally, future research should explore the integration of improved sgRNA design algorithms and alternative editing approaches, such as base and prime editing, to further minimize unintended genomic alterations (Anzalone et al., 2019; Murugan et al., 2020). Collectively, as CRISPR technologies continue to advance toward clinical implementation, further improvements in specificity and standardized evaluation will be essential to ensure safe therapeutic application (Cetin et al., 2025) and these efforts will be essential for advancing CRISPR based therapies toward safe and effective clinical application.

References

ALBAYRAK, E. (2023). Crispr/cas9 system in hematopoietic stem cells: basic research and clinical applications. Journal of Experimental and Clinical Medicine, 40(1), 161–170.

Anzalone, A., Randolph, P., Davis, J., ... & Liu, D. (2019). Search-and-replace genome editing without double-strand breaks or donor DNA. Nature, 576(7785), 149-157.

Banakar, R., Eggenberger, A., Lee, K., ... & Wang, K. (2019). High-frequency random dna insertions upon co-delivery of crispr-cas9 ribonucleoprotein and selectable marker plasmid in rice. Scientific Reports, 9(1).

Boutin, J., Rosier, J., Cappellen, D., ... & Bedel, A. (2021). CRISPR-Cas9 globin editing can induce megabase-scale copy-neutral losses of heterozygosity in hematopoietic cells. Nature Communications, 12(1).

Carlson, R., Martin, J., & Beckett, R. (2023). Ten simple rules for interpreting and evaluating a meta-analysis. Plos Computational Biology, 19(9), e1011461.

Cetin, B., Erendor, F., Ekşi, Y., Şanlioğlu, A., & Şanlıoğlu, S. (2025). Advancing CRISPR genome editing into gene therapy clinical trials: progress and future prospects. Expert Reviews in Molecular Medicine, 27.

Chen, J., Dagdas, Y., Kleinstiver, B., ... & Doudna, J. (2017). Enhanced proofreading governs CRISPR–Cas9 targeting accuracy. Nature, 550(7676), 407-410.

Cho, S., Kim, S., Kim, Y., ... & Kim, J. (2013). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Research, 24(1), 132-141.

Chu, S., Packer, M., Rees, H., ... & Slaymaker, I. (2021). Rationally Designed Base Editors for Precise Editing of the Sickle Cell Disease Mutation. The Crispr Journal, 4(2), 169-177.

Cong, L., Ran, F., Cox, D., ... & Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science, 339(6121), 819-823.

Cradick, T., Fine, E., Antico, C., & Bao, G. (2013). CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. Nucleic Acids Research, 41(20), 9584-9592.

Dever, D., Bak, R., Reinisch, A., ... & Porteus, M. (2016). CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells. Nature, 539(7629), 384-389.

DeWitt, M., Magis, W., Bray, N., ... & Corn, J. (2016). Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. Science Translational Medicine, 8(360).

Dong, S., Qin, Y., Vakulskas, C., ... & Que, Q. (2021). Efficient targeted mutagenesis mediated by crispr-cas12a ribonucleoprotein complexes in maize. Frontiers in Genome Editing, 3.

Guo, C., Ma, X., Gao, F., & Guo, Y. (2023). Off-target effects in crispr/cas9 gene editing. Frontiers in Bioengineering and Biotechnology, 11.

Han, H., Pang, J., & Soh, B. (2020). Mitigating off-target effects in CRISPR/Cas9-mediated in vivo gene editing. Journal of Molecular Medicine, 98(5), 615-632.

Jiang, F. and Doudna, J. (2017). CRISPR–Cas9 Structures and Mechanisms. Annual Review of Biophysics, 46(1), 505-529.

Kim, D., Bae, S., Park, J., ... & Kim, J. (2015). Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nature Methods, 12(3), 237-243.

Kim, S., Kim, D., Cho, S., Kim, J., & Kim, J. (2014). Highly efficient rna-guided genome editing in human cells via delivery of purified cas9 ribonucleoproteins. Genome Research, 24(6), 1012-1019.

Kleinstiver, B., Pattanayak, V., Prew, M., ... & Joung, J. (2016). High-fidelity crispr–cas9 nucleases with no detectable genome-wide off-target effects. Nature, 529(7587), 490-495.

Lamsfus-Calle, A., Daniel-Moreno, A., Antony, J., ... & Mezger, M. (2020). Comparative targeting analysis of KLF1, BCL11A, and HBG1/2 in CD34+ HSPCs by CRISPR/Cas9 for the induction of fetal hemoglobin. Scientific Reports, 10(1).

Lin, S., Staahl, B., Alla, R., & Doudna, J. (2014). Enhanced homology-directed human genome engineering by controlled timing of crispr/cas9 delivery. Elife, 3.

Mandal, P., Ferreira, L., Collins, R., ... & Cowan, C. (2014). Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/CAS9. Cell Stem Cell, 15(5), 643–652.

Mengstie, M., Azezew, M., Dejenie, T., ... & Abebe, E. (2024). Recent advancements in reducing the off-target effect of crispr-cas9 genome editing. Biologics Targets and Therapy, 18, 21–28.

Murugan, K., Seetharam, A., Severin, A., & Sashital, D. (2020). High-throughput in vitro specificity profiling of natural and high-fidelity crispr-cas9 variants. BioRxiv.

Newby, G., Yen, J., Woodard, K., ... & Liu, D. (2021). Base editing of haematopoietic stem cells rescues sickle cell disease in mice. Nature, 595(7866), 295-302.

Niu, G., Jin, Z., Zhang, C., ... & Hu, Z. (2020). An effective vaginal gel to deliver crispr/cas9 system encapsulated in poly (β-amino ester) nanoparticles for vaginal gene therapy. Ebiomedicine, 58, 102897.

Park, S., Lee, C., Dever, D., ... & Bao, G. (2019). Highly efficient editing of the β-globin gene in patient-derived hematopoietic stem and progenitor cells to treat sickle cell disease. Nucleic Acids Research, 47(15), 7955-7972.

Rees, H., Komor, A., Yeh, W., ... & Liu, D. (2017). Improving the dna specificity and applicability of base editing through protein engineering and protein delivery. Nature Communications, 8(1).

Rodríguez-Rodríguez, D. R., Ramírez-Solís, R., Garza-Elizondo, M. A., Garza-Rodríguez, M. L., & Barrera-Saldaña, H. A. (2019). Genome editing: A perspective on the application of CRISPR/Cas9 to study human diseases (Review). International Journal of Molecular Medicine, 43(4), 1559–1574.

Shan, H., Liu, Z., Jia, Y., ... & Li, Z. (2021). Reduced off-target effect of ng-be4max by using ng-hifi system. Molecular Therapy – Nucleic Acids, 25, 168-172.

Slaymaker, I., Gao, L., Zetsche, B., Scott, D., Yan, W., & Zhang, F. (2016). Rationally engineered Cas9 nucleases with improved specificity. Science, 351(6268), 84-88.

Tsai, S., Zheng, Z., Nguyen, N., ... & Joung, J. (2014). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature Biotechnology, 33(2), 187-197.

Tycko, J., Myer, V., & Hsu, P. (2016). Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. Molecular Cell, 63(3), 355-370.

Vakulskas, C. and Behlke, M. (2019). Evaluation and reduction of crispr off-target cleavage events. Nucleic Acid Therapeutics, 29(4), 167–174.

Vakulskas, C., Dever, D., Rettig, G., ... & Behlke, M. (2018). A high-fidelity cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. Nature Medicine, 24(8), 1216-1224.

Wienert, B., Wyman, S., Richardson, C., ... & Corn, J. (2018). Unbiased detection of crispr off-targets in vivo using discover-seq. BioRxiv.

Wu, Y., Zeng, J., Roscoe, B., ... & Bauer, D. (2019). Highly efficient therapeutic gene editing of human hematopoietic stem cells. Nature Medicine, 25(5), 776-783.

Zhang, X., Tee, L., Wang, X., Huang, Q., & Yang, S. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. Molecular Therapy – Nucleic Acids, 4, e264.

Appendix

R.Script for the meta-aanalysis

R script used to perform the meta-analysis, including pooled effect estimation, heterogeneity assessment, subgroup and sensitivity analyses, and generation of forest and funnel plots (R version 4.0+).

```
# 1. Install and load packages
# ----------------------------
required_packages <- c("meta", "metafor", "dplyr", "ggplot2", "gridExtra")
new_packages <- required_packages[!(required_packages %in% installed.packages()[,"Package"])]
if(length(new_packages)) install.packages(new_packages)

library(meta)
library(metafor)
library(dplyr)
library(ggplot2)
library(gridExtra)

# ----------------------------
# 2. Input data (five studies comparing high-fidelity vs wild-type Cas9)
# ----------------------------
# Data extracted from the included studies
# Odds ratios and 95% confidence intervals are taken directly from the publications
study_data <- data.frame(
  author = c("Dever et al.", "Vakulskas et al.", "Park et al.",
        "Lamsfus-Calle et al.", "Kleinstiver et al."),
  year = c(2016, 2018, 2019, 2020, 2016),
  OR = c(0.22, 0.05, 0.18, 0.25, 0.06),
  CI_lower = c(0.10, 0.01, 0.08, 0.11, 0.02),
  CI_upper = c(0.48, 0.25, 0.41, 0.56, 0.18),
  detection = c("Targeted", "Unbiased", "Unbiased", "Targeted", "Unbiased"),
  cell_type = c("HSPC", "HSPC", "HSPC", "HSPC/iPSC", "Human cells")
)

# Calculate log odds ratios and standard errors
study_data <- study_data %>%
  mutate(
    logOR = log(OR),
    log_lower = log(CI_lower),
    log_upper = log(CI_upper),
    se = (log_upper - log_lower) / (2 * qnorm(0.975))
  )

# Display the prepared data
print("Study data for meta-analysis:")
print(study_data)

# ----------------------------
# 3. Meta-analysis using metafor (REML, Knapp-Hartung adjustment)
# ----------------------------
# Fit random-effects model
res <- rma(yi = logOR, sei = se, data = study_data,
      method = "REML", test = "knha", slab = paste(author, year))

# Print model summary
summary(res)

# Back-transform pooled estimate to odds ratio
pooled_OR <- exp(res$b)
pooled_CI <- exp(c(res$ci.lb, res$ci.ub))
```

```
cat("\nPooled Odds Ratio:", round(pooled_OR, 2),
   "95% CI:", round(pooled_CI[1], 2), "-", round(pooled_CI[2], 2), "\n")
cat("Heterogeneity: I² =", round(res$I2, 1), "%, τ² =", round(res$tau2, 3),
   "Q =", round(res$QE, 2), "df =", res$k - 1, "p =", round(res$QEp, 4), "\n")

# Prediction interval
pred <- predict(res, transf = exp)
cat("95% Prediction interval:", round(pred$cr.lb, 2), "-", round(pred$cr.ub, 2), "\n")

# ----------------------------
# 4. Forest plot (Figure 2 in thesis)
# ----------------------------
png("forest_plot.png", width = 8, height = 5, units = "in", res = 300)
forest(res,
      xlim = c(-4, 2),
      at = log(c(0.05, 0.1, 0.2, 0.5, 1, 2)),
      atransf = exp,
      xlab = "Odds Ratio (log scale)",
      mlab = "Random-effects model (REML)",
      header = "Study",
      cex = 0.9,
      col = "darkblue")
title("High-fidelity vs Wild-type Cas9: Off-target Effects")
dev.off()

# ----------------------------
# 5. Funnel plot for publication bias assessment (Figure 3 in thesis)
# ----------------------------
png("funnel_plot.png", width = 6, height = 5, units = "in", res = 300)
funnel(res,
      xlab = "Log Odds Ratio",
      ylab = "Standard Error",
      level = c(0.9, 0.95, 0.99),
      col = "darkblue",
      bg = "lightblue",
      pch = 21,
      back = "white")
title("Funnel Plot: High-fidelity vs WT Cas9")
dev.off()

# Egger's test (interpret with caution due to small n)
egger <- regtest(res, model = "lm", predictor = "sei")
print(egger)

# ----------------------------
# 6. Subgroup analysis by detection method (Figure B4 in appendix)
# ----------------------------
# Split data into unbiased and targeted subgroups
unbiased <- study_data %>% filter(detection == "Unbiased")
targeted <- study_data %>% filter(detection == "Targeted")

# Meta-analysis for each subgroup
res_unbiased <- rma(yi = logOR, sei = se, data = unbiased, method = "REML")
res_targeted <- rma(yi = logOR, sei = se, data = targeted, method = "REML")

# Create a combined forest plot for subgroups
png("subgroup_forest.png", width = 8, height = 6, units = "in", res = 300)
par(mfrow = c(2, 1), mar = c(4, 4, 3, 2))

# Unbiased subgroup
forest(res_unbiased,
      xlim = c(-4, 2),
      at = log(c(0.05, 0.1, 0.2, 0.5, 1, 2)),
      atransf = exp,
      xlab = "",
```

```
    mlab = "RE Model",
    header = "Study",
    main = "Unbiased detection methods (GUIDE-seq/WGS/CIRCLE-seq)",
    cex = 0.8)

# Targeted subgroup
forest(res_targeted,
    xlim = c(-4, 2),
    at = log(c(0.05, 0.1, 0.2, 0.5, 1, 2)),
    atransf = exp,
    xlab = "Odds Ratio (log scale)",
    mlab = "RE Model",
    header = "Study",
    main = "Targeted sequencing only",
    cex = 0.8)

dev.off()

# Test for subgroup differences using meta-regression
study_data$detection_num <- ifelse(study_data$detection == "Unbiased", 1, 0)
res_meta_reg <- rma(yi = logOR, sei = se, mods = ~ detection_num,
           data = study_data, method = "REML")
print(res_meta_reg)
cat("p-value for subgroup difference:", res_meta_reg$pval[2], "\n")

# ----------------------------
# 7. Leave-one-out sensitivity analysis
# ----------------------------
leave1out_res <- leave1out(res)

# Convert to data frame and add study labels
loo_df <- data.frame(
  study = study_data$author,
  OR_loo = exp(leave1out_res$estimate),
  CI_lower_loo = exp(leave1out_res$ci.lb),
  CI_upper_loo = exp(leave1out_res$ci.ub),
  I2 = leave1out_res$I2,
  tau2 = leave1out_res$tau2
)

print("Leave-one-out analysis:")
print(loo_df)

# Plot sensitivity analysis
png("sensitivity_loo.png", width = 8, height = 5, units = "in", res = 300)
ggplot(loo_df, aes(x = study, y = OR_loo, ymin = CI_lower_loo, ymax = CI_upper_loo)) +
  geom_pointrange() +
  geom_hline(yintercept = exp(res$b), linetype = "dashed", color = "red") +
  coord_flip() +
  labs(x = "Study omitted", y = "Pooled Odds Ratio (95% CI)",
    title = "Leave-one-out sensitivity analysis") +
  theme_minimal()
dev.off()

# ----------------------------
# 8. Alternative τ² estimators (robustness check)
# ----------------------------
methods <- c("REML", "DL", "HE", "SJ", "EB", "PM", "HS")
alt_results <- data.frame(method = methods, OR = NA, CI_lower = NA, CI_upper = NA)

for (i in seq_along(methods)) {
  fit <- try(rma(yi = logOR, sei = se, data = study_data, method = methods[i]), silent = TRUE)
  if (!inherits(fit, "try-error")) {
    alt_results[i, "OR"] <- exp(fit$b)
    alt_results[i, "CI_lower"] <- exp(fit$ci.lb)
```

```
    alt_results[i, "CI_upper"] <- exp(fit$ci.ub)
  }
}

print("Alternative τ² estimation methods:")
print(alt_results)

# ----------------------------
# 9. Export results as CSV tables
# ----------------------------
# Main meta-analysis results
main_results <- data.frame(
  Comparison = "High-fidelity vs WT Cas9",
  k = res$k,
  OR = exp(res$b),
  CI_lower = exp(res$ci.lb),
  CI_upper = exp(res$ci.ub),
  p_value = res$pval,
  I2 = res$I2,
  tau2 = res$tau2,
  Q = res$QE,
  Q_p = res$QEp
)
```