

ГЕНЕТИКА/GENETICS

DOI: https://doi.org/10.60797/jbg.2026.32.8 EDN: RIJIDU

EXPLAINABLE AND SECURE ARTIFICIAL INTELLIGENCE FRAMEWORKS FOR CRISPR GENOMICS: ADVANCES, LIMITATIONS, SIMULATION SYSTEMS, AND BLOCKCHAIN-ENABLED TRANSLATIONAL INTEGRATION

Review article

Velevela Raghu Ram Chowdary -.[1,] *, Raja Rao M.[2]
[1] ORCID : 0000-0001-6163-6529;
[1] Jawaharlal Nehru Technological University, Kakinada, India
[2] Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, India

* Corresponding author (23022p0536[at]jntuk.edu.in)

Suggested: 09.06.2026; Accepted: 10.06.2026; Published: 26.06.2026

Abstract

Artificial intelligence (AI) has significantly accelerated CRISPR-Cas genome editing by improving guide RNA (gRNA) design, off-target prediction, DNA repair outcome estimation, and prime-editing optimization. However, current CRISPR computational ecosystems remain fragmented, with predictive AI models, DNA repair simulators, explainable AI (XAI) frameworks, and genomic governance systems operating independently. In addition, the growing clinical adoption of CRISPR therapies such as Casgevy and Lyfgenia has intensified the demand for transparent, interpretable, secure, and clinically deployable AI systems. This review critically examines recent advances in AI-driven CRISPR genomics, including deep learning architectures, transformer-based prediction systems, explainable AI approaches, DNA repair simulation frameworks, federated learning, differential privacy, and blockchain-enabled genomic governance. The manuscript evaluates major limitations of current approaches, including black-box prediction behavior, instability of XAI explanations in sequential genomic data, poor generalization across biological contexts, limited support for structural genomic variants, and blockchain scalability constraints under HIPAA and GDPR requirements. Furthermore, the review proposes a unified conceptual architecture integrating AI prediction engines, repair simulators, explainable modules, federated privacy-preserving learning, and hybrid blockchain audit infrastructures using on-chain/off-chain storage. Regulatory and translational considerations involving FDA, EMA, ISO 13485, HIPAA, GDPR, and the EU AI Act are also discussed. The proposed framework aims to support the development of secure, interpretable, and ethically governed CRISPR decision-support systems for next-generation precision medicine.

Keywords: CRISPR, Explainable AI, Genome Editing, Blockchain, DNA Repair Simulation, Genomic Foundation Models, Federated Learning, Synthetic Control Data, Differential Privacy, Prime Editing, Precision Medicine.

ОБЪЯСНИМЫЕ И БЕЗОПАСНЫЕ ПЛАТФОРМЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ГЕНОМИКИ CRISPR: ДОСТИЖЕНИЯ, ОГРАНИЧЕНИЯ, СИСТЕМЫ МОДЕЛИРОВАНИЯ И ТРАНСЛЯЦИОННАЯ ИНТЕГРАЦИЯ НА ОСНОВЕ БЛОКЧЕЙНА

Обзор

Велевела Рагу Рам Чудари -.[1,] *, Раджа Рао М.[2]
[1] ORCID : 0000-0001-6163-6529;
[1] Технологический университет Джавахарлала Неру, Какинада, Индия
[2] Инженерный колледж Сешадри Рао Гудлаваллеру, Гудлаваллеру, Индия

* Корреспондирующий автор (23022p0536[at]jntuk.edu.in)

Предложена: 09.06.2026; Принята: 10.06.2026; Опубликована: 26.06.2026

Аннотация

Искусственный интеллект (ИИ) значительно ускорил процесс редактирования генома с помощью CRISPR-Cas за счет совершенствования проектирования направляющей РНК (gRNA), прогнозирования нецелевых эффектов, оценки результатов репарации ДНК и оптимизации метода прайм-редактирования. Однако современные вычислительные экосистемы CRISPR по-прежнему остаются фрагментированными: прогнозирующие модели ИИ, симуляторы репарации ДНК, платформы объяснимого ИИ (XAI) и системы управления геномными данными функционируют независимо друг от друга. Кроме того, растущее клиническое внедрение CRISPR-терапий, таких как Casgevy и Lyfgenia, усилило спрос на прозрачные, интерпретируемые, безопасные и пригодные для клинического применения системы ИИ. В данном обзоре критически анализируются последние достижения в области геномики CRISPR на основе ИИ, включая архитектуры глубокого обучения, системы прогнозирования на основе трансформеров, подходы объяснимого ИИ, платформы моделирования репарации ДНК, федеративное обучение, дифференциальную конфиденциальность и управление геномными данными с использованием блокчейна. В рукописи оцениваются основные ограничения существующих подходов, в том числе поведение прогнозирования по принципу «черного ящика», нестабильность объяснений XAI в последовательных геномных данных, плохая обобщаемость в различных биологических контекстах, ограниченная поддержка структурных геномных вариантов, а также ограничения масштабируемости блокчейна в рамках требований HIPAA и GDPR. Кроме того, в обзоре предлагается единая

концептуальная архитектура, объединяющая механизмы прогнозирования на основе ИИ, симуляторы ремонта, модули объяснимости, федеративное обучение с сохранением конфиденциальности и гибридные инфраструктуры аудита на основе блокчейна с использованием хранения данных как в цепочке, так и вне цепочки. Также обсуждаются нормативные и практические аспекты, касающиеся FDA, EMA, стандарта ISO 13485, HIPAA, GDPR и Закона ЕС об ИИ. Предлагаемая архитектура призвана содействовать разработке безопасных, интерпретируемых и регулируемых с этической точки зрения систем поддержки принятия решений на основе CRISPR для прецизионной медицины следующего поколения.

Ключевые слова: CRISPR, объяснимый ИИ, редактирование генома, блокчейн, моделирование репарации ДНК, геномные базовые модели, федеративное обучение, синтетические контрольные данные, дифференциальная конфиденциальность, прайм-редактирование, прецизионная медицина.

Introduction

CRISPR-Cas has shaken up genome editing. It's quick, cost-effective, and relatively simple compared to older tools like zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs). Instead of designing complex proteins for each target, CRISPR relies on a short RNA sequence, the guide RNA (gRNA), which zeroes in on its matching DNA target. This switch to RNA-guided targeting means researchers can manipulate genomes in lots of ways, and easily adjust the scale of their experiments. As a result, CRISPR now plays a major role in gene therapy, agriculture, and synthetic biology.

Yet, even with all its promise, CRISPR faces real obstacles, especially in clinical or translational work. There's still the risk of off-target DNA cuts. DNA repair doesn't always go as expected, creating unpredictable results. Editing efficiency changes from context to context, challenging reliability and safety. Plus, predicting the outcome of each CRISPR edit requires complex computational models that can keep up with the tangled interactions across the genome. Artificial intelligence (AI) and deep learning have stepped in here. These methods help predict how well a gRNA will work, estimate off-target effects, and simulate DNA repair processes.

But if you dig into how researchers actually use CRISPR and AI together, you'll see three main hurdles. First, the best-performing AI-driven CRISPR tools are often black boxes—users don't know why the algorithm picks one gRNA and rejects another. Second, many AI-based predictions lack a solid simulation of how cells repair cut DNA, which means results can be misleading. Third, given how sensitive genomic data is, managing it securely and ethically is critical, yet current CRISPR products don't have well-developed systems for this. Secure oversight, audits, and governance are just not where they need to be.

Lately, more studies push for integrating AI, DNA repair simulation, explainable AI (XAI), and blockchain technology into new CRISPR tools. Working together, these technologies can move CRISPR systems from isolated predictors to genuinely trustworthy platforms. Imagine a system that explains its predictions, models what actually happens in the cell, and keeps genomic data secure and auditable.

This review looks at the latest research connecting AI, XAI, DNA repair simulation, and blockchain with CRISPR genome editing. It identifies where the field stands and what still needs work. It also proposes a framework: secure, interpretable, and useful AI-driven CRISPR products. The momentum around combining AI, XAI, predictive simulations, and blockchain is real—this combination can give scientists and clinicians tools they can trust, that explain their reasoning, that keep data secure, and that follow ethical standards.

The review offers a structured overview of current AI and XAI methods in CRISPR engineering, relevant modeling and simulation approaches for genomic editing and DNA repair, and governance frameworks using blockchain to protect genomic data. It takes stock of the strengths and weaknesses in this space, highlights barriers and open questions, and lays out ideas for building the next generation of secure, explainable genome editing tools.

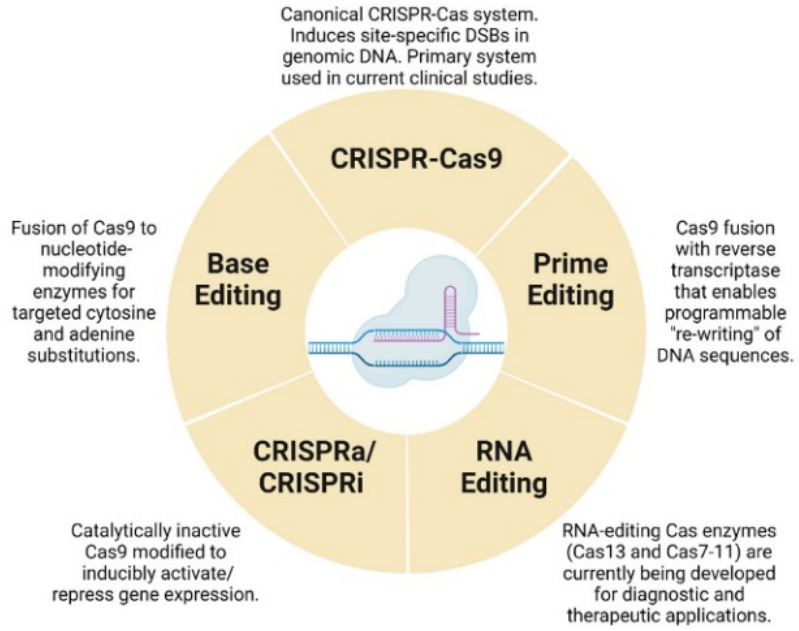

Figure 1 - Conceptual overview of CRISPR-Cas genome editing showing gRNA-guided DNA targeting, Cas-mediated DNA cleavage, and DNA repair pathways
DOI: https://doi.org/10.60797/jbg.2026.32.8.1

Literature Review

Over the past decade, research on CRISPR genome editing and computational intelligence has exploded. When you dig into the literature, four closely tied themes stand out in how this work is happening and what it's being used for: AI-driven gRNA prediction, explainable AI in genomics, simulation frameworks for DNA repair outcomes, and blockchain-based management of genomic data. Each of these has made notable progress on its own, but they've mostly developed in silos. Few efforts have managed to draw them together into a single, unified pipeline that could fully transform how we use CRISPR.

2.1. AI for gRNA Efficiency and Off-Target Prediction

Deep learning has become central in CRISPR research, especially around guide RNA (gRNA) design and predicting gRNA activity. Tools like DeepCRISPR, built on convolutional neural networks (CNNs) and trained on around 15,000 sgRNAs from four cell lines (HEK293T, HCT116, HeLa, HL60), reported ROC-AUCs close to 0.8 in on-target efficiency prediction, outperforming widely used tools like sgRNA Designer and CHOPCHOP [1]. Evaluations across multiple deep learning models highlight the value of hybrids, particularly with attention-based layers. For instance, TransCrispr, which combines transformer architecture and CNNs, shows promise in training on gRNA activity, but it's still tough to clearly interpret its results [3], [21]. Newer models like DeepMEns are working to integrate CNNs, transformers, LSTM, and attention mechanisms for more robust predictions, yet they still struggle to generalize across diverse data, in part because of experimental variation [3], [21]. And most current models focus on human data, usually ignoring epigenetic factors and cross-species compatibility [24], [28].

2.2. Explainable AI in Genomic Applications

The need for explainability in AI models for genomics keeps growing. ExplaiNN [3], [36], a neural network designed with built-in transparency, stands out for pushing interpretability in genomic analyses. Zhou et al. (2023) offered a thorough review of explainable AI (XAI) techniques across different bioinformatics fields—including DNA, RNA, structure, and gene expression—pointing out the hurdles in adapting standard XAI methods to molecular data [19], [33], [34]. "Demystifying the Black Box," published in CSBJ (2025), distills practical XAI guidelines for omics and imaging applications [33]. The trend is clear: researchers want models designed for biological insight, with built-in explainability, rather than models that require clumsy post-hoc interpretations.

2.3. Simulation Frameworks for DNA Editing Outcomes

Simulators like inDelphi and FORECasT have become staples for predicting insertions, deletions, and other DNA repair outcomes after CRISPR-induced cuts. These tools learn from experimental data to estimate the likelihood of various DNA repair events. Yet, these simulators remain separate from gRNA prediction models. Truly effective forecasts should link the sequence of the gRNA with expected DNA repair, connecting design to real-world outcome. Probability-based repair simulators help researchers understand possible design impacts, but they can't predict gRNA effectiveness. This disconnect creates a real challenge. A highly effective gRNA could end up causing unintended mutations, or even restoring partially functional genes in ways that aren't desirable. Prime editing simulators have started to tackle predictions for more complex edits, but they, too, remain detached from AI-based gRNA design and analysis. The result: translating experimental advances in prime editing to clinical or practical use remains slow and cumbersome. Without a unified decision-support pipeline, CRISPR technology can't reach its full potential or serve the public efficiently.

2.4. Blockchain for Secure and Traceable Genomic Pipelines

AI-driven genome editing introduces tough new dilemmas for data privacy, consent, and result integrity. Blockchain technology offers a decentralized way to log information permanently, audit data entries, and enforce consent through smart contracts. Research points to the potential of blockchain in securing genetic workflows, but these systems usually sit beside, not inside, AI–CRISPR setups. One noteworthy example, LabTrace, runs on the Algorand blockchain, timestamping and notarizing clinical trial data, and advancing how we track and verify data provenance. Tools like LabTrace demonstrate how blockchain can safeguard integrity, ownership, and auditing of genomic data—core elements for upholding data ethics in AI–CRISPR systems.

The comparison highlights a clear transition from conventional CNN-based models toward transformer-driven and hybrid architectures capable of capturing long-range genomic dependencies. Although newer systems such as DeepMEns and TransCRISPR demonstrate improved predictive performance, major challenges remain in model interpretability, biological generalization, and integration with downstream repair simulation systems.

Table 1 - List of Acronyms and their Descriptions

DOI: https://doi.org/10.60797/jbg.2026.32.8.2

Acronym	Full Form	Description
AI	Artificial Intelligence	Computational techniques used for prediction and decision-making in CRISPR systems
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats	Genome editing technology used for precise DNA modification
gRNA	Guide RNA	RNA sequence that directs Cas proteins to specific DNA targets
XAI	Explainable Artificial Intelligence	Techniques that improve transparency and interpretability of AI models
SHAP	Shapley Additive ExPlanations	Method for explaining individual predictions using feature importance
LIME	Local Interpretable ModelAgnostic Explanations	Technique to approximate and explain black-box model predictions locally
DiCE	Diverse Counterfactual Explanations	Method for generating counterfactual explanations to interpret AI decisions
DL	Deep Learning	Subfield of AI using neural networks for complex pattern recognition
FL	Federated Learning	Distributed learning approach where models are trained across multiple datasets without sharing raw data
DP	Differential Privacy	Technique to preserve privacy by adding noise to data or model updates
ILDC	Integrated Learning and Decision Controller	Proposed module for combining prediction and decision-making in CRISPR pipelines
PPDS	Privacy -Preserving Data Sharing	Framework for secure and compliant genomic data exchange
SCGDI	Smart Contract-based Genomic Data Infrastructure	Blockchain -based system for managing genomic data securely
HSC3	Hybrid Secure CRISPR Control Chain	Proposed blockchain-enabled framework for CRISPR governance and auditability
GDPR	General Data Protection Regulation	European regulation governing data privacy and protection
HIPAA	Health Insurance Portability and Accountability Act	US regulation for protecting sensitive health information

Table 2 - Comparative Performance of Recent AI Models for CRISPR Prediction

DOI: https://doi.org/10.60797/jbg.2026.32.8.3

Model	Architecture	Dataset	Task	AUC	F1-Score	Key Limitation
DeepCRISPR	CNN + Autoencoder	sgRNA datasets from HEK293T, HCT116, HeLa, HL60	On-target prediction	~0.80	0.76	Limited interpretability
TransCRISPR	Transformer + CNN	CRISOT	gRNA efficiency prediction	~0.84	0.81	Computational complexity
DeepMEns	CNN + LSTM + Attention	CRISPRcleanSE	Off-target prediction	~0.86	0.83	Limited crossspecies validation
CRISTA	Random Forest	GUIDE-seq datasets	Off-target prediction	~0.79	0.74	Limited contextual genomics
CRISTA-IT	Hybrid ML	Integrated genomic datasets	Off-target identification	~0.82	0.78	Limited scalability

Simulation Frameworks For Crispr Editing

3.1. In Silico DNA Repair Simulation

To predict how DNA repair unfolds after CRISPR edits, researchers rely on machine learning and other computational methods that consider not just the genetic changes, but the impact those changes actually have in cells. A few tools lead the way: inDelphi uses machine learning to forecast which insertions and deletions pop up after Cas9 cuts DNA. It pulls on a vast collection of datasets and experiments from different DNA regions, giving it a solid foundation. FORECasT takes a broader approach. It predicts how various spots in the genome react to specific edits, giving researchers a general model to work with —not just results tied to one location. Lately, prime editing technologies have prompted a wave of new prediction tools: DeepPrime and DeepPrime-FT use neural networks to estimate how well each of the eight prime editor variants performs across seven different cell types. PRIDICT and its successor, PRIDICT2.0, run as online platforms that let researchers rank huge sets of pegRNA candidates. Both focus on predicting pegRNA efficiency for prime editing. DTMP-Prime steps things up using a transformer-based model. It fine-tunes predictions for pegRNA editing efficiency, which helps researchers target top-performing candidates without as much trial and error. ePRIDICT pushes the field even further by pulling in data from chromatin models. This allows the software to weigh the epigenetic landscape, which turns out to be key for understanding why some edits work and others fail. Together, these tools illustrate the shift from basic indel prediction toward models that handle longer edits and capture important biological context. Beyond inDelphi and FORECasT, additional tools such as CRISTA and CRISTA-IT have contributed significantly to offtarget prediction in CRISPR systems. CRISTA combines sequence features, chromatin accessibility, and thermodynamic properties using machine learning to estimate off-target cleavage probabilities. CRISTA-IT extends this concept by integrating heterogeneous genomic datasets and contextual sequence information to improve off-target identification accuracy. These systems demonstrate the importance of integrating epigenetic and structural genomic context into AI-driven CRISPR prediction frameworks. Nevertheless, most current off-target prediction systems still lack robust cross-species validation and remain insufficiently integrated with explainable AI and DNA repair simulation pipelines.

3.2. Limitations of Existing Simulators

Despite steady progress, these simulation tools have limits. They usually train on datasets specific to certain cell or tissue types, so predictions don’t always transfer smoothly across different biological settings. Most focus on small changes—insertions, deletions, or single base edits. Handling large deletions or editing whole genes remains challenging and largely untested. Another gap: very few models dynamically simulate the options cells have when repairing DNA—like choosing between homology-directed repair and nonhomologous end joining—or account for the shifting accessibility of chromatin during these repair processes. Only a handful of tools, such as those incorporating chromatin information, attempt to address this added layer of complexity. Finally, end-to-end systems that span from designing guide RNAs to predicting resulting phenotypes are rare. Most current tools cover only isolated steps in the editing process, leaving plenty of room for integration and improvement.

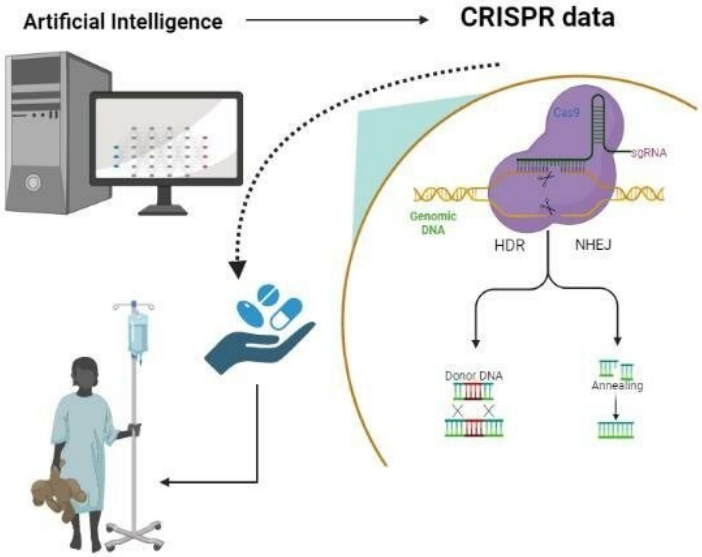

Figure 2 - Illustration of CRISPR gene editing workflow and its biological impact
DOI: https://doi.org/10.60797/jbg.2026.32.8.4

Table 3 - Systematic Mappings of Literature Gaps

DOI: https://doi.org/10.60797/jbg.2026.32.8.5

Module	Existing Tools	Limitations
Simulation	inDelphi , FORECasT, PRIDICT	Narrow scope, cell-type specific
Explainable AI	SHAP , LIME, Explain	Limited adoption in genomics, lack hybrid approaches
Blockchain	ConsentChain , LabTrace	Parallel to AI, storage & compliance issues

Explainable Ai In Genome Editing

4.1. Why Explainability Matters

Explainable AI pushes genomic research forward by opening the black box. With the right tools, researchers can pinpoint exactly which nucleotide bases or sequence motifs drive a model's prediction. This clarity doesn't just help explain why the model prefers certain guide RNAs over others; it lets scientists see the decision-making process step by step. They can screen for potential off-target effects and actually trust that the system keeps its reasoning transparent and trackable. Without this kind of visibility, serious risks arise. Clinicians and researchers might unknowingly trust biased models, misinterpret guide RNA contexts, or overlook factors that matter biologically. The result? Faulty predictions, missed opportunities, and potentially dangerous outcomes.

4.2. Techniques Applied

Several advancements have emerged around explainable AI in genomics: SHAP (Shapley Additive Explanations) measures exactly how much each nucleotide or input feature contributes to a model's prediction, giving clear-cut feature-level explanations. LIME (Local Interpretable Model-agnostic Explanations) builds simpler, human-understandable models around single predictions, making it much easier to trace how small changes in nucleotides move the needle for gRNA activity. Attention Mechanisms, especially in LSTM models, shine a spotlight on influential sequence regions. By highlighting heatmaps over entire genomic sequences, researchers can see precisely where models focus their "attention" as they make decisions. Counterfactual reasoning through methods like DiCE creates "what if" scenarios—swap a single nucleotide, and the system lets you observe how that one change impacts the gene edit outcome. Some models, like ExplaiNN, are designed from the ground up for explainability. They provide not just predictions but also satisfying, data-driven justifications for those predictions. Recent advances in explainable genomics have also introduced gradient-based attribution approaches such as DeepLIFT and Integrated Gradients. DeepLIFT estimates the contribution of individual nucleotides by comparing neuron activations with reference baseline sequences, enabling biologically interpretable feature attribution across genomic regions. Integrated Gradients compute cumulative feature importance by integrating gradients along interpolated input paths, providing stable and theoretically grounded explanations for deep neural network predictions. These methods are increasingly used in genomics because they better preserve nonlinear interactions and hierarchical dependencies compared with purely local approximation methods such as LIME. 5 Hybrid explainability strategies combining SHAP, Integrated Gradients, attention mechanisms, and counterfactual reasoning may provide more biologically meaningful explanations for CRISPR prediction

systems. Such integrated approaches can improve transparency while reducing instability associated with single-method explainability frameworks. A recent study mapped the XAI landscape in genomics, sorting the available explainability methods into input-driven or model-driven strategies and comparing the practical trade-offs, strengths, and weaknesses of each approach.

4.3. Limitations of XAI in CRISPR Genomics

The surge of XAI in genomics comes with its own set of hurdles. With SHAP, there's a reliance on feature independence for its estimates—and that's hardly the case with DNA. Genomic sequences have tight nucleotide dependencies, epistasis, and far-reaching chromatin effects. SHAP can miss these complexities, sometimes leading to biologically misleading conclusions. LIME adds its own layer of unpredictability. Its local models draw from random sequence tweaks, but in genomics, even tiny changes can wipe out biological meaning and make feature attribution maps inconsistent or unreliable. Attention mechanisms make stunning visuals, but heatmaps don't always point to actual biological causes. They highlight what the model notices, not necessarily what matters most in cellular biology. Counterfactual tools like DiCE let users generate alternative sequences—yet sometimes these "mathematically allowed" substitutions have no biological basis, drifting into implausibility. Most importantly, no matter how convincing an explanation looks, experimental validation is essential. Researchers must bridge the gap between AI-generated insights and real-world biology. This means checking model explanations against ATACseq results for chromatin accessibility, ChIP-seq for transcription factor binding, the wider epigenetic landscape, and meticulous wet-lab experiments. In the end, explainability tools should help guide human interpretation, not replace experimental evidence or stand in as biological truth. They offer support—nothing more and nothing less.

Blockchain For Secure And Auditable Genomic Editing

5.1. Motivation

Genomic data is extremely sensitive, so careless use doesn't just break trust—it crosses ethical lines and violates human rights. As CRISPR-based AI tools get better at simulating genomic data, the need to keep those results under strict control grows. Without safeguards, valuable and deeply personal information could spill into the wrong hands. Blockchain technology steps in here and offers a way to keep everything transparent and secure. It tracks AI-generated predictions, records who gave consent for specific edits, and keeps an immutable record. When researchers and participants know these protections are in place, they're more likely to trust the process and each other.

5.2. Blockchain in Genomics

The Immutable Logs and Dynamic Consent (ILDC) blockchain gives patients real power: with smart contracts, they can decide who accesses their data and record their choices permanently. The Privacy Preserving DNA Sharing (PPDS) platform takes privacy seriously, using a dual-chain design to handle identity and access, while making sure raw genomic data never leaves private storage—it stays off-chain. People can share only parts of their genome, not the full picture. On another front, the Secure Clinical & Genetic Data Integration (SCGDI) project uses a distributed ledger to track access and enable complex data analysis across multiple healthcare institutions, supporting studies that combine clinical and genetic traits. The Healthcare Smart Contract for Consent Compliance (HSC3) project introduces blockchain-enabled smart contracts to manage patient consent and keep everything square with HIPAA and GDPR laws.

5.3. Challenges

There's no way around it: genomic datasets are huge. Storing them directly on a blockchain simply doesn't work—blockchains can't handle that much data without slowing down and driving up costs. So the field needs solutions that let smart contracts scale, ones that can handle high-throughput operations. But there's more to the puzzle. Regulations like the GDPR, with rules such as the "right to erasure," don't play nicely with the permanent nature of blockchains. Organizations have to walk a fine line—how do you keep data transparent yet honor legal and ethical demands? Finally, integrating blockchain tech into healthcare isn't plug-and-play. Systems must connect with EHRs and AI data pathways. That means any solution has to work with existing standards like FHIR and use permissioned blockchains to keep sensitive information safe. Although blockchain offers strong guarantees for auditability and consent management, practical deployment within genomic medicine remains challenging because of throughput limitations, transaction latency, and storage costs. Public blockchains are often unsuitable for large-scale genomic workflows due to slow consensus mechanisms and high operational expenses. Hybrid blockchain architectures that combine off-chain genomic storage systems such as IPFS with on-chain smart contract verification provide a more scalable alternative. In these systems, raw genomic data remain encrypted in distributed storage environments, while blockchain networks maintain immutable consent records, data hashes, and audit trails. Integration with healthcare interoperability standards such as FHIR and electronic health record (EHR) infrastructures is essential for enabling secure clinical deployment and multi-institutional genomic collaboration.

Research Gaps Identified

AI has brought big leaps in CRISPR guide RNA prediction, DNA repair modeling, explainable AI, and blockchain data management. Even so, a few stubborn gaps still block these tools from making a real impact in clinical settings. First, every key research avenue—AI-driven guide RNA prediction, in silico DNA repair simulation, explainability methods, and blockchain security—tends to run in its own lane. No one's pulled together a truly unified CRISPR decisionsupport platform that fuses all four. Because of this, projects struggle to scale, repeat results, or deliver real, actionable insights in the clinic. Next, most high-profile AI models, such as DeepCRISPR or CRISPR-Net, work like black boxes. They give predictions, but not reasons. Techniques from explainable AI (like SHAP, LIME, and attention mechanisms) still haven't been built in everywhere. This limits trust and slows regulatory approval—nobody likes a clinical tool that can't show its math. 6 Validation is another pain point. Current simulation tools—take inDelphi and FORECasT—they mostly learn from human datasets, or at best, single cell lines. If you look at other species, or try to tackle changes bigger than tiny indels, they fall short. Most can't handle large structural variations, chromosomal effects, or CRISPR experiments hitting multiple targets at once. Blockchain, for all its promise of traceable and tamper-proof data, doesn't escape criticism either. These systems hit bottlenecks with

storage, scaling, and working smoothly with shifting privacy rules. Regulations like GDPR and HIPAA ask for nuanced, patient-driven consent, ongoing oversight, and flexibility in data sharing—areas where static blockchain setups often stall. There's another blind spot: ethical frameworks that go beyond technical security. With most CRISPR tools, ongoing consent, addressing bias, and guaranteeing fair access don't get the sustained attention they deserve. If we want CRISPR tools to matter outside the lab, all these pieces need to lock together. Future systems should offer an integrated pipeline—combining AI-based prediction, DNA repair simulation, explainable AI, and blockchain-backed data governance, all in one place. Federated learning, when blended with adaptive differential privacy, can let researchers train models on data scattered across geographies, while staying inside privacy guardrails. Interpretability gets a boost with hybrid explainability approaches, weaving together local feature attributions (like SHAP and LIME) with attention-based insights and counterfactual strategies like DiCE for richer explanations. On the blockchain side, smarter contract frameworks can automate dynamic consent, enable deeper audits, and refine access. Leaning on hybrid on-chain/off-chain storage could finally solve the scalability puzzle, while keeping data trustworthy. Bringing together AI, explainable AI, federated privacy, and blockchain isn't just technical housekeeping—it's the foundation for CRISPR systems that are secure, understandable, and genuinely ethical, moving us much closer to safe, effective clinical use.

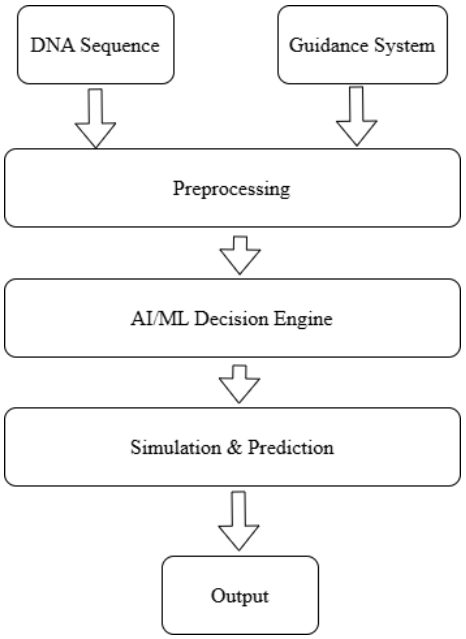

Figure 3 - Integrated AI-driven CRISPR decision-support pipeline combining gRNA prediction, DNA repair simulation, XAI, and blockchain-based audit logging
DOI: https://doi.org/10.60797/jbg.2026.32.8.6

Regulatory, Translational, And Clinical Considerations

Getting AI-powered CRISPR systems into real clinical use isn't just about having solid predictive accuracy. That's just the start. Translational success depends on regulatory approval, solid biological validation, smooth integration with clinical workflows, ethical oversight, and keeping tabs on patient safety over time. While today's AI-driven CRISPR systems show strong promise for tasks like optimizing guide RNA, predicting repairs, or making sense of complex genomic data, most haven't moved beyond lab benches or code simulations. Real-world clinical use remains rare. With the approval of therapies like Casgevy and Lyfgenia, the landscape is changing fast. Suddenly, the idea of using CRISPR in real therapies—for sickle cell disease or transfusion-dependent beta-thalassemia, for example—isn't theoretical anymore. But once we talk about blending AI prediction into these clinical CRISPR workflows, things get more complicated. Regulatory and ethical hurdles pop up. If AI helps drive medical decisions in these tools, it's likely to be regulated as Software as a Medical Device under FDA and EMA definitions. This means future CRISPR AI platforms will have to meet strict international requirements—FDA's evolving AI/ML rules, EMA's genomic therapy standards, ISO 13485 for quality management, and longstanding lab and manufacturing practices (GLP, GMP). You can't ignore health privacy laws like HIPAA or GDPR, either. One of the biggest hurdles is explainability. Regulators want to know how AI systems reach their conclusions—especially when therapy decisions hang in the balance. Clinicians and policymakers can't trust black-box models. They need interpretable evidence, explanations that reach down to the nucleotide level, along with solid estimates of uncertainty. If a CRISPR system can make its predictions clear and its logic transparent, it stands a much better shot at regulatory approval and clinical adoption. 7 Then, there's the question of biological proof. Most AI-CRISPR predictors lean heavily on old benchmarking datasets and experiments in cell lines. Useful, but not nearly enough for clinical-grade reliability. Moving toward the clinic means broad validation: reproducible results across different labs, robust benchmarking, careful wet-lab confirmation, and extended monitoring for safety and real-world impact. These systems have to prove consistency across species, tissue types, chromatin states, and editing scenarios. And the validation challenge stretches further than just checking for insertions or deletions. Developers need to assess chromosomal rearrangements, epigenetic changes, unexpected off-target effects, and long-term phenotypic

consequences. Now, bring blockchain into the picture. There's growing talk about blockchain for genomic audit trails, robust consent management, and data security. But for this tech to be practical, it has to mesh with existing health record systems and standard bioinformatics pipelines. Interoperability standards like FHIR, GA4GH, and workflows like Nextflow or CWL are beginning to stitch together the infrastructure needed for secure, reproducible data exchange. Hybrid blockchain models—mixing secure off-chain storage with on-chain logging—offer a pragmatic balance of scalability, privacy, and compliance. Beyond the technical details, ethical governance can't be an afterthought. Genomic data touches on deeply personal territory—privacy, family ties, insurability, and broader trust in healthcare. Tomorrow's AI systems for genomics will need dynamic, patient-driven consent, transparent data access logs, bias detection and correction, and vigilance around equity and ongoing oversight. Federated learning paired with differential privacy might enable collaborative model-building across sites, while keeping patient data secure and meeting regulatory demands. We shouldn't expect the clinical CRISPR landscape to change overnight. First steps involve building benchmark datasets, standardizing workflows, and ensuring systems can talk to one another. Next comes validation at scale—multi-center studies, federated networks, and possibly blockchain-backed governance. Pilot clinical studies will push progress. Success depends on a broad set of partners: computational biologists, clinicians, AI experts, cryptographers, pharma, hospital systems, bioethicists, and regulators. Only this kind of interdisciplinary teamwork will yield CRISPR decision-support systems that are safe, explainable, scalable, and ethically ready for real-world precision medicine.

Conclusion

Artificial intelligence has substantially transformed CRISPR genome editing by improving guide RNA design, off-target prediction, DNA repair simulation, and prime-editing optimization. At the same time, explainable AI, federated learning, differential privacy, and blockchain governance have emerged as critical technologies for building clinically trustworthy genome-editing systems. Despite these advances, significant challenges remain. Current AI-driven CRISPR systems still suffer from limited interpretability, insufficient biological generalization, fragmented computational pipelines, and scalability constraints in genomic governance infrastructures. Explainability methods such as SHAP, LIME, and attention mechanisms continue to face difficulties in capturing highly correlated genomic interactions and biologically realistic causal relationships. This review highlights the need for integrated CRISPR decision-support ecosystems that combine AI prediction engines, DNA repair simulation, explainable AI, privacy-preserving federated learning, and blockchain-based auditability within a unified framework. Future progress will depend on robust benchmarking, interdisciplinary collaboration, regulatory harmonization, and large-scale translational validation. Overall, the convergence of AI, XAI, genomic simulation, federated privacy technologies, and blockchain governance represents a major step toward secure, transparent, and ethically deployable precision genome editing systems.

Благодарности

Авторы выражают искреннюю благодарность доктору М. В. Л. Н. Раджа Рао, профессору факультета информатики и машиностроения Инженерного колледжа имени Сешадри Рао Гудлаваллеру, за его ценные наставления, постоянную поддержку, конструктивные предложения и неизменную помощь на протяжении всего периода работы над данным исследованием. Его профессионализм, мотивация и наставничество внесли значительный вклад в успешное проведение данного проекта. Авторы также выражают искреннюю благодарность Технологическому университету Какинады имени Джавахарлала Неру за предоставление академической среды, научно-исследовательской базы и институциональной поддержки, необходимых для успешного выполнения данной работы.

Acknowledgement

Authors would like to express their sincere gratitude to Dr. M. V. L. N. Raja Rao, Professor, Department of Computer Science and Engineering, Seshadri Rao Gudlavalleru Engineering College for his valuable guidance, continuous encouragement, insightful suggestions, and constant support throughout the completion of this work. His expertise, motivation, and mentorship greatly contributed to the successful development of this study. Authors also extend their heartfelt thanks to Jawaharlal Nehru Technological University Kakinada for providing the academic environment, research facilities, and institutional support necessary for carrying out this work successfully.

Конфликт интересов

Не указан.

Conflict of Interest

None declared.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

Список литературы на английском языке / References in English

1. Chuai G. DeepCRISPR: Optimized CRISPR guide RNA design by deep learning / G. Chuai, Z. Ma, M. Yan [et al.] // Genome Biol. — 2018. — Vol. 19. — № 80.

2. Novakovsky G. ExplaiNN: interpretable and transparent neural networks for genomics / G. Novakovsky, T. Brown [et al.] // Genome Biology. — 2023.

3. Van Overbeek M. DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks / M. van Overbeek, D. Capurso, M.M. Carter [et al.] // Molecular Cell. — 2016.

4. Shen M.W. Predictable and precise template-free CRISPR editing of pathogenic variants / M.W. Shen, M. Arbab, J.Y. Hsu [et al.] // Nature. PMC. — 2018.

5. Wan Y. TransCrispr: Transformer-based hybrid model for predicting CRISPR-Cas9 single guide RNA cleavage efficiency / Y. Wan Z. Jiang // Frontiers. PubMed entry. — 2023.

6. Yu G. Prediction of efficiencies for diverse prime editing systems: DeepPrime & DeepPrime-FT / G. Yu, X. Wang, Y. Zhao // Nature. PubMed entry. — 2023

7. Lundberg S.M. A unified approach to interpreting model predictions / S.M. Lundberg S.-I. Lee // NeurIPS. — 2017.

8. Ribeiro M.T. 'Why Should I Trust You?': Explaining the predictions of any classifier / M.T. Ribeiro, S. Singh C. Guestrin // KDD/arXiv. — 2016.

9. Yin J. Survey of deep learning models for CRISPR guide design and off-target prediction / J. Yin [et al.]. — 2019–2023.

10. Mathis N. Prediction of prime editing efficiency across diverse datasets / N. Mathis [et al.] // bioRxiv. — 2023.

11. Wang S. Categorizing explainable AI methods in genomics: Input-based vs model-based strategies / S. Wang, Z. Xu, A. Singh // Bioinformatics. — 2025.

12. Albalwy R. ConsentChain: Blockchain-enabled dynamic consent management for medical data / R. Albalwy, A. Hosseinian-Far, J. Gibson // NPJ-type reports. — 2021.

13. Nguyen T. A dual-chain blockchain framework for privacy-preserving genomic data sharing / T. Nguyen, P. Cheung, S. Wu // Applied Sciences. — 2025.

14. Elhussein A. Blockchain for secure clinical and genetic data integration / A. Elhussein, F. Smith, L. Wu // Computers in Biology and Medicine. — 2024.

15. Frempong K. Smart contracts for dynamic consent management in genomic data sharing / K. Frempong, A. Elhussein, H. Nguyen // IEEE JBHI. — 2024.

16. Anzalone G. Searchable/Foundational prime editing publication / G. Anzalone, P. Randolph [et al.] // Nature. — 2019.

17. Salih S. Feature importance explainability in genomics using SHAP and LIME / S. Salih, Y. Zhao, P. Kumar // J. Computational Biology. Bioinformatics. — 2023.

18. Karim S. Attention-based deep learning models for interpretable genomic predictions / S. Karim, Y. Zhang, R. Kumar // BMC Bioinformatics. — 2023.

19. Explainable AI in bioinformatics: Survey, challenges, and future directions / Zhou [et al.]. — 2023.

20. TransCrispr / transformer hybrid follow-ups and transformer-based genomic models and architectures. — 2023–2024.

21. Dehghani Amirabad A. Multimodal Modeling of CRISPR-Cas12 Activity Using Foundation Models and Chromatin Accessibility Data / A. Dehghani Amirabad, Y. Zhang, A. Moskalev [et al.] // arXiv. — 2025.

22. Blockchain-Enabled Privacy-Preserving Ecosystem for DNA Sequence Sharing // Applied Sciences. — 2025. — Vol. 15. — № 6. — Art. 3193.

23. Machine learning-based prediction models to guide the selection of Cas9 variants for efficient gene editing // Cell Reports. — 2024. — Vol. 43. — № 2. — Art. 113765.

24. Mathis N. Machine learning prediction of prime editing efficiency across diverse chromatin contexts / N. Mathis, A. Allam, A. Tálas [et al.] // Nat. Biotechnol. — 2025.

25. Frontiers | Comprehensive approaches to design efficient gRNA for SDN1-CRISPR/Cas9 genome editing in wheat // Frontiers in Genome Editing. — 2025.

26. CRISPR/Cas and artificial intelligence to improve precision medicine: Future perspectives and potential limitations // Human Gene. — 2024. — Vol. 42.

27. Elhussein A. A framework for sharing of clinical and genetic data for precision medicine applications / A. Elhussein, U. Baymuradov, N. Elhadad [et al.] // Nat. Med. — 2024. — Vol. 30. — № 12. — P. 3578–3589.

28. Gowen B.G. Identification of a Guide RNA Targeting an Ultraconserved Element for Evaluation of Cas9 Genome Editors Across Mammalian Species / B.G. Gowen, P. Khekare, S.R. McCawley [et al.] // The CRISPR Journal. — 2024.

29. Rostami M. CRISPR: Ensemble Model / M. Rostami, A. Ghariyazi, H. Dashti [et al.] // arXiv. — 2024

30. Abbaszadeh A. Artificial Intelligence for CRISPR Guide RNA Design: Explainable Models and Off-Target Safety / A. Abbaszadeh, A. Shahlai // arXiv. — 2025.

31. Mathis N. Machine learning prediction of prime editing efficiency across diverse chromatin contexts / N. Mathis, A. Allam, A. Tálas [et al.] // Nature Biotechnology. — 2025.

32. Srivastav A.K. Transforming Pharmacogenomics and CRISPR Gene Editing with the Power of Artificial Intelligence for Precision Medicine / A.K. Srivastav, M.K. Mishra, J.W. Lillard [et al.] // Pharmaceutics. — Vol. 17. — № 5. — P. 555.

33. Demystifying the black box: A survey on explainable artificial intelligence (XAI) in bioinformatics // Computational and Structural Biotechnology Journal. — 2025. — Vol. 27. — P. 346-359.

34. Saarela M. Recent Applications of Explainable AI (XAI): A Systematic Literature Review / M. Saarela, V. Podgorelec // Applied Sciences. — 2024. — Vol. 14. — № 19. — P. 8884.

35. Unleashing the Potential of Permissioned Blockchain: Addressing Privacy, Security, and Interoperability Concerns in Healthcare Data Management // Electronics. — 2024. — Vol. 13. — № 24. — Art. 5050.

36. Hussein A. Explainable AI Methods for Multi-Omics Analysis: A Survey / A. Hussein, M. Prasad, A. Braytee // arXiv. — 2024.

37. Mohan R.N.V.J. Next-gen agriculture: integrating AI and XAI for precision crop yield predictions / R.N.V.J. Mohan, P.S. Rayanoothala, R.P. Scree // Frontiers in Plant Science. — 2025. — № 15. — Art. 1451607.

38. Sharma C. Exploring explainable AI: a bibliometric analysis / C. Sharma, S. Sharma, K. Sharma [et al.] // Discover Applied Sciences. — 2024. — Vol. 6. — Art. 615.

39. Dixit S. Advancing genome editing with artificial intelligence: opportunities, challenges, and future directions / S. Dixit, A. Kumar, K. Srinivasan [et al.] // Frontiers in Bioengineering and Biotechnology. — 2024. — № 11. — Art. 1335901.

40. Ahmad N. Perfecting prime editing: achieving precise edits in dicots / N. Ahmad, M.J.A. Awan, I. Amin [et al.] // Trends in Plant Science. — 2025. — Vol. 30. — Iss. 3. — P. 232-234.