Review

Transgene Mapping in Animals: What to Choose?

Alexander Smirnov [1,*] , Maksim Makarenko [2] and Anastasia Yunusova [1]

1 Institute of Cytology and Genetics, Siberian Branch of Russian Academy of Sciences (SB RAS), Novosibirsk 630090, Russia
2 Department of Genetics and Life Sciences, Sirius University of Science and Technology, Sirius Federal Territory, Sochi 354340, Russia
* Correspondence: smirnovskaven@bionet.nsc.ru

Abstract: The phenomenal progress in biotechnology and genomics is both inspiring and overwhelming—a classic curse of choice, particularly when it comes to selecting methods for mapping transgene DNA integration sites. Transgene localization remains a crucial task for the validation of transgenic mouse or other animal models generated by pronuclear microinjection. Due to the inherently random nature of DNA integration, reliable characterization of the insertion site is essential. Over the years, a vast number of mapping methods have been developed, and new approaches continue to emerge, making the choice of the most suitable technique increasingly complex. Factors such as cost, required reagents, and the nature of the generated data require careful consideration. In this review, we provide a structured overview of current transgene mapping techniques, which we have broadly classified into three categories: classic PCR-based methods (such as inverse PCR and TAIL-PCR), next-generation sequencing with target enrichment, and long-read sequencing platforms (PacBio and Oxford Nanopore). To aid in decision-making, we include a comparative table summarizing approximate costs for the methods. While each approach has its own advantages and limitations, we highlight our top four recommended methods, which we believe offer the best balance of cost-effectiveness, reliability, and simplicity for identifying transgene integration sites.

Keywords: transgene; TAIL-PCR; long-read sequencing; genome walking; pronuclear microinjection

1. Introduction

Transgenic animals are the backbone of modern biology. It is nothing short of a scientific marvel that foreign DNA can integrate into a genome without direct assistance. However, in many cases, transgene insertion remains a black box—unless we can precisely determine the integration site. While early transgene mapping methods were often laborious and technically challenging, the excitement of discovering an integration site, especially when something was unexpectedly misplaced, was undeniable. One integration might have landed within a coding gene, another—next to a non-coding RNA that had only recently been annotated, sometimes the host gene would interact with transgene and form a hybrid transcript. With the explosion of biotechnology and genomics, a vast array of transgene mapping methods has emerged. The field has progressed tremendously. Between 1990 and 2010, before affordable whole-genome sequencing (WGS) became widely available, scientists had to get creative in their quest to locate transgene insertion sites. Numerous PCR-based methods were invented, falling under the umbrella of "genome walking" [1,2]. In an outdated but impressive 2011 review, Leoni et al. catalogued 53 different

genome walking methods [3]. The number of such methods likely exceeds a few hundred by now. Modern transgene mapping involves long-read sequencing with target enrichment and multi-omics approaches [1]. Today, integration sites, local chromatin states, and expression levels can all be analyzed in parallel with unprecedented precision and throughput [4]. Although most reviews adopt a historical, archivist perspective, we will deviate from this approach and instead focus on the practical appeal of mapping methods. Despite the wealth of available catalogs, including a recent comprehensive overview [1], researchers who simply want to sequence their transgenic mouse (or other creature) may find themselves overwhelmed by the sheer variety of available techniques.

The choice of mapping method is not just a matter of financial constraints, but also depends on the type of data one seeks to obtain, including read length, on-target (transgene) coverage, and discovery of accompanying genome rearrangements. So, what is the best approach? Is it expensive to just apply WGS to your mice and what coverage would be enough? Would targeted locus amplification (TLA) be optimal to resolve multicopy concatemers? Should one save costs and rely on thermal asymmetric interlaced PCR (TAIL-PCR), leaving the results to sheer luck (quite literally)? We aim to share our experience and recommendations, focusing on sequencing-based approaches for random transgene insertions in animals and, to some extent, in cultured cells.

1.1. Features of Random Transgenic Insertion in Animals

In this review, we focus on transgenic animals in which random integration of DNA is typically achieved via pronuclear injection. A natural question arises: if genotyping can be easily performed using qPCR to distinguish between heterozygotes and homozygotes, why bother identifying the exact integration locus at all (Figure 1)?

From a practical perspective, knowing the integration site can prevent downstream complications, especially in cases involving multiple insertions. A large-scale analysis of F0 mouse founder lines showed that approximately 20% had more than one integration site [5]. Multiple transgene loci may lead to unexpected segregation patterns, complicating both genotyping and phenotype interpretation.

Transgene integration is often influenced by position effect variegation (PEV)—the insertion site can significantly affect transgene expression levels. This is well illustrated in Chinese hamster ovary (CHO) cells, widely used in industrial protein production. Studies have shown that transgene insertions often occur in transcriptionally active regions, which are also prone to structural instability, including rearrangements over time [6–8]. As reviewed by Cabrera et al., integration into such regions may enhance expression but can also interfere with endogenous gene regulation [9]. Transgenes may be influenced by regulatory sequences located at considerable distances [10], emphasizing the importance of identifying their integration sites.

Furthermore, studies in mice have shown that nearly half of random integrations could potentially disrupt host gene function, either by inserting into introns or causing deletions of coding exons—for example, 45% (17/38) in report of Yan et al. [11], and 53% (21/40) in another work [12]. A recent study of the widely used Ucp1-Cre mouse line—which exhibits lethality in homozygous animals—revealed that the integration of a BAC transgene resulted in a large deletion and inversion affecting four genes, with potential additional effects on seven neighboring genes [13]. Notably, the presence of an active Ucp1 gene copy, which should not exist in the experimental model, influenced fat tissue homeostasis. Such cases are frequent, and genomic sequencing of established mouse strains often resembles archaeological investigation. According to the Mouse Genome Database, only 5% of over 8000 documented mouse transgenic lines have had their integration sites mapped [12].

Another unanticipated feature of random integration is the cointegration of unrelated DNA fragments. Initially considered rare, such events are now frequently observed thanks to deep genome sequencing in both cell lines and animals [14–16]. New quantitative methods analyzing CRISPR/Cas9-induced DNA breaks have shown that DNA is often incorporated at double-stranded break (DSB) ends—at frequencies of 0.1–1% per DSB [17,18]. This includes not only cotransfected DNA (which is expected to be abundant) but also genomic segments, repetitive elements, and regulatory sequences. For example, Geng et al. reported a striking "insertional bingo" event, discovering a ~200 bp fragment of *E. coli* DNA, a ~6 kb Cas9 plasmid backbone, and a local genomic duplication at the Cas9 target site [16].

Following pronuclear microinjection, the DNA repair machinery recognizes linear transgene ends and attempts to resolve DSBs by ligating whatever DNA is available [19]. Most commonly, transgene fragments are joined into concatemers, but integrations can also include plasmid backbones, bacterial genomic DNA, or even telomeric repeats (see recent review [20]). In the well-known hornless cattle case, a 200 bp "Celtic" allele was introduced using transcription activator-like effector nucleases (TALENs), but a plasmid backbone fragment was later discovered during U.S. Food and Drug Administration (FDA) re-evaluation [21]. This contamination could have been identified early using plasmid-specific primers—a practice that should become standard in long-term projects. Another illustrative case is the mouse line described by Chiang et al., in which the transgene was fragmented and inserted into host genome with a 168 bp segment of *Corynebacterium* DNA [22]. This sequence likely originated from the lab environment during DNA preparation. Cointegrations of *E. coli* fragments are very common as well [12,15]. Curiously, Hussmann et al. even identified a 165 bp bovine DNA fragment integrated into a CRISPR/Cas9 reporter in human cells—presumably captured from fetal bovine serum in the culture medium [23]. These findings highlight that the nucleus is a crowded environment, and the risk of foreign DNA integration at DSBs is non-negligible and should be carefully considered during mapping. These risks can potentially be minimized by treating plasmid preps with exonucleases to remove bacterial contaminants and carefully performing gel extraction steps during DNA preparation for microinjections. Better be safe than risk commemorating your sloppiness in a genome of transgenic animal.

Random integration is also frequently accompanied by large-scale structural rearrangements, including deletions, inversions, tandem duplications, and chromosomal translocations. Goodwin et al. found that over 50% of analyzed mouse lines carried chromosomal deletions, while 15 out of 40 also harbored duplications [12]. Similarly, Cain-Hom et al. reported two chromosomal translocations, two cointegrations, and three duplications near the insertion sites in Cre-deleter rodent lines [14]. Numerous other cases involving large tandem duplications have also been described [24–27]. The underlying reasons for the high frequency of duplications near integration sites remain to be fully elucidated.

Even when such structural changes do not directly affect the phenotype, they can interfere with genotyping, copy number analysis, and transgene detection. Therefore, high-resolution mapping—such as through long-read sequencing (LRS) or TLA—is strongly recommended, even for supposedly "well-characterized" transgenic lines.
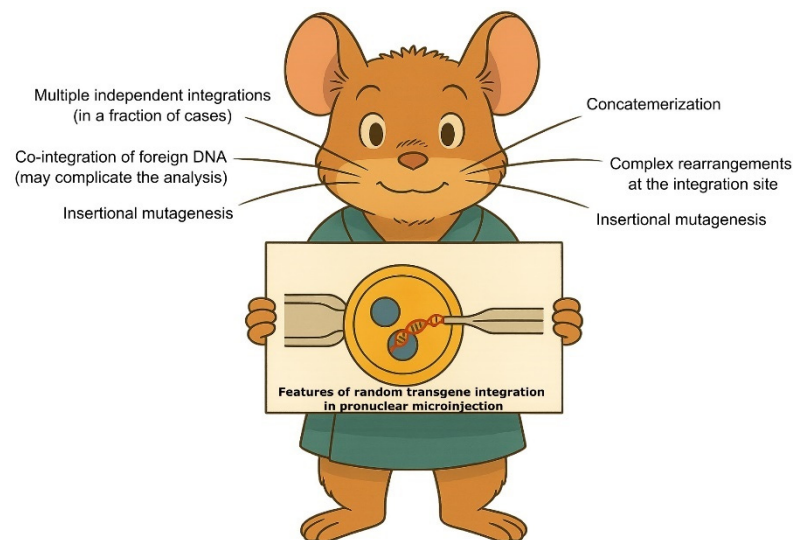
Figure 1. Features of the DNA integration in the pronuclear microinjection to be considered during transgene mapping. These features could complicate the transgene mapping and data analysis (details in the main text). The image was generated with ChatGPT, version GPT4o.

1.2. PCR-Based Methods for Transgene Mapping

Classic genome walking approaches include: ligation of universal adapters to linearized genomic DNA (LM-PCR), linear amplification using biotinylated primers (LAM-PCR), circularization of restriction fragments (Inverse PCR (iPCR)), annealing of semi-random primers (e.g., TAIL-PCR, PST-PCR) [1,28]. While these basic principles remain unchanged, many clever modifications have since appeared—ranging from improved degenerate primer designs for TAIL-PCR, to tagmentation-assisted adapter ligation [29], or sonication-based approaches replacing enzymatic restriction in inverse PCR [30]. All of them can be effective for transgene mapping in animals, but without direct meta-analysis under similar conditions, it's not very useful to discuss each one in detail.

Inverse PCR (iPCR) is one of the earliest and most widely used PCR-based approaches for mapping transgene integration sites [31,32]. Genomic DNA is first digested with restriction enzymes. The resulting fragments, including transgene–genome junctions, are self-ligated to form circularized DNA molecules. Outward-facing primers complementary to the transgene sequence amplify the unknown flanking region. Its efficiency remains remarkably high. For example, in the TRIP-Cas9 project, hundreds of transposon insertions were successfully mapped using iPCR [33]. We have also used iPCR to excise hundreds of transgene copies from a single embryo sample in order to analyze concatemer structures [34].

LM-PCR also involves restriction digestion of genomic DNA, followed by ligation with universal adapters [35]. The transgene-genome junction is amplified via nested PCR using a combination of gene-specific and adapter primers. Unlike iPCR, this method does not require digestion inside the transgene. However, its efficiency is affected by the need for adapter sets tailored to different restriction enzymes. Newer modifications introduce an additional digestion step to eliminate non-specific ligation products, but this requires precise restriction site planning and custom adapter preparation [36]. A recent version uses A-tailing, biotinylated primers, streptavidin capture, and secondary amplification [37]. Splinkerette PCR uses a specially designed hairpin adapter (formed from two ~48/61 nt annealed oligos), which provides greater specificity compared to simple ligation or circularization [38,39]. It has been used in mapping transposon and viral

insertion sites [40–42] and was recently adapted for mapping integrations in CHO cells with high efficiency [43].

Originally developed for mapping lentiviral integrations, linear amplification mediated PCR (LAM-PCR) uses a biotinylated primer to linearly amplify single-stranded DNA, which is then captured by streptavidin beads [44]. A second strand is synthesized with random primers and digested with restriction enzymes to create a ligation site for PCR adapters. This enrichment strategy improves specificity over standard LM-PCR. Later improvements replaced the restriction step: after capturing the ssDNA, a single-stranded adapter is ligated, and amplification proceeds with two primers [45]. LAM-PCR has also been combined with sonication for deep profiling of viral integration sites [46].

TAIL-PCR remains one of the most accessible transgene mapping tools, especially for beginners. Unlike other methods, it does not require restriction digestion, primer biotinylation, or commercial kits. All that is needed is a few gene-specific primers and a set of arbitrary degenerate (AD) primers, such as 5′-NGTCGASWGANAWGAA-3′. First reaction of TAIL-PCR typically involves the following steps: high-stringency cycles with high annealing temperature to let sequence-specific (SS) primers generate single-stranded DNA, low-stringency cycle (~25 °C) where AD primers bind randomly to genomic DNA, and normal amplification with nested SS and AD primers to enrich transgene–genome junctions. This is followed by nested PCR to improve specificity (Figure 2A).

Originally developed for T-DNA insertion mapping in plants, TAIL-PCR had a 50–70% success rate [47,48]. Later, hiTAIL-PCR improved specificity by optimizing primer structure and PCR cycling [49]. Some reports noted only 20–30% [50] or 39–69% [51] efficiency of this method. Authors demonstrated that pooling classic AD primers in various combinations or designing new AD primers with lower degeneracy levels improved efficiency two-fold. Additional factors that help to improve outcomes include novel processive polymerases, optimizing PCR annealing temperatures, and stronger dilution of the first reaction [52]. Another group observed up to 83% success of TAIL-PCR in mouse transgene mapping even when using original protocol [11]. Compared to alternative methods, TAIL-PCR has a broader range of applications and high efficiency for mapping random insertions in transgenic animals [11,53,54], cell cultures [55], zebrafish [56], and plants [57,58]. Dozens of related methods have emerged based on the same thermal asymmetry principle, including Wristwatch PCR [59], Fork PCR [60], PER-PCR [61], PST-PCR [28]. These modifications aim to reduce non-specific products or extend amplicons beyond 3–4 kb to capture structural rearrangements flanking the integration site.

We recommend classical hiTAIL-PCR using multiple long AD primer pools to minimize the risk of amplifying transgene–transgene junctions (Table 1) [49]. In our own experience, this method worked in over 80% of cases [62,63], later we reanalyzed the uncharted cases with another transgene-specific primers and found end truncations [64]. That said, genomic rearrangements at transgene ends can reduce the reliability of all PCR-based approaches—sometimes, there's just no primer-binding site at all [65]. Chimeric products due to PCR [66] and ambiguous bands where parts of the transgene map to different chromosomes [67] are not uncommon, so always confirm results with alternative methods like long-distance PCR or LRS.

While PCR-based mapping is not the gold standard anymore in the next-generation sequencing (NGS) era, it still offers valuable solutions for small-scale, cost-sensitive projects. Among them, TAIL-PCR remains our go-to for locating transgene insertions—requiring little more than a few PCRs and Sanger reads.
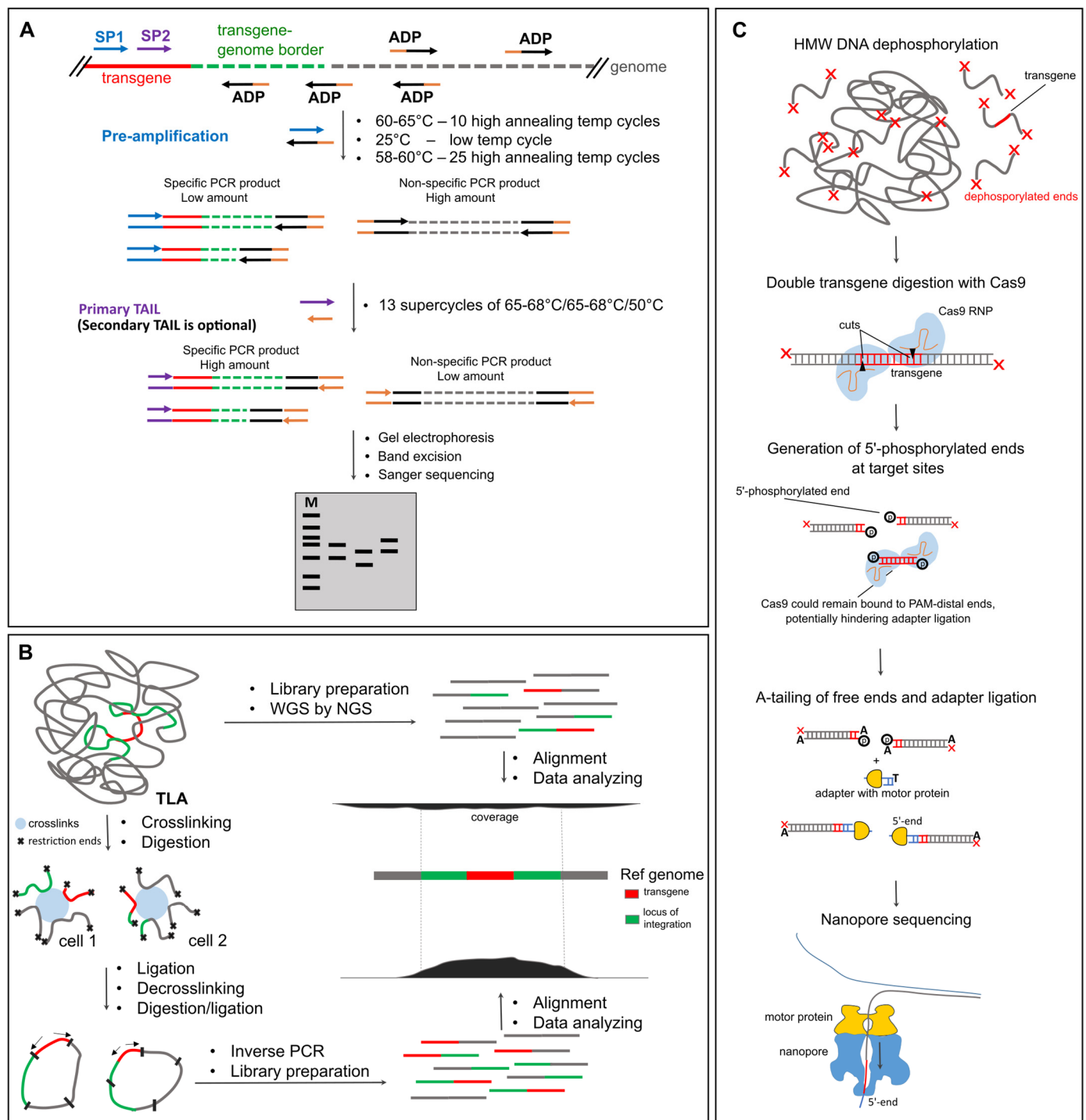
Figure 2. Selected methods for transgene mapping in animals. (**A**) hiTAIL-PCR. The schematic overview of the method shows the main steps which may differ between alternative TAIL-PCR approaches. SP—sequence-specific primer, ADP—arbitrary degenerate primer. (**B**) NGS-based methods: WGS and TLA. (**C**) Nanopore with enrichment by Cas9 digestion.

1.3. Next-Generation Sequencing and Target Enrichment

NGS has become an essential part of genomics research [68,69]. Objectively, the most efficient way to identify transgene integration sites is through WGS at sufficient coverage (Figure 2B). But what is the optimal genome coverage for reliable transgene detection? Several studies of transgenic mouse lines have shown that a haploid genome coverage as low as 8× [70] or 11.5× [71] may be sufficient for mapping small insertions. Srivastava

et al. reported unsuccessful mapping using standard paired-end sequencing at 18× coverage, and applied mate-pair sequencing instead [72]. WGS is also commonly used for mapping insertions in transgenic farm animals. Zhang et al. sequenced a transgenic cow carrying a human lactoferrin BAC insert. Although the bovine genome was sequenced at ~10× coverage, the effective coverage of the BAC insert reached 20–50× due to concatemerization [54]. However, internal rearrangements made the structure too complex for short-read NGS to resolve. Another study used ultra-deep sequencing (~268×) to analyze a 3.1 kb SRY-GFP construct knock-in [73]. Despite the high coverage, two alleles with complex structural variants had to be resolved using PacBio. Interestingly, the Cas9-linearized vector caused concatemerization but no random integrations outside of the intended "safe harbor" locus [73]. The same group successfully sequenced F1 offspring of a hornless bull with notorious backbone integration [74] at 20× coverage [75]. Transgenic crops are generally sequenced at ~13–14× [76], 21× [77], 29× [78], or even 70× [79], although T-DNA insertions are usually less repetitive and easier to map. These examples illustrate the approximate sequencing depth needed to identify insertion sites. On modern Illumina platforms (e.g., 150 bp paired-end reads), such coverage can still be relatively expensive, especially for large-scale screening (Table 2).

Table 1. Cost-effective transgene mapping methods. The efficiency/scalability/cost are given as subjective close estimates based on the literature and personal experience. Cost calculations are presented in Table 2.

	Method 1: hiTAIL-PCR
Efficiency/Scalability/Cost	~80%/low/$80 per line
Advantages	Simple and cost-effective protocol with high efficiency. Only requires standard PCR, gel electrophoresis and Sanger sequencing. Universal AD primers are compatible with most genomes. Generates relatively long PCR products (300–2000 bp), which improve alignment accuracy over short NGS reads. The hiTAIL-PCR design suppresses non-specific short amplicons [49]. Amplicon length can be extended further with protocol modifications [52].
Problems	Requires an intact primer binding site: As with any genome-walking PCR method, successful amplification depends on a functional transgene-specific primer site. If initial attempts fail, new primers spaced every 300–400 bp along the transgene may be required. Non-specific amplification: based on the conditions (transgene copy number, genome complexity, degenerate primer sequence), non-specific amplification may represent a problem. Transgene–transgene junctions are also efficiently amplified and give the misleading characteristic size shift at the gel after the secondary TAIL-PCR. This can be countered by using different AD primers or restriction digestion of the transgene–transgene regions.
Perfect for	Single-copy, intact transgene insertions
	Method 2: WGS by NGS
Efficiency/Scalability/Cost	~95%/average/$250–2400
Advantages	Not linked to a specific sequence, making it effective regardless of transgene truncations.
Problems	Costly: Achieving 10–30× genome coverage for reliable mapping typically costs over $1000, depending on sequencing provider and genome size. Short read length limits ability to resolve complex integration sites, such as flanking duplications or inversions (which are relatively frequent).
Perfect for	Urgent low-scale mapping experiments

Table 1. *Cont.*

	Method 3: TLA		
Efficiency/Scalability/Cost	100%/average/$150–2000		
Advantages	Uses proximity ligation to enrich for sequences near a known transgene region, increasing the chance of capturing insertion breakpoints with short reads		
	The crosslinking protocol could be established in the lab to enrich NGS data [80], making it one of the most cost-effective mapping methods.		
Problems	Large constructs (e.g., BACs) or insertions with unknown elements may require multiple primer pairs causing additional expenses.		
	Less accessible than other methods: the protocol involves complex sample preparation and may be more practical through commercial services, which can be expensive and non-transparent.		
Perfect for	Most cases		
	Method 4: Nanopore LRS + Cas9 enrichment		
Efficiency/Scalability/Cost	100%/average/$350–1000		
Advantages	Long reads enable unambiguous mapping: sequencing reads spanning thousands of base pairs can cover entire integration loci and flanking rearrangements.		
	Cas9-based enrichment could use multiple gRNA increasing coverage efficiency. Protocols for large scale in vitro gRNA synthesis from PCR templates are simple and fast [81].		
Problems	Requires high-molecular-weight (HMW) DNA: Extraction protocols are technically demanding and require fresh or well-preserved samples. Degraded short-length DNA or overly viscous samples can ruin flow cell performance.		
	Even with enrichment, coverage may be limited. Nanopore error rates (~1%) can be problematic for distinguishing barcoded or repetitive sequences [82]		
Perfect for	Multicopy concatemers, complex insert sites		

To improve detection sensitivity and reduce sequencing costs, target enrichment techniques have been developed to increase the proportion of reads covering transgene-genome junctions. Since an insertion site represents only a tiny fraction of the mouse genome, sequencing a 10 kb transgene at high coverage (>10×) requires just a few thousand reads—an insignificant portion of a typical NGS dataset (Table 2). Many enrichment methods have emerged (see recent review [1]), some based on earlier molecular biology strategies such as LAM-PCR [83], TAIL-PCR [84], inverse PCR [85], while others involve newer approaches like chromatin crosslinking or Cas9-mediated enrichment [86]. All these are used to enrich sequencing libraries prior to high-throughput sequencing [87]. In this section, we briefly describe several popular enrichment methods applicable to transgene mapping: biotinylated probe capture (hybrid capture), chromatin-crosslinking (TLA), and others. The final choice depends on user expertise and available resources.

Perhaps the most widely used enrichment technique for mapping transgene integration is TLA [80]. TLA builds on the principles of chromatin conformation capture (3C/Hi-C). In the first step, formaldehyde crosslinks chromatin, fixing together DNA regions that are physically close—including the transgene and flanking genomic sequences. Next, the DNA is digested with a frequent-cutting enzyme (e.g., NlaIII), followed by religation under dilute conditions to promote intramolecular ligation. After reverse crosslinking, a second round of digestion and religation produces circular DNA molecules enriched in ligation products near the transgene. PCR with outward-facing transgene-specific primers amplifies these circles, allowing selective enrichment of flanking genomic regions. The resulting fragments are subjected to standard library preparation and sequencing (Figure 2B). Although the TLA protocol appears complex, it can be performed in any lab with modest resources [80,88–90]. However, data analysis requires proficiency in interpreting chromatin

ligation-based datasets. For this reason, many researchers outsource TLA mapping to commercial providers like Cergentis [13,91–94]. TLA has proven particularly valuable in large-scale transgenic mouse studies [12].

Typically, the region with the highest coverage—often exceeding 100 kb—indicates the most likely insertion site. Large constructs like BACs may require 5–6 primer sets and rounds of TLA to achieve sufficient coverage [13,93]. A major advantage of TLA is that the resulting amplicons contain not only flanking sequences but also the entire transgene. Also, because homologous chromosomes occupy distinct nuclear territories, TLA is also capable of haplotyping, detecting SNVs, and identifying large structural variants near integration sites. However, the use of short Illumina reads (~150 bp) limits resolution in repetitive regions and fails to fully resolve complex concatemers. Combining TLA with LRS can improve structural resolution: transgene flanks identified by TLA can guide Cas9 digestion and Nanopore-based enrichment [95,96].

Another widely used method is hybrid target capture [1], including solid-state microarrays [97] and magnetic beads. The latter approach is more convenient. Biotinylated DNA or RNA probes anneal to denatured, fragmented genomic DNA. Hybridized molecules are captured using streptavidin-coated magnetic beads. The captured DNA is then extended by polymerase to complete sequencing templates. A major advantage of hybrid capture is that overlapping 60–120 nt probes can cover an entire transgene sequence—especially useful for random integration mapping, because the borders of the insert could be truncated. This method has been used successfully in multiple studies [22,98], and is considered cost-effective once probes are synthesized (Table 2). For instance, Magembe et al. used a pool of 413 xGen Lockdown probes to tile an 18 kb T-DNA region in plants. They found around 10–20% of target reads in the NGS data. Although probe coverage was uneven, 30 and 27 of each of the T-DNA ends from 34 lines were successfully mapped [98]. In another study, capture probes targeted bovine leukemia virus (BLV) insertions with modest enrichment: 10.2% of the total reads mapped to the target proviral genome [99]. Iwase et al. used hybrid enrichment to detect HIV-1 integration sites and generated around 5% of the target reads of the total data [100].

An intriguing and recent addition to the toolbox is T7-based transcriptional mapping, used by Li et al. for locating transposon insertions [101]. This method requires addition of a ~20 bp T7 promoter near the end of the transgene. Genomic DNA is subjected to in vitro transcription, followed by cDNA synthesis using random primers—eliminating the need for restriction digestion or ligation. Although the effective read length depends on the transcription reaction, cDNAs can exceed 1 kb, enabling efficient transgene-genome junction recovery. This approach is promising but may suffer from loss of the T7 sequence during random integration events.

In summary, NGS is a powerful tool for mapping transgene insertions. For many applications, WGS or commercial TLA remains the best choice (Figure 2B), depending on budget and available expertise (Table 1). However, the limited read length of short-read platforms often complicates mapping—especially for rearranged or repetitive regions. For example, in the study by Siddique et al., only one end of a T-DNA insertion was resolved even at $36\times$ coverage [102]. In another report, Peng et al. mapped a complex insertion in a repetitive region of the maize genome but even $41\times$ WGS and TAIL-PCR failed to identify the region, which required long-read sequencing [103]. For transgenic core facilities or large-scale mouse projects, implementing enrichment protocols such as hybrid capture or TLA can significantly improve mapping outcomes. In this review, we only scratched the surface of available tools. While many protocols are low-cost, they require significant optimization and bench skills. Still, for researchers who can manage custom biotinylated probe synthesis or chromatin crosslinking, the results are often worth the effort.

As one colleague once remarked, during yet another transgene mapping crisis: "What am I supposed to do with these 100 bp snippets? Give me long reads or I'm out!".

Table 2. Comparison of costs and labor time for transgene mapping methods. Estimates are based on a hypothetical 10 kb transgene and should be adjusted according to the expected insert size. Pricing and time estimates exclude DNA isolation and do not account for delivery time, which may vary significantly depending on geographic location. High-throughput sequencing using platforms such as Revio (PacBio), PromethION (Nanopore), and NovaSeq 6000 (Illumina) is typically outsourced to specialized service providers rather than conducted in individual laboratories. Therefore, when planning such experiments, it is essential to consider additional factors, including probe design and synthesis time, shipping logistics, and service turnaround—each of which can substantially affect the overall project cost and timeline.

	Preparation Price per Sample *	Sequencing Price per Sample	Sufficient Sequence Data (Gb)/On-Target Data (%)/On-Target Coverage (Reads)	Preparation/Run Time
Inverse PCR [32]	$20–$30	$10–$30 (Sanger)	<0.001 Gb/NA/NA	~9–12/3–4 h
TAIL-PCR [49]	$40–$50	$10–$30 (Sanger)	<0.001 Gb/NA/NA	~8–12/3–4 h
WGS by NGS (Illumina paired-end 150 bp) [54,70,73,75]	$75–$135	NGS Option A: NovaSeq 6000 S4 ~ $160–$250 NGS Option B: NextSeq 500/550 ~$1900–$2400	30 Gb/<0.01%/>10	NGS Option A: ~3–5/45 h NGS Option B: ~3–5/35 h
NGS + TLA (commercial) [12,13,93]	$1000–$2000		NA	Weeks
NGS + TLA (lab) [88–90]	$50–$75	NGS Option A: ~ $35–$70 NGS Option B: ~ $200–$250	3 Gb/~30–70%/>30	NGS Option A: ~36–48/35 h NGS Option B: ~36–48/45 h
NGS + hybrid capture (using 120 nt commercial tiling probes) [73,74]	$180–$250	NGS Option A: ~ $10–$20 NGS Option B: ~ $75–$150	1 Gb/~40–80%, up to 95% **/>30	NGS Option A: ~24–36/45 h NGS Option B: ~24–36/35 h
NGS + hybrid capture (probes made in the lab)	$50–$60	NGS Option A: ~ $10–$20 NGS Option B: ~ $75–$150	1 Gb/~80–90%, up to 93% **/>50	NGS Option A: ~50/45 h NGS Option B: ~50/35 h
NGS + T7 In vitro transcription [101]	$50–$70	NGS Option A: ~ $35–$70 NGS Option B: ~ $200–$250	3 Gb/~35–70%/>30	NGS Option A: ~6–9/45 h NGS Option B: ~6–9/35 h
PacBio WGS [34,104]	~ $100–$150	$900–$1600	45–90 Gb/>0.01%/>15–25	6–10/24–36 h
PacBio + hybrid capture (using 120nt commercial probes) [105]	~ $350–$500	$125–$200	5–10 Gb/40–60%/>30	30–40/24–36 h

Table 2. *Cont.*

	Preparation Price per Sample *	Sequencing Price per Sample	Sufficient Sequence Data (Gb)/On-Target Data (%)/On-Target Coverage (Reads)	Preparation/Run Time
Oxford Nanopore Technologies (ONT) WGS [15,106,107]	~ $100–$150	ONT Option A: MinION, 2–3 flow cells $1200–$2400 ONT Option B: PromethION (shared) $300–$600	60–90 Gb/>0.01%/20–30	ONT Option A: ~5–7/24–60 h ONT Option B: ~5–7/48–72 h
ONT + nCATs [26,96,108]	~ $160–$200	ONT Option A (1 flow cell): $600–$800 ONT Option B: $100–$150	30 Gb/10–40% ***/20–30	ONT Option A: ~7–10/24–60 h ONT Option B: ~7–10/48–72 h
ONT + internal cuts (AFIS-seq, CRISPR-LRS) [27,46,109]	$150–$200	ONT Option A (1 flow cell): $600–$800 ONT Option B: $100–$150	30 Gb/5–40% ***/>30	ONT Option A: ~7–10/24–60 h ONT Option B: ~7–10/48–72 h
Nanopore + Xdrop (commercial) [16,110,111]	$650–$900	ONT Option A (1 flow cell): $600–$800 ONT Option B: $100–$150	10 Gb/~60–90%/>30	ONT Option A: ~4–5 days/24–60 h ONT Option B: ~4–5 days/48–72 h

* The price includes NGS library preparation, along with quality and quantity control. For Sanger-based methods, the price includes enzymatic reactions and dideoxynucleotide triphosphates labeled with fluorescent dyes. ** Depends on multiple parameters related to probe quality. *** Depends on gRNA efficiency.

1.4. Long-Read Sequencing

In recent years, two independent platforms—PacBio (Pacific Biosciences) and Oxford Nanopore Technologies—have developed third-generation sequencing (TGS), also referred to as single-molecule sequencing (SMS) or LRS [82,112]. These technologies routinely produce reads in the 10–100 kb range and avoid PCR-associated artifacts. LRS has been successfully applied for genome polishing [107], sequencing of repetitive chromosome regions [113], and even for whole-genome assembly from single sandflies [114]. Novel applications include RNA isoform sequencing and epigenetic modifications measurements, combined with single-cell sequencing approaches [112,115,116]. Here, we focus on the use of LRS for transgene mapping and concatemer structure analysis.

The PacBio platform is based on single-molecule real-time (SMRT) sequencing. Fragmented DNA is ligated to single-stranded hairpin adapters from both sides, and a sequencing primer anneals to the hairpin region. Fluorescently labeled nucleotides allow base detection as docked polymerase molecule replicates the circularized DNA in a special well (SMRT cell). PacBio reads are typically limited to 25–30 kb, so that the circular consensus sequencing (CCS) strategy enables multiple polymerase passes over the same molecule, greatly increasing accuracy [117]. PacBio has been used for transgene mapping in mice [34,118] and plants [105], although it is less frequently chosen than Nanopore. When comparing the two platforms, the CCS mode of PacBio offers superior fidelity (>99%) compared to earlier generations of Nanopore sequencing (~90–95%) [112,117]. Moreover, Nanopore sequencing is particularly prone to errors in homopolymer regions [119]. However, the error rate is not of primal importance for transgene mapping, because long read length compensates for errors. The cost of both LRS platforms continues to fall and is now broadly comparable [82], depending on the specific instrument (Table 2). A comprehensive and critical comparison of the two LRS methods is provided in a recent review of Schell et al. [117].

Oxford Nanopore sequencing works by measuring ionic current changes as DNA moves through a biological nanopore embedded in a membrane [120]. This enables extremely long (megabase) reads, although average read lengths are typically similar to PacBio. Different authors casually report long reads around 200 kb [27], 238 kb [106], or 351 kb [121]. Occasionally such reads could contain transgenes and provide valuable insight into concatemer structure.

Below are selected examples to guide Nanopore-based experimental planning. Technology and chemistry improvements are ongoing, but for most transgene mapping experiments, a single MinION flow cell (typically R9 series) can produce 5–10 Gb of data—sufficient for a typical animal or plant transgenic line. In one early study, Nicholls et al. generated 4.88 Gb (1.8× haploid genome coverage) using a MinION run that yielded 611,279 reads with an N50 of 28 kb [15]. Among these, 25 reads contained transgene fragments, but only one 5.5 kb read spanned the genome–transgene junction within a 450 kb concatemer [15]. Suzuki et al. used a single MinION flow cell to sequence a transgenic mouse, obtaining 3 Gb of data (1× hgc; 922,210 reads; N50 = 7.6 kb). A 21.5 kb read covering one and a half copies of the transgene allowed successful integration mapping [106]. Another group investigated Cre-deleter mouse lines that failed to yield homozygotes in PCR screenings. TLA identified a 95 kb tandem duplication close to the floxed cassette in the gene of interest with unedited coding sequence. Three Nanopore runs produced 13 Gb (4.4× hgc; 699,343 reads; N50 = 40.7 kb), identifying 9 on-target reads and unambiguously resolving the rearrangement [25]. Giraldo et al. sequenced transgenic crops using one flow cell per sample and obtained 7.3–10.4 Gb with sufficient on-target coverage, though average read lengths varied from 1.6 to 12 kb [122]. In a soybean study, Li et al. generated 2.8 Gb (2.5× hgc; 1,061,117 reads) and found two reads spanning transgene–genome junctions. The results confirmed the site previously mapped by TAIL-PCR, highlighting the latter's cost-efficiency [121].

These examples illustrate that running a single MinION may yield only a few useful reads and become a costly endeavor as transgenes represent only ~0.01% of the genome. Enrichment strategies are often necessary when working with transgene mapping. In contrast to NGS-based enrichment methods, LRS approaches must preserve long DNA fragments. Two commonly used strategies—hybrid capture and Cas9 digestion—are compatible with LRS [82,123].

For PacBio, DNA is usually fragmented and size-selected to ~10–20 kb, while Nanopore sequencing often uses high-molecular-weight DNA [124]. Biotinylated probe enrichment for PacBio has been used to enrich symbiont genomes by 11–200× [125] or blood group system loci by 737× [124]. Biotin-based PacBio enrichment method, LIFE-seq, was introduced by Zhang et al. [105]. This method uses 75 nt tiling probes to cover known plasmid sequences (~99% coverage). Seven transgenic crop samples were enriched and sequenced, yielding 1.8–2.7 Gb per sample. On average, 17,000–25,000 unique CCS reads (average length ~6 kb, N50 ~17 kb) were obtained [105]. These data enabled mapping of insertion sites and partial concatemer reconstruction. Biotin enrichment was also applied to Nanopore sequencing. In the soybean study mentioned earlier, enrichment allowed identification of 51 transposon integration sites from a single Nanopore flow cell [121]. Although probe synthesis is costly and may reduce read length during sample preparation [125], this strategy avoids transgene fragmentation and does not require preservation of transgene ends. Other enrichment strategies for LRS include sonication-based inverse PCR (SIP) [30] and TLA-seq [126], although these are complex and less standardized than Cas9-based methods.

The CRISPR/Cas9 system has become a favored tool for target enrichment. In this approach, guide RNAs define cleavage points in the genome or transgene, producing

ligation-compatible ends. Though PacBio-compatible [127,128], most applications in transgene mapping use Nanopore. One widely adopted method is nCATS (Nanopore Cas9-Targeted Sequencing), where high-molecular-weight DNA is dephosphorylated, treated with Cas9–gRNA RNPs, and only the phosphorylated cut ends are ligated to Nanopore adapters [108]. nCATS method has become very popular for human diagnostics with enrichment of targeted regions of 25× [129], 665× [130], >100× [131]. Enrichment is especially useful for LRS in clinical samples with heterogenous cell populations or low target DNA quantity [119,132].

nCATS has been successfully applied to transgene mapping in various organisms [26,95,133]. Low et al. used nCATS to confirm site-specific integration of human ACE2 transgene into the Rosa26 locus via Bxb1-mediated recombination. With one flow cell they achieved 195× coverage of an 8.5 kb cassette [26]. In the same study, they sequenced mouse line with random multicopy integration of a similar transgene, and two 70–80 kb contigs were identified which contained the transgene-genome borders [26]. Other group compared nCATS, TLA, and Southern blotting to map transgene insertions in CHO cells [96]. For small transgenes (3–6 copies), nCATS produced contigs up to 41.6 kb from 22 reads and successfully resolved rearrangements. Notably, this allowed confirmation of peculiar Southern blot results obtained earlier—demonstrating the continuity of two mapping technologies [96]. nCATS is now supported by an official Nanopore protocol, but it requires prior knowledge of flanking sequences and would not be useful for initial transgene mapping.

Alternative Cas9 enrichment method is based on the same principle but DNA is digested inside the transgene region (Figure 2C). Funnily enough, this otherwise straightforward approach still lacks a definitive and concise acronym. The method is inconsistently named across publications and is referred to as "Targeted Cas9 sequencing" in the official Nanopore protocol—a term easily confused with nCATS, which, unlike this method, requires prior knowledge of the flanking sequences. For clarity, we propose a temporary name: CHAD (CRISPR-based Homing for Anchored Detection). Given how much scientists enjoy inventing acronyms—see the many creative efforts for TAIL-PCR modifications—it might be time to standardize the terminology, especially considering the growing popularity of the CHAD approach. One of the first applications of this strategy was AFIS-seq (Amplification-Free Integration Site sequencing), which mapped lentiviral integrations using paired Cas9 cuts inside the transgene. Enrichment ranged from 285–1612×, with average read lengths of ~12 kb [46]. In comparison to NGS-based S-EPTS/LM-PCR method, AFIS-seq provided fewer ambiguous reads thanks to longer sequencing length. McDonald et al. applied CHAD with a single cut to human samples to study mobile elements. One flow cell yielded ~110,000 reads, 31% of which were on-target (54× enrichment; N50 = 25 kb) [134]. Similarly, Hertel et al. used dual cuts flanking eGFP transgene in CHO cells, achieving 86–244× enrichment and revealing unplanned random integrations [135]. Bryant et al. used CRISPR-LRS with paired gRNAs to map several transgenes in mice. For a 217 kb BAC, 9 reads (0.03%) spanned transgene-genome borders [27]. With extra guides, enrichment improved to 0.15–0.35%. However, internal concatemer structure was lost due to Cas9 fragmentation: in the Sm22-Cre mouse line where qPCR detected ~20 copies, Nanopore only detected a max of 4 per read [27]. Ironically, WGS of the ultra-high molecular weight (HMW) DNA with Nanopore generated more useful detail in a few reads (6 selected reads, 89 kb average read length) than Cas9 enrichment due to the longer read sizes [27]. We also applied CHAD to a 5 kb hACE2 concatemer (~70 copies). Nanopore WGS (0.25× genome coverage) yielded 15 transgene reads, while CHAD produced 864 reads longer than 3 kb, mapping one border at the cost of losing internal concatemer structure [109]. We suspect that reads with the second transgene-

genome border were lost because we enriched with only one Cas9 site instead of two (Figure 2C). Importantly, Cas9 often blocks the protospacer adjacent motif (PAM)-distal end [136], hindering adapter ligation and reducing coverage in the respective direction by 2–10× [129,130,137]. Thermolabile Proteinase K treatment [137] or using Cpf1, which does not block ends, may help to improve nuclease-based targeting [138].

Finally, a novel Nanopore-compatible method, Xdrop, offers an original approach to target enrichment [110,139]. In this technique, the target locus is captured indirectly using a short PCR amplicon that is designed to lie within or near the region of interest, such as a transgene. HMW genomic DNA is mixed with PCR reagents and primers, and encapsulated in droplets using an oil emulsion system. During the droplet PCR fluorescence is triggered by an intercalating dye only in droplets that contain the specific target DNA. Typically, only about 0.01% of the double emulsion droplets will contain the desired fragment. These fluorescent droplets are then isolated via fluorescence-activated cell sorting (FACS) () and subjected to single-molecule multiple displacement amplification (dMDA) to amplify the enriched genomic DNA. The resulting product is then sequenced using the Nanopore platform [110,139].

Early publications have already demonstrated the successful use of this method to map transgenes in mice [110] and plants [111], as well as to detect complex genomic rearrangements in human cells [16,110]. These studies show that indirect targeting by droplet PCR provides very high enrichment levels (100× to 3000×) and enables detailed resolution of internal rearrangements, albeit at the cost of reduced average read length (around 5 kb) [110]. Given the technical complexity and specialized instrumentation involved, it is unlikely that Xdrop will be used routinely for mapping transgenes in animal models. However, one clear advantage is that indirect enrichment preserves the internal structure of concatemers, which is often lost in Cas9-based methods.

Ultimately, we would recommend the CHAD approach for most transgene mapping scenarios (Table 1). While a typical Nanopore run on a standard flow cell may yield only 3–5 reads per million reads covering a transgene border—sometimes with no guarantee of successful mapping—Cas9 enrichment offers a more targeted and controlled strategy, and it is not especially difficult to implement. One full run using this method requires a single flow cell (~$800) and a library prep kit (~$200), both of which can potentially be reused, making it cost-effective for many labs (Table 2). Unfortunately, CHAD destroys the internal concatemer structure unlike the original nCATS, where the cuts are introduced in the flanking sequences and preserve the concatemer structure, with up to 30–100 kp inside concatemer, which could be enough to assemble the whole insert, depending on the transgene size [96].

Compared to PCR-based techniques, there are few disadvantages to LRS, aside from the requirement for larger quantities of high-molecular-weight genomic DNA (in the microgram range), which is usually not a problem when working with animal tissue, but may be a problem with valuably founders or tiny model animals. At the same time, it's important to note that current Cas9 enrichment workflows generally lead to low sequencing coverage, making them unsuitable for applications requiring single-nucleotide resolution, such as precise indel detection or barcode identification.

2. Conclusions

Transgene mapping remains a critical yet technically diverse task, with no universal solution. In this review, we evaluated a range of available methods—from classic PCR-based genome walking to advanced enrichment protocols for NGS and LRS (Figure 3). For small-scale projects or initial screening, we recommend hiTAIL-PCR as a low-cost and accessible method. It requires minimal optimization and demonstrates high success rates,

especially when transgene ends are preserved. For reliable integration site search, WGS and TLA allow high-throughput mapping, though they could be costly and typically require access to sequencing facilities and bioinformatics support. When long-range information is essential, particularly in concatemer inserts or rearranged regions, Nanopore sequencing combined with Cas9-based enrichment (e.g., CHAD) is currently the most promising approach, because it enables sequencing of long DNA fragments for easy alignments. However, LRS methods are still evolving and can be technically demanding, with variable enrichment efficiency and sensitivity to DNA quality. Looking ahead, the future of transgene mapping is promising. Perhaps in five years emerging techniques such as adaptive sampling [82], AI-enhanced base calling [140], and real-time alignment filtering [141] will likely make LRS accessible and targeted to specific regions. This will signify the end of the old genome walking era, but until then we have to keep walking.
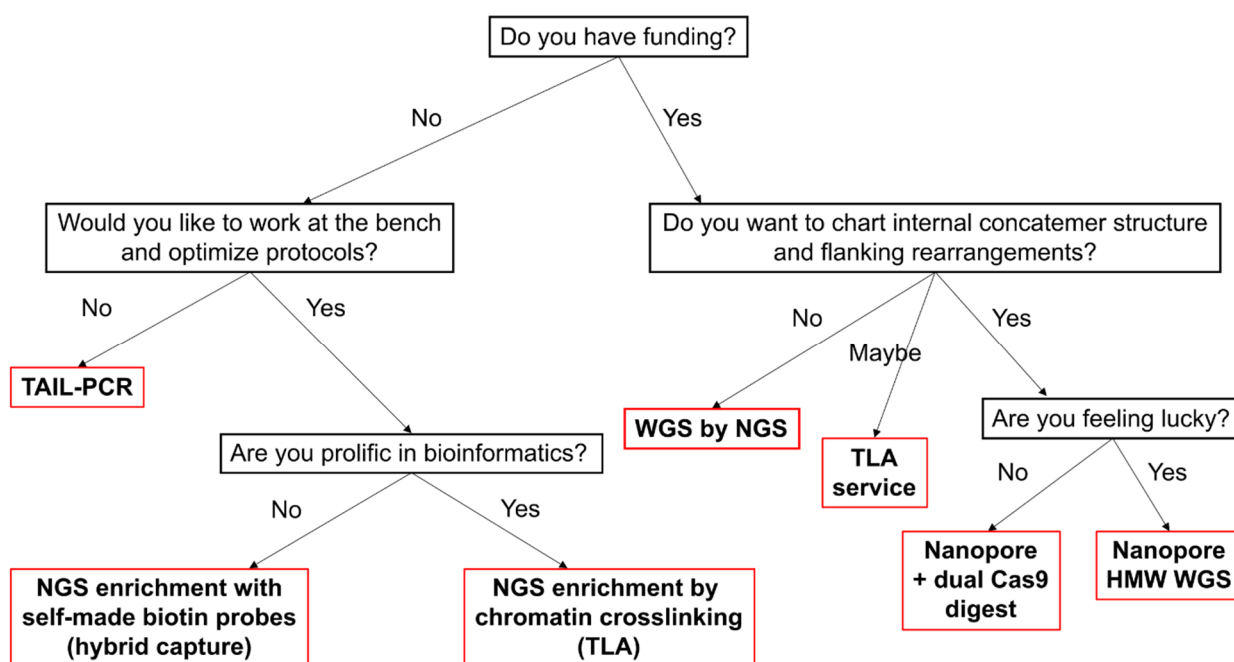
Figure 3. Decision tree for choosing a mapping method.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

NGS   Next-generation sequencing
LRS   Long-read sequencing
TLA   Targeted locus amplification
WGS   Whole-genome sequencing

References

1. Zhu, Z.; Lu, S.; Wang, H.; Wang, F.; Xu, W.; Zhu, Y.; Xue, J.; Yang, L. Innovations in Transgene Integration Analysis: A Comprehensive Review of Enrichment and Sequencing Strategies in Biotechnology. *ACS Appl. Mater. Interfaces* **2025**, *17*, 2716–2735. [CrossRef] [PubMed]
2. Kalendar, R.; Shustov, A.V.; Seppänen, M.M.; Schulman, A.H.; Stoddard, F.L. Palindromic Sequence-Targeted (PST) PCR: A Rapid and Efficient Method for High-Throughput Gene Characterization and Genome Walking. *Sci. Rep.* **2019**, *9*, 17707. [CrossRef]
3. Leoni, C.; Volpicella, M.; De Leo, F.; Gallerani, R.; Ceci, L.R. Genome Walking in Eukaryotes. *FEBS J.* **2011**, *278*, 3953–3977. [CrossRef] [PubMed]
4. Vandereyken, K.; Sifrim, A.; Thienpont, B.; Voet, T. Methods and Applications for Single-Cell and Spatial Multi-Omics. *Nat. Rev. Genet.* **2023**, *24*, 494–515. [CrossRef] [PubMed]
5. Nakanishi, T.; Kuroiwa, A.; Yamada, S.; Isotani, A.; Yamashita, A.; Tairaka, A.; Hayashi, T.; Takagi, T.; Ikawa, M.; Matsuda, Y.; et al. FISH Analysis of 142 EGFP Transgene Integration Sites into the Mouse Genome. *Genomics* **2002**, *80*, 564–574. [CrossRef]
6. Bandyopadhyay, A.A.; O'Brien, S.A.; Zhao, L.; Fu, H.; Vishwanathan, N.; Hu, W. Recurring Genomic Structural Variation Leads to Clonal Instability and Loss of Productivity. *Biotechnol. Bioeng.* **2019**, *116*, 41–53. [CrossRef]
7. Lee, J.S.; Kildegaard, H.F.; Lewis, N.E.; Lee, G.M. Mitigating Clonal Variation in Recombinant Mammalian Cell Lines. *Trends Biotechnol.* **2019**, *37*, 931–942. [CrossRef]
8. Dhiman, H.; Campbell, M.; Melcher, M.; Smith, K.D.; Borth, N. Predicting Favorable Landing Pads for Targeted Integrations in Chinese Hamster Ovary Cell Lines by Learning Stability Characteristics from Random Transgene Integrations. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 3632–3648. [CrossRef]
9. Cabrera, A.; Edelstein, H.I.; Glykofrydis, F.; Love, K.S.; Palacios, S.; Tycko, J.; Zhang, M.; Lensch, S.; Shields, C.E.; Livingston, M.; et al. The Sound of Silence: Transgene Silencing in Mammalian Cell Engineering. *Cell Syst.* **2022**, *13*, 950–973. [CrossRef]
10. Laboulaye, M.A.; Duan, X.; Qiao, M.; Whitney, I.E.; Sanes, J.R. Mapping Transgene Insertion Sites Reveals Complex Interactions Between Mouse Transgenes and Neighboring Endogenous Genes. *Front. Mol. Neurosci.* **2018**, *11*, 385. [CrossRef]
11. Yan, B.-W.; Zhao, Y.-F.; Cao, W.-G.; Li, N.; Gou, K.-M. Mechanism of Random Integration of Foreign DNA in Transgenic Mice. *Transgenic Res.* **2013**, *22*, 983–992. [CrossRef] [PubMed]
12. Goodwin, L.O.; Splinter, E.; Davis, T.L.; Urban, R.; He, H.; Braun, R.E.; Chesler, E.J.; Kumar, V.; Van Min, M.; Ndukum, J.; et al. Large-Scale Discovery of Mouse Transgenic Integration Sites Reveals Frequent Structural Variation and Insertional Mutagenesis. *Genome Res.* **2019**, *29*, 494–505. [CrossRef]
13. Halurkar, M.S.; Inoue, O.; Singh, A.; Mukherjee, R.; Ginugu, M.; Ahn, C.; Bonatto Paese, C.L.; Duszynski, M.; Brugmann, S.A.; Lim, H.-W.; et al. The Widely Used Ucp1-Cre Transgene Elicits Complex Developmental and Metabolic Phenotypes. *Nat. Commun.* **2025**, *16*, 770. [CrossRef]
14. Cain-Hom, C.; Splinter, E.; van Min, M.; Simonis, M.; van de Heijning, M.; Martinez, M.; Asghari, V.; Cox, J.C.; Warming, S. Efficient Mapping of Transgene Integration Sites and Local Structural Changes in Cre Transgenic Mice Using Targeted Locus Amplification. *Nucleic Acids Res.* **2017**, *45*, e62. [CrossRef]
15. Nicholls, P.K.; Bellott, D.W.; Cho, T.-J.; Pyntikova, T.; Page, D.C. Locating and Characterizing a Transgene Integration Site by Nanopore Sequencing. *G3 Genes|Genomes|Genet.* **2019**, *9*, 1481–1486. [CrossRef]
16. Geng, K.; Merino, L.G.; Wedemann, L.; Martens, A.; Sobota, M.; Sanchez, Y.P.; Søndergaard, J.N.; White, R.J.; Kutter, C. Target-Enriched Nanopore Sequencing and de Novo Assembly Reveals Co-Occurrences of Complex on-Target Genomic Rearrangements Induced by CRISPR-Cas9 in Human Cells. *Genome Res.* **2022**, *32*, 1876–1891. [CrossRef]
17. Giannoukos, G.; Ciulla, D.M.; Marco, E.; Abdulkerim, H.S.; Barrera, L.A.; Bothmer, A.; Dhanapal, V.; Gloskowski, S.W.; Jayaram, H.; Maeder, M.L.; et al. UDiTaS™, a Genome Editing Detection Method for Indels and Genome Rearrangements. *BMC Genom.* **2018**, *19*, 212. [CrossRef] [PubMed]
18. Bi, C.; Yuan, B.; Zhang, Y.; Wang, M.; Tian, Y.; Li, M. Prevalent Integration of Genomic Repetitive and Regulatory Elements and Donor Sequences at CRISPR-Cas9-Induced Breaks. *Commun. Biol.* **2025**, *8*, 94. [CrossRef] [PubMed]
19. Guirouilh-Barbat, J.; Lambert, S.; Bertrand, P.; Lopez, B.S. Is Homologous Recombination Really an Error-Free Process? *Front. Genet.* **2014**, *5*, 175. [CrossRef]

20. Smirnov, A.; Battulin, N. Concatenation of Transgenic DNA: Random or Orchestrated? *Genes* **2021**, *12*, 1969. [CrossRef]

21. Norris, A.L.; Lee, S.S.; Greenlees, K.J.; Tadesse, D.A.; Miller, M.F.; Lombardi, H.A. Template Plasmid Integration in Germline Genome-Edited Cattle. *Nat. Biotechnol.* **2020**, *38*, 163–164. [CrossRef] [PubMed]

22. Chiang, C.; Jacobsen, J.C.; Ernst, C.; Hanscom, C.; Heilbut, A.; Blumenthal, I.; Mills, R.E.; Kirby, A.; Lindgren, A.M.; Rudiger, S.R.; et al. Complex Reorganization and Predominant Non-Homologous Repair Following Chromosomal Breakage in Karyotypically Balanced Germline Rearrangements and Transgenic Integration. *Nat. Genet.* **2012**, *44*, 390–397. [CrossRef] [PubMed]

23. Hussmann, J.A.; Ling, J.; Ravisankar, P.; Yan, J.; Cirincione, A.; Xu, A.; Simpson, D.; Yang, D.; Bothmer, A.; Cotta-Ramusino, C.; et al. Mapping the Genetic Landscape of DNA Double-Strand Break Repair. *Cell* **2021**, *184*, 5653–5669.e25. [CrossRef] [PubMed]

24. Ohigashi, I.; Yamasaki, Y.; Hirashima, T.; Takahama, Y. Identification of the Transgenic Integration Site in Immunodeficient Tgε26 Human CD3ε Transgenic Mice. *PLoS ONE* **2010**, *5*, e14391. [CrossRef]

25. Sailer, S.; Coassin, S.; Lackner, K.; Fischer, C.; McNeill, E.; Streiter, G.; Kremser, C.; Maglione, M.; Green, C.M.; Moralli, D.; et al. When the Genome Bluffs: A Tandem Duplication Event during Generation of a Novel Agmo Knockout Mouse Model Fools Routine Genotyping. *Cell Biosci.* **2021**, *11*, 54. [CrossRef]

26. Low, B.E.; Hosur, V.; Lesbirel, S.; Wiles, M.V. Efficient Targeted Transgenesis of Large Donor DNA into Multiple Mouse Genetic Backgrounds Using Bacteriophage Bxb1 Integrase. *Sci. Rep.* **2022**, *12*, 5424. [CrossRef]

27. Bryant, W.B.; Yang, A.; Griffin, S.H.; Zhang, W.; Rafiq, A.M.; Han, W.; Deak, F.; Mills, M.K.; Long, X.; Miano, J.M. CRISPR-Cas9 Long-Read Sequencing for Mapping Transgenes in the Mouse Genome. *CRISPR J.* **2023**, *6*, 163–175. [CrossRef]

28. Kalendar, R.; Shustov, A.V.; Schulman, A.H. Palindromic Sequence-Targeted (PST) PCR, Version 2: An Advanced Method for High-Throughput Targeted Gene Characterization and Transposon Display. *Front. Plant Sci.* **2021**, *12*, 691940. [CrossRef]

29. Hamada, M.; Nishio, N.; Okuno, Y.; Suzuki, S.; Kawashima, N.; Muramatsu, H.; Tsubota, S.; Wilson, M.H.; Morita, D.; Kataoka, S.; et al. Integration Mapping of piggyBac-Mediated CD19 Chimeric Antigen Receptor T Cells Analyzed by Novel Tagmentation-Assisted PCR. *EBioMedicine* **2018**, *34*, 18–26. [CrossRef]

30. Alquezar-Planas, D.E.; Löber, U.; Cui, P.; Quedenau, C.; Chen, W.; Greenwood, A.D. DNA Sonication Inverse PCR for Genome Scale Analysis of Uncharacterized Flanking Sequences. *Methods Ecol. Evol.* **2021**, *12*, 182–195. [CrossRef]

31. Triglia, T.; Peterson, M.G.; Kemp, D.J. A Procedure for in Vitro Amplification of DNA Segments That Lie Outside the Boundaries of Known Sequences. *Nucleic Acids Res.* **1988**, *16*, 8186. [CrossRef]

32. Ochman, H.; Gerber, A.S.; Hartl, D.L. Genetic Applications of an Inverse Polymerase Chain Reaction. *Genetics* **1988**, *120*, 621–623. [CrossRef] [PubMed]

33. Schep, R.; Leemans, C.; Brinkman, E.K.; Van Schaik, T.; Van Steensel, B. Protocol: A Multiplexed Reporter Assay to Study Effects of Chromatin Context on DNA Double-Strand Break Repair. *Front. Genet.* **2022**, *12*, 785947. [CrossRef]

34. Smirnov, A.; Fishman, V.; Yunusova, A.; Korablev, A.; Serova, I.; Skryabin, B.V.; Rozhdestvensky, T.S.; Battulin, N. DNA Barcoding Reveals That Injected Transgenes Are Predominantly Processed by Homologous Recombination in Mouse Zygote. *Nucleic Acids Res.* **2019**, *48*, 719–735. [CrossRef] [PubMed]

35. O'Malley, R.C.; Alonso, J.M.; Kim, C.J.; Leisse, T.J.; Ecker, J.R. An Adapter Ligation-Mediated PCR Method for High-Throughput Mapping of T-DNA Inserts in the Arabidopsis Genome. *Nat. Protoc.* **2007**, *2*, 2910–2917. [CrossRef]

36. Yu, D.; Zhou, T.; Sun, X.; Sun, Z.; Sheng, X.; Tan, Y.; Liu, L.; Ouyang, N.; Xu, K.; Shi, K.; et al. Cyclic Digestion and Ligation-Mediated PCR Used for Flanking Sequence Walking. *Sci. Rep.* **2020**, *10*, 3434. [CrossRef] [PubMed]

37. Lung, J.; Hung, M.-S.; Chen, C.-Y.; Yang, T.-M.; Lin, C.-K.; Fang, Y.-H.; Jiang, Y.-Y.; Liao, H.-F.; Lin, Y.-C. An Optimized Ligation-Mediated PCR Method for Chromosome Walking and Fusion Gene Chromosomal Breakpoints Identification. *Biol. Methods Protoc.* **2024**, *9*, bpae037. [CrossRef]

38. Uren, A.G.; Mikkers, H.; Kool, J.; Van Der Weyden, L.; Lund, A.H.; Wilson, C.H.; Rance, R.; Jonkers, J.; Van Lohuizen, M.; Berns, A.; et al. A High-Throughput Splinkerette-PCR Method for the Isolation and Sequencing of Retroviral Insertion Sites. *Nat. Protoc.* **2009**, *4*, 789–798. [CrossRef]

39. Potter, C.J.; Luo, L. Splinkerette PCR for Mapping Transposable Elements in Drosophila. *PLoS ONE* **2010**, *5*, e10168. [CrossRef]

40. Dambrot, C.; Buermans, H.P.J.; Varga, E.; Kosmidis, G.; Langenberg, K.; Casini, S.; Elliott, D.A.; Dinnyes, A.; Atsma, D.E.; Mummery, C.L.; et al. Strategies for Rapidly Mapping Proviral Integration Sites and Assessing Cardiogenic Potential of Nascent Human Induced Pluripotent Stem Cell Clones. *Exp. Cell Res.* **2014**, *327*, 297–306. [CrossRef]

41. Jia, W.; Guan, Z.; Shi, S.; Xiang, K.; Chen, P.; Tan, F.; Ullah, N.; Diaby, M.; Guo, M.; Song, C.; et al. The Annotation of Zebrafish Enhancer Trap Lines Generated with PB Transposon. *Curr. Issues Mol. Biol.* **2022**, *44*, 2614–2621. [CrossRef] [PubMed]

42. Sato, M.; Inada, E.; Saitoh, I.; Nakamura, S.; Watanabe, S. In Vivo Piggybac-Based Gene Delivery towards Murine Pancreatic Parenchyma Confers Sustained Expression of Gene of Interest. *Int. J. Mol. Sci.* **2019**, *20*, 3116. [CrossRef] [PubMed]

43. Han, H.-J.; Kim, D.H.; Baik, J.Y. A Splinkerette PCR-Based Genome Walking Technique for the Identification of Transgene Integration Sites in CHO Cells. *J. Biotechnol.* **2023**, *371–372*, 1–9. [CrossRef]

44. Schmidt, M.; Schwarzwaelder, K.; Bartholomae, C.; Zaoui, K.; Ball, C.; Pilz, I.; Braun, S.; Glimm, H.; Von Kalle, C. High-Resolution Insertion-Site Analysis by Linear Amplification–Mediated PCR (LAM-PCR). *Nat. Methods* **2007**, *4*, 1051–1057. [CrossRef]

45. Gabriel, R.; Eckenberg, R.; Paruzynski, A.; Bartholomae, C.C.; Nowrouzi, A.; Arens, A.; Howe, S.J.; Recchia, A.; Cattoglio, C.; Wang, W.; et al. Comprehensive Genomic Access to Vector Integration in Clinical Gene Therapy. *Nat. Med.* **2009**, *15*, 1431–1436. [CrossRef] [PubMed]

46. van Haasteren, J.; Munis, A.M.; Gill, D.R.; Hyde, S.C. Genome-Wide Integration Site Detection Using Cas9 Enriched Amplification-Free Long-Range Sequencing. *Nucleic Acids Res.* **2021**, *49*, e16. [CrossRef]

47. Singer, T.; Burke, E. High-Throughput TAIL-PCR as a Tool to Identify DNA Flanking Insertions. In *Plant Functional Genomics*; Humana Press: Totowa, NJ, USA, 2003; Volume 236, pp. 241–272. ISBN 978-1-59259-413-9.

48. Liu, Y.-G.; Whittier, R.F. Thermal Asymmetric Interlaced PCR: Automatable Amplification and Sequencing of Insert End Fragments from P1 and YAC Clones for Chromosome Walking. *Genomics* **1995**, *25*, 674–681. [CrossRef]

49. Liu, Y.-G.; Chen, Y. High-Efficiency Thermal Asymmetric Interlaced PCR for Amplification of Unknown Flanking Sequences. *BioTechniques* **2007**, *43*, 649–656. [CrossRef] [PubMed]

50. Zhang, H.; Xu, W.; Feng, Z.; Hong, Z. A Low Degenerate Primer Pool Improved the Efficiency of High-Efficiency Thermal Asymmetric Interlaced PCR to Amplify T-DNA Flanking Sequences in Arabidopsis Thaliana. *3 Biotech* **2018**, *8*, 14. [CrossRef]

51. Wu, L.; Di, D.-W.; Zhang, D.; Song, B.; Luo, P.; Guo, G.-Q. Frequent Problems and Their Resolutions by Using Thermal Asymmetric Interlaced PCR (TAIL-PCR) to Clone Genes in *Arabidopsis* T-DNA Tagged Mutants. *Biotechnol. Biotechnol. Equip.* **2015**, *29*, 260–267. [CrossRef]

52. Jia, X.; Lin, X.; Chen, J. Linear and Exponential TAIL-PCR: A Method for Efficient and Quick Amplification of Flanking Sequences Adjacent to Tn5 Transposon Insertion Sites. *AMB Expr.* **2017**, *7*, 195. [CrossRef] [PubMed]

53. Luo, W.; Li, Z.; Huang, Y.; Han, Y.; Yao, C.; Duan, X.; Ouyang, H.; Li, L. Generation of AQP2-Cre Transgenic Mini-Pigs Specifically Expressing Cre Recombinase in Kidney Collecting Duct Cells. *Transgenic Res.* **2014**, *23*, 365–375. [CrossRef] [PubMed]

54. Zhang, R.; Yin, Y.; Zhang, Y.; Li, K.; Zhu, H.; Gong, Q.; Wang, J.; Hu, X.; Li, N. Molecular Characterization of Transgene Integration by Next-Generation Sequencing in Transgenic Cattle. *PLoS ONE* **2012**, *7*, e50348. [CrossRef]

55. Zelensky, A.N.; Schimmel, J.; Kool, H.; Kanaar, R.; Tijsterman, M. Inactivation of Pol θ and C-NHEJ Eliminates off-Target Integration of Exogenous DNA. *Nat. Commun.* **2017**, *8*, 66. [CrossRef]

56. Kondrychyn, I.; Garcia-Lecea, M.; Emelyanov, A.; Parinov, S.; Korzh, V. Genome-Wide Analysis of Tol2 Transposon Reintegration in Zebrafish. *BMC Genom.* **2009**, *10*, 418. [CrossRef]

57. Johansson, O.N.; Töpel, M.; Pinder, M.I.M.; Kourtchenko, O.; Blomberg, A.; Godhe, A.; Clarke, A.K. Skeletonema Marinoi as a New Genetic Model for Marine Chain-Forming Diatoms. *Sci. Rep.* **2019**, *9*, 5391. [CrossRef] [PubMed]

58. Gong, W.; Zhou, Y.; Wang, R.; Wei, X.; Zhang, L.; Dai, Y.; Zhu, Z. Analysis of T-DNA Integration Events in Transgenic Rice. *J. Plant Physiol.* **2021**, *266*, 153527. [CrossRef]

59. Wang, L.; Jia, M.; Li, Z.; Liu, X.; Sun, T.; Pei, J.; Wei, C.; Lin, Z.; Li, H. Wristwatch PCR: A Versatile and Efficient Genome Walking Strategy. *Front. Bioeng. Biotechnol.* **2022**, *10*, 792848. [CrossRef]

60. Pan, H.; Guo, X.; Pan, Z.; Wang, R.; Tian, B.; Li, H. Fork PCR: A Universal and Efficient Genome-Walking Tool. *Front. Microbiol.* **2023**, *14*, 1265580. [CrossRef]

61. Li, H.; Lin, Z.; Guo, X.; Pan, Z.; Pan, H.; Wang, D. Primer Extension Refractory PCR: An Efficient and Reliable Genome Walking Method. *Mol. Genet. Genom.* **2024**, *299*, 27. [CrossRef]

62. Burkov, I.A.; Serova, I.A.; Battulin, N.R.; Smirnov, A.V.; Babkin, I.V.; Andreeva, L.E.; Dvoryanchikov, G.A.; Serov, O.L. Expression of the Human Granulocyte–Macrophage Colony Stimulating Factor (hGM-CSF) Gene under Control of the 5′-Regulatory Sequence of the Goat Alpha-S1-Casein Gene with and without a MAR Element in Transgenic Mice. *Transgenic Res.* **2013**, *22*, 949–964. [CrossRef] [PubMed]

63. Serova, I.A.; Dvoryanchikov, G.A.; Andreeva, L.E.; Burkov, I.A.; Dias, L.P.B.; Battulin, N.R.; Smirnov, A.V.; Serov, O.L. A 3,387 Bp 5′-Flanking Sequence of the Goat Alpha-S1-Casein Gene Provides Correct Tissue-Specific Expression of Human Granulocyte Colony-Stimulating Factor (hG-CSF) in the Mammary Gland of Transgenic Mice. *Transgenic Res.* **2012**, *21*, 485–498. [CrossRef] [PubMed]

64. Smirnov, A.V.; Kontsevaya, G.V.; Feofanova, N.A.; Anisimova, M.V.; Serova, I.A.; Gerlinskaya, L.A.; Battulin, N.R.; Moshkin, M.P.; Serov, O.L. Unexpected Phenotypic Effects of a Transgene Integration Causing a Knockout of the Endogenous Contactin-5 Gene in Mice. *Transgenic Res.* **2018**, *27*, 1–13. [CrossRef]

65. Le Saux, A.; Houdebine, L.-M.; Jolivet, G. Chromosome Integration of BAC (Bacterial Artificial Chromosome): Evidence of Multiple Rearrangements. *Transgenic Res.* **2010**, *19*, 923–931. [CrossRef]

66. Won, M.; Dawid, I.B. PCR Artifact in Testing for Homologous Recombination in Genomic Editing in Zebrafish. *PLoS ONE* **2017**, *12*, e0172802. [CrossRef]

67. Pillai, M.M.; Venkataraman, G.M.; Kosak, S.; Torok-Storb, B. Integration Site Analysis in Transgenic Mice by Thermal Asymmetric Interlaced (TAIL)-PCR: Segregating Multiple-Integrant Founder Lines and Determining Zygosity. *Transgenic Res.* **2008**, *17*, 749–754. [CrossRef]

68. Brlek, P.; Bulić, L.; Bračić, M.; Projić, P.; Škaro, V.; Shah, N.; Shah, P.; Primorac, D. Implementing Whole Genome Sequencing (WGS) in Clinical Practice: Advantages, Challenges, and Future Perspectives. *Cells* **2024**, *13*, 504. [CrossRef]

69. Giani, A.M.; Gallo, G.R.; Gianfranceschi, L.; Formenti, G. Long Walk to Genomics: History and Current Approaches to Genome Sequencing and Assembly. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 9–19. [CrossRef] [PubMed]

70. Ji, Y.; Abrams, N.; Zhu, W.; Salinas, E.; Yu, Z.; Palmer, D.C.; Jailwala, P.; Franco, Z.; Roychoudhuri, R.; Stahlberg, E.; et al. Identification of the Genomic Insertion Site of Pmel-1 TCR α and β Transgenes by Next-Generation Sequencing. *PLoS ONE* **2014**, *9*, e96650. [CrossRef]

71. Yong, C.S.M.; Sharkey, J.; Duscio, B.; Venville, B.; Wei, W.-Z.; Jones, R.F.; Slaney, C.Y.; Mir Arnau, G.; Papenfuss, A.T.; Schröder, J.; et al. Embryonic Lethality in Homozygous Human Her-2 Transgenic Mice Due to Disruption of the Pds5b Gene. *PLoS ONE* **2015**, *10*, e0136817. [CrossRef]

72. Srivastava, S.K.; Wolinski, P.; Pereira, A. A Strategy for Genome-Wide Identification of Gene Based Polymorphisms in Rice Reveals Non-Synonymous Variation and Functional Genotypic Markers. *PLoS ONE* **2014**, *9*, e105335. [CrossRef] [PubMed]

73. Owen, J.R.; Hennig, S.L.; McNabb, B.R.; Mansour, T.A.; Smith, J.M.; Lin, J.C.; Young, A.E.; Trott, J.F.; Murray, J.D.; Delany, M.E.; et al. One-Step Generation of a Targeted Knock-in Calf Using the CRISPR-Cas9 System in Bovine Zygotes. *BMC Genom.* **2021**, *22*, 118. [CrossRef] [PubMed]

74. Carlson, D.F.; Lancto, C.A.; Zang, B.; Kim, E.-S.; Walton, M.; Oldeschulte, D.; Seabury, C.; Sonstegard, T.S.; Fahrenkrug, S.C. Production of Hornless Dairy Cattle from Genome-Edited Cell Lines. *Nat. Biotechnol.* **2016**, *34*, 479–481. [CrossRef]

75. Young, A.E.; Mansour, T.A.; McNabb, B.R.; Owen, J.R.; Trott, J.F.; Brown, C.T.; Van Eenennaam, A.L. Genomic and Phenotypic Analyses of Six Offspring of a Genome-Edited Hornless Bull. *Nat. Biotechnol.* **2020**, *38*, 225–232. [CrossRef]

76. Niu, L.; He, H.; Zhang, Y.; Yang, J.; Zhao, Q.; Xing, G.; Zhong, X.; Yang, X. Efficient Identification of Genomic Insertions and Flanking Regions through Whole-Genome Sequencing in Three Transgenic Soybean Events. *Transgenic Res.* **2021**, *30*, 1–9. [CrossRef]

77. Guo, B.; Guo, Y.; Hong, H.; Qiu, L.-J. Identification of Genomic Insertion and Flanking Sequence of G2-EPSPS and GAT Transgenes in Soybean Using Whole Genome Sequencing Method. *Front. Plant Sci.* **2016**, *7*, 1009. [CrossRef]

78. Xu, W.; Zhang, H.; Zhang, Y.; Shen, P.; Li, X.; Li, R.; Yang, L. A Paired-End Whole-Genome Sequencing Approach Enables Comprehensive Characterization of Transgene Integration in Rice. *Commun. Biol.* **2022**, *5*, 667. [CrossRef]

79. Kovalic, D.; Garnaat, C.; Guo, L.; Yan, Y.; Groat, J.; Silvanovich, A.; Ralston, L.; Huang, M.; Tian, Q.; Christian, A.; et al. The Use of Next Generation Sequencing and Junction Sequence Analysis Bioinformatics to Achieve Molecular Characterization of Crops Improved Through Modern Biotechnology. *Plant Genome* **2012**, *5*, 149–163. [CrossRef]

80. De Vree, P.J.P.; De Wit, E.; Yilmaz, M.; Van De Heijning, M.; Klous, P.; Verstegen, M.J.A.M.; Wan, Y.; Teunissen, H.; Krijger, P.H.L.; Geeven, G.; et al. Targeted Sequencing by Proximity Ligation for Comprehensive Variant Detection and Local Haplotyping. *Nat. Biotechnol.* **2014**, *32*, 1019–1025. [CrossRef]

81. Gilpatrick, T.; Wang, J.Z.; Weiss, D.; Norris, A.L.; Eshleman, J.; Timp, W. IVT Generation of guideRNAs for Cas9-Enrichment Nanopore Sequencing. *bioRxiv* **2023**. [CrossRef]

82. Hook, P.W.; Timp, W. Beyond Assembly: The Increasing Flexibility of Single-Molecule Sequencing Technology. *Nat. Rev. Genet.* **2023**, *24*, 627–641. [CrossRef] [PubMed]

83. Volpicella, M.; Leoni, C.; Costanza, A.; Fanizza, I.; Placido, A.; Ceci, L.R. Genome Walking by Next Generation Sequencing Approaches. *Biology* **2012**, *1*, 495–507. [CrossRef] [PubMed]

84. Zhao, S.; Wang, Y.; Zhu, Z.; Chen, P.; Liu, W.; Wang, C.; Lu, H.; Xiang, Y.; Liu, Y.; Qian, Q.; et al. Streamlined Whole-Genome Genotyping through NGS-Enhanced Thermal Asymmetric Interlaced (TAIL)-PCR. *Plant Commun.* **2024**, *5*, 100983. [CrossRef]

85. Salnikov, P.A.; Khabarova, A.A.; Koksharova, G.S.; Mungalov, R.V.; Belokopytova, P.S.; Pristyazhnuk, I.E.; Nurislamov, A.R.; Somatich, P.; Gridina, M.M.; Fishman, V.S. Here and There: The Double-Side Transgene Localization. *Vavilov J. Genet. Breed.* **2021**, *25*, 607–612. [CrossRef]

86. Malekshoar, M.; Azimi, S.A.; Kaki, A.; Mousazadeh, L.; Motaei, J.; Vatankhah, M. CRISPR-Cas9 Targeted Enrichment and Next-Generation Sequencing for Mutation Detection. *J. Mol. Diagn.* **2023**, *25*, 249–262. [CrossRef]

87. Singh, R.R. Target Enrichment Approaches for Next-Generation Sequencing Applications in Oncology. *Diagnostics* **2022**, *12*, 1539. [CrossRef] [PubMed]

88. Wang, G.; Zhang, C.; Kambara, H.; Dambrot, C.; Xie, X.; Zhao, L.; Xu, R.; Oneglia, A.; Liu, F.; Luo, H.R. Identification of the Transgene Integration Site and Host Genome Changes in MRP8-Cre/Ires-EGFP Transgenic Mice by Targeted Locus Amplification. *Front. Immunol.* **2022**, *13*, 875991. [CrossRef] [PubMed]

89. Stadermann, A.; Gamer, M.; Fieder, J.; Lindner, B.; Fehrmann, S.; Schmidt, M.; Schulz, P.; Gorr, I.H. Structural Analysis of Random Transgene Integration in CHO Manufacturing Cell Lines by Targeted Sequencing. *Biotechnol. Bioeng.* **2022**, *119*, 868–880. [CrossRef]

90. Lefferts, J.W.; Boersma, V.; Hagemeijer, M.C.; Hajo, K.; Beekman, J.M.; Splinter, E. Targeted Locus Amplification and Haplotyping. In *Haplotyping*; Peters, B.A., Drmanac, R., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2023; Volume 2590, pp. 31–48. ISBN 978-1-07-162818-8.

91. Tosh, J.L.; Rickman, M.; Rhymes, E.; Norona, F.E.; Clayton, E.; Mucke, L.; Isaacs, A.M.; Fisher, E.M.C.; Wiseman, F.K. The Integration Site of the APP Transgene in the J20 Mouse Model of Alzheimer's Disease. *Wellcome Open Res.* **2018**, *2*, 84. [CrossRef]

92. Hinteregger, B.; Loeffler, T.; Flunkert, S.; Neddens, J.; Birner-Gruenberger, R.; Bayer, T.A.; Madl, T.; Hutter-Paier, B. Transgene Integration Causes RARB Downregulation in Homozygous Tg4–42 Mice. *Sci. Rep.* **2020**, *10*, 6377. [CrossRef]

93. Wong, A.M.; Patel, T.P.; Altman, E.K.; Tugarinov, N.; Trivellin, G.; Yanovski, J.A. Characterization of the Adiponectin Promoter + Cre Recombinase Insertion in the Tg(Adipoq-Cre)1Evdr Mouse by Targeted Locus Amplification and Droplet Digital PCR. *Adipocyte* **2021**, *10*, 21–27. [CrossRef]

94. Fan, Y.; Chen, W.; Wei, R.; Qiang, W.; Pearson, J.D.; Yu, T.; Bremner, R.; Chen, D. Mapping Transgene Insertion Sites Reveals the α-Cre Transgene Expression in Both Developing Retina and Olfactory Neurons. *Commun. Biol.* **2022**, *5*, 411. [CrossRef] [PubMed]

95. Leitner, K.; Motheramgari, K.; Borth, N.; Marx, N. Nanopore Cas9-targeted Sequencing Enables Accurate and Simultaneous Identification of Transgene Integration Sites, Their Structure and Epigenetic Status in Recombinant Chinese Hamster Ovary Cells. *Biotechnol. Bioeng.* **2023**, *120*, 2403–2418. [CrossRef] [PubMed]

96. Clappier, C.; Böttner, D.; Heinzelmann, D.; Stadermann, A.; Schulz, P.; Schmidt, M.; Lindner, B. Deciphering Integration Loci of CHO Manufacturing Cell Lines Using Long Read Nanopore Sequencing. *New Biotechnol.* **2023**, *75*, 31–39. [CrossRef]

97. DuBose, A.J.; Lichtenstein, S.T.; Narisu, N.; Bonnycastle, L.L.; Swift, A.J.; Chines, P.S.; Collins, F.S. Use of Microarray Hybrid Capture and Next-Generation Sequencing to Identify the Anatomy of a Transgene. *Nucleic Acids Res.* **2013**, *41*, e70. [CrossRef] [PubMed]

98. Magembe, E.M.; Li, H.; Taheri, A.; Zhou, S.; Ghislain, M. Identification of T-DNA Structure and Insertion Site in Transgenic Crops Using Targeted Capture Sequencing. *Front. Plant Sci.* **2023**, *14*, 1156665. [CrossRef] [PubMed]

99. Ohnuki, N.; Kobayashi, T.; Matsuo, M.; Nishikaku, K.; Kusama, K.; Torii, Y.; Inagaki, Y.; Hori, M.; Imakawa, K.; Satou, Y. A Target Enrichment High Throughput Sequencing System for Characterization of BLV Whole Genome Sequence, Integration Sites, Clonality and Host SNP. *Sci. Rep.* **2021**, *11*, 4521. [CrossRef]

100. Iwase, S.C.; Miyazato, P.; Katsuya, H.; Islam, S.; Yang, B.T.J.; Ito, J.; Matsuo, M.; Takeuchi, H.; Ishida, T.; Matsuda, K.; et al. HIV-1 DNA-Capture-Seq Is a Useful Tool for the Comprehensive Characterization of HIV-1 Provirus. *Sci. Rep.* **2019**, *9*, 12326. [CrossRef]

101. Li, X.; Chen, W.; Martin, B.K.; Calderon, D.; Lee, C.; Choi, J.; Chardon, F.M.; McDiarmid, T.A.; Daza, R.M.; Kim, H.; et al. Chromatin Context-Dependent Regulation and Epigenetic Manipulation of Prime Editing. *Cell* **2024**, *187*, 2411–2427.e25. [CrossRef]

102. Siddique, K.; Wei, J.; Li, R.; Zhang, D.; Shi, J. Identification of T-DNA Insertion Site and Flanking Sequence of a Genetically Modified Maize Event IE09S034 Using Next-Generation Sequencing Technology. *Mol. Biotechnol.* **2019**, *61*, 694–702. [CrossRef]

103. Peng, C.; Mei, Y.; Ding, L.; Wang, X.; Chen, X.; Wang, J.; Xu, J. Using Combined Methods of Genetic Mapping and Nanopore-Based Sequencing Technology to Analyze the Insertion Positions of G10evo-EPSPS and Cry1Ab/Cry2Aj Transgenes in Maize. *Front. Plant Sci.* **2021**, *12*, 690951. [CrossRef] [PubMed]

104. Sheehan, M.; Kumpf, S.W.; Qian, J.; Rubitski, D.M.; Oziolor, E.; Lanz, T.A. Comparison and Cross-Validation of Long-Read and Short-Read Target-Enrichment Sequencing Methods to Assess AAV Vector Integration into Host Genome. *Mol. Ther. Methods Clin. Dev.* **2024**, *32*, 101352. [CrossRef] [PubMed]

105. Zhang, H.; Li, R.; Guo, Y.; Zhang, Y.; Zhang, D.; Yang, L. LIFE-Seq: A Universal Large Integrated DNA Fragment Enrichment Sequencing Strategy for Deciphering the Transgene Integration of Genetically Modified Organisms. *Plant Biotechnol. J.* **2022**, *20*, 964–976. [CrossRef] [PubMed]

106. Suzuki, O.; Koura, M.; Uchio-Yamada, K.; Sasaki, M. Analysis of the Transgene Insertion Pattern in a Transgenic Mouse Strain Using Long-Read Sequencing. *Exp. Anim.* **2020**, *69*, 279–286. [CrossRef]

107. Adams, P.E.; Thies, J.L.; Sutton, J.M.; Millwood, J.D.; Caldwell, G.A.; Caldwell, K.A.; Fierst, J.L. Identifying Transgene Insertions in *Caenorhabditis Elegans* Genomes with Oxford Nanopore Sequencing. *PeerJ* **2024**, *12*, e18100. [CrossRef]

108. Gilpatrick, T.; Lee, I.; Graham, J.E.; Raimondeau, E.; Bowen, R.; Heron, A.; Downs, B.; Sukumar, S.; Sedlazeck, F.J.; Timp, W. Targeted Nanopore Sequencing with Cas9-Guided Adapter Ligation. *Nat. Biotechnol.* **2020**, *38*, 433–438. [CrossRef]

109. Smirnov, A.; Nurislamov, A.; Koncevaya, G.; Serova, I.; Kabirova, E.; Chuyko, E.; Maltceva, E.; Savoskin, M.; Zadorozhny, D.; Svyatchenko, V.A.; et al. Characterizing a Lethal CAG-ACE2 Transgenic Mouse Model for SARS-CoV-2 Infection Using Cas9-Enhanced Nanopore Sequencing. *Transgenic Res.* **2024**, *33*, 453–466. [CrossRef]

110. Blondal, T.; Gamba, C.; Møller Jagd, L.; Su, L.; Demirov, D.; Guo, S.; Johnston, C.M.; Riising, E.M.; Wu, X.; Mikkelsen, M.J.; et al. Verification of CRISPR Editing and Finding Transgenic Inserts by Xdrop Indirect Sequence Capture Followed by Short- and Long-Read Sequencing. *Methods* **2021**, *191*, 68–77. [CrossRef]

111. Zarka, K.A.; Jagd, L.M.; Douches, D.S. T-DNA Characterization of Genetically Modified 3-R-Gene Late Blight-Resistant Potato Events with a Novel Procedure Utilizing the Samplix Xdrop® Enrichment Technology. *Front. Plant Sci.* **2024**, *15*, 1330429. [CrossRef]

112. Warburton, P.E.; Sebra, R.P. Long-Read DNA Sequencing: Recent Advances and Remaining Challenges. *Annu. Rev. Genom. Hum. Genet.* **2023**, *24*, 109–132. [CrossRef]

113. Jain, M.; Koren, S.; Miga, K.H.; Quick, J.; Rand, A.C.; Sasani, T.A.; Tyson, J.R.; Beggs, A.D.; Dilthey, A.T.; Fiddes, I.T.; et al. Nanopore Sequencing and Assembly of a Human Genome with Ultra-Long Reads. *Nat. Biotechnol.* **2018**, *36*, 338–345. [CrossRef] [PubMed]

114. Huang, M.; Kingan, S.; Shoue, D.; Nguyen, O.; Froenicke, L.; Galvin, B.; Lambert, C.; Khan, R.; Maheshwari, C.; Weisz, D.; et al. Improved High Quality Sand Fly Assemblies Enabled by Ultra Low Input Long Read Sequencing. *Sci. Data* **2024**, *11*, 918. [CrossRef] [PubMed]

115. Gordon, M.G.; Kathail, P.; Choy, B.; Kim, M.C.; Mazumder, T.; Gearing, M.; Ye, C.J. Population Diversity at the Single-Cell Level. *Annu. Rev. Genom. Hum. Genet.* **2024**, *25*, 27–49. [CrossRef] [PubMed]

116. Liu, T.; Conesa, A. Profiling the Epigenome Using Long-Read Sequencing. *Nat. Genet.* **2025**, *57*, 27–41. [CrossRef]

117. Schell, T.; Greve, C.; Podsiadlowski, L. Establishing Genome Sequencing and Assembly for Non-Model and Emerging Model Organisms: A Brief Guide. *Front. Zool.* **2025**, *22*, 7. [CrossRef]

118. Meier, M.J.; Beal, M.A.; Schoenrock, A.; Yauk, C.L.; Marchetti, F. Whole Genome Sequencing of the Mutamouse Model Reveals Strain- and Colony-Level Variation, and Genomic Features of the Transgene Integration Site. *Sci. Rep.* **2019**, *9*, 13775. [CrossRef]

119. Wongsurawat, T.; Jenjaroenpun, P.; De Loose, A.; Alkam, D.; Ussery, D.W.; Nookaew, I.; Leung, Y.-K.; Ho, S.-M.; Day, J.D.; Rodriguez, A. A Novel Cas9-Targeted Long-Read Assay for Simultaneous Detection of IDH1/2 Mutations and Clinically Relevant MGMT Methylation in Fresh Biopsies of Diffuse Glioma. *Acta Neuropathol. Commun.* **2020**, *8*, 87. [CrossRef]

120. Wang, Y.; Zhao, Y.; Bollas, A.; Wang, Y.; Au, K.F. Nanopore Sequencing Technology, Bioinformatics and Applications. *Nat. Biotechnol.* **2021**, *39*, 1348–1365. [CrossRef]

121. Li, S.; Jia, S.; Hou, L.; Nguyen, H.; Sato, S.; Holding, D.; Cahoon, E.; Zhang, C.; Clemente, T.; Yu, B. Mapping of Transgenic Alleles in Soybean Using a Nanopore-Based Sequencing Strategy. *J. Exp. Bot.* **2019**, *70*, 3825–3833. [CrossRef]

122. Giraldo, P.A.; Shinozuka, H.; Spangenberg, G.C.; Smith, K.F.; Cogan, N.O.I. Rapid and Detailed Characterization of Transgene Insertion Sites in Genetically Modified Plants via Nanopore Sequencing. *Front. Plant Sci.* **2021**, *11*, 602313. [CrossRef]

123. Leung, A.W.-S.; Leung, H.C.-M.; Wong, C.-L.; Zheng, Z.-X.; Lui, W.-W.; Luk, H.-M.; Lo, I.F.-M.; Luo, R.; Lam, T.-W. ECNano: A Cost-Effective Workflow for Target Enrichment Sequencing and Accurate Variant Calling on 4800 Clinically Significant Genes Using a Single MinION Flowcell. *BMC Med. Genom.* **2022**, *15*, 43. [CrossRef] [PubMed]

124. Steiert, T.A.; Fuß, J.; Juzenas, S.; Wittig, M.; Hoeppner, M.P.; Vollstedt, M.; Varkalaite, G.; ElAbd, H.; Brockmann, C.; Görg, S.; et al. High-Throughput Method for the Hybridisation-Based Targeted Enrichment of Long Genomic Fragments for PacBio Third-Generation Sequencing. *NAR Genom. Bioinform.* **2022**, *4*, lqac051. [CrossRef]

125. Lefoulon, E.; Vaisman, N.; Frydman, H.M.; Sun, L.; Voland, L.; Foster, J.M.; Slatko, B.E. Large Enriched Fragment Targeted Sequencing (LEFT-SEQ) Applied to Capture of Wolbachia Genomes. *Sci. Rep.* **2019**, *9*, 5939. [CrossRef]

126. Tilleman, L.; Rubben, K.; Van Criekinge, W.; Deforce, D.; Van Nieuwerburgh, F. Haplotyping Pharmacogenes Using TLA Combined with Illumina or Nanopore Sequencing. *Sci. Rep.* **2022**, *12*, 17734. [CrossRef] [PubMed]

127. Hafford-Tear, N.J.; Tsai, Y.-C.; Sadan, A.N.; Sanchez-Pintado, B.; Zarouchlioti, C.; Maher, G.J.; Liskova, P.; Tuft, S.J.; Hardcastle, A.J.; Clark, T.A.; et al. CRISPR/Cas9-Targeted Enrichment and Long-Read Sequencing of the Fuchs Endothelial Corneal Dystrophy–Associated TCF4 Triplet Repeat. *Genet. Med.* **2019**, *21*, 2092–2102. [CrossRef]

128. Tsai, Y.-; Brown, K.; Bernardi, M.; Harting, J.; Clelland, C. Single-Molecule Sequencing of the *C9orf72* Repeat Expansion in Patient iPSCs. *Bio-Protocol* **2024**, *14*, e5060. [CrossRef] [PubMed]

129. Watson, C.M.; Crinnion, L.A.; Lindsay, H.; Mitchell, R.; Camm, N.; Robinson, R.; Joyce, C.; Tanteles, G.A.; Halloran, D.J.O.; Pena, S.D.J.; et al. Assessing the Utility of Long-Read Nanopore Sequencing for Rapid and Efficient Characterization of Mobile Element Insertions. *Lab. Investig.* **2021**, *101*, 442–449. [CrossRef]

130. Stangl, C.; De Blank, S.; Renkens, I.; Westera, L.; Verbeek, T.; Valle-Inclan, J.E.; González, R.C.; Henssen, A.G.; Van Roosmalen, M.J.; Stam, R.W.; et al. Partner Independent Fusion Gene Detection by Multiplexed CRISPR-Cas9 Enrichment and Long Read Nanopore Sequencing. *Nat. Commun.* **2020**, *11*, 2861. [CrossRef]

131. Xu, S.; Shiomi, H.; Yamashita, Y.; Koyama, S.; Horie, T.; Baba, O.; Kimura, M.; Nakashima, Y.; Sowa, N.; Hasegawa, K.; et al. CRISPR-Cas9-Guided Amplification-Free Genomic Diagnosis for Familial Hypercholesterolemia Using Nanopore Sequencing. *PLoS ONE* **2024**, *19*, e0297231. [CrossRef]

132. Cottingham, H.; Judd, L.M.; Wisniewski, J.A.; Wick, R.R.; Stanton, T.D.; Vezina, B.; Macesic, N.; Peleg, A.Y.; Okeke, I.N.; Holt, K.E.; et al. Targeted Sequencing of *Enterobacterales* Bacteria Using CRISPR-Cas9 Enrichment and Oxford Nanopore Technologies. *mSystems* **2025**, *10*, e01413-24. [CrossRef]

133. López-Girona, E.; Davy, M.W.; Albert, N.W.; Hilario, E.; Smart, M.E.M.; Kirk, C.; Thomson, S.J.; Chagné, D. CRISPR-Cas9 Enrichment and Long Read Sequencing for Fine Mapping in Plants. *Plant Methods* **2020**, *16*, 121. [CrossRef] [PubMed]

134. McDonald, T.L.; Zhou, W.; Castro, C.P.; Mumm, C.; Switzenberg, J.A.; Mills, R.E.; Boyle, A.P. Cas9 Targeted Enrichment of Mobile Elements Using Nanopore Sequencing. *Nat. Commun.* **2021**, *12*, 3586. [CrossRef] [PubMed]

135. Hertel, O.; Neuss, A.; Busche, T.; Brandt, D.; Kalinowski, J.; Bahnemann, J.; Noll, T. Enhancing Stability of Recombinant CHO Cells by CRISPR/Cas9-Mediated Site-Specific Integration into Regions with Distinct Histone Modifications. *Front. Bioeng. Biotechnol.* **2022**, *10*, 1010719. [CrossRef]

136. Reginato, G.; Dello Stritto, M.R.; Wang, Y.; Hao, J.; Pavani, R.; Schmitz, M.; Halder, S.; Morin, V.; Cannavo, E.; Ceppi, I.; et al. HLTF Disrupts Cas9-DNA Post-Cleavage Complexes to Allow DNA Break Processing. *Nat. Commun.* **2024**, *15*, 5789. [CrossRef] [PubMed]

137. Keraite, I.; Becker, P.; Canevazzi, D.; Frias-López, C.; Dabad, M.; Tonda-Hernandez, R.; Paramonov, I.; Ingham, M.J.; Brun-Heath, I.; Leno, J.; et al. A Method for Multiplexed Full-Length Single-Molecule Sequencing of the Human Mitochondrial Genome. *Nat. Commun.* **2022**, *13*, 5902. [CrossRef]

138. Lu, W.; Lan, X.; Zhang, T.; Sun, H.; Ma, S.; Xia, Q. Precise Characterization of Bombyx Mori Fibroin Heavy Chain Gene Using Cpf1-Based Enrichment and Oxford Nanopore Technologies. *Insects* **2021**, *12*, 832. [CrossRef]

139. Madsen, E.B.; Höijer, I.; Kvist, T.; Ameur, A.; Mikkelsen, M.J. Xdrop: Targeted Sequencing of Long DNA Molecules from Low Input Samples Using Droplet Sorting. *Hum. Mutat.* **2020**, *41*, 1671–1679. [CrossRef]

140. Abdelwahab, O.; Torkamaneh, D. Artificial Intelligence in Variant Calling: A Review. *Front. Bioinform.* **2025**, *5*, 1574359. [CrossRef]

141. Kovaka, S.; Hook, P.W.; Jenike, K.M.; Shivakumar, V.; Morina, L.B.; Razaghi, R.; Timp, W.; Schatz, M.C. Uncalled4 Improves Nanopore DNA and RNA Modification Detection via Fast and Accurate Signal Alignment. *Nat. Methods* **2025**, *22*, 681–691. [CrossRef]