genes

MDPI

Review

Concatenation of Transgenic DNA: Random or Orchestrated?

Alexander Smirnov [1] and Nariman Battulin [1,2,*]

[1] Laboratory of Developmental Genetics, Institute of Cytology and Genetics SB RAS, 630090 Novosibirsk, Russia; smirnovskaven@bionet.nsc.ru
[2] Institute of Genetic Technologies, Novosibirsk State University, 630090 Novosibirsk, Russia
* Correspondence: battulin@bionet.nsc.ru

Abstract: Generation of transgenic organisms by pronuclear microinjection has become a routine procedure. However, while the process of DNA integration in the genome is well understood, we still do not know much about the recombination between transgene molecules that happens in the first moments after DNA injection. Most of the time, injected molecules are joined together in head-to-tail tandem repeats—the so-called concatemers. In this review, we focused on the possible concatenation mechanisms and how they could be studied with genetic reporters tracking individual copies in concatemers. We also discuss various features of concatemers, including palindromic junctions and repeat-induced gene silencing (RIGS). Finally, we speculate how cooperation of DNA repair pathways creates a multicopy concatenated insert.

Keywords: transgenic animals; concatemers; double-strand breaks (DSBs); homologous recombination (HR); non-homologous end-joining (NHEJ); palindromes; repeat-induced gene silencing (RIGS); pronuclear microinjection

check for updates

Citation: Smirnov, A.; Battulin, N. Concatenation of Transgenic DNA: Random or Orchestrated?. *Genes* **2021**, *12*, 1969. https://doi.org/10.3390/genes12121969

Academic Editor: Miroslav Plohl

Received: 10 November 2021
Accepted: 9 December 2021
Published: 10 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Historical Overview of the Concatenation Studies

Concatenation is a common term to describe a process of linking monomers together. In nature, DNA concatenation could be used by viruses to amplify genome copy numbers [1,2]. Alternative applications of this principle include mitochondrial DNA replication [3], telomere maintenance [4], and multimerization of human artificial chromosomes [5] or replicative plasmids [6]. We will use the term "concatenation" in the context of transgene delivery, when tandemly arrayed transgene copies co-integrated into a genomic site. This process differs from plasmid multimerization, because after injection of hundreds of molecules their ends are recognized as double-strand breaks (DSBs) in a swift and chaotic fashion, determined by a complex interaction of DNA repair pathways.

Concatenated DNA was found in all model organisms, including mammals (pronuclear microinjection) [7], plants (microprojectile bombardment/T-DNA transformation) [8,9], or nematodes (gonad injection) [10]. Concatenation is a constant companion of transgenesis in cell culture and embryos and was reported for all types of DNA donors, including lentiviral backbones [11], bacterial artificial chromosomes [12,13], transposons [14], Cas9-cleaved minicircles [15], and single-stranded ODNs [16,17]. In this review, we will mainly focus on the mouse pronuclear microinjection as one of the most studied and standardized methods.

Microinjection of DNA into the pronucleus of zygote is the main route for obtaining transgenic animals [18]. The effectiveness of this method was first demonstrated almost 40 years ago [19,20]. During microinjection, around 1000 copies of the linear transgene molecules enter the pronucleus. Until the moment of integration, they recombine into tandem repeats (concatemers) containing up to several hundred copies of the transgene and in the overwhelming majority of situations directed in tandem ("head-to-tail") (Figure 1A). The fact that in most cases, the transgenic DNA is found in the form of multiple repeated monomers (tandems) located at a single locus [21,22] predicts the existence of extrachromosomal concatenation mechanisms. This phenomenon has been extensively studied by many research groups in embryos [23–26] and injected [27,28] and transfected cells [29,30].

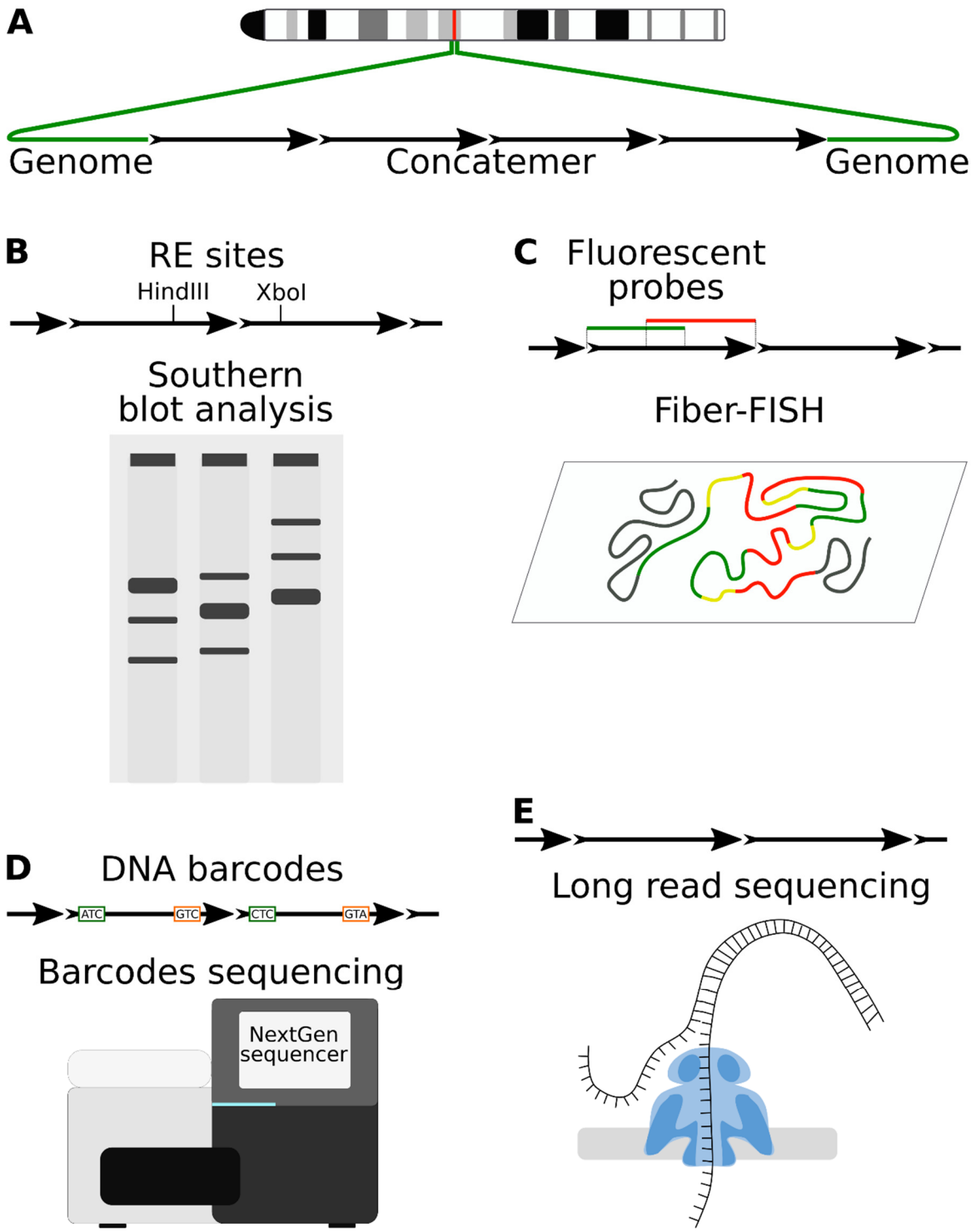

Figure 1. Transgene concatenation and recombination reporters. (**A**) Schematic representation of the concatenated tandem repeat at the chromosomal locus. (**B**) Transgene copies could be labeled with alternative restriction sites to study concatemer by Southern blot. (**C**) FISH analysis on extended DNA fibers demonstrates tandem organization of concatemers. (**D**) Barcoded plasmid library could be used to investigate concatenation. Barcodes from the transgenic copies are amplified with PCR and analyzed by NGS. Connections between barcodes are used to model continuous chains of copies. (**E**) Oxford Nanopore technology, a real-time single-molecule sequencing method, is becoming popular for concatemer analysis.

Several explanations for the phenomenon of concatenation have been proposed. One model predicts that a concatemer is built from one or more circular copies through de novo amplification by a rolling circle amplification or other means. This model could be

inferred from the high efficiency of concatenation, the notion of occasionally observed identical transgene-transgene junctions [31], and the fact that cells possess mechanisms for gene amplification [5,32,33]. However, the amplification hypothesis was disproved by the studies with concatenation reporters, which demonstrated that copies in the concatemers are unique (Figure 1B–D) (see below). The second hypothetical mechanism of concatenation can be referred to as the "overlapping fragment model" [7,21]. Injection of the supercoiled plasmid molecules can produce concatemers [28]. In theory, occasional breakage of circular molecules will result in a pool of overlapping homologous DNA fragments that would efficiently recombine in the zygote. Finally, the most "down-to-earth" hypothesis postulates that concatemers are formed by homologous recombination (HR) between the linear ends of transgenic molecules with some contribution from circular copies [32,33]. However, this model does not explain why linear ends are preferentially processed by HR factors rather than subjected to random non-homologous end-joining (NHEJ). None of the proposed models of transgene concatenation had been unequivocally confirmed, even decades after their appearance, because analysis of repetitive DNA recombination was difficult with the existing methods.

Understanding the structure of concatemers is important and could be achieved only by labeling copies with individual signatures. There were only a few experiments with such reporters. One notable work from the laboratory of the future Nobel prize winner Mario Capecchi came out almost 40 years ago [27] but is often cited today. The researchers injected cultured cell nuclei with varying numbers and shapes of DNA molecules (linearized and circular). The number of molecules varied from 2 to 1000. To study the concatenation mechanism, they marked the injected molecules with primitive "barcodes" (HSVtk gene in two orientations) [27]. Molecules with alternative HSVtk variants (A/B variants) generated different signals on the Southern blot after restriction (Figure 1B). Analysis of concatemers from cell clones showed that they consisted of two mixed variants of molecules. This implied that concatemers assemble from several linear copies, although the resolution of the method did not allow to reject the amplification hypothesis with confidence. One of the main conclusions of this pioneering work was that the copy number of inserts correlated with the concentration of injected DNA and that the presence of linear DNA ends stimulated concatenation, which also hinted at the assembly of concatemers from individual molecules rather than one amplified circular copy. The follow-up analysis with similar constructs demonstrated that there is a narrow time window, when DNA molecules could recombine, because separating A/B injections by one hour results in a biased ligation of copy variants into uniform repeats rather than interspersed concatemers [28].

After some time, this approach was re-explored with a different reporter. In the updated assay, a direct repeat of two LINE-1 elements was injected in the pronuclei [34]. The two copies were identical except the restriction sites, which were used for Southern blot analysis. The amplified construct would have provided a single band, while recombination between copies would lead to a change in the restriction site order in concatemer. Initial Southern blot analysis indicated that injected molecules exchanged sites during concatenation. In addition, thorough analysis of the concatenated inserts was undertaken after plasmid rescue. The authors described several complex recombination patterns, including deletions and duplications inside the tandem LINE1 reporter, indicating the intermolecular recombination [34].

Concatenation was considered as a promising tool for assembling lengthy transgenes from fragments directly in the zygote. Transgenic mouse lines carrying human albumin (hSA), assembled from 7 + 13 + 17 kb fragments, or human adenosine deaminase (ADA) (29 + 39 kb fragments), were created by pronuclear microinjection [35,36]. In the first case, the overlapping regions of the fragments were about 2–3 kb, and in the second, 19.5 kb. Southern blot analysis confirmed that correct copies of the gene were present in most of the transgenic offspring, although incomplete copies or rearrangements were also encountered [35]. Of note, there was little concatenation between identical fragments. In a similar work, the human lipoprotein receptor (VLDLR) gene was successfully assembled

from two fragments of 40 and 35 kb, with an overlap of 8 kb [37]. The authors used FISH analysis on extended DNA fibers (Fiber-FISH) to visualize a stretch of head-to-tail tandems in the genome of transgenic animals (Figure 1C). Later, this concatenation approach was abandoned, because microinjection of BACs into the pronucleus proved to be effective [38,39].

We have suggested improving the concatenation reporter by combining DNA barcoding and next-generation (NGS) sequencing (Figure 1D). We created a plasmid library with thousands of unique barcode pairs separated by a restriction site. Note that in this assay, barcodes are located close to the exposed ends of the construct, rather than inside the backbone. Linearized copies with barcoded ends are injected into the pronucleus, which causes concatenation and ligation of barcoded ends together. Barcodes from transgenic animals are amplified with PCR and identified by NGS sequencing (Figure 1D), providing plenty of data for analysis. This reporter can discover copy amplification cases. Connections between the barcodes, observed in the NGS data, are used to build a chain of molecules, and understand the order of copies in concatemer. Unfortunately, only direct head-to-tail repeats could be detected because PCR amplification of inverted repeats is ineffective [40].

Recently, single-molecule sequencing technologies, e.g., Oxford Nanopore technology (Figure 1E), made a revolutionary step forward for genome analysis (see the corresponding chapter of this review). These methods allow study of the complex structural variations and palindromes, because PCR amplification is not required in sample preparation. Ironically, the story of concatenation began way back before the invention of PCR and made a comeback when PCR went out of fashion.

2. Information from the Internal Junctions

Sequencing internal junctions between transgene copies (Figure 2) was a primal source of data about concatemer structure before invention of the methods that allow the study of concatenated repeats, such as long-range sequencing or DNA barcoding. In the nucleus, linear transgene ends are recognized as DSBs by DNA repair machinery. Non-homologous end-joining is a primary repair pathway in embryos and embryonic stem (ES) cells, while microhomology-mediated end-joining MMEJ serves as a backup plan and uses small homologies (5–25) revealed from DNA resection [38,41]. MMEJ is responsible for 10–20% of DSB repair and competes with HR at MRN-resected ends [39]. Experiments with genetic reporters in mouse ES cells show that in the absence of classical NHEJ (Ku80−/−, LigIV−/−), 93% of breaks are repaired with microhomologies (>2 nucleotides). Normally, this value is 2–3 times less, depending on the specific locus [42].

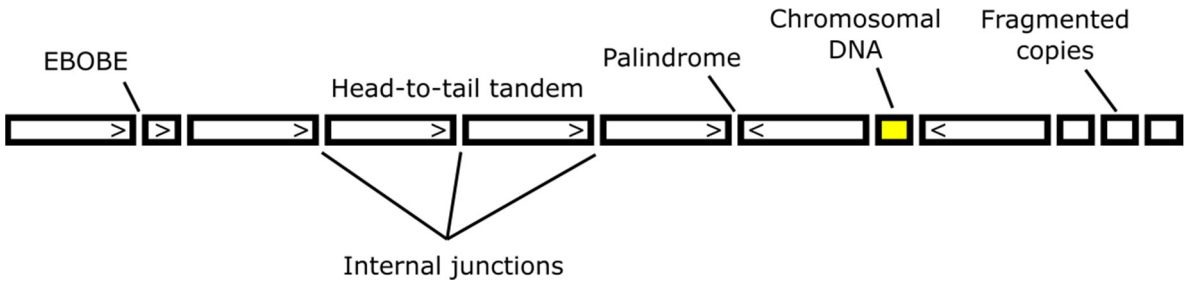

Figure 2. Concatemer elements that could be present at the integration site (see explanations in the text).

The injection of embryos with molecules having different restriction digested ends (5′, 3′, blunt) did not reveal any bias in end repair [32,43]. Most of the time, internal junctions from the injected embryos had only small deletions: clippings of the transgene ends prior to ligation [33,44]. In rare cases, transgenes are re-ligated by sticky ends formed from restriction endonuclease digestion [45]. Thus, most of the internal junctions in concatemers represent NHEJ-ligation products.

The incorporation of concatemers into the genome also occurs due to NHEJ/MMEJ activity. Inactivation of the two pathways through inhibition of Ku70/Ku80/ Lig4 (NHEJ)

and polymerase θ (TMEJ, the main MMEJ pathway) reduces the frequency of random insertion of transgenes by five orders of magnitude [46]. The MMEJ signatures are frequently seen when analyzing the nucleotide sequences of the transgenic-genomic boundaries in cells and embryos [47–49].

In some cases, additional DNA co-integrates with concatemers. The fragments include chromosomal and mitochondrial DNA, mobile elements, telomere repeats, and bacterial DNA, and are typically found at the transgene-genome junctions but could also localize inside the concatemer [13,50–53].

The hierarchy of the fragment assembly during concatenation that can be devised from the NHEJ/MMEJ signatures at the junctions is an interesting topic for speculation. The frequency of microhomology-mediated joining at the transgene-genome borders seems to be higher than at the internal junctions. This fact might indicate that integration in the genome occurs at the final stages of concatemer formation when most of the ends are resected. Systematic analysis of the contiguous concatemer inserts with single-molecule sequencing will be required to clarify this issue and offer a road map of concatenation.

3. Palindromes

Palindromic sequences are broadly represented in the vertebrate genomes [54] and are associated with repetitive sequences, such as short microsatellite repeats [55], mobile elements [56], de novo structural variations in cancer cells [57,58], and sex chromosomes [59,60]. Some prominent palindromes have been extensively studied. For example, in the human genome, palindromic AT-rich repeat (PATRR) is responsible for the frequent reciprocal translocation of regions at chromosome 11 и22 during spermatogenesis [61], which causes Emanuel syndrome [62].

The presence of palindromes in the genome provokes rearrangements, but there is no consensus on the mechanism of their occurrence. According to the most popular model, replication and transcription of genome regions induce the appearance of single-stranded DNA regions, which leads to the formation of special cruciform structures and hairpins [63,64]. Interestingly, post-mitotic neurons that halted replication can destroy palindromes [55], and introducing replication origin close to the palindrome structure does not increase the frequency of rearrangements [65]. It is also known that the DSB adjacent to the palindromic construct initiates resection, leading to the emergence of single-stranded hairpin-forming regions in DNA [66]. The existence of cruciforms and hairpins in palindromes has been shown experimentally [61,65]. These structures are recognized and bound by many DNA damage response factors [67,68] and cell endonucleases, which process these topologically constrained regions into DSBs. Possible candidates include endonucleases Mre11 [69,70], GEN1 and Artemis [71], ERCC4 (Xpf) [72], or MutL complex [73].

Recombination of palindromic junctions is studied using special reporter constructs [56]. The authors obtained insertions in the yeast genome with two types of palindromic fusions: with a large (1000 bp) or short (12 bp) spacer. A spacer is a unique piece of DNA that separates two inverted repeats (palindrome). The palindromic regions themselves were about 2000 bp. This and other experiments demonstrate that even short spacers are capable of stabilizing palindromes, possibly due to prevention of the denaturation at the central point between two inverted repeats [56,74].

In another work, an episomal reporter construct was used, in which the GFP gene was interrupted by the insertion of a palindromic region (40 + 40 bp). Experiments were carried out in several types of cells (HeLa, HEK293T, and COS-7) [65]. Positive recombination events (deletions in the palindrome) were detected as a result of the restoration of the GFP gene using a donor plasmid (full-length GFP gene without a start codon). Meganuclease I-SceI and a similar reporter plasmid were used as controls, where an I-SceI site was inserted instead of the palindromic region. It was found that the presence of a palindromic region in the plasmid provokes DSB with a frequency comparable to targeted digestion with nuclease (1–2% of cells). At the same time, the insertion of a small 35bp spacer between inverted copies notably reduced the frequency of breaks, down to the background level [65].

A similar study with palindromic Alu repeats showed similar results and estimated the size of the spacer that blocks rearrangements (52 bp) [63].

So, how can this knowledge translate to concatenation studies?

Concatemers frequently contain palindromic segments (that is, "head-to-head" or "tail-to-tail" inverted repeats) as a consequence of random transgene ligation [42,53,56]. To understand the process of concatenation, it is important to study palindromic variants of junctions between copies, and their frequency and stability in the animal genome. Concatemers are stably inherited by the descendants of the transgenic founder animal, but the fate of palindromic junctions between copies has hardly been studied in the generations. Long-term observation of palindromic fusion in a transgenic mouse strain has helped to better assess the stability of this structure. In the laboratory of Maria Jasin, experiments were carried out to create mice with a reporter construct from two consecutive copies of the LacZ gene [75]. One of the transgenic lines harbored a palindromic insert of two transgenes equal to 15.4 kb in length. This mouse line represented a rare case of a perfect palindromic junction with no deletions inside the internal junction [75–77]. The authors studied the inheritance of the transgene and documented the high frequency of small deletions and more complex recombination products at the junction point in the palindrome in about 35% of the offspring [75,76]. An immortalized cell culture was obtained from mice of this strain [77]. Long-term cultivation of cells with a palindromic structure showed that the average frequency of rearrangements within the palindrome was 5.5×10^{-3}, which is approximately 0.5% of cells per passage. Most of the rearrangements were small deletions in the palindromic junction, which broke the central symmetry (of the 118 rearrangements analyzed, 93% were deletions). The destruction of the symmetry increased the stability of the palindrome in the subclone by a factor of 25 [77] (Figure 3). In general, the frequency and patterns of rearrangements within the palindrome in mice and cells were very similar.

Unfortunately, palindrome studies in transgenic embryos or mouse lines are rare, because common molecular methods based on PCR are not suitable for detection and sequencing of inverted repeats. This is probably due to the formation of a hairpin in the PCR product, which prevents amplification ([78]; our unpublished data). Single-molecule sequencing methods do not suffer from these limitations and could be used to investigate complex transgene junctions [40,52]. We used the PacBio approach to analyze concatemer with dozens of palindromes. Most of the observed inverted junctions (head-to-head or tail-to-tail) carried asymmetric deletions, which are believed to stabilize the palindrome [40]. An alternative stabilization mechanism (spacer inclusion) was uncommon, probably because the chance of incorporating a DNA fragment between two inverted copies is not high during concatenation.

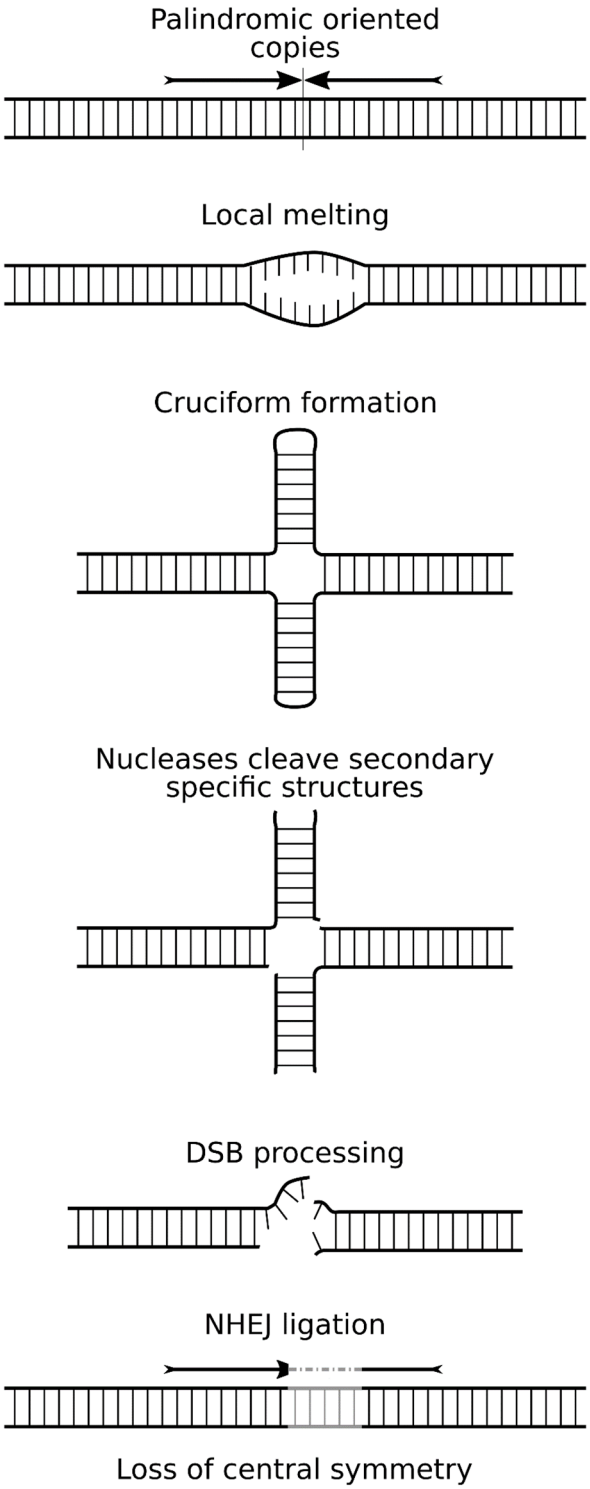

Figure 3. Cells stabilize palindromic junctions by introducing asymmetrical deletions. The general scheme of the appearance of cruciforms and their transformation into DSB. The reasons for denaturation can be different: torsion stress, replication, transcription, close DSB.

4. Molecular Mechanisms of Concatenation

To understand the molecular mechanisms resulting in concatenation, we should first establish if there really is a preference in copy orientation. In the case of random copy ligation by NHEJ, a distribution of 25%: 50%: 25% can be expected for the three possible types of orientations (50% corresponds to a head-to-tail orientation). Numerous descriptions of concatemers in pronuclear microinjection provide compelling evidence that head-to-tail

orientation has an observed frequency over 90%. Importantly, the calculations were made by different methods: Southern blotting [23,24,41,48,79], whole-genome sequencing [49], Oxford Nanopore technology (ONT) [52], and PacBio [40]. There are also cases of very long concatemers (~267 copies) with exclusively tandem orientation of transgenes [80]. Although in some reports observed transgenes are ligated randomly [81], it is safe to assume that head-to-tail concatenation is a characteristic feature of pronuclear microinjection and there must be a mechanism to organize the direction of the copies.

Two main HR mechanisms, synthesis-dependent strand annealing (SDSA) and double-strand break repair (DSBR), participate in the assembly of transgenes into the concatemer tandem (Figure 4C–F). The joint step for both pathways involves initial 3′-resection of double-stranded ends of linear molecules and a subsequent search for a homologous region for invasion [82]. Apparently, this would be the junction site between the two copies. In the SDSA scenario, extension of the end is carried out to achieve overlapping homologies that are patched together in a head-to-tail order. For DSBR, capture of the second resected end is required, then a double Holliday junction (dHJ) is formed, which can be resolved in two ways (with or without crossing over). Cells typically strive to prevent crossovers with the Sgs1 (BLM) helicase activity, which eliminates dHJ by active branch migration and dissolution [79,83].

The DNA ends participating in HR are very active, and some molecules could be copied three to five times before finally being integrated into the concatemer [40]. The data from both LINE1 and barcoded reporters show that most of the copies participate in recombination (50% and 80%, respectively) [34,40]. The addition of biotin to the transgene ends blocked concatenation [84,85], demonstrating that linear ends are indeed initiators of recombination.

Another indication of end recombination was discovered accidentally in the barcoded concatemers and was named Elongation Beyond Original Broken End (EBOBE) [40]. The EBOBE pattern represents intermediate DNA synthesis metabolite, an elongation of the initial resected transgene end that copied the junction up to the barcode (Figure 4H). This heterologous region at the 3′-end prevents the copy from returning for a new round of synthesis, so it joins the concatemer via NHEJ or MMEJ. Such mechanisms are known for some other HR metabolites (non-canonical HR termination) [86] in combination with the activity of theta-mediated end-joining (TMEJ) polymerase θ [87].

Retrospective analysis of the published cases provides examples for the EBOBE pattern. For instance, several reports described a single copy of transgene that was flanked by a truncated copy with almost no clipping at the internal junctions [35,51,88]. These cases have similar traits: minor clipping inside the transgene-transgene junction and a short elongation fragment nearby.

Unfortunately, it is still impossible to exclude the possibility that two copies were ligated independently. We have revisited this problem with our barcoding approach [40]. In this case, the barcode serves as an EBOBE “trap” that helps to detect these cases. Using this logic, we annotated four cases of EBOBE in long PacBio reads (around 56 copies in total) and three cases, accidentally caught at the transgene-genome junctions in other embryos.

Some but not all EBOBE fragments had microhomology and were probably processed by MMEJ as the single-stranded HR.

Similar recombination patterns could be found in the cell culture experiments with targeted DNA integration. It is a common knowledge that homologous arms of the donor vector participate in SDSA-directed integration [89]. Sometimes, the donor cannot be processed completely, and DNA synthesis is halted inside the construct, leaving truncated insertion. In the case of transgene concatenation, HR is initiated not between the genome and the homologous arms, but between the pool of identical transgene copies, and the truncated intermediates are preserved inside the concatemer.

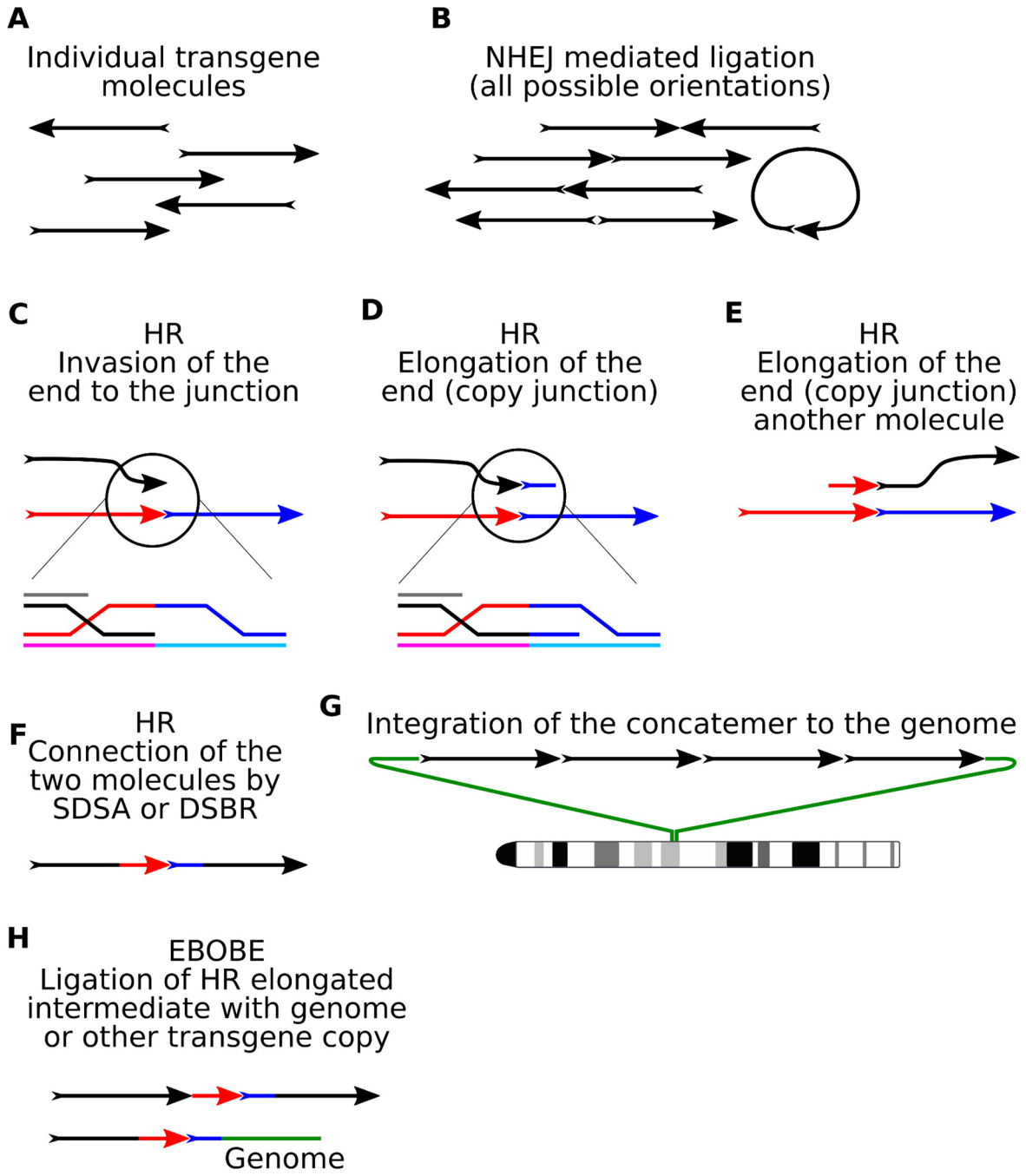

Figure 4. DNA repair pathways that contribute to transgene concatenation. (**A**) Injected transgene molecules. (**B**) Initial legation of copies is carried out by NHEJ, creating templates for HR. (**C**) Resected ends invade donor templates and (**D**,**E**) copy a portion of the junction (blue). (**F**) Re-annealing of the complementary synthesized ends leads to connection of the copies into tandem repeats. (**G**) Integration of the concatemer into the genome is facilitated by NHEJ or MMEJ after concatenation. (**H**) EBOBE pattern—a product of DNA end synthesis that is incorporated into the concatemer after HR termination.

The contribution of other HR repair mechanisms, such as single-strand annealing (SSA), microhomology-mediated break-induced replication (MM-BIR), and multi-invasion recombination (MIR), is unknown. We did not find a significant impact of these pathways on concatenation in our assay. MM-BIR occurs upon invasion of non-homologous regions of the genome due to microhomology at the 3′end of the molecule. MM-BIR would lead to copying of an extended concatemer region (from several thousand to many hundreds of thousand base pairs), which would lead to duplication of a large region [90]. MIR is

a novel potentially mutagenic DSB repair pathway [91]. Recent work by Wolf-Dietrich Heyer's group has demonstrated that resected 3′-filaments can use internal single-stranded regions far from the 3′-end of the filament to search for homology. This leads to the fact that one filament can invade two independent regions of the genome and combine them into dHJ [92], which, when resolving dHJ with crossing over, leads to translocations of chromosomal fragments. Most likely, these mutagenic pathways (SSA, MM-BIR, MIR) are suppressed during concatenation, because their initiation requires extended end resection (>1 kb) [90,91]. In conditions where homologous sites for invasion are in excess, SDSA/DSBR should utilize normal resected ends <1 kb [93] and effectively compete with the aforementioned pathways. In theory, the multi-invasion frequency and template switching during HR in concatenation could be studied with a uniformly barcoded vector library. Instead of barcoding transgene ends, it would require the insertion of multiple barcodes or SNPs every ~1000 bp of the backbone to analyze gene conversion between copies.

In this chapter, we discussed how high HR activity in the embryo organizes molecules into tandem repeats. Curiously, there is not much evidence about transgene copy orientation in cell cultures. In theory, copy arrangements there would be more random due to NHEJ/MMEJ domination through the cell cycle [88], although non-random head-to-tail concatemers [85] and end recombination (our unpublished data) suggest that concatenated ends could be processes by HR in cells. In addition, transgenic cell clones are obtained by lipofection, electroporation, or viral delivery [94] in the asynchronous cell cultures. It would be interesting to compare concatemers in embryos/cultured cells obtained by the same delivery method (microinjection).

5. Repeat-Induced Gene Silencing

Repeat-induced gene silencing (RIGS) leads to expression silencing in high-copy transgene insertions and is established gradually with time in cell passages, embryo development, or in later generations. RIGS exists across all multicellular organisms, including Drosophila [95], fungi [96], plants [97], and mammals [98,99]. In animal biotechnology, it is a general rule to avoid multicopy founders during initial transgene screening in fear of RIGS, but the molecular mechanisms of this process remain unknown in mammals.

The most perplexing feature of RIGS is that, unlike in gene position effect variegation that is determined by the surrounding regulatory elements [100,101], RIGS occurs regardless of the integration site or specific transgene sequences and can onset later in development [102]. To make things even more complicated, sometimes multicopy concatemers evade RIGS (different constructs show different results), and a positive correlation between the copy number and expression is not rare [12,103].

Another observed peculiarity that could be attributed to RIGS is "expression plasticity", the notion that the expression of transgenes at a single locus negatively affects each other [104] and that consequent integration of two highly expressed genes can decrease their expression [105]. It was also shown that reactivation of transgene expression by antibiotic selection can activate neighbor transgenes in cis [106].

It is possible to directly interfere with RIGS by Cre-mediated reduction of the transgene copy number at the site (from 8 to 1–2 copies), which led to an increase in the expression [107] in mice, but other experiments using the same Cre-based approach did not find a positive correlation between the expression and copy number reduction at another locus [108]. The fact that even a single copy integration can be silenced with passages adds more complexity to the picture [109]. Strikingly, the expression of the constructs inside safe-harbor loci, such as human AAVS1 or mouse Rosa26, could be silenced in some cases as well [110–112].

How can RIGS be explained? In plants, small RNAs generated from the expressed transgenes guide homologous mRNA cleavage for post-transcriptional gene silencing (PTGS) [113]. In addition, plants can silence transgenes through RNA-directed DNA methylation (RdDM), which acts at transcriptional level (TGS) [114]. In Drosophila, a similar process is guided by Piwi-interacting RNAs [115]. Studies in the mouse cells

suggested the possibility of RNAi gene silencing associated with convergent transcription dsRNA of palindromic copies, which are common in the concatemers [115]. However, later experiments with a huge 1000 copy tandem reporter [116] unequivocally showed through Dicer knock-down that the RNAi pathway does not establish silencing [117].

Alternatively, there is a possibility that RIGS is provoked by the structural changes in the local DNA. Some observations suggest that RIGS could be caused by direct DNA-DNA pairing at repeats and that such recombination-independent recognition of DNA homology could be common in different organisms [118].

The correlation of RIGS and repressive epigenetic chromatin marks is well established. DNA methylation is an extensively characterized epigenetic modification (Jones, 2012). The levels of methylation of the CpG residues in the promoter of repetitive sequences generally correlate with the copy number. High levels of methylations are associated with lower gene expression [107,119–123] and silenced transgenes are less sensitive to DNase I and carry repressive chromatin marks [97,124]. Experiments with large silenced repeats have proven that HP1/p150/CAF-1 establish condensed heterochromatin at the multicopy transgenes. This effect was observed in several cell lines [125,126].

The most extensive study of the tandemly amplified plasmids, which have similarities to transgene concatemers, originates in the laboratory of Noriaki Shimizu. The group has investigated the causes of RIGS in the IR/MAR plasmid amplification system that lead to extrachromosomal or chromosomal multimerization. Notably, they discovered that inhibition or knock-out of a broad spectrum of histone deacetylases (HDAC1-2, SIRT1) alleviates the RIGS phenomenon [127]. Interestingly, the HDAC inhibition effect only had an impact during initial multimerization of the IR/MAR plasmids.

It should be noted that this system is not identical to the transgene concatenation during microinjection, because IR/MAR plasmids carry Dhfr selection cassette and Ori sequences that facilitate an alternative multimerization mechanism. Nevertheless, findings from the IR/MAR assay will be relevant to the study of RIGS in concatemers.

Cell culture experiments aiming at high gene expression employ transposable elements to achieve stable single-copy integrations or induce high-copy amplification in combination with measures to avoid RIGS. These measures may include chromatin relaxation by inhibitors (trichostatin A, butyrate [119]; 5-azacytidine (5-AzaC) [128] or attaching functional elements to the transgene (BAC genomic regions, matrix attachment regions (MARs)) [12,109,129,130].

Obviously, vertebrate RIGS is not a simple phenomenon, and is likely established at multiple levels, including transcription, DNA methylation, and heterochromatinization. Other probable molecular mechanisms include a deficit of transcription factors, epigenetic silencing in meiosis, DNA structural folding, nuclear lamina retention [131], and transcription-induced silencing.

We intend to employ an expressed barcoded reporter that would be instrumental in studying several aspects of RIGS in cultured cells and transgenic embryos (Figure 5). The reporter is based on a randomly integrated concatemer, consisting of several types of barcoded monomers. LoxP monomers will be utilized to change the order and number of copies [107,108]. Inducible monomers that carry Tet-On promoters will be used to study how reactivation of expression affects RIGS in cis. The inclusion of insulated monomers will show how MAR-elements [103,132] behave in a silenced concatemer. Since the order of copies in the repeat could have an effect on expression [133], it would be interesting to examine if the flanking copies avoid silencing. During concatenation, transgene copies are joined randomly, and the distribution of monomer variants in multicopy integrants will depend on the proportion of monomers in the delivered DNA (Figure 5). Information about individual barcode expression from the RNA seq data could be combined with other modern molecular techniques, such as nanopore methylation sequencing [134], ChIP-seq analysis of transcription factors and chromatin proteins, and dynamic analysis of expression with SLAMseq [135] after HDAC/HP1/DNA methyltransferases (DNMTs)

inhibition. This will help to better understand molecular drivers of RIGS and the timeline of events that leads to expression silencing.

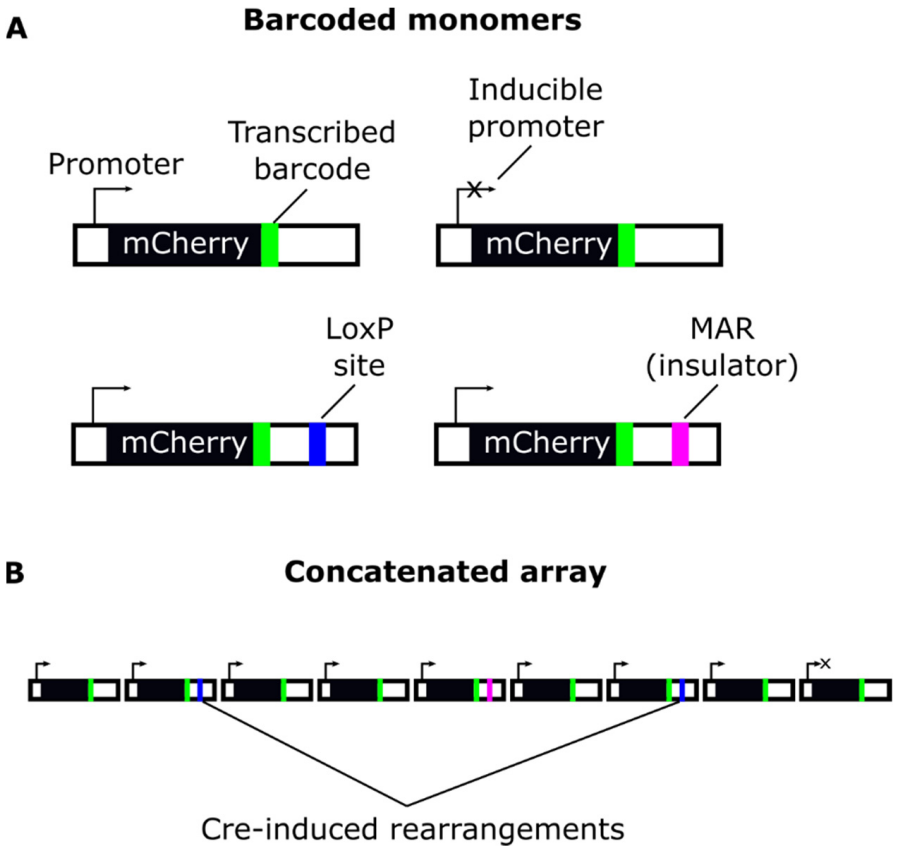

Figure 5. Barcoded reporter for concatenation of expressed monomers to study RIGS mechanisms. (**A**) Types of barcoded monomers. (**B**) Transfection of monomers in specific proportions (e.g., 10:1:1:1, where 10 is the standard monomer) could be used to create unique concatenated arrays for RIGS analysis by RNAseq and other methods. Transfection of the cells with Cre-recombinase would change the order and number of copies.

6. Novel Methods for Studying Transgene Concatenation

Careful readers should have grasped the idea that concatemers are not mere "repeats" of transgene monomers, but they have a complex internal composition scaling with the copy number. Until recently, cataloguing these genetic patterns was impossible because the repetitive nature of concatemers obstructed sequence assembly by next-generation sequencing (NGS). Fortunately, a plethora of novel genomic methods for transgene DNA analysis appeared at the end of the last decade. One of them, targeted locus amplification (TLA), based on proximity DNA ligation is specifically designed to map transgenes [136–138]. In this review, we will not focus on it, because TLA mostly provides information about the integration site and the surrounding genomic locus. Undoubtedly, this method will be important for systematic analysis of the thousands of archived transgenic mouse lines [52] to chart integration site preference and analyze collateral genomic damage, including the pervasive presence of the flanking chromosomal duplications that could be found in random and targeted integration approaches [80,129,130,139,140].

Single-molecule sequencing (SMS) technologies will have a huge impact on genomic research, particularly for the study of transgene localization and the internal concatemer structure. Two distinctive approaches are currently being developed: Oxford Nanopore Technology (ONT) and Pacific Bioscience Single-molecule Real-time sequencing (PacBio) [141] (Figure 1D). In ONT, protein pores are integrated into a polymer membrane and an ionic current is passed through the nanopore by setting a voltage across the membrane. A

helicase molecule is attached to DNA (or RNA) fragments during sample preparation, which allows molecules to translocate through the pore, generating a change in the current. The signal is deciphered into corresponding nucleotides or their modifications in real time. PacBio uses a different method based on fluorescence detection in the special flow cell picolitre-sized wells. A single polymerase molecule is localized. When a circularized DNA molecule is added to the well, polymerase binds a single-stranded region in the adapter. During complementary strand synthesis with labeled nucleotides, the fluorescence signals are captured through the transparent bottoms of the cells in real time.

Both methods offer great utility for concatemer analysis, as they are suitable for palindrome detection and generate long reads, with the median length ranging from 10 to 100 kb [141,142], but are also capable of achieving read lengths up to hundreds and thousands of kbs [143], so that one read could contain >20 concatenated transgene copies. ONT was recently employed to sequence transgenic mouse lines with interesting integration patterns [56,88,144], signifying a start of a new era in mouse genome analysis. We have used both methods to study concatenation of barcoded library in the embryos [40]; unpublished data) and discovered plenty of complex internal concatemer structures.

It should be noted that SMS technologies suffer from two critical limitations that stall their utility for transgene concatenation analysis. First of all, these methods are still very expensive for analyzing transgene inserts. Even large concatemers spanning 0.1–1 Mb in size correspond to only about 0.02% of the mouse diploid genome, which will make up a small fraction of the total sequencing data. The results of a typical ONT/PacBio run usually constitute 1–5x mouse genome coverage [56,88,144] and it could be difficult to assemble the whole concatemer from these data. We have a performed transgene sequences experiment for concatemers with high copy numbers (>100). The ONT MiniON device produced about 8 Gb, which contained around 0.9 Mb of transgene sequences, equal to 0.02%, and contained about 50 unique copies (50%) (unpublished data). In an alternative PacBio experiment with another transgenic line, we obtained 16.1 Gb, of which 1.16 Mb (0.07%) corresponded to the transgene sequence and contained around 90 unique copies (80%) [40].

For this reason, SMS methods are frequently paired with target enrichment to increase the transgene read percentage. Target DNA capture with sequence-specific biotinylated probes can provide extremely high enrichment ratios (10^3–10^5) at the cost of simplicity [145]. Today, most common techniques for ONT and PacBio involve cleaving target transgene sites in dephosphorylated genomic DNA samples with CRISPR/Cas9 nuclease and ligating adapter for sequencing to the phosphorylated ends [136–138]. This is a powerful approach that could increase transgene read counts 50–60× [138,146] with up to 1600× in some cases [144], but it is not applicable for concatemer studies, because fragmenting of the internal structure is undesirable. With this in mind, we could appreciate DNA barcodes as unique Cas9 target sites that could be used for transgene enrichment. In theory, concatemer structure loss could also be avoided by mapping integration sites first [147], using these flanking sites as Cas9 targets, and subsequent DNA fragmentation, consequently decreasing the size of the enriched concatemer fragment to affordable lengths.

The second SMS limitation is the high error rate. Current routine applications achieve around 90–95% accuracy for both technologies [148,149], which might be enough for general transgene mapping, but is absolutely deadly for the short barcode identification. In the aforementioned experiments, we observed an 83% accuracy for ONT and 92% for PacBio reads; however, it did not help to reliably identify barcodes from long reads and we had to rely on the NGS data for validation. In addition, current error rates obstruct interpretation of nucleotide sequences at transgene-transgene junctions, limiting the value of sequencing data. Again, simply increasing the sequencing depth by upscaling or target enrichment could be useful as both technologies profit from consensus from multiple reads and reach >99% accuracy [150–152], but it could be costly.

Furthermore, new evidence warns that palindromes might represent a challenge for ONT sequencers, because the secondary DNA structure affects strand transfer dynamics

through the pore [153]. This is arguably a minor drawback but has to be addressed to study palindromic junctions in detail. How the PacBio approach treats palindromic regions has not been systematically analyzed yet. In our own PacBio data, we did not see a palindrome-associated drop of quality during concatemer sequencing [40].

In conclusion, SMS methods offer unprecedented opportunities to study internal concatemer structure when combined with DNA barcoding but require improvements to be useful in the high-throughput analysis of transgenic inserts.

7. Conclusions and Future Direction

Repair of double-strand breaks during zygotic division is very poorly characterized [154,155]. In general, it is considered that the early embryo has high baseline HR rates, because the cell cycle is shifted towards S/G2 phases [156].

Many known integrated DNA reporters could be combined with pronuclear injection to study the dynamics of DNA repair during zygotic division. For example, barcoded transposons [157] or fluorescence-based reporters [158] are used to calculates outcomes of DSB repair (HR, NHEJ, MMEJ) in cells and could be adapted for generation of transgenic mice for such experiments. Alternatively, droplet digital PCR-based assays could provide a convenient tool to study Cas9 indel signatures at any genomic sites in the embryo [159,160]. These reporters could also be used in ES cells or other cell culture types to directly compare repair efficiencies at specific sites with the embryo.

Concatenation assays represent a convenient alternative to integrated reporters and could provide plenty of data for the fundamental studies of DNA repair. The advantages of the pronuclear microinjection over cell culture experiments include an opportunity to choose the timing of zygote development. For example, injections in pre-replicative early pronucleus, late G2-stage pronucleus, or in the two-cell stage embryos show distinct outcomes in gene-targeting experiments [156,161]. A barcoded concatemer system would be a convenient way to evaluate HR/NHEJ activities at these stages.

Finally, barcoded DNA injections could be combined with powerful methods, such as auxin-dependent protein degradation [162], co-injections of the DNA repair factors (Rad51, Rad18, CtIP, Ku) [155,163–165], or small molecule inhibitors [166], to dynamically alter the DNA damage balance and discover factors responsible for transgene end recombination.

From the angle of biotechnology, understanding concatenation mechanisms is highly desired to control the outcome of the experiments both in manufacturing cell lines and in animal transgenesis. For commercial cell lines, rearranged concatenated transgene integrants often result in genetically unstable loci [14], potentially prone to RIGS. Identifying epigenetic factors responsible for RIGS will help to achieve reliable transgene expression through chemical inhibition, as was shown in the extrachromosomal circles model with HDAC inhibitors [6]. Moreover, preventing concatenation altogether by limiting end recombination with biotin [84] or inhibiting HR [167] are promising strategies to avoid rearrangements in the gene-targeting experiments. Lastly, one may even fantasize that better control over concatenation will help to assemble and integrate megabase-sized transgenes directly in an injected zygote in a Gibson cloning style [37,168], possibly using long single-stranded DNA fragments [169,170] that are preferred substrates for HR and less susceptible to random ligation.

Concatenation studies is a nice example of resurrection of the suspended mystery with new genomic technologies, which will undoubtedly bring many new discoveries about DNA recombination in the nucleus.

Author Contributions: Conceptualization, A.S. and N.B.; writing—original draft preparation: A.S.; writing—review and editing, A.S. and N.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Russian Foundation for Basic Research, grant number 20-34-70087. N.B. supported from the Ministry of Education and Science of Russian Federation, grant

#2019-0546 (FSUS-2020-0040). Data analysis performed on computational nodes of the Institute of Cytology and Genetics (budget project no. 0259-2021-0016).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

References

1. Yang, J.; Zhou, W.; Zhang, Y.; Zidon, T.; Ritchie, T.; Engelhardt, J.F. Concatamerization of adeno-associated virus circular genomes occurs through intermolecular recombination. *J. Virol.* **1999**, *73*, 9468–9477. [CrossRef]
2. Maurer, A.C.; Weitzman, M.D. Adeno-Associated Virus Genome Interactions Important for Vector Production and Transduction. *Hum. Gene Ther.* **2020**, *31*, 499–511. [CrossRef] [PubMed]
3. Chen, X.J.; Clark-Walker, G.D. Unveiling the mystery of mitochondrial DNA replication in yeasts. *Mitochondrion* **2018**, *38*, 17–22. [CrossRef]
4. Tomaska, L.; Nosek, J.; Kar, A.; Willcox, S.; Griffith, J.D. A New View of the T-Loop Junction: Implications for Self-Primed Telomere Extension, Expansion of Disease-Related Nucleotide Repeat Blocks, and Telomere Evolution. *Front. Genet.* **2019**, *10*, 792. [CrossRef]
5. Pesenti, E.; Liskovykh, M.; Okazaki, K.; Mallozzi, A.; Reid, C.; Abad, M.A.; Jeyaprakash, A.A.; Kouprina, N.; Larionov, V.; Masumoto, H.; et al. Analysis of Complex DNA Rearrangements during Early Stages of HAC Formation. *ACS Synth. Biol.* **2020**, *9*, 3267–3287. [CrossRef]
6. Shimizu, N. Gene Amplification and the Extrachromosomal Circular DNA. *Genes* **2021**, *12*, 1533. [CrossRef] [PubMed]
7. Bishop, J.O. Chromosomal insertion of foreign DNA. *Reprod. Nutr. Dev.* **1996**, *36*, 607–618.
8. Svitashev, S.K.; Pawlowski, W.P.; Makarevitch, I.; Plank, D.W.; Somers, D.A. Complex transgene locus structures implicate multiple mechanisms for plant transgene rearrangement. *Plant J.* **2002**, *32*, 433–445. [CrossRef]
9. Jupe, F.; Rivkin, A.C.; Michael, T.P.; Zander, M.; Motley, S.T.; Sandoval, J.P.; Keith Slotkin, R.; Chen, H.; Castanon, R.; Nery, J.R.; et al. The complex architecture and epigenomic impact of plant T-DNA insertions. *PLoS Genet.* **2019**, *15*, e1007819. [CrossRef] [PubMed]
10. Mello, C.C.; Kramer, J.M.; Stinchcomb, D.; Ambros, V. Efficient gene transfer in C.elegans: Extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **1991**, *10*, 3959–3970. [CrossRef] [PubMed]
11. Wang, Y.; Wang, Y.; Chang, T.; Huang, H.; Yee, J.-K. Integration-defective lentiviral vector mediates efficient gene editing through homology-directed repair in human embryonic stem cells. *Nucleic Acids Res.* **2017**, *45*, e29. [CrossRef]
12. Chandler, K.J.; Chandler, R.L.; Broeckelmann, E.M.; Hou, Y.; Southard-Smith, E.M.; Mortlock, D.P. Relevance of BAC transgene copy number in mice: Transgene copy number variation across multiple transgenic lines and correlations with transgene integrity and expression. *Mamm. Genome* **2007**, *18*, 693–708. [CrossRef]
13. 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] [PubMed]
14. Rajendran, S.; Balasubramanian, S.; Webster, L.; Lee, M.; Vavilala, D.; Kulikov, N.; Choi, J.; Tang, C.; Hunter, M.; Wang, R.; et al. Accelerating and de-risking CMC development with transposon-derived manufacturing cell lines. *Biotechnol. Bioeng.* **2021**, *118*, 2301–2311. [CrossRef] [PubMed]
15. Danner, E.; Lebedin, M.; de la Rosa, K.; Kühn, R. A homology independent sequence replacement strategy in human cells using a CRISPR nuclease: Replace Targeting. *Open Biol.* **2021**, *11*, 200283. [CrossRef]
16. Boel, A.; De Saffel, H.; Steyaert, W.; Callewaert, B.; De Paepe, A.; Coucke, P.J.; Willaert, A. CRISPR/Cas9-mediated homology-directed repair by ssODNs in zebrafish induces complex mutational patterns resulting from genomic integration of repair-template fragments. *Dis. Model. Mech.* **2018**, *11*, dmm035352. [CrossRef] [PubMed]
17. Skryabin, B.V.; Kummerfeld, D.-M.; Gubar, L.; Seeger, B.; Kaiser, H.; Stegemann, A.; Roth, J.; Meuth, S.G.; Pavenstädt, H.; Sherwood, J.; et al. Pervasive head-to-tail insertions of DNA templates mask desired CRISPR-Cas9–mediated genome editing events. *Sci. Adv.* **2020**, *6*, eaax2941. [CrossRef]
18. Pu, X.; Young, A.P.; Kubisch, H.M. Production of transgenic mice by pronuclear microinjection. In *Methods in Molecular Biology*; Humana Press: New York, NY, USA, 2019; Volume 1874, pp. 17–41.
19. Gordon, J.W.; Scangos, G.A.; Plotkin, D.J.; Barbosa, J.A.; Ruddle, F.H. Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc. Natl. Acad. Sci. USA* **1980**, *77*, 7380–7384. [CrossRef] [PubMed]
20. Brinster, R.L.; Chen, H.Y.; Trumbauer, M.; Senear, A.W.; Warren, R.; Palmiter, R.D. Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. *Cell* **1981**, *27*, 223–231. [CrossRef]
21. Smith, K. Theoretical mechanisms in targeted and random integration of transgene DNA. *Reprod. Nutr. Dev.* **2001**, *41*, 465–485. [CrossRef]
22. 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]
23. Palmiter, R.D.; Chen, H.Y.; Brinster, R.L. Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring. *Cell* **1982**, *29*, 701–710. [CrossRef]

24. Brinster, R.L.; Ritchie, K.A.; Hammer, R.E.; O'Brien, R.L.; Arp, B.; Storb, U. Expression of a microinjected immunoglobulin gene in the spleen of transgenic mice. *Nature* **1983**, *306*, 332–336. [CrossRef] [PubMed]
25. Khillan, J.S.; Overbeek, P.A.; Westphal, H. Drosophila P element integration in the mouse. *Dev. Biol.* **1985**, *109*, 247–250. [CrossRef]
26. Overbeek, P.A.; Lai, S.-P.; Van Quill, K.R.; Westphal, H. Tissue-Specific Expression in Transgenic Mice of a Fused Gene Containing RSV Terminal Sequences. *Science* **1986**, *231*, 1574–1577. [CrossRef]
27. Folger, K.R.; Wong, E.A.; Wahl, G.; Capecchi, M.R. Patterns of integration of DNA microinjected into cultured mammalian cells: Evidence for homologous recombination between injected plasmid DNA molecules. *Mol. Cell. Biol.* **1982**, *2*, 1372–1387. [CrossRef]
28. Folger, K.R.; Thomas, K.; Capecchi, M.R. Nonreciprocal exchanges of information between DNA duplexes coinjected into mammalian cell nuclei. *Mol. Cell. Biol.* **1985**, *5*, 59–69. [CrossRef] [PubMed]
29. Brouillette, S.; Chartrand, P. Intermolecular recombination assay for mammalian cells that produces recombinants carrying both homologous and nonhomologous junctions. *Mol. Cell. Biol.* **1987**, *7*, 2248–2255. [CrossRef]
30. Gening, L.; Takeshita, M.; Levine, R.L.; Peden, K.W.; Grollman, A.P. Extrachromosomal unequal homologous recombination and gene conversion in simian kidney cells: Effects of UV damage. *Mutat. Res.-DNA Repair* **1998**, *407*, 11–24. [CrossRef]
31. Rohan, R.M.; King, D.; Frels, W.I. Direct sequencing of PCR-amplified junction fragments from tandemly repeated transgenes. *Nucleic Acids Res.* **1990**, *18*, 6089–6095. [CrossRef]
32. Brinster, R.L.; Chen, H.Y.; Trumbauer, M.E.; Yagle, M.K.; Palmiter, R.D. Factors affecting the efficiency of introducing foreign DNA into mice by microinjecting eggs. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 4438–4442. [CrossRef] [PubMed]
33. Hamada, T.; Sasaki, H.; Seki, R.; Sakaki, Y. Mechanism of chromosomal integration of transgenes in microinjected mouse eggs: Sequence analysis of genome-transgene and transgene-transgene junctions at two loci. *Gene* **1993**, *128*, 197–202. [CrossRef]
34. Würtele, H.; Gusew, N.; Lussier, R.; Chartrand, P. Characterization of in vivo recombination activities in the mouse embryo. *Mol. Genet. Genomics* **2005**, *273*, 252–263. [CrossRef]
35. Pieper, F.R.; Wit, I.C.M.D.; Pronk, A.C.J.; Kooiman, P.M.; Strijker, R.; Krimpenfort, P.J.A.; Nuyens, J.H.; Boer, H.A.D. Efficient generation of functional transgenes by homologous recombination in murine zygotes. *Nucleic Acids Res.* **1992**, *20*, 1259–1264. [CrossRef]
36. Migchielsen, A.A.J.; Breuer, M.L.; Hershfield, M.S.; Valerio, D. Full genetic rescue of adenosine deaminase-deficient mice through introduction of the human gene. *Hum. Mol. Genet.* **1996**, *5*, 1523–1532. [CrossRef] [PubMed]
37. Tacken, P.J.; Zee, A.V.D.; Beumer, T.L.; Florijn, R.J.; Gijpels, M.J.J.; Havekes, L.M.; Frants, R.R.; Dijk, K.W.V.; Hofker, M.H. Effective generation of very low density lipoprotein receptor transgenic mice by overlapping genomic DNA fragments: High testis expression and disturbed spermatogenesis. *Transgenic Res.* **2001**, *10*, 211–221. [CrossRef] [PubMed]
38. Ramsden, D.A.; Carvajal-Garcia, J.; Gupta, G.P. Mechanism, cellular functions and cancer roles of polymerase-theta-mediated DNA end joining. *Nat. Rev. Mol. Cell Biol.* **2021**. [CrossRef] [PubMed]
39. Truong, L.N.; Li, Y.; Shi, L.Z.; Hwang, P.Y.-H.; He, J.; Wang, H.; Razavian, N.; Berns, M.W.; Wu, X. Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 7720–7725. [CrossRef] [PubMed]
40. 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]
41. McVey, M.; Lee, S.E. MMEJ repair of double-strand breaks (director's cut): Deleted sequences and alternative endings. *Trends Genet.* **2008**, *24*, 529–538. [CrossRef]
42. Schimmel, J.; Kool, H.; van Schendel, R.; Tijsterman, M. Mutational signatures of non-homologous and polymerase theta-mediated end-joining in embryonic stem cells. *EMBO J.* **2017**, *36*, 3634–3649. [CrossRef]
43. Dai, J.; Cui, X.; Zhu, Z.; Hu, W. Non-Homologous End Joining Plays a Key Role in Transgene Concatemer Formation in Transgenic Zebrafish Embryos. *Int. J. Biol. Sci.* **2010**, *6*, 756–768. [CrossRef] [PubMed]
44. Pawlik, K.M.; Sun, C.-W.; Higgins, N.P.; Townes, T.M. End joining of genomic DNA and transgene DNA in fertilized mouse eggs. *Gene* **1995**, *165*, 173–181. [CrossRef]
45. Suemizu, H.; Muguruma, K.; Maruyama, C.; Tomisawa, M.; Kimura, M.; Hioki, K.; Shimozawa, N.; Ohnishi, Y.; Tamaoki, N.; Nomura, T. Transgene stability and features of rasH2 mice as an animal model for short-term carcinogenicity testing. *Mol. Carcinog.* **2002**, *34*, 1–9. [CrossRef]
46. 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] [PubMed]
47. Yan, B.; Li, D.; Gou, K. Homologous illegitimate random integration of foreign DNA into the X chromosome of a transgenic mouse line. *BMC Mol. Biol.* **2010**, *11*, 58. [CrossRef]
48. Nakade, S.; Tsubota, T.; Sakane, Y.; Kume, S.; Sakamoto, N.; Obara, M.; Daimon, T.; Sezutsu, H.; Yamamoto, T.; Sakuma, T.; et al. Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun.* **2014**, *5*, 5560. [CrossRef]
49. Masumura, K.; Sakamoto, Y.; Kumita, W.; Honma, M.; Nishikawa, A.; Nohmi, T. Genomic integration of lambda EG10 transgene in gpt delta transgenic rodents. *Genes Environ.* **2015**, *37*, 24. [CrossRef]
50. Mark, W.H.; Signorelli, K.; Blum, M.; Kwee, L.; Lacy, E. Genomic structure of the locus associated with an insertional mutation in line 4 transgenic mice. *Genomics* **1992**, *13*, 159–166. [CrossRef]

51. Murnane, J.P.; Yu, L.C. Acquisition of telomere repeat sequences by transfected DNA integrated at the site of a chromosome break. *Mol. Cell. Biol.* **1993**, *13*, 977–983. [CrossRef]
52. 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] [PubMed]
53. Kang, Y.K.; Park, J.S.; Lee, C.S.; Yeom, Y.I.; Chung, A.S.; Lee, K.K. Efficient integration of short interspersed element-flanked foreign DNA via homologous recombination. *J. Biol. Chem.* **1999**, *274*, 36585–36591. [CrossRef]
54. Ganapathiraju, M.K.; Subramanian, S.; Chaparala, S.; Karunakaran, K.B. A reference catalog of DNA palindromes in the human genome and their variations in 1000 Genomes. *Hum. Genome Var.* **2020**, *7*, 40. [CrossRef]
55. Ditch, S.; Sammarco, M.C.; Banerjee, A.; Grabczyk, E. Progressive GAA·TTC repeat expansion in human cell lines. *PLoS Genet.* **2009**, *5*, e1000704. [CrossRef] [PubMed]
56. Ramakrishnan, S.; Kockler, Z.; Evans, R.; Downing, B.D.; Malkova, A. Single-strand annealing between inverted DNA repeats: Pathway choice, participating proteins, and genome destabilizing consequences. *PLoS Genet.* **2018**, *14*, e1007543. [CrossRef] [PubMed]
57. Tanaka, S.S.; Mitsuda, S.H.; Shimizu, N. How a replication origin and matrix attachment region accelerate gene amplification under replication stress in mammalian cells. *PLoS ONE* **2014**, *9*, e103439. [CrossRef] [PubMed]
58. Hasty, P.; Montagna, C. Chromosomal rearrangements in cancer. *Mol. Cell. Oncol.* **2014**, *1*, e29904. [CrossRef]
59. Trombetta, B.; Cruciani, F. Y chromosome palindromes and gene conversion. *Hum. Genet.* **2017**, *136*, 605–619. [CrossRef] [PubMed]
60. Jackson, E.K.; Bellott, D.W.; Cho, T.J.; Skaletsky, H.; Hughes, J.F.; Pyntikova, T.; Page, D.C. Large palindromes on the primate X Chromosome are preserved by natural selection. *Genome Res.* **2021**, *31*, 1337–1352. [CrossRef]
61. Inagaki, H.; Ohye, T.; Kogo, H.; Tsutsumi, M.; Kato, T.; Tong, M.; Emanuel, B.S.; Kurahashi, H. Two sequential cleavage reactions on cruciform DNA structures cause palindrome-mediated chromosomal translocations. *Nat. Commun.* **2013**, *4*, 1592. [CrossRef]
62. Inagaki, H.; Kato, T.; Tsutsumi, M.; Ouchi, Y.; Ohye, T.; Kurahashi, H. Palindrome-mediated translocations in humans: A new mechanistic model for gross chromosomal rearrangements. *Front. Genet.* **2016**, *7*, 125. [CrossRef]
63. Voineagu, I.; Narayanan, V.; Lobachev, K.S.; Mirkin, S.M. Replication stalling at unstable inverted repeats: Interplay between DNA hairpins and fork stabilizing proteins. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 9936–9941. [CrossRef] [PubMed]
64. Wang, G.; Vasquez, K.M. Effects of replication and transcription on DNA Structure-Related genetic instability. *Genes* **2017**, *8*, 17. [CrossRef]
65. Holkers, M.; De Vries, A.A.F.; Gonçalves, M.A.F.V. Nonspaced inverted DNA repeats are preferential targets for homology-directed gene repair in mammalian cells. *Nucleic Acids Res.* **2012**, *40*, 1984–1999. [CrossRef]
66. Chen, H.; Lisby, M.; Symington, L. RPA Coordinates DNA End Resection and Prevents Formation of DNA Hairpins. *Mol. Cell* **2013**, *50*, 589–600. [CrossRef] [PubMed]
67. Brázda, V.; Hároníková, L.; Liao, J.C.C.; Fridrichová, H.; Jagelská, E.B. Strong preference of BRCA1 protein to topologically constrained non-B DNA structures. *BMC Mol. Biol.* **2016**, *17*, 14. [CrossRef] [PubMed]
68. Brázda, V.; Coufal, J. Recognition of local DNA structures by p53 protein. *Int. J. Mol. Sci.* **2017**, *18*, 375. [CrossRef]
69. Lobachev, K.S.; Gordenin, D.A.; Resnick, M.A. The Mre11 complex is required for repair of hairpin-capped double-strand breaks and prevention of chromosome rearrangements. *Cell* **2002**, *108*, 183–193. [CrossRef]
70. Ait Saada, A.; Costa, A.B.; Sheng, Z.; Guo, W.; Haber, J.E.; Lobachev, K.S. Structural parameters of palindromic repeats determine the specificity of nuclease attack of secondary structures. *Nucleic Acids Res.* **2021**, *49*, 3932–3947. [CrossRef]
71. Rass, U.; Compton, S.A.; Matos, J.; Singleton, M.R.; Ip, S.C.Y.; Blanco, M.G.; Griffith, J.D.; West, S.C. Mechanism of Holliday junction resolution by the human GEN1 protein. *Genes Dev.* **2010**, *24*, 1559–1569. [CrossRef] [PubMed]
72. Kirschner, K.; Melton, D.W. Multiple roles of the ERCC1-XPF endonuclease in DNA repair and resistance to anticancer drugs. *Anticancer Res.* **2010**, *30*, 3223–3232.
73. Halabi, A.; Fuselier, K.T.B.; Grabczyk, E. GAA•TTC repeat expansion in human cells is mediated by mismatch repair complex MutL and depends upon the endonuclease domain in MLH3 isoform one. *Nucleic Acids Res.* **2018**, *46*, 4022–4032. [CrossRef]
74. Svetec Miklenić, M.; Gatalica, N.; Matanović, A.; Žunar, B.; Štafa, A.; Lisnić, B.; Svetec, I.K. Size-dependent antirecombinogenic effect of short spacers on palindrome recombinogenicity. *DNA Repair* **2020**, *90*, 102848. [CrossRef]
75. Zhou, Z.H.; Akgün, E.; Jasin, M. Repeat expansion by homologous recombination in the mouse germ line at palindromic sequences. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 8326–8333. [CrossRef]
76. Akgün, E.; Zahn, J.; Baumes, S.; Brown, G.; Liang, F.; Romanienko, P.J.; Lewis, S.; Jasin, M. Palindrome resolution and recombination in the mammalian germ line. *Mol. Cell. Biol.* **1997**, *17*, 5559–5570. [CrossRef]
77. Cunningham, L.A.; Coté, A.G.; Cam-Ozdemir, C.; Lewis, S.M. Rapid, Stabilizing Palindrome Rearrangements in Somatic Cells by the Center-Break Mechanism. *Mol. Cell. Biol.* **2003**, *23*, 8740–8750. [CrossRef]
78. Mikhailov, K.V.; Efeykin, B.D.; Panchin, A.Y.; Knorre, D.A.; Logacheva, M.D.; Penin, A.A.; Muntyan, M.S.; Nikitin, M.A.; Popova, O.V.; Zanegina, O.N.; et al. Coding palindromes in mitochondrial genes of Nematomorpha. *Nucleic Acids Res.* **2019**, *47*, 6858–6870. [CrossRef]
79. Daley, J.M.; Gaines, W.A.; Kwon, Y.; Sung, P. Regulation of DNA Pairing in Homologous Recombination. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a017954. [CrossRef]

80. 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]
81. 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]
82. Ranjha, L.; Howard, S.M.; Cejka, P. Main steps in DNA double-strand break repair: An introduction to homologous recombination and related processes. *Chromosoma* **2018**, *127*, 187–214. [CrossRef] [PubMed]
83. Bizard, A.H.; Hickson, I.D. The dissolution of double Holliday junctions. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a016477. [CrossRef]
84. Gutierrez-Triana, J.A.; Tavhelidse, T.; Thumberger, T.; Thomas, I.; Wittbrodt, B.; Kellner, T.; Anlas, K.; Tsingos, E.; Wittbrodt, J. Efficient single-copy HDR by 5′ modified long dsDNA donors. *Elife* **2018**, *7*, e39468. [CrossRef]
85. Canaj, H.; Hussmann, J.A.; Li, H.; Beckman, K.A.; Goodrich, L.; Cho, N.H.; Li, Y.J.; Santos, D.A.; McGeever, A.; Stewart, E.M.; et al. Deep profiling reveals substantial heterogeneity of integration outcomes in CRISPR knock-in experiments. *bioRxiv* **2019**. [CrossRef]
86. Hartlerode, A.J.; Willis, N.A.; Rajendran, A.; Manis, J.P.; Scully, R. Complex Breakpoints and Template Switching Associated with Non-canonical Termination of Homologous Recombination in Mammalian Cells. *PLOS Genet.* **2016**, *12*, e1006410. [CrossRef]
87. Wyatt, D.W.; Feng, W.; Conlin, M.P.; Yousefzadeh, M.J.; Roberts, S.A.; Mieczkowski, P.; Wood, R.D.; Gupta, G.P.; Ramsden, D.A. Essential Roles for Polymerase θ-Mediated End Joining in the Repair of Chromosome Breaks. *Mol. Cell* **2016**, *63*, 662–673. [CrossRef]
88. Shibata, A. Regulation of repair pathway choice at two-ended DNA double-strand breaks. *Mutat. Res. Mol. Mech. Mutagen.* **2017**, *803–805*, 51–55. [CrossRef] [PubMed]
89. Paix, A.; Folkmann, A.; Goldman, D.H.; Kulaga, H.; Grzelak, M.J.; Rasoloson, D.; Paidemarry, S.; Green, R.; Reed, R.R.; Seydoux, G. Precision genome editing using synthesis-dependent repair of Cas9-induced DNA breaks. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E10745–E10754. [CrossRef]
90. Verma, P.; Greenberg, R.A. Noncanonical views of homology-directed DNA repair. *Genes Dev.* **2016**, *30*, 1138–1154. [CrossRef]
91. Piazza, A.; Heyer, W.-D. Homologous Recombination and the Formation of Complex Genomic Rearrangements. *Trends Cell Biol.* **2019**, *29*, 135–149. [CrossRef] [PubMed]
92. Piazza, A.; Wright, W.D.; Heyer, W.-D. Multi-invasions Are Recombination Byproducts that Induce Chromosomal Rearrangements. *Cell* **2017**, *170*, 760–773.e15. [CrossRef]
93. Zhou, Y.; Caron, P.; Legube, G.; Paull, T.T. Quantitation of DNA double-strand break resection intermediates in human cells. *Nucleic Acids Res.* **2014**, *42*, e19. [CrossRef] [PubMed]
94. Fus-Kujawa, A.; Prus, P.; Bajdak-Rusinek, K.; Teper, P.; Gawron, K.; Kowalczuk, A.; Sieron, A.L. An Overview of Methods and Tools for Transfection of Eukaryotic Cells in vitro. *Front. Bioeng. Biotechnol.* **2021**, *9*, 634. [CrossRef]
95. Dorer, D.R.; Henikoff, S. Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. *Cell* **1994**, *77*, 993–1002. [CrossRef]
96. Catalanotto, C.; Nolan, T.; Cogoni, C. Homology effects in Neurospora crassa. *FEMS Microbiol. Lett.* **2006**, *254*, 182–189. [CrossRef]
97. Ye, F.; Signer, E.R. RIGS (repeat-induced gene silencing) in Arabidopsis is transcriptional and alters chromatin configuration. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 10881–10886. [CrossRef]
98. Henikoff, S. Conspiracy of silence among repeated transgenes. *BioEssays* **1998**, *20*, 532–535. [CrossRef]
99. Garrick, D.; Fiering, S.; Martin, D.I.K.; Whitelaw, E. Repeat-induced gene silencing in mammals. *Nat. Genet.* **1998**, *8*, 56–59. [CrossRef]
100. Girton, J.R.; Johansen, K.M. Chapter 1 Chromatin Structure and the Regulation of Gene Expression: The Lessons of PEV in Drosophila. *Adv. Genet.* **2008**, *61*, 1–43.
101. Sidorenko, L.V.; Lee, T.F.; Woosley, A.; Moskal, W.A.; Bevan, S.A.; Merlo, P.A.O.; Walsh, T.A.; Wang, X.; Weaver, S.; Glancy, T.P.; et al. GC-rich coding sequences reduce transposon-like, small RNA-mediated transgene silencing. *Nat. Plants* **2017**, *3*, 875–884. [CrossRef]
102. Fu, X.; Kohli, A.; Twyman, R.M.; Christou, P. Alternative silencing effects involve distinct types of non-spreading cytosine methylation at a three-gene, single-copy transgenic locus in rice. *Mol. Gen. Genet.* **2000**, *263*, 106–118. [CrossRef]
103. Kostyrko, K.; Neuenschwander, S.; Junier, T.; Regamey, A.; Iseli, C.; Schmid-Siegert, E.; Bosshard, S.; Majocchi, S.; Le Fourn, V.; Girod, P.-A.; et al. MAR-Mediated transgene integration into permissive chromatin and increased expression by recombination pathway engineering. *Biotechnol. Bioeng.* **2017**, *114*, 384–396. [CrossRef]
104. Eszterhas, S.K.; Bouhassira, E.E.; Martin, D.I.K.; Fiering, S. Transcriptional Interference by Independently Regulated Genes Occurs in Any Relative Arrangement of the Genes and Is Influenced by Chromosomal Integration Position. *Mol. Cell. Biol.* **2002**, *22*, 469–479. [CrossRef]
105. Carver, J.; Ng, D.; Zhou, M.; Ko, P.; Zhan, D.; Yim, M.; Shaw, D.; Snedecor, B.; Laird, M.W.; Lang, S.; et al. Maximizing antibody production in a targeted integration host by optimization of subunit gene dosage and position. *Biotechnol. Prog.* **2020**, *36*, e2967. [CrossRef]

106. Zimak, J.; Wagoner, Z.W.; Nelson, N.; Waechtler, B.; Schlosser, H.; Kopecky, M.; Wu, J.; Zhao, W. Epigenetic silencing directs expression heterogeneity of stably integrated multi-transcript unit genetic circuits. *Sci. Rep.* **2021**, *11*, 2424. [CrossRef]
107. Rosser, J.M.; An, W. Repeat-induced gene silencing of L1 transgenes is correlated with differential promoter methylation. *Gene* **2010**, *456*, 15–23. [CrossRef]
108. Williams, A.; Harker, N.; Ktistaki, E.; Veiga-fernandes, H.; Roderick, K.; Tolaini, M.; Norton, T.; Williams, K.; Kioussis, D. Position effect variegation and imprinting of transgenes in lymphocytes. *Nucleic Acids Res.* **2008**, *36*, 2320–2329. [CrossRef]
109. Alonso-González, L.; Couldrey, C.; Meinhardt, M.W.; Cole, S.A.; Wells, D.N.; Laible, G. Primary transgenic bovine cells and their rejuvenated cloned equivalents show transgene-specific epigenetic differences. *PLoS ONE* **2012**, *7*, e35619. [CrossRef]
110. Ordovás, L.; Boon, R.; Pistoni, M.; Chen, Y.; Wolfs, E.; Guo, W.; Sambathkumar, R.; Bobis-Wozowicz, S.; Helsen, N.; Vanhove, J.; et al. Efficient recombinase-mediated cassette exchange in hPSCs to study the hepatocyte lineage reveals AAVS1 locus-mediated transgene inhibition. *Stem Cell Rep.* **2015**, *5*, 918–931. [CrossRef]
111. Gödecke, N.; Zha, L.; Spencer, S.; Behme, S.; Riemer, P.; Rehli, M.; Hauser, H.; Wirth, D. Controlled re-activation of epigenetically silenced Tet promoter-driven transgene expression by targeted demethylation. *Nucleic Acids Res.* **2017**, *45*, e147. [CrossRef]
112. Klatt, D.; Cheng, E.; Hoffmann, D.; Santilli, G.; Thrasher, A.J.; Brendel, C.; Schambach, A. Differential Transgene Silencing of Myeloid-Specific Promoters in the AAVS1 Safe Harbor Locus of Induced Pluripotent Stem Cell-Derived Myeloid Cells. *Hum. Gene Ther.* **2020**, *31*, 199–210. [CrossRef]
113. Luo, Z.; Chen, Z. Improperly terminated, unpolyadenylated mRNA of sense transgenes is targeted by RDR6-mediated RNA silencing in Arabidopsis. *Plant Cell* **2007**, *19*, 943–958. [CrossRef]
114. Pérez-González, A.; Caro, E. Effect of transcription terminator usage on the establishment of transgene transcriptional gene silencing. *BMC Res. Notes* **2018**, *11*, 511. [CrossRef]
115. Klattenhoff, C.; Xi, H.; Li, C.; Lee, S.; Xu, J.; Khurana, J.S.; Zhang, F.; Schultz, N.; Koppetsch, B.S.; Nowosielska, A.; et al. The Drosophila HP1 Homolog Rhino Is Required for Transposon Silencing and piRNA Production by Dual-Strand Clusters. *Cell* **2009**, *138*, 1137–1149. [CrossRef]
116. Tsukamoto, T.; Hashiguchi, N.; Janicki, S.M.; Tumbar, T.; Belmont, A.S.; Spector, D.L. Visualization of gene activity in living cells. *Nat. Cell Biol.* **2000**, *2*, 871–878. [CrossRef]
117. Wang, F.; Koyama, N.; Nishida, H.; Haraguchi, T.; Reith, W.; Tsukamoto, T. The Assembly and Maintenance of Heterochromatin Initiated by Transgene Repeats Are Independent of the RNA Interference Pathway in Mammalian Cells. *Mol. Cell. Biol.* **2006**, *26*, 4028–4040. [CrossRef]
118. Mazur, A.K.; Gladyshev, E. Partition of Repeat-Induced Point Mutations Reveals Structural Aspects of Homologous DNA-DNA Pairing. *Biophys. J.* **2018**, *115*, 605–615. [CrossRef]
119. McBurney, M.W.; Mai, T.; Yang, X.; Jardine, K. Evidence for Repeat-Induced Gene Silencing in Cultured Mammalian Cells: Inactivation of Tandem Repeats of Transfected Genes. *Exp. Cell Res.* **2002**, *274*, 1–8. [CrossRef]
120. Calero-Garcia, M.; Gaspar, H.B. Gene-ectomy: Gene ablation with CRISPR/Cas9 in human hematopoietic cells. *Cell Stem Cell* **2014**, *15*, 529–530. [CrossRef]
121. Osterlehner, A.; Simmeth, S.; Göpfert, U. Promoter methylation and transgene copy numbers predict unstable protein production in recombinant chinese hamster ovary cell lines. *Biotechnol. Bioeng.* **2011**, *108*, 2670–2681. [CrossRef] [PubMed]
122. Jones, P.A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **2012**, *13*, 484–492. [CrossRef] [PubMed]
123. Sanchez-Luque, F.J.; Kempen, M.J.H.C.; Gerdes, P.; Vargas-Landin, D.B.; Richardson, S.R.; Troskie, R.L.; Jesuadian, J.S.; Cheetham, S.W.; Carreira, P.E.; Salvador-Palomeque, C.; et al. LINE-1 Evasion of Epigenetic Repression in Humans. *Mol. Cell* **2019**, *75*, 590–604.e12. [CrossRef] [PubMed]
124. Fukuma, M.; Ganmyo, Y.; Miura, O.; Ohyama, T.; Shimizu, N. Cloning and characterization of a human genomic sequence that alleviates repeat-induced gene silencing. *PLoS ONE* **2016**, *11*, e0153338. [CrossRef] [PubMed]
125. Yan, H.; Xiang, X.; Chen, Q.; Pan, X.; Cheng, H.; Wang, F. HP1 cooperates with CAF-1 to compact heterochromatic transgene repeats in mammalian cells. *Sci. Rep.* **2018**, *8*, 14141. [CrossRef]
126. Janicki, S.M.; Tsukamoto, T.; Salghetti, S.E.; Tansey, W.P.; Sachidanandam, R.; Prasanth, K.V.; Ried, T.; Shav-Tal, Y.; Bertrand, E.; Singer, R.H.; et al. From silencing to gene expression: Real-time analysis in single cells. *Cell* **2004**, *116*, 683–698. [CrossRef]
127. Taniguchi, R.; Utani, K.; Thakur, B.; Ishine, K.; Aladjem, M.I.; Shimizu, N. SIRT1 stabilizes extrachromosomal gene amplification and contributes to repeat-induced gene silencing. *J. Biol. Chem.* **2021**, *296*, 100356. [CrossRef] [PubMed]
128. Wen, J.; Wu, J.; Cao, T.; Zhi, S.; Chen, Y.; Aagaard, L.; Zhen, P.; Huang, Y.; Zhong, J.; Huang, J. Methylation silencing and reactivation of exogenous genes in lentivirus-mediated transgenic mice. *Transgenic Res.* **2021**, *30*, 63–76. [CrossRef]
129. 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*, gkw1329. [CrossRef]
130. 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]
131. Brueckner, L.; Zhao, P.A.; Schaik, T.; Leemans, C.; Sima, J.; Peric-Hupkes, D.; Gilbert, D.M.; Steensel, B. Local rewiring of genome–nuclear lamina interactions by transcription. *EMBO J.* **2020**, *39*, e103159. [CrossRef]

132. Maksimenko, O.; Gasanov, N.B.; Georgiev, P. Regulatory Elements in Vectors for Efficient Generation of Cell Lines Producing Target Proteins. *Acta Nat.* **2015**, *7*, 15–26. [CrossRef]
133. Lau, S.; Jardine, K.; McBurney, M.W. DNA methylation pattern of a tandemly repeated lacZ transgene indicates that most copies are silent. *Dev. Dyn.* **1999**, *215*, 126–138. [CrossRef]
134. Xu, L.; Seki, M. Recent advances in the detection of base modifications using the Nanopore sequencer. *J. Hum. Genet.* **2020**, *65*, 25–33. [CrossRef]
135. Herzog, V.A.; Reichholf, B.; Neumann, T.; Rescheneder, P.; Bhat, P.; Burkard, T.R.; Wlotzka, W.; von Haeseler, A.; Zuber, J.; Ameres, S.L. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* **2017**, *14*, 1198–1204. [CrossRef]
136. 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]
137. Watson, C.M.; Crinnion, L.A.; Hewitt, S.; Bates, J.; Robinson, R.; Carr, I.M.; Sheridan, E.; Adlard, J.; Bonthron, D.T. Cas9-based enrichment and single-molecule sequencing for precise characterization of genomic duplications. *Lab. Investig.* **2020**, *100*, 135–146. [CrossRef]
138. 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]
139. Wilkie, T.M.; Palmiter, R.D. Analysis of the integrant in MyK-103 transgenic mice in which males fail to transmit the integrant. *Mol. Cell. Biol.* **1987**, *7*, 1646–1655. [CrossRef] [PubMed]
140. 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]
141. Van Dijk, E.L.; Jaszczyszyn, Y.; Naquin, D.; Thermes, C. The Third Revolution in Sequencing Technology. *Trends Genet.* **2018**, *34*, 666–681. [CrossRef]
142. 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]
143. Payne, A.; Holmes, N.; Rakyan, V.; Loose, M. Bulkvis: A graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* **2019**, *35*, 2193–2198. [CrossRef] [PubMed]
144. 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]
145. 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] [PubMed]
146. Slesarev, A.; Viswanathan, L.; Tang, Y.; Borgschulte, T.; Achtien, K.; Razafsky, D.; Onions, D.; Chang, A.; Cote, C. CRISPR/Cas9 targeted CAPTURE of mammalian genomic regions for characterization by NGS. *Sci. Rep.* **2019**, *9*, 3587. [CrossRef]
147. 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]
148. Buck, D.; Weirather, J.L.; de Cesare, M.; Wang, Y.; Piazza, P.; Sebastiano, V.; Wang, X.J.; Au, K.F. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Research* **2017**, *6*, 100. [CrossRef]
149. Delahaye, C.; Nicolas, J. Sequencing DNA with nanopores: Troubles and biases. *PLoS ONE* **2021**, *16*, e0257521. [CrossRef]
150. Ardui, S.; Ameur, A.; Vermeesch, J.R.; Hestand, M.S. Single molecule real-time (SMRT) sequencing comes of age: Applications and utilities for medical diagnostics. *Nucleic Acids Res.* **2018**, *46*, 2159–2168. [CrossRef]
151. Wenger, A.M.; Peluso, P.; Rowell, W.J.; Chang, P.C.; Hall, R.J.; Concepcion, G.T.; Ebler, J.; Fungtammasan, A.; Kolesnikov, A.; Olson, N.D.; et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nat. Biotechnol.* **2019**, *37*, 1155–1162. [CrossRef]
152. Wick, R.R.; Judd, L.M.; Holt, K.E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* **2019**, *20*, 129. [CrossRef]
153. Spealman, P.; Burrell, J.; Gresham, D. Inverted duplicate DNA sequences increase translocation rates through sequencing nanopores resulting in reduced base calling accuracy. *Nucleic Acids Res.* **2020**, *48*, 4940–4945. [CrossRef] [PubMed]
154. Khokhlova, E.V.; Fesenko, Z.S.; Sopova, J.V.; Leonova, E.I. Features of dna repair in the early stages of mammalian embryonic development. *Genes* **2020**, *1*, 1138. [CrossRef]
155. Wilde, J.J.; Aida, T.; del Rosario, R.C.H.; Kaiser, T.; Qi, P.; Wienisch, M.; Zhang, Q.; Colvin, S.; Feng, G. Efficient embryonic homozygous gene conversion via RAD51-enhanced interhomolog repair. *Cell* **2021**, *184*, 3267–3280.e18. [CrossRef] [PubMed]
156. Abe, T.; Inoue, K.-I.; Furuta, Y.; Kiyonari, H. Pronuclear Microinjection during S-Phase Increases the Efficiency of CRISPR-Cas9-Assisted Knockin of Large DNA Donors in Mouse Zygotes. *Cell Rep.* **2020**, *31*, 107653. [CrossRef]
157. Schep, R.; Brinkman, E.K.; Leemans, C.; Vergara, X.; van der Weide, R.H.; Morris, B.; van Schaik, T.; Manzo, S.G.; Peric-Hupkes, D.; van den Berg, J.; et al. Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol. Cell* **2021**, *81*, 2216–2230.e10. [CrossRef]
158. Roidos, P.; Sungalee, S.; Benfatto, S.; Serçin, Ö.; Stütz, A.M.; Abdollahi, A.; Mauer, J.; Zenke, F.T.; Korbel, J.O.; Mardin, B.R. A scalable CRISPR/Cas9-based fluorescent reporter assay to study DNA double-strand break repair choice. *Nat. Commun.* **2020**, *1*, 1–15. [CrossRef]

159. Miyaoka, Y.; Mayerl, S.J.; Chan, A.H.; Conklin, B.R. Detection and Quantification of HDR and NHEJ Induced by Genome Editing at Endogenous Gene Loci Using Droplet Digital PCR. *Methods Mol. Biol.* **2018**, *1768*, 349–362. [CrossRef]
160. Watry, H.L.; Feliciano, C.M.; Gjoni, K.; Takahashi, G.; Miyaoka, Y.; Conklin, B.R.; Judge, L.M. Rapid, precise quantification of large DNA excisions and inversions by ddPCR. *Sci. Rep.* **2020**, *10*, 14896. [CrossRef]
161. Gu, B.; Posfai, E.; Rossant, J. Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. *Nat. Biotechnol.* **2018**, *36*, 632–637. [CrossRef]
162. Yesbolatova, A.; Saito, Y.; Kitamoto, N.; Makino-Itou, H.; Ajima, R.; Nakano, R.; Nakaoka, H.; Fukui, K.; Gamo, K.; Tominari, Y.; et al. The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat. Commun.* **2020**, *11*, 5701. [CrossRef]
163. Morozov, V.; Wawrousek, E.F. Single-strand DNA-mediated targeted mutagenesis of genomic DNA in early mouse embryos is stimulated by Rad51/54 and by Ku70/86 inhibition. *Gene Ther.* **2008**, *15*, 468–472. [CrossRef]
164. Canny, M.D.; Moatti, N.; Wan, L.C.K.; Fradet-Turcotte, A.; Krasner, D.; Mateos-Gomez, P.A.; Zimmermann, M.; Orthwein, A.; Juang, Y.C.; Zhang, W.; et al. Inhibition of 53BP1 favors homology-dependent DNA repair and increases CRISPR-Cas9 genome-editing efficiency. *Nat. Biotechnol.* **2018**, *36*, 95–102. [CrossRef]
165. Nambiar, T.S.; Billon, P.; Diedenhofen, G.; Hayward, S.B.; Taglialatela, A.; Cai, K.; Huang, J.-W.; Leuzzi, G.; Cuella-Martin, R.; Palacios, A.; et al. Stimulation of CRISPR-mediated homology-directed repair by an engineered RAD18 variant. *Nat. Commun.* **2019**, *10*, 3395. [CrossRef]
166. Cao, X.; Kouyama-Suzuki, E.; Pang, B.; Kurihara, T.; Mori, T.; Yanagawa, T.; Shirai, Y.; Tabuchi, K. Inhibition of DNA ligase IV enhances the CRISPR/Cas9-mediated knock-in efficiency in mouse brain neurons. *Biochem. Biophys. Res. Commun.* **2020**, *533*, 449–457. [CrossRef]
167. Yu, C.; Liu, Y.; Ma, T.; Liu, K.; Xu, S.; Zhang, Y.; Liu, H.; La Russa, M.; Xie, M.; Ding, S.; et al. Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell Stem Cell* **2015**, *16*, 142–147. [CrossRef]
168. Gibson, D.G.; Young, L.; Chuang, R.Y.; Venter, J.C.; Hutchison, C.A.; Smith, H.O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **2009**, *6*, 343–345. [CrossRef]
169. Minev, D.; Guerra, R.; Kishi, J.Y.; Smith, C.; Krieg, E.; Said, K.; Hornick, A.; Sasaki, H.M.; Filsinger, G.; Beliveau, B.J.; et al. Rapid in vitro production of single-stranded DNA. *Nucleic Acids Res.* **2019**, *47*, 11956–11962. [CrossRef]
170. Inoue, Y.U.; Morimoto, Y.; Yamada, M.; Kaneko, R.; Shimaoka, K.; Oki, S.; Hotta, M.; Asami, J.; Koike, E.; Hori, K.; et al. An optimized preparation method for long ssDNA donors to facilitate quick knock-in mouse generation. *Cells* **2021**, *10*, 1076. [CrossRef]