

Article

Systematic Comparison of Computational Tools for Sanger Sequencing-Based Genome Editing Analysis

Kanae Aoki, Mai Yamasaki [†], Riku Umezono [†], Takanori Hamamoto [†] and Yusuke Kamachi ^{*ID}

School of Engineering Science, Kochi University of Technology, Kami 782-8502, Japan; 275086d@gs.kochi-tech.ac.jp (M.Y.)

^{*} Correspondence: kamachi.yusuke@kochi-tech.ac.jp; Tel.: +81-887-57-2510

[†] These authors contributed equally to this work.

Abstract: Successful genome editing depends on the cleavage efficiency of programmable nucleases (PNs) such as the CRISPR–Cas system. Various methods have been developed to assess the efficiency of PNs, most of which estimate the occurrence of indels caused by PN-induced double-strand breaks. In these methods, PN genomic target sites are amplified through PCR, and the resulting PCR products are subsequently analyzed using Sanger sequencing, high-throughput sequencing, or mismatch detection assays. Among these methods, Sanger sequencing of PCR products followed by indel analysis using online web tools has gained popularity due to its user-friendly nature. This approach estimates indel frequencies by computationally analyzing sequencing trace data. However, the accuracy of these computational tools remains uncertain. In this study, we compared the performance of four web tools, TIDE, ICE, DECODR, and SeqScreener, using artificial sequencing templates with predetermined indels. Our results demonstrated that these tools were able to estimate indel frequency with acceptable accuracy when the indels were simple and contained only a few base changes. However, the estimated values became more variable among the tools when the sequencing templates contained more complex indels or knock-in sequences. Moreover, although these tools effectively estimated the net indel sizes, their capability to deconvolute indel sequences exhibited variability with certain limitations. These findings underscore the importance of judiciously selecting and using an appropriate tool with caution, depending on the type of genome editing being performed.

Keywords: ICE; CRISPR–Cas; DECODR; indel analysis; SeqScreener; TIDE



Citation: Aoki, K.; Yamasaki, M.; Umezono, R.; Hamamoto, T.; Kamachi, Y. Systematic Comparison of Computational Tools for Sanger Sequencing-Based Genome Editing Analysis. *Cells* **2024**, *13*, 261. <https://doi.org/10.3390/cells13030261>

Academic Editor: Peter Droge

Received: 25 December 2023

Revised: 28 January 2024

Accepted: 28 January 2024

Published: 30 January 2024



Copyright: © 2024 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. Introduction

Genome editing using CRISPR–Cas systems (clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins) has revolutionized modern biological research due to its high efficiency and ease of use compared to other programmable nucleases (PNs) such as ZFNs and TALENs [1,2]. The two main components of the CRISPR–Cas system are the Cas proteins, which are responsible for cleaving DNA, and CRISPR RNAs (crRNAs, also called guide RNAs or gRNAs), which direct catalytic activity against target DNA through complementarity between the spacer segment of the crRNA and target DNA. CRISPR–Cas systems are highly programmable because they allow for the replacement of spacer sequences of crRNAs. However, the cleavage efficiency of the CRISPR–Cas ribonucleoprotein (RNP) complex can vary among gRNAs [3,4]. Therefore, successful genome editing relies on the careful selection of gRNAs, and it is crucial to pre-evaluate the cleavage efficiency of candidate gRNAs for efficient genome editing. This is typically achieved by measuring the degree of insertions and deletions (indels) induced by the CRISPR–Cas complex at the target sites [5].

Several strategies have been developed to analyze indel frequencies; however, each strategy has specific limitations [5–7]. Mismatch cleavage assays, high-resolution melting

analysis (HRMA), and heteroduplex mobility assays are based on the cleavage or modified migration of heteroduplexes formed by mutated and wild-type (wt) DNA strands [8–10]. These methods have been widely used for the preliminary screening of gRNA activity due to their simplicity and low cost. The most popular among these techniques utilizes T7 endonuclease 1 (T7E1) or Surveyor nuclease to cleave mismatches formed between modified and unmodified DNA strands [11,12]. Despite these advantages, mismatch cleavage assay signals have been shown to be more strongly associated with indel complexity than with indel frequency, resulting in underestimation, especially for samples with a single dominant indel [6]. Thus, indel detection methods that depend on sequencing technologies have recently been developed. Among them, Sanger sequencing of subcloned DNAs derived from PCR amplicons was initially used [13,14]. High-throughput sequencing of PCR amplicons has also been employed successfully [15,16]; however, this approach is expensive and requires more time than other approaches. Additionally, a method called Indel Detection by Amplicon Analysis (IDAA) has been developed to detect indels, in which labeled PCR amplicons are subjected to capillary electrophoresis for fragment analysis [14], although nucleotide sequence change information could not be obtained.

Concurrently, a computational tool called Tracking the Indels by Decomposition (TIDE) was developed. TIDE analyzes Sanger sequencing trace data of PCR amplicons to estimate the frequency of editing efficiency and indel distribution [17]. Subsequently, similar computational tools have been developed including ICE (Inference of CRISPR Edits) [18], DECODR (Deconvolution of Complex DNA Repair) [19], and the SeqScreener Gene Edit Confirmation App (an online tool provided by Thermo Fisher Scientific). These computational tools rely on decomposition algorithms to identify the indel spectrum such as the number of indels, indel size, and total indel frequency. They achieve this by comparing Sanger sequencing trace data from the wild-type control and genome-edited sample sequences. While the algorithms share many features, each tool has specific modifications that may result in different outputs.

These computational tools have been shown to outperform mismatch cleavage assays in terms of accuracy and quantitative capability [6] and have been successfully used in numerous studies [5]. Although these tools have been shown to provide indel profiles similar to those obtained by high-throughput sequencing [6,18,19], the accuracy of indel frequency estimation has not been extensively investigated. Furthermore, a recent report indicated that TIDE, ICE, and DECODR produced widely divergent indel frequency data from the same samples derived from CRISPR–Cas9-induced indels in mouse tumor models [20]. Therefore, in this study, we quantitatively compared the performances of these four computational tools using artificial sequencing templates with defined indels. For this purpose, we first cloned various indels that had been induced by CRISPR–Cas9 or CRISPR–Cas12a at several zebrafish gene loci. By using Sanger sequencing trace data obtained from various combinations of predetermined indels, we quantitatively assessed the performance of these computational tools. We demonstrated that these tools were able to estimate indel frequencies with reasonable accuracy when indels had a few base changes and indel frequencies were in the midrange. However, variations in estimated values became more pronounced among the tools when the samples contained more complicated indels or when the indel frequencies were in a low or high range. Among the four tools, DECODR provided the most accurate estimations of indel frequencies for the majority of samples, which was consistent with the findings of a recent report [20]. While all four tools accurately estimated the net indel sizes, DECODR was found to be the most useful when one wishes to identify indel sequences. We also assessed the performance of these computational tools for estimating the knock-in efficiency of a short epitope tag sequence using a similar strategy and found that TIDE-based TIDER outperformed the other tools for this purpose. Taken together, these findings suggest that appropriate tools should be used depending on the type of genome editing.

2. Materials and Methods

2.1. Preparation of CRISPR–Cas9 and Cas12a RNP Complexes

The CRISPR–Cas9 RNP complexes were prepared as previously described [21]. The crRNAs (Alt-R CRISPR–Cas9 crRNA; listed in Supplementary Table S1), tracrRNAs (Alt-R CRISPR–Cas9 tracrRNA-ATTO 550), and the Cas9 protein (Alt-R S.p. Cas9 Nuclease V3) were purchased from IDT (Coralville, IA, USA). In brief, 3 μ M grRNA composed of crRNA and tracrRNA was combined with an equal volume of 3 μ M Cas9 to assemble a 1.5 μ M RNP complex.

CRISPR–Cas12a RNP complexes were prepared using crRNAs (Alt-R CRISPR–Cas12a crRNA; listed in Supplementary Table S1) and the Cas12a protein (Alt-R A.s. Cas12a Nuclease Ultra) purchased from IDT. The 3 μ M crRNA solution was combined with an equal volume of 3 μ M Cas12a diluted in Cas9 working buffer (20 mM HEPES pH 7.5, 150 mM KCl) and incubated at 37 °C for 10 min to assemble the RNP complex. Just before microinjection, the RNP complex was mixed with mRNA encoding the Venus fluorescent protein to monitor the success of the microinjection, resulting in an injection solution consisting of 1 μ M CRISPR–Cas12a RNP complex and 40 ng/ μ L Venus mRNA. Venus mRNA was prepared as previously described [22].

2.2. Microinjection and Genomic DNA Preparation

At the 1-cell stage, 1 nL of Cas9 or 1.5 nL of the Cas12a RNP complex (1.5 fmol RNP complex) was microinjected into the yolk of Tüpfel long-fin (TL) zebrafish embryos. Embryos at 1 day post-fertilization were lysed in 20 μ L/embryo of genomic DNA extraction buffer (10 mM Tris-HCl pH 8, 0.1 mM EDTA, 0.2% [v/v] Triton-X, 200 mM NaCl, and 0.2 mg/mL proteinase K) and heated at 55 °C for 2–3 h to dissolve the embryos. After heating at 95 °C for 10 min to inactivate protease K, the crude genomic DNA solution was directly subjected to polymerase chain reaction (PCR).

2.3. Cloning of Indels Induced by Cas9 and Cas12a

Genomic DNA fragments encompassing the target sites of the crRNAs (Cas9 otx2b_AA/Cas12a otx2b_02, Cas9 pax2a_AB, Cas9 pou2_AG, Cas12a sox2_01, Cas9 sox3_AA, Cas9 sox11a_AB, Cas12a sox11b_01, and Cas12a sox19b_01) were amplified using KOD One PCR Master Mix (Toyobo, Osaka, Japan) and the primers listed in Supplementary Table S2. These fragments were cloned into the pUC19 vector using the restriction sites included in the primers. Sanger sequencing was performed to identify the indel sequences.

2.4. Construction of Knock-In Donor Plasmids

To construct the knock-in donor plasmids, the DNA fragments of the left and right homology arms were amplified using specific primers for the *pax2a*, *pou2*, *sox11a*, and *sox3* genes (the primer sequences are listed in Supplementary Table S2) and the pUC clones containing the wild-type sequence of the respective genes serving as templates. For *pax2a*, *pou2*, and *sox3*, the reverse primers used for amplifying the 5' arm included a FLAG-tag encoding sequence for C-terminal tagging. For *sox11a*, the forward primer used for amplifying the 3' arm included a FLAG-tag encoding sequence for N-terminal tagging. The 5' and 3' arm fragments were simultaneously cloned into a pUC19 vector to obtain the donor plasmid (e.g., pUC19-[pax2a-5' arm]-FLAG-[pax2a-3' arm]).

2.5. Preparation of Indel Mixtures for the Artificial Sanger Sequencing Template

To prepare artificial template samples for Sanger sequencing, we initially amplified CRISPR–Cas target sequences for the wild-type, various indels, and the knock-in templates using KOD One PCR Master Mix, the pUC universal primers listed in Supplementary Table S2, and corresponding pUC19 clones as templates. The PCR products were purified using NucleoSpin Gel and PCR Clean-up columns (MACHEREY-NAGEL, Düren, Germany). The purified PCR products were subsequently combined in various ratios to mimic real genome-edited samples. These artificial samples were then subjected to Sanger

sequencing by a commercial vendor (Eurofin Genomics, Tokyo, Japan) using the primers listed in Supplementary Table S2.

2.6. Estimation of Indel Frequencies by Computational Tools

The Sanger sequencing raw data were processed using KB Basecaller (Thermo Fisher Scientific, Waltham, MA, USA) to determine the base calls. The resulting trace data, in the ab1 file format, were subsequently analyzed using TIDE (version 3.3.0; indel size range, 15 and 35 for indels induced by Cas9 and Cas12a, respectively; <https://tide.nki.nl> (accessed on 1 November 2023)), ICE (v3; <https://ice.synthego.com> (accessed on 1 November 2023)), DECODR (<https://decodr.org> (accessed on 1 November 2023)), and SeqScreener (<https://apps.thermofisher.com/apps/gea-web/#/setup> (accessed on 1 November 2023)) to estimate the indel frequencies. It is important to note that the Sanger sequencing trace data base-called using PeakTrace should not be used as it tends to treat low-level indels as sequencing noise, resulting in an underestimation of the indel frequency.

2.7. Zebrafish Husbandry

Zebrafish (*Danio rerio*) were bred and maintained under standard laboratory conditions following a 14-h light/10-h dark cycle. All zebrafish experiments were conducted in compliance with the Fundamental Guidelines for Proper Conduct of Animal Experiments and Related Activities in Academic Research Institutions under the jurisdiction of the Ministry of Education, Culture, Sports, Science and Technology of Japan, using protocols approved by the Animal Experiments Committee of Kochi University of Technology.

3. Results and Discussion

3.1. Design of a Quantitative Assessment of Computational Tools for Indel Analysis

We aimed to quantitatively assess the performance of computational tools for indel analysis by using predetermined indels as artificial genome-edited samples (Figure 1). To achieve this, we first collected various indels that were induced by CRISPR–Cas systems in several zebrafish genes. Since the type of indels may affect the performance of computational tools, we utilized CRISPR–Cas9 and CRISPR–Cas12a as these two systems are known to generate different types of indels. CRISPR–Cas9 primarily induces short indels, whereas CRISPR–Cas12a tends to produce larger indels, typically with deletions exceeding 10 bp in length [23,24]. To prepare artificial template samples for Sanger sequencing, we mixed wild-type (wt) sequences with one to six indels in various ratios to mimic real samples. The resulting trace data, which were base-called using KB Basecaller, served as input for the four computational tools TIDE, ICE, DECODR, and SeqScreener. Subsequently, the outputted indel frequencies were compared to the actual values. The performance was evaluated using the root mean square error (RMSE) between the estimated and actual values of the indel frequencies for each gRNA and indel mixture type.

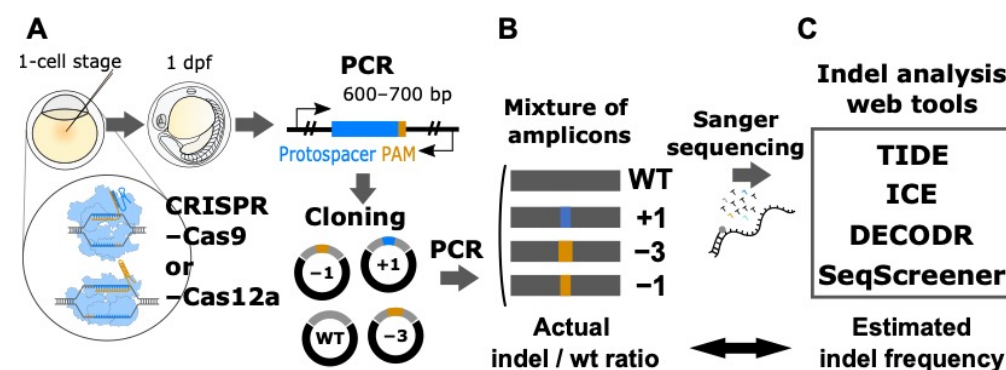


Figure 1. Experimental design for the quantitative assessment of Sanger sequencing-based computational tools for indel analysis. (A) Cloning of genomic DNA fragments containing indels induced by CRISPR–Cas9 and CRISPR–Cas12a. Various indels that were induced by CRISPR–Cas9 and CRISPR–Cas12a

in several zebrafish genes were PCR-amplified and subsequently cloned into the pUC19 plasmid. (B) Preparation of artificial genome-edited samples. The insert DNAs containing predefined indels or the wild-type (wt) sequence were PCR-amplified and purified. These purified PCR products were mixed at varying indel/wt ratios and used as artificial genome-edited samples for Sanger sequencing. (C) Computational tool analysis. The resulting Sanger sequencing trace data that were base-called by the KB Basecaller were used as input for the four computational tools TIDE, ICE, DECODR, and SeqScreener, after which the outputted indel frequencies were compared to the actual values.

3.2. Cloning Genomic DNA Fragments Containing Indels Induced by Genome Editing

We first cloned various indels that were induced by CRISPR–Cas9 or CRISPR–Cas12a in several zebrafish genes. To achieve this, we injected CRISPR–Cas9 RNP complexes containing one of the gRNAs targeting *pax2a*, *otx2b*, *pou2*, *sox3*, or *sox11a* (Supplementary Table S1) into zebrafish embryos at the one-cell stage. For CRISPR–Cas12a, we used gRNAs targeting *otx2b*, *sox2*, *sox11b*, or *sox19b* (Supplementary Table S1). Genomic DNA was extracted from the injected embryos at 1 day post-fertilization, and the surrounding sequences were amplified through PCR (Figure 1). The PCR amplicons were subsequently cloned into the pUC19 vector and subjected to Sanger sequencing. As expected, the indels induced by CRISPR–Cas9 were mostly short deletions and/or insertions, whereas the majority of the indels induced by CRISPR–Cas12a contained larger deletions and/or insertions, where typical indels contained more than 10 base deletions, with the largest one being a 134-base deletion (Figures 2–4). The PCR-amplified and purified DNAs containing these various indels and the wild-type (wt) sequence were mixed at varying indel/wt ratios and used as artificial genome-edited samples for Sanger sequencing.

3.3. Performance Evaluation of Computational Tools for Indel Estimation

To evaluate the performance of indel analysis tools for indels induced by CRISPR–Cas9, we used indels obtained from four Cas9 gRNAs targeting the *otx2b*, *pou2*, *sox3*, and *sox11a* genes (Figure 2). We first tested the performance of these tools by mixing one of the indel sequences with their wild-type (wt) sequences at various ratios, representing the simplest scenario. The tested indel percentages were 0%, 2.5%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, 97.5%, and 100%. After Sanger sequencing of these mixtures, the resulting trace data were inputted into the indel analysis tools using the wt trace data as a reference. The output data, representing the estimated indel frequency values, were plotted against the actual indel frequency values (Figure 2). These plots revealed that the four indel analysis tools generally performed well, with some exceptions, but they provided different indel frequency values, especially for samples with low or high indel frequencies. This observation is consistent with recent findings suggesting significant variability in indel frequency values among TIDE, ICE, and DECODR [20]. The RMSE values for all samples indicated that DECODR estimated the indel frequencies most accurately, followed by ICE, SeqScreener, and TIDE. TIDE and SeqScreener yielded erroneous values for larger or compound indels, as exemplified by the indels *otx2b* #8, *sox3* #4, and *sox11a* #10. In these cases, TIDE reported indel frequencies that were much lower than expected, likely due to a unique calculation process that excluded mis-decomposed indels with low *p*-values from the total indel frequency values. In fact, when the total indel frequencies were recalculated by assigning the remaining part of the wt percentages, where TIDE (100 – WT) values were obtained by subtracting the wt percentages from 100%, they aligned better with the actual values, indicating that the wt percentage values were more reliable in TIDE outputs. Interestingly, these recalculated values were similar to those obtained by SeqScreener, suggesting that these two tools employ a similar algorithm. In contrast, TIDE and SeqScreener produced more accurate values for samples with low indel frequencies (2.5% and 5%) than ICE or DECODR, where ICE and DECODR often overlooked low-level indels (Figure 2, Supplementary Table S3). This also suggests that the detection limit of indels may be a few percent when using TIDE and SeqScreener.

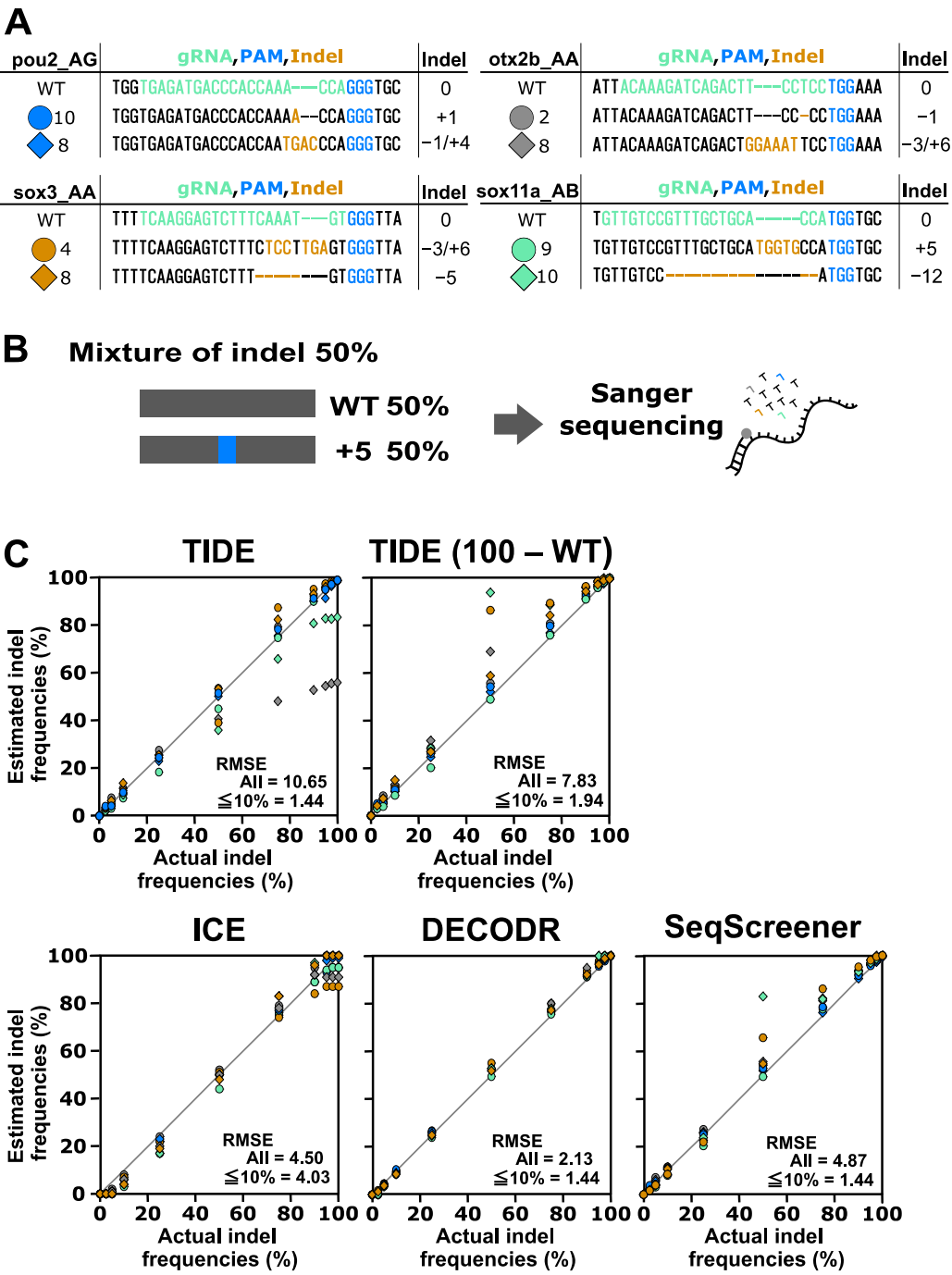


Figure 2. Performance of computational tools for simple mixtures containing an indel sequence induced using Cas9. (A) Indel sequences used in this analysis that were induced using Cas9 at the *otx2b*, *pou2*, *sox3*, and *sox11a* genes complexed with corresponding gRNAs, *otx2b*_AA, *pou2*_AG, *sox3*_AA, and *sox11a*_AB, respectively. The gRNA spacer sequences and the protospacer adjacent motif (PAM) sequence NGG are shown in green and blue, respectively. Inserted nucleotides are shown in orange, and deleted nucleotides are represented by dashed marks in orange. (B) An example of an artificial genome-edited sample representing 50% indels. (C) Scatter plots comparing indel frequencies (in percent) between the actual values and values estimated by TIDE, ICE, DECODR, and SeqScreener. In the TIDE (100 – WT) plot, values obtained by subtracting the wt percentage from 100% were used. The root mean square error (RMSE) values are shown for all data points as well as specifically for the subset of data points within the 2.5–10% indel range in each plot. The RMSE values for each indel type and all the datasets used for the plots are provided in Supplementary Table S3.

Next, we conducted a performance test on the indel analysis tools using a more complex scenario (mixture A) where two or three indel sequences were mixed with the corresponding wt sequence at various ratios (Figure 3). The indel sequence percentages used in this test were 0%, 5%, 10%, 25%, 50%, 75%, 90%, 95%, and 100%. To achieve these percentages, equal proportions of the respective indels were included in each point (a 75% indel mixture is shown in Figure 3 as an example). After Sanger sequencing and computational analysis, the estimated indel frequency values were plotted against the actual indel frequency values (Figure 3). The plots revealed that the four indel analysis tools generally performed well; however, TIDE reported indel frequencies that were much lower than expected for samples with high indel frequencies.

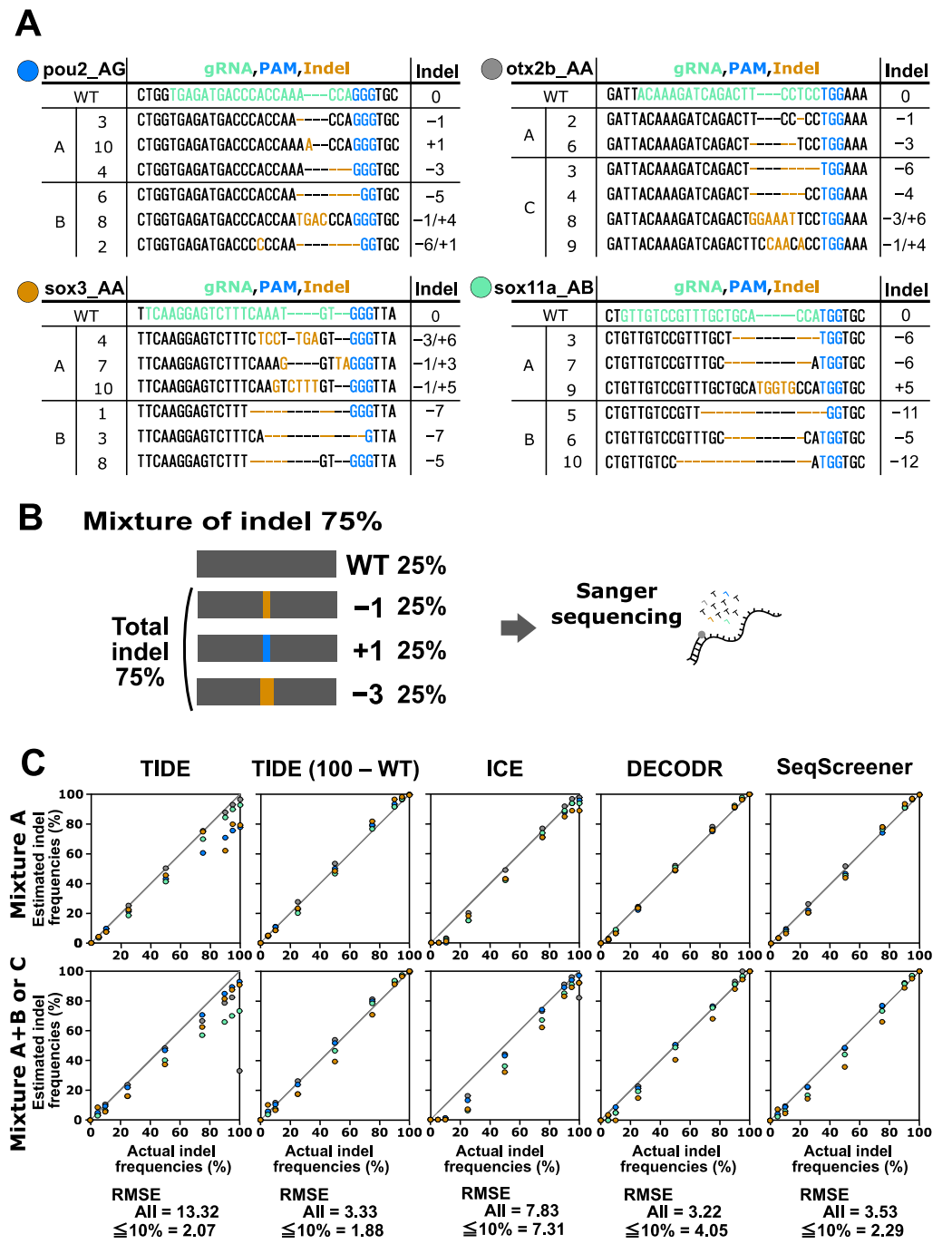


Figure 3. Performance of computational tools for complex mixtures containing two to six indel sequences induced by Cas9. (A) Indel sequences induced by Cas9 as

described in Figure 2. Mixtures of group A contained two to three indels. Mixtures of group B that contained three indels were used in combination with mixtures of group A. Mixture C for *otx2b*_AA was used separately. (B) An example of an artificial genome-edited sample representing 75% indels. (C) Scatter plots comparing indel frequencies (in percent) between the actual values and values estimated by TIDE, ICE, DECODR, and SeqScreener. In the TIDE (100 – WT) plot, values obtained by subtracting the wt percentage from 100% were used. The upper and lower panels show plots for mixture A and mixture A+B or C, respectively. The RMSE values are shown for all data points as well as specifically for the subset of data points within the 5–10% indel range in each tool. The RMSE values for each mixture type and all the datasets used for the plots are provided in Supplementary Table S3.

We then conducted a performance test on the indel analysis tools using a mixture of four to six indel sequences with their corresponding wt sequences at various ratios, which represented the most complicated scenarios (mixture A+B and mixture C) (Figure 3). The indel sequence percentages used in these complex samples were the same as those used in the mixture A samples. It was observed that all tools tended to underestimate indel frequencies, particularly for the *sox3* A+B and *sox11a* A+B samples, indicating that these tools struggled with accurately identifying larger deletions. The RMSE values for all of the samples indicated that DECODR once again exhibited the best performance, followed by SeqScreener, ICE, and TIDE. ICE often overlooked low-level indels when the expected indel percentages were 5% and 10%.

Finally, we performed a performance test on the indel analysis tools for indels induced by CRISPR–Cas12a. In this test, the indels obtained from the four Cas12a gRNAs for the *otx2b*, *sox2*, *sox11b*, and *sox19b* genes were used (Figure 4). These indels were mixed with the corresponding wt sequences in various ratios. Two types of mixtures were prepared for this experiment: mixtures containing two or three indel sequences (mixture A) and mixtures containing four to six indel sequences (mixture A+B and mixture C). The indel sequence percentages used were the same as those used for the CRISPR–Cas9 multi-indel samples. For the CRISPR–Cas12a samples containing larger indels, these tools tended to provide more divergent values compared to the CRISPR–Cas9 samples. The RMSE values for all samples indicated that DECODR performed the best for the Cas12a samples, followed by SeqScreener, TIDE, and ICE. However, the differences among the four tools were smaller than those observed in the CRISPR–Cas9 experiments.

Overall, our data indicated that DECODR consistently provided the most accurate indel frequency values for most of the indel samples among the four indel analysis tools. This observation positions DECODR as the preferred choice for Sanger sequencing-based indel analysis. Furthermore, DECODR has no limitations on the inference window sizes for deletions or insertions, whereas TIDE and ICE have limited window sizes of –50/+50 and –30/+14, respectively (window size information is unavailable for SeqScreener), suggesting an additional advantage of DECODR. These findings are consistent with a recent report indicating that DECODR's estimation of in vivo generated indels is more strongly correlated with high-throughput sequencing analysis than TIDE and ICE [20]. The lower performance of TIDE in the high indel frequency range may be attributable to its calculation procedure, as described above, in which mis-decomposed indels with low *p*-values are excluded from the total indel frequencies, despite the wild-type percentage values closely approximating the actual values. ICE tended to underestimate indel frequencies in the low indel frequency range possibly due to the treatment of low signal peaks as sequencing noise. It is worth mentioning that in the lower indel frequency range, TIDE and SeqScreener appear to provide more accurate indel frequency values than DECODR, as evidenced by their lower RMSE values when the calculation was restricted to samples with 2.5–10% indels (Figures 2–4).

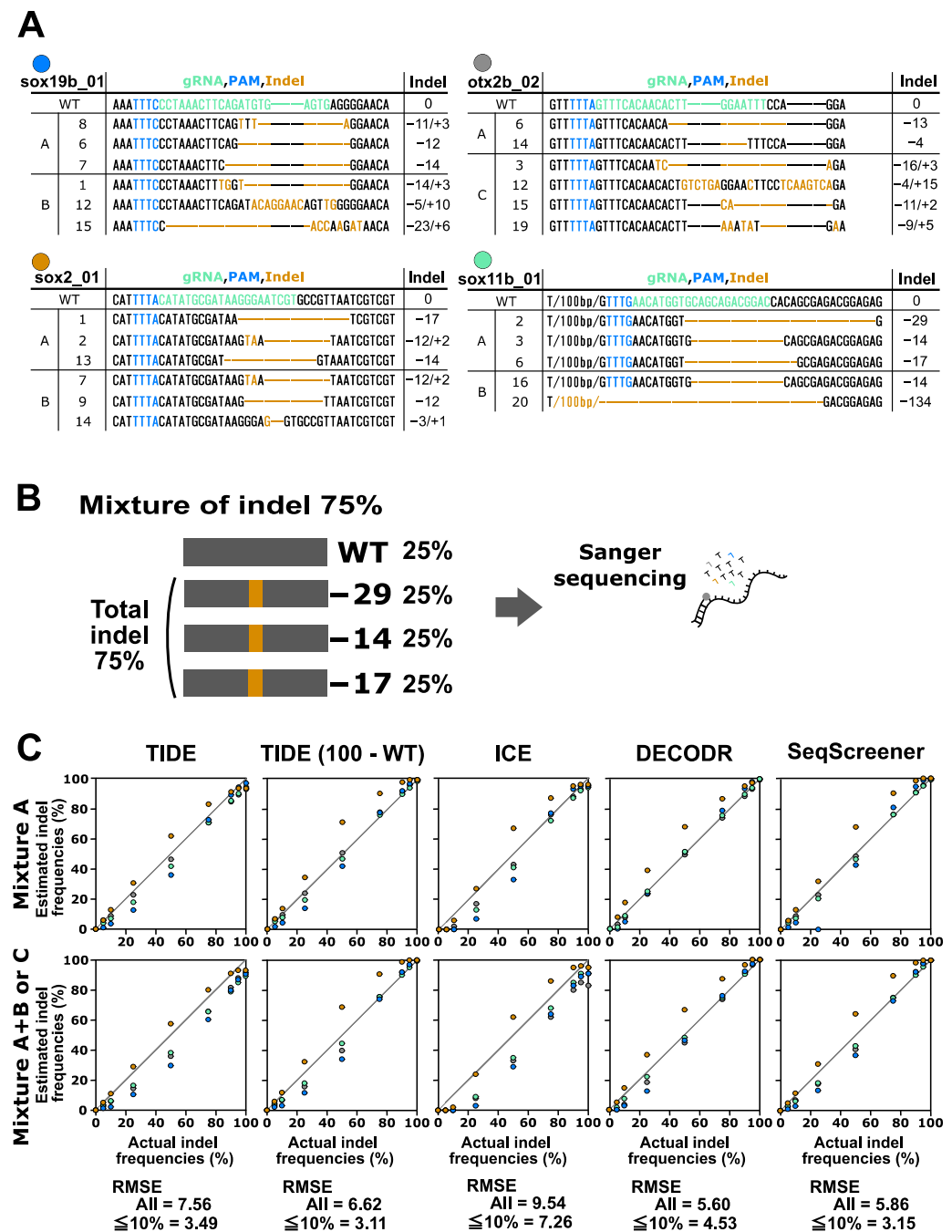


Figure 4. Performance of computational tools for complex mixtures containing two to six indel sequences induced by Cas12a. (A) Indel sequences that were induced by Cas12a at the *otx2b*, *sox2*, *sox11b*, and *sox19b* genes complexed with the corresponding gRNAs *otx2b_02*, *sox2_01*, *sox11b_01*, and *sox19b_01*, respectively. The gRNA spacer sequences and the PAM sequence TTTV are shown in green and blue, respectively. Inserted nucleotides are shown in orange, and deleted nucleotides are represented by dashed marks in orange. Mixtures A, B, and C are as described in Figure 3. (B) An example of an artificial genome-edited sample representing 75% indels. (C) Scatter plots comparing indel frequencies (in percent) between the actual values and values estimated by TIDE, ICE, DECODR, and SeqScreener. In the TIDE (100 – WT) plot, values obtained by subtracting the wt percentage from 100% were used. The upper and lower panels show plots for mixture A and mixture A+B or C, respectively. The RMSE values are shown for all data points as well as specifically for the subset of data points within the 5–10% indel range in each tool. The RMSE values for each mixture type and all the datasets used for the plots are provided in Supplementary Table S3.

3.4. Performance Evaluation of Computational Tools for Deconvoluting Indel Sizes and Compositions

The indel analysis tools provide information about the sizes, compositions, and frequencies of indels as well as the total indel frequencies. To assess the effectiveness of these tools in determining indel sizes in complex mixtures containing Cas9 or Cas12a-induced indels, we examined their accuracy in identifying the constituents of artificial samples, specifically mixtures A+B or C, each comprising a total of 75% indels (refer to Figures 3 and 4). For the sake of simplicity in data analysis, we only considered the net size changes resulting from simultaneous insertions and deletions.

As outlined in Table 1, the four computational tools effectively decomposed all constituents of the mixtures, except for the 134-bp deletion induced by Cas12a (*sox11b_01*, clone #20), indicating their high proficiency in deconvolution even when handling complex indel sequences. We observed discrepancies in the frequencies of individual indels, which appear to stem from inaccuracies in estimating the quantities of individual indels and/or the extent of mis-decomposed sequences. This discrepancy was particularly noticeable in the Cas9 *pou2* A+B and Cas12a *otx2b* C samples, where TIDE and ICE reported significantly lower percentages for specific indels (Cas9 *pou2* clone #4, Cas12a *otx2b* clone #12). The RMSE values for all samples indicated that DECODR exhibited superior performance in deciphering indels induced by both Cas9 and Cas12a, while TIDE, ICE, and SeqScreener showed similar performances.

Table 1. Comparison of the deconvolution accuracy of indel sizes and their frequencies.

Cas	gRNA	Indel Mixture	Clone #	Deletion	Insertion	Net Change	Actual Indel Frequency (%)	Estimated Indel Frequency (%)			
								TIDE	ICE	DECODR	Seq Screener
Cas9	<i>otx2b</i> _AA	C	wt				25.0	18.9	22.0	23.7	23.5
			8	−3	6	3	37.5	23.4	36.0	38.0	28.6
			9	−1	4	3	18.8	17.3	19.0	19.4	21.5
			4	−4		−4	18.8	16.0	18.0	18.8	18.6
			3	−6		−6	0.0	9.9	0.0	0.0	0.1
						Others					
	<i>pou2</i> _AG	A+B	wt				25.0	20.2	23.0	23.2	23.0
			3	−1		−1	12.5	8.5	9.0	9.4	9.4
			10		1	1	12.5	8.1	9.0	10.7	10.6
			4	−3		−3	12.5	4.9	11.0	11.6	7.5
			8	−1	4	3	12.5	9.5	9.0	11.5	11.0
			6	−5		−5	25.0	33.9	36.0	33.4	35.2
			2	−6	1	−5	0.0	4.5	0.0	0.0	3.4
						Others					
	<i>sox3</i> _AA	A+B	wt				25.0	29.3	32.0	31.8	33.9
			7	−1	3	2	12.5	6.4	9.0	10.4	9.1
			4	−3	6	3	12.5	16.8	18.0	17.7	16.7
			10	−1	5	4	12.5	14.5	13.0	15.1	13.8
			8	−5		−5	12.5	13.3	11.0	12.0	12.7
			1	−7		−7	25.0	11.5	11.0	13.0	13.7
			3	−7		−7	0.0	0.0	0.0	0.0	0.1
						Others					
	<i>sox11a</i> _AB	A+B	wt				25.0	21.6	25.0	24.4	26.6
			6	−5		−5	12.5	12.1	18.0	17.2	18.4
			9		5	5	12.5	11.4	10.0	14.4	13.5
			3	−6		−6	25.0	13.1	15.0	14.2	14.5
			7	−6		−6	12.5	5.5	7.0	8.9	8.5
			5	−11		−11	12.5	12.2	15.0	16.5	15.1
			10	−12		−12	0.0	2.5	2.0	4.3	3.5
						Others					
						RMSE		6.09	4.79	4.27	4.89

Table 1. Cont.

Cas	gRNA	Indel Mixture	Clone #	Deletion	Insertion	Net Change	Actual Indel Frequency (%)	Estimated Indel Frequency (%)			
								TIDE	ICE	DECODR	Seq Screener
Cas12a	otx2b_02	C	wt				25.0	24.5	26.0	26.6	24.8
			19	−9	5	−4	18.8	17.4	18.0	20.6	19.9
			15	−11	2	−9	18.8	17.8	21.0	21.3	20.7
			12	−4	15	11	18.8	15.6	7.0	15.2	15.2
			3	−16	3	−13	18.8	14.0	15.0	16.4	16.0
						Others	0.0	0.0	1.0	0.0	4.6
	sox2_01	A+B	wt				25.0	9.3	9.0	12.8	10.3
			14	−3	1	−2	12.5	11.5	10.0	12.1	12.1
			2	−12	2	−10	25.0	29.7	34.0	31.8	35.1
			7	−12	2	−10	25.0	29.7	34.0	31.8	35.1
			9	−12		−12	12.5	10.9	13.0	12.1	12.3
			13	−14		−14	12.5	14.4	16.0	16.3	16.1
			1	−17		−17	12.5	13.7	13.0	14.9	14.2
						Others	0.0	0.0	0.0	0.0	0.5
	sox11b_01	A+B	wt				25.0	24.5	23.0	25.4	25.0
			3	−14		−14	30.0	34.2	35.0	35.3	37.7
			16	−14		−14	30.0	34.2	35.0	35.3	37.7
			6	−17		−17	15.0	15.9	17.0	19.3	19.2
			2	−29		−29	15.0	15.6	16.0	20.0	17.1
			20	−134		−134	15.0	0.0	0.0	0.0	0.0
						Others	0.0	0.0	0.0	0.0	0.8
	sox19b_01	A+B	wt				25.0	26.0	27.0	24.1	27.0
			12	−5	10	5	12.5	12.5	12.0	13.4	14.7
			8	−11	3	−8	12.5	10.6	9.0	12.3	13.9
			1	−14	3	−11	12.5	12.6	15.0	12.3	12.9
			6	−12		−12	12.5	9.7	12.0	14.0	12.7
			7	−14		−14	12.5	11.5	11.0	14.0	13.1
			15	−23	6	−17	12.5	3.5	5.0	9.9	5.8
						Others	0.0	0.0	0.0	0.0	0.0
							RMSE	4.82	5.53	4.46	5.27

When applying these computational tools to clonal cells derived from genome editing, one might wish to identify exact indel sequences. Among the four tools, only DECODR is designed to comprehensively deconvolute both inserted and deleted sequences. ICE and SeqScreener are designed to deconvolute deleted sequences, but they output extra nucleotide sequences as unknown nucleotides (Ns). TIDE, on the other hand, does not provide indel sequence information except for a single base-pair insertion, and hence, it was excluded from this comparison. To simulate clonal cell analysis, we used 50/50 mixtures of wt/indel and indel/indel sequences, representing monoallelic and biallelic editing events, respectively, and evaluated the effectiveness of these tools in determining indel sequences (Figure 5). When using 50/50 mixtures of wt/indel, all three tools successfully estimated simple deletions, as expected (indels *otx2b* #2, *sox3* #8, and *sox11a* #10). Simple insertions were also well decomposed by these tools, with DECODR accurately predicting inserted sequences (indels *pou2* #10 and *sox11a* #9). However, for composite indels resulting from combined insertion and deletion events, even DECODR outputted erroneous sequences (all the other indels). In ICE and SeqScreener, the inserted sequences were incorrectly presented as the wt sequence along with additional inserted nucleotides represented as unknown nucleotides (Ns). When using 50/50 mixtures of indel/indel, the accuracy of sequence prediction by DECODR decreased in the composite indel sample *pou2* #8/#10 but rather increased in *sox19b* #12/#15. These data indicate that the current ability of these tools to estimate indel sequences is limited, suggesting that the decomposition algorithm may need improvement to fully deconvolute complex indels.

A

otx2b_AA		(gRNA, PAM, Indel)	WT / Indel (%)
Sanger	WT	ATTACAAGATCAGACTTCC--- TCCTGGAAA	50 / 50
	template	ATTACAAGATCAGACTTCC ---CCTGGAAA	
Tool output	ICE	ATTACAAGATCAGACTTCC ---CCTGGAAA	38 / 42
	DECODR	ATTACAAGATCAGACTTCC ---CCTGGAAA	47.1 / 52.9
	SeqScreener	ATTACAAGATCAGACTTCC ---CCTGGAAA	45.8 / 52.1

otx2b_AA		(gRNA, PAM, Indel)	WT / Indel (%)
Sanger	WT	ATTACAAGATCAGACTTCC--- TCCTGGAAA	50 / 50
	template	ATTACAAGATCAGACTGGAAAT TCCTGGAAA	
Tool output	ICE	ATTACAAGATCAGACTTCCNNN TCCTGGAAA	43 / 49
	DECODR	ATTACAAGATCAGACTTCCAAT TCCTGGAAA	47.2 / 52.8
	SeqScreener	ATTACAAGATCAGACTTCCNNN TCCTGGAAA	44.4 / 27.9

otx2b_AA		(PAM, Indel)	Indel (%)
Sanger	2	ATTACAAGATCAGACTTCC--- ---CCTGGAAA	50 / 50
	template	ATTACAAGATCAGACTGGAAAT TCCTGGAAA	
Tool output	ICE	ATTACAAGATCAGACTTCCNNN ---CCTGGAAA	46
	DECODR	ATTACAAGATCAGACTTCCAAT ---CCTGGAAA	56.2
	SeqScreener	ATTACAAGATCAGACTTCCNNN ---CCTGGAAA	43.8
		ATTACAAGATCAGACTTCCNNN TCCTGGAAA	41.0
		ATTACAAGATCAGACTTCCNNN TCCTGGAAA	20.1

C

sox3_AA		(gRNA, PAM, Indel)	WT / Indel (%)
Sanger	WT	TTTCAAGGAGTCTTCAAA T---GTGGTTA	50 / 50
	template	TTTCAAGGAGTCTTTCCTCC TTGAGTGGTTA	
Tool output	ICE	TTTCAAGGAGTCTTTCCTCC NNNTGTGGTTA	50 / 40
	DECODR	TTTCAAGGAGTCTTTCCTCC TTATGTGGTTA	44.9 / 55.1
	SeqScreener	TTTCAAGGAGTCTTTCCTCC NNNTGTGGTTA	34.3 / 47.0

sox3_AA		(gRNA, PAM, Indel)	WT / Indel (%)
Sanger	WT	TTTCAAGGAGTCTTTCCTCC TTGAGTGGTTA	50 / 50
	template	TTTCAAGGAGTCTTTCCTCC ---GTGGTTA	
Tool output	ICE	TTTCAAGGAGTCTTTCCTCC ---GTGGTTA	45 / 48
	DECODR	TTTCAAGGAGTCTTTCCTCC ---GTGGTTA	48.1 / 51.9
	SeqScreener	TTTCAAGGAGTCTTTCCTCC ---GTGGTTA	45.3 / 51.2

sox3_AA		(PAM, Indel)	Indel (%)
Sanger	4	TTTCAAGGAGTCTTTCCTCC TTGAGTGGTTA	50 / 50
	template	TTTCAAGGAGTCTTTCCTCC ---GTGGTTA	
Tool output	ICE	TTTCAAGGAGTCTTTCCTCC NNNTGTGGTTA	44
	DECODR	TTTCAAGGAGTCTTTCCTCC TTGAGTGGTTA	49.1
	SeqScreener	TTTCAAGGAGTCTTTCCTCC NNNTGTGGTTA	45.8
		TTTCAAGGAGTCTTTCCTCC ---GTGGTTA	50.9
		TTTCAAGGAGTCTTTCCTCC ---GTGGTTA	50.5

E

sox19b_01		(gRNA, PAM, Indel)	WT / Indel (%)
Sanger	WT	AAATTCCCTAAACTTCAGATGTGA--- GTGAG GGG	50 / 50
	template	AAATTCCCTAAACTTCAGATACAGGAACAG TTGG GGG	
Tool output	ICE	AAATTCCCTAAACTTCAGATGTGANNNG GTGAG GGG	50 / 39
	DECODR	AAATTCCCTAAACTTCAGATGTGAGTGGAG TTGG GGG	54.2 / 45.8
	SeqScreener	AAATTCCCTAAACTTCAGATGTGANNNG GTGAG GGG	53.6 / 42.7

sox19b_01		(gRNA, PAM, Indel)	WT / Indel (%)
Sanger	WT	AAATTCCCTAAACTTCAGATGTGAG TGAG GGGAAC	50 / 50
	template	AAATTCCCTAAACTTCAGATCCAA GATTAAC	
Tool output	ICE	AAATTCCCTAAACTTCAGATGTGAG GGGAAC	18 / 18
	DECODR	AAATTCCCTAAACTTCAGATGTGAG CTTAAC	7
	SeqScreener	AAATTCCCTAAACTTCAGATGTGAG CTTAAC	24.7 / 33.9
		AAATTCCCTAAACTTCAGATGTGAG CTTAAC	14.3 / 9.3
		AAATTCCCTAAACTTCAGATGTGAG GGGAAC	6.6

B

pou2_AG		(gRNA, PAM, Indel)	WT / Indel (%)
Sanger	WT	TGGTGAGATGACCCACCAAA--- CCAGGGTGC	50 / 50
	template	TGGTGAGATGACCCACCAATGAC CCAGGGTGC	
Tool output	ICE	TGGTGAGATGACCCACCAANNN CCAGGGTGC	46 / 50
	DECODR	TGGTGAGATGACCCACCAATGAC CCAGGGTGC	47.3 / 52.7
	SeqScreener	TGGTGAGATGACCCACCAANNN CCAGGGTGC	47.1 / 51.7

pou2_AG		(gRNA, PAM, Indel)	WT / Indel (%)
Sanger	WT	TGGTGAGATGACCCACCAAA--- CCAGGGTGC	50 / 50
	template	TGGTGAGATGACCCACCAAAA CCAGGGTGC	
Tool output	ICE	TGGTGAGATGACCCACCAAAA CCAGGGTGC	47 / 50
	DECODR	TGGTGAGATGACCCACCAAAAN CCAGGGTGC	48.1 / 51.9
	SeqScreener	TGGTGAGATGACCCACCAAAAN CCAGGGTGC	48.1 / 51.8

pou2_AG		(PAM, Indel)	Indel (%)
Sanger	8	TGGTGAGATGACCCACCAATGAC CCAGGGTGC	50 / 50
	template	TGGTGAGATGACCCACCAAAA--- CCAGGGTGC	
Tool output	ICE	TGGTGAGATGACCCACCAANNN CCAGGGTGC	36
	DECODR	TGGTGAGATGACCCACCAAAAN CCAGGGTGC	60.3
	SeqScreener	TGGTGAGATGACCCACCAAAAN CCAGGGTGC	38.3
		TGGTGAGATGACCCACCAAAAN CCAGGGTGC	61.5
		TGGTGAGATGACCCACCAAAAN CCAGGGTGC	61.5

D

sox11a_AB		(gRNA, PAM, Indel)	WT / Indel (%)
Sanger	WT	TGTTGTCGTTTGTGCA ---CCATGGTGC	50 / 50
	template	TGTTGTCGTTTGTGCA TGTTGCCATGGTGC	
Tool output	ICE	TGTTGTCGTTTGTGCA NNNNCCATGGTGC	48 / 44
	DECODR	TGTTGTCGTTTGTGCA TGTTGCCATGGTGC	50.5 / 49.5
	SeqScreener	TGTTGTCGTTTGTGCA NNNNCCATGGTGC	50.7 / 46.8

sox11a_AB		(gRNA, PAM, Indel)	WT / Indel (%)
Sanger	WT	TGTTGTCGTTTGTGCA CCATGGTGC	50 / 50
	template	TGTTGTCC--- ---ATGGTGC	
Tool output	ICE	TGTTGTCC--- ---ATGGTGC	38 / 42
	DECODR	TGTTGTCC--- ---ATGGTGC	47.1 / 52.9
	SeqScreener	TGTTGTCC--- ---ATGGTGC	17.1 / 23.3

sox11a_AB		(PAM, Indel)	Indel (%)
Sanger	9	TGTTGTCGTTTGTGCA TGTTGCCATGGTGC	50 / 50
	template	TGTTGTCC--- ---ATGGTGC	
Tool output	ICE	TGTTGTCGTTTGTGCA NNNNCCATGGTGC	59
	DECODR	TGTTGTCGTTTGTGCA TGTTGCCATGGTGC	59.5
	SeqScreener	TGTTGTCGTTTGTGCA NNNNCCATGGTGC	61.7
		TGTTGTCC--- ---ATGGTGC	40.5
		TGTTGTCC--- ---ATGGTGC	36.3

Figure 5. Comparison of the deconvolution capability of indel sequences in clonal cell analysis. Sanger sequencing was performed on 50/50 mixtures of wt/indel and indel/indel sequences, mimicking monoallelic and biallelic editing events, respectively. The indels used were as follows: (A) *otx2b* Cas9

indels #2 and #8, (B) *pou2* Cas9 indels #8 and #10, (C) *sox3* Cas9 indels #4 and #8, (D) *sox11a* Cas9 indels #9 and #10, (E) *sox19b* Cas12a indels #12 and #15. The resulting trace data were analyzed using ICE, DECODR, and SeqScreener, with the clonal sample type option selected in DECODR and SeqScreener. Each panel displays the alignment of actual sequences with the outputted indel sequences and their frequencies (minor alleles with low frequencies are not shown). In the Sanger template sequences, the gRNA spacer and the PAM sequences are shown in light green and light blue, respectively. Inserted nucleotides are shown in orange, and deleted nucleotides are represented by dashed marks in orange. In the tool output sequences, correctly and incorrectly identified inserted/deleted nucleotides are shown in bright blue and red, respectively. The inserted unknown nucleotides (Ns) are shown in bright green. In the wt/indel outputs, only the estimated indel sequences are shown, as the wt sequences were all correct.

3.5. Performance Evaluation of Computational Tools for Estimating Knock-in Frequency

The indel analysis tools evaluated above were also designed to estimate knock-in frequencies. A TIDE-based tool modified for knock-in analysis is called TIDER [25]. We therefore tested the performance of the computational tools for knock-in analysis using a similar strategy, where artificial samples that mimic genome editing outcomes of the knock-in procedure were prepared by mixing knock-in sequences with wt and indel sequences (Figure 6). In this experiment, a 51 bp sequence containing the FLAG-tag-encoding sequence was inserted at the beginning or end of the coding sequences of four zebrafish genes: *pax2a*, *pou2*, *sox11a*, and *sox3*. First, these knock-in sequences were mixed with only the wt sequence at various ratios as a simple case, although such mixtures (i.e., mixtures without indels) would not be produced in real knock-in experiments. The knock-in sequence percentages used were 0%, 1.25%, 2.5%, 5%, 10%, 20%, 50%, and 100%. We focused on a low range because knock-in frequencies are generally low. After Sanger sequencing of these mixtures, the resulting trace data were inputted into computational tools with Sanger trace data (for TIDER) or simple nucleotide sequences (for the other three) of the knock-in sequences as the second reference. The output data of the estimated knock-in frequency values were plotted against the actual knock-in frequency values (Figure 6). For these artificial knock-in samples, only TIDER was able to estimate reasonable values close to the expected values. Next, the knock-in sequences were mixed with the wt and Cas9-induced indel sequences in varying ratios to mimic a realistic sample (Figure 6). Again, only TIDER provided estimated indel values that were close to the expected values, although it generated more divergent values than those obtained with the indel-omitted mixtures, especially when the samples contained less than 10% knock-in sequences. The superior performance of TIDER for knock-in analysis was confirmed by its lower RMSE values. This variable performance is likely caused by differences in the tool workflow. Only TIDER uses additional Sanger trace data from the knock-in sequence as the second reference, whereas the other three tools only use nucleotide sequences in the text format. Taken together, these findings indicate that only TIDER can be reliably used for knock-in frequency analysis when the knock-in insertion length is more than several dozen base pairs, although detection in a low knock-in range may not be sufficiently accurate.

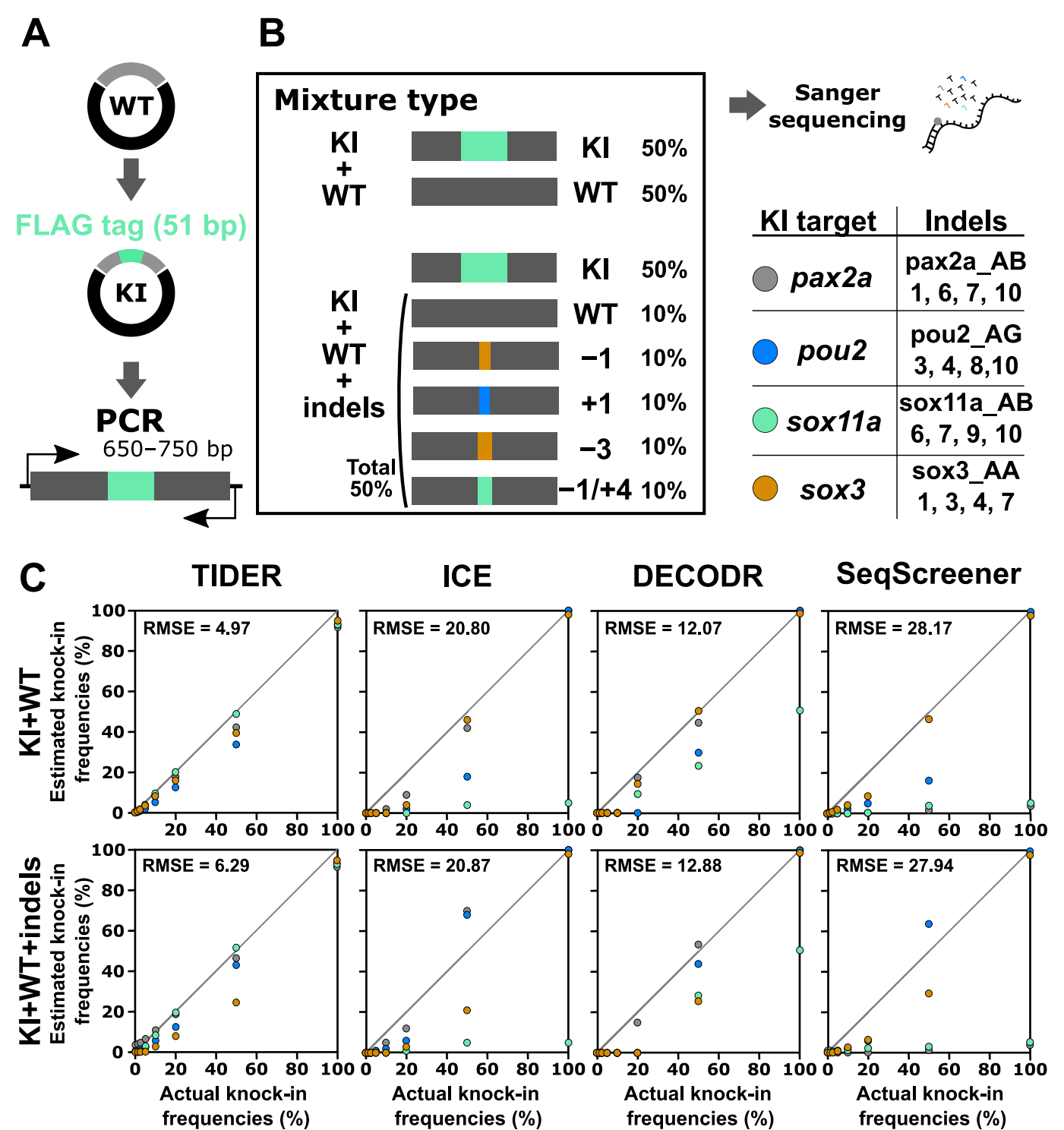


Figure 6. Performance of computational tools for mixtures containing a knock-in sequence with or without indel sequences induced by Cas9. **(A)** A schematic illustration of the knock-in sequences for the *pax2a*, *pou2*, *sox11a*, and *sox3* genes, which contain the FLAG-tag sequence at the beginning or end of the coding sequences. **(B)** Examples of artificial genome-edited samples representing 50% knock-in with or without indels. The indel sequences used in this analysis are shown as clone numbers, the indel sequences of which are shown in Supplementary Table S4. **(C)** Scatter plots comparing knock-in frequencies (in percent) between the actual values and values estimated by TIDER, ICE, DECODR, and SeqScreener. The upper and lower panels show plots for mixtures without indels and mixtures with indels, respectively. The RMSE values are shown in each plot. The RMSE values for each indel type and all the datasets used for the plots are provided in Supplementary Table S3.

4. Conclusions

In this study, we conducted a comprehensive comparison of four web tools: TIDE, ICE, DECODR, and SeqScreener. We used artificial sequencing templates with predetermined indels to assess their performance in indel frequency estimation and indel sequence deconvolution. Our systematic investigation revealed that these four computational tools consistently provided indel frequency estimates within an acceptable range for the majority of the tested indel mixtures. Notably, DECODR exhibited superior prediction accuracy across most samples, particularly when the samples contained a high percentage of indels. Consequently, DECODR is currently the preferred choice for Sanger sequencing-based indel frequency analysis, especially when the expected indels exceed the detection size limits of TIDE (−50/+50) and ICE (−30/+14). However, in scenarios where a very low indel frequency range (<10% total indels) is anticipated, TIDE and SeqScreener may offer more accurate indel frequency values than DECODR. Moreover, DECODR is particularly valuable when indel sequence information is necessary for clonal cell analysis, as it has the unique capability to predict inserted sequences, despite some limitations in accuracy for composite indels. In the context of knock-in frequency analysis, TIDER can be considered the primary choice, with a note of caution when interpreting low estimated values for knock-in percentages. Based on these findings, we recommend the thoughtful selection of a tool according to the specific requirements of the genome editing approach.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cells13030261/s1>, Supplementary Table S1: crRNAs and their protospacer target sequences used in this study; Supplementary Table S2: Lists of primers used in this study; Supplementary Table S3: The datasets of computational tool outputs and their analysis used for scatter plots; Supplementary Table S4: Indel sequences induced by CRISPR–Cas9 for artificial knock-in samples. References [21,22] are cited in Supplementary Materials.

Author Contributions: Conceptualization, Y.K.; Methodology, Y.K.; Formal Analysis, K.A. and Y.K.; Investigation, K.A., M.Y., R.U., T.H. and Y.K.; Data Curation, K.A., M.Y., R.U., T.H. and Y.K.; Writing—Original Draft Preparation, K.A. and Y.K.; Writing—Review and Editing, Y.K.; Visualization, K.A.; Supervision, Y.K.; Project Administration, Y.K.; Funding Acquisition, Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Grant-in-Aid for Scientific Research of Japan Society for the Promotion of Science [JSPS KAKENHI 21K06202 to YK].

Institutional Review Board Statement: The animal study protocol was approved by the Animal Experiments Committee of Kochi University of Technology (approval number: 2019–001).

Informed Consent Statement: Not applicable.

Data Availability Statement: All of the data including the Sanger sequencing raw data will be shared upon reasonable request to the corresponding author.

Acknowledgments: We are grateful to current and previous members of Y.K.'s laboratory for their valuable input and technical assistance.

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

References

1. Barman, A.; Deb, B.; Chakraborty, S. A Glance at Genome Editing with CRISPR–Cas9 Technology. *Curr. Genet.* **2020**, *66*, 447–462. [[CrossRef](#)] [[PubMed](#)]
2. Liu, G.; Lin, Q.; Jin, S.; Gao, C. The CRISPR-Cas Toolbox and Gene Editing Technologies. *Mol. Cell* **2022**, *82*, 333–347. [[CrossRef](#)] [[PubMed](#)]
3. Haeussler, M.; Schönig, K.; Eckert, H.; Eschstruth, A.; Mianné, J.; Renaud, J.-B.; Schneider-Maunoury, S.; Shkumatava, A.; Teboul, L.; Kent, J.; et al. Evaluation of Off-Target and on-Target Scoring Algorithms and Integration into the Guide RNA Selection Tool CRISPOR. *Genome Biol.* **2016**, *17*, 148. [[CrossRef](#)] [[PubMed](#)]
4. Doench, J.G.; Fusi, N.; Sullender, M.; Hegde, M.; Vaimberg, E.W.; Donovan, K.F.; Smith, I.; Tothova, Z.; Wilen, C.; Orchard, R.; et al. Optimized sgRNA Design to Maximize Activity and Minimize Off-Target Effects of CRISPR-Cas9. *Nat. Biotechnol.* **2016**, *34*, 184–191. [[CrossRef](#)] [[PubMed](#)]

5. Bennett, E.P.; Petersen, B.L.; Johansen, I.E.; Niu, Y.; Yang, Z.; Chamberlain, C.A.; Met, Ö.; Wandall, H.H.; Frödin, M. INDEL Detection, the ‘Achilles Heel’ of Precise Genome Editing: A Survey of Methods for Accurate Profiling of Gene Editing Induced Indels. *Nucleic Acids Res.* **2020**, *48*, 11958–11981. [[CrossRef](#)] [[PubMed](#)]
6. Sentmanat, M.F.; Peters, S.T.; Florian, C.P.; Connelly, J.P.; Pruett-Miller, S.M. A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Sci. Rep.* **2018**, *8*, 888. [[CrossRef](#)] [[PubMed](#)]
7. Carrington, B.; Bishop, K.; Sood, R. A Comprehensive Review of Indel Detection Methods for Identification of Zebrafish Knockout Mutants Generated by Genome-Editing Nucleases. *Genes* **2022**, *13*, 857. [[CrossRef](#)]
8. Dahlem, T.J.; Hoshijima, K.; Jurynek, M.J.; Gunther, D.; Starker, C.G.; Locke, A.S.; Weis, A.M.; Voytas, D.F.; Grunwald, D.J. Simple Methods for Generating and Detecting Locus-Specific Mutations Induced with TALENs in the Zebrafish Genome. *PLoS Genet.* **2012**, *8*, e1002861. [[CrossRef](#)]
9. Zhu, X.; Xu, Y.; Yu, S.; Lu, L.; Ding, M.; Cheng, J.; Song, G.; Gao, X.; Yao, L.; Fan, D.; et al. An Efficient Genotyping Method for Genome-Modified Animals and Human Cells Generated with CRISPR/Cas9 System. *Sci. Rep.* **2014**, *4*, 6420. [[CrossRef](#)]
10. Ota, S.; Hisano, Y.; Muraki, M.; Hoshijima, K.; Dahlem, T.J.; Grunwald, D.J.; Okada, Y.; Kawahara, A. Efficient Identification of TALEN-Mediated Genome Modifications Using Heteroduplex Mobility Assays. *Genes Cells* **2013**, *18*, 450–458. [[CrossRef](#)]
11. Qiu, P.; Shandilya, H.; D’Alessio, J.M.; O’Connor, K.; Durocher, J.; Gerard, G.F. Mutation Detection Using Surveyor™ Nuclease. *BioTechniques* **2004**, *36*, 702–707. [[CrossRef](#)] [[PubMed](#)]
12. Mashal, R.D.; Koontz, J.; Sklar, J. Detection of Mutations by Cleavage of DNA Heteroduplexes with Bacteriophage Resolvases. *Nat. Genet.* **1995**, *9*, 177–183. [[CrossRef](#)] [[PubMed](#)]
13. Moore, F.E.; Reyon, D.; Sander, J.D.; Martinez, S.A.; Blackburn, J.S.; Khayter, C.; Ramirez, C.L.; Joung, J.K.; Langenau, D.M. Improved Somatic Mutagenesis in Zebrafish Using Transcription Activator-Like Effector Nucleases (TALENs). *PLoS ONE* **2012**, *7*, e37877. [[CrossRef](#)] [[PubMed](#)]
14. Yang, Z.; Steentoft, C.; Hauge, C.; Hansen, L.; Thomsen, A.L.; Niola, F.; Vester-Christensen, M.B.; Frödin, M.; Clausen, H.; Wandall, H.H.; et al. Fast and Sensitive Detection of Indels Induced by Precise Gene Targeting. *Nucleic Acids Res.* **2015**, *43*, e59. [[CrossRef](#)] [[PubMed](#)]
15. Güell, M.; Yang, L.; Church, G.M. Genome Editing Assessment Using CRISPR Genome Analyzer (CRISPR-GA). *Bioinformatics* **2014**, *30*, 2968–2970. [[CrossRef](#)] [[PubMed](#)]
16. Park, J.; Lim, K.; Kim, J.-S.; Bae, S. Cas-Analyzer: An Online Tool for Assessing Genome Editing Results Using NGS Data. *Bioinformatics* **2017**, *33*, 286–288. [[CrossRef](#)] [[PubMed](#)]
17. Brinkman, E.K.; Chen, T.; Amendola, M.; van Steensel, B. Easy Quantitative Assessment of Genome Editing by Sequence Trace Decomposition. *Nucleic Acids Res.* **2014**, *42*, e168. [[CrossRef](#)]
18. Conant, D.; Hsiao, T.; Rossi, N.; Oki, J.; Maures, T.; Waite, K.; Yang, J.; Joshi, S.; Kelso, R.; Holden, K.; et al. Inference of CRISPR Edits from Sanger Trace Data. *CRISPR J.* **2022**, *5*, 123–130. [[CrossRef](#)]
19. Bloh, K.; Kanchana, R.; Bialk, P.; Banas, K.; Zhang, Z.; Yoo, B.-C.; Kmiec, E.B. Deconvolution of Complex DNA Repair (DECODR): Establishing a Novel Deconvolution Algorithm for Comprehensive Analysis of CRISPR-Edited Sanger Sequencing Data. *CRISPR J.* **2021**, *4*, 120–131. [[CrossRef](#)]
20. Brockman, Q.R.; Scherer, A.; McGivney, G.R.; Gutierrez, W.R.; Rytlewski, J.; Sheehan, A.; Warriar, A.; Lavery, E.A.; Roughton, G.; Carnevale, N.C.; et al. Discrepancies in Indel Software Resolution with Somatic CRISPR/Cas9 Tumorigenesis Models. *Sci. Rep.* **2023**, *13*, 14798. [[CrossRef](#)]
21. Okada, K.; Aoki, K.; Tabei, T.; Sugio, K.; Imai, K.; Bonkohara, Y.; Kamachi, Y. Key Sequence Features of CRISPR RNA for Dual-Guide CRISPR-Cas9 Ribonucleoprotein Complexes Assembled with Wild-Type or HiFi Cas9. *Nucleic Acids Res.* **2022**, *50*, 2854–2871. [[CrossRef](#)]
22. Ranawakage, D.C.; Takada, T.; Kamachi, Y. HiBiT-qIP, HiBiT-Based Quantitative Immunoprecipitation, Facilitates the Determination of Antibody Affinity under Immunoprecipitation Conditions. *Sci. Rep.* **2019**, *9*, 6895. [[CrossRef](#)]
23. Meshalkina, D.A.; Glushchenko, A.S.; Kysil, E.V.; Mizgirev, I.V.; Frolov, A. SpCas9- and LbCas12a-Mediated DNA Editing Produce Different Gene Knockout Outcomes in Zebrafish Embryos. *Genes* **2020**, *11*, 740. [[CrossRef](#)]
24. Swarts, D.C.; Jinek, M. Cas9 versus Cas12a/Cpf1: Structure–Function Comparisons and Implications for Genome Editing. *Wiley Interdiscip. Rev. RNA* **2018**, *9*, e1481. [[CrossRef](#)]
25. Brinkman, E.K.; Kousholt, A.N.; Harmsen, T.; Leemans, C.; Chen, T.; Jonkers, J.; van Steensel, B. Easy Quantification of Template-Directed CRISPR/Cas9 Editing. *Nucleic Acids Res.* **2018**, *46*, e58. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.