Computational Tools and Resources Supporting CRISPR-Cas Experiments

Pawel Sledzinski [ID]**, Mateusz Nowaczyk and Marta Olejniczak *** [ID]

Department of Genome Engineering, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland; psledzinski@ibch.poznan.pl (P.S.); matnowaczyk@ibch.poznan.pl (M.N.)
* Correspondence: marta.olejniczak@ibch.poznan.pl; Tel.: +48 61 852 8503

Abstract: The CRISPR-Cas system has become a cutting-edge technology that revolutionized genome engineering. The use of Cas9 nuclease is currently the method of choice in most tasks requiring a specific DNA modification. The rapid development in the field of CRISPR-Cas is reflected by the constantly expanding ecosystem of computational tools aimed at facilitating experimental design and result analysis. The first group of CRISPR-Cas-related tools that we review is dedicated to aid in guide RNA design by prediction of their efficiency and specificity. The second, relatively new group of tools exploits the observed biases in repair outcomes to predict the results of CRISPR-Cas edits. The third class of tools is developed to assist in the evaluation of the editing outcomes by analysis of the sequencing data. These utilities are accompanied by relevant repositories and databases. Here we present a comprehensive and updated overview of the currently available CRISPR-Cas-related tools, from the perspective of a user who needs a convenient and reliable means to facilitate genome editing experiments at every step, from the guide RNA design to analysis of editing outcomes. Moreover, we discuss the current limitations and challenges that the field must overcome for further improvement in the CRISPR-Cas endeavor.

Keywords: genome editing; CRISPR; database; off-target; sgRNA; microhomology; NHEJ; HDR; predictive algorithms; sgRNA design; NGS analysis

1. Introduction

The CRISPR-Cas (clustered regularly interspaced short palindromic repeats-CRISPR-associated protein) system has become a cutting edge technology that has revolutionized the field of genome engineering. The use of Cas9 nuclease is currently the method of choice in most tasks requiring a specific DNA cleavage due to its low cost, relative ease of use and straightforward applications. In its natural context, the CRISPR-Cas9 system builds an anti-mobile genetic element adaptive immunity of bacteria and archaea. It stores the record of, e.g., viral infections in the prokaryotic genome in the form of CRISPR arrays composed of acquired fragments of viral DNA (called spacers) separated by palindromic repeats. Transcription of the array and subsequent processing of the transcript leads to formation of a CRISPR-associated RNA (crRNA) or, simply, guide RNA (gRNA). Its distinguishing feature is a strict complementarity of its part containing the spacer to the original fragment of the viral genome (protospacer). crRNA anneals to the separately transcribed trans-activating crRNA (tracrRNA) and then forms a ribonucleoprotein complex (RNP complex) with the Cas nuclease. The crRNA guides the complex to any complementary sequence with a directly adjacent trinucleotide protospacer-adjacent motif (PAM, NGG in case of SpCas9) in the genome of an invading element. PAM is located directly downstream of the protospacer and plays a critical role in the process of self vs. non-self discrimination

preventing autoimmune response. Subsequently, the Cas9 nuclease introduces a double-strand break (DSB) 3 bp upstream of the PAM, leading to inactivation of an invader [1].

To simplify the system in the laboratory setting, a chimeric RNA molecule containing crRNA and tracrRNA, which is transcribed as a single guide RNA (sgRNA), is commonly applied [2]. Therefore, the basic version of the system requires only two components, namely, sgRNA and Cas9 nuclease, which both can be expressed from a single plasmid in cells to be edited. Alteration of the nucleotide composition of the gRNA allows the system to target virtually any sequence in the genome and specifically introduce a DSB [1]. The break is subsequently repaired either by nonhomologous end joining (NHEJ) or by homology-directed repair (HDR). The NHEJ pathway results in small insertions or deletions (indels) traditionally considered random, but recent evidence suggests that the outcomes are predictable to a certain extent [3–7]. This pathway is frequently utilized to induce frameshift mutations that lead, consequently, to the knockout (KO) of a gene. The HDR route requires a homologous repair template, either endogenous, such as a sister chromatid, or exogenous, in the form of plasmid or single-stranded oligodeoxyribonucleotide (ssODN), and leads to precise repair or modification, such as a point mutation or gene fragment insertion.

As outlined above, the specificity of the system relies on the base pairing between the gRNA and the target sequence. The intrinsic problem of the system is, therefore, the off-target activity at *loci* exhibiting only partial complementarity to the gRNA. Indeed, it has been demonstrated that even several mismatches can be tolerated at some *loci*, generating considerable nonspecific mutational activity [8–12]. Moreover, even in the case of perfect complementarity, the cleavage efficiency varies significantly across the genome [13–19]. An additional problem is that the most widely used nuclease from *Streptococcus pyogenes* (SpCas9) also recognizes non-canonical PAMs such as NAG and NGA [10,12]. Many modifications of the system have been developed to circumvent these problems, including the use of rationally engineered Cas9 nucleases, paired nickases (Cas9n), orthologous nucleases, different delivery modalities and inducible systems, but no approach to date has shown a complete absence of the off-target effect [20–28]. Consequently, the essential task that the researcher must face in most of the CRISPR-Cas experiments is to design an experiment and to choose tools that lead to an acceptable ratio between on-target and off-target activity.

The rapid development in the field of genome engineering is reflected by the constantly expanding ecosystem of computational tools aimed at facilitating experimental design and result analysis. The main group of CRISPR-Cas computational tools is dedicated to aid in gRNA design by prediction of their efficiency (on-target activity) and specificity (off-target activity). These tools historically range from simple sequence alignment-based approaches to elaborate scorings of mismatches in any position of the gRNA and evaluation of the sequence features and are expanding progressively towards incorporation of the chromatin context and genetic variation data into the scoring algorithms. The relatively new group of tools exploits the observed biases in repair outcomes to predict the most likely results of a given CRISPR-Cas edit. The third class of tools is developed to assist in the evaluation of the editing results by analysis of the Sanger/next-generation sequencing (NGS) data. We present a schematic overview of CRISPR-Cas genome editing experiment as well as main groups of involved computational tools in Figure 1.

The large number of divergent tools inevitably leads to difficulty in finding the appropriate one for a specific type of experiment. There are several excellent reviews summarizing advancements in the field [29–31], but they are limited in their scope mainly to gRNA design. Furthermore, the rapid evolution and divergence in the area of CRISPR-Cas-related tools and in genome editing itself generates the need for constant updates.

Here, we present an updated overview of the currently available CRISPR-Cas-related tools from the perspective of a user who needs a convenient and reliable means to facilitate genome editing experiments at every step, from gRNA design to analysis of editing outcomes.
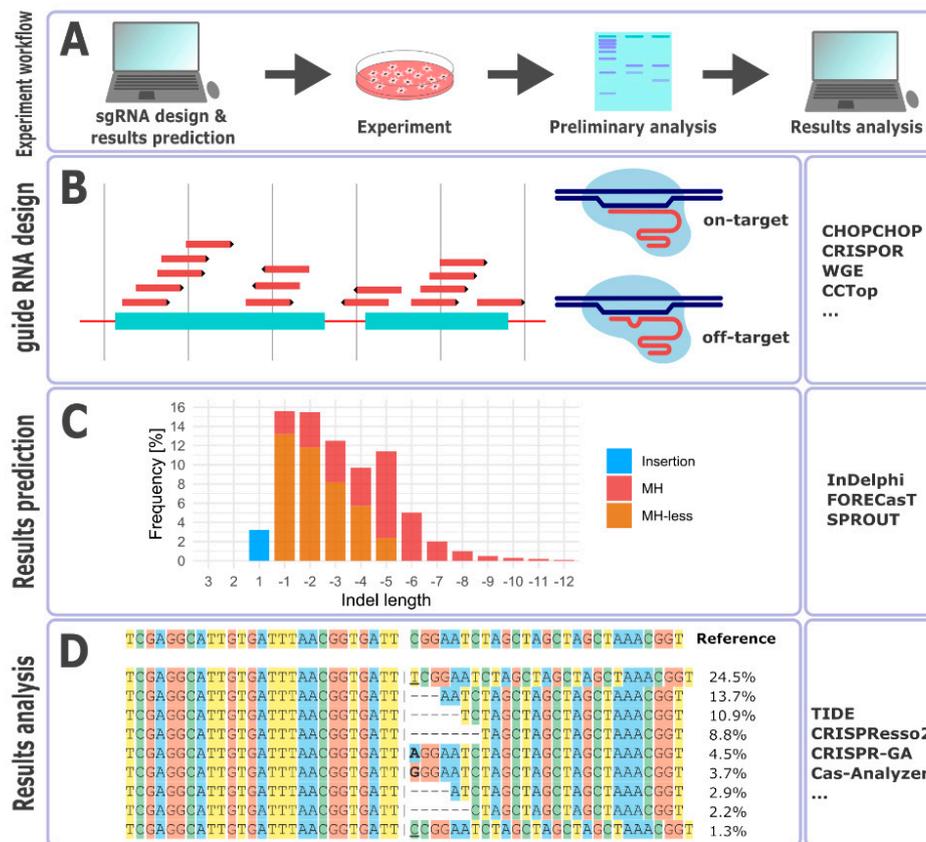
Figure 1. (**A**). The overview of a CRISPR-Cas experiment. Preliminary analysis involves screening of gRNA activity. Results analysis refers to NGS data analysis. (**B**). The gRNA design step involves the prediction of the gRNA efficiency (on-target activity) at a given *locus* and specificity (off-target activity); (**C**). The most probable results of CRISPR-Cas edits can be predicted by the analysis of microhomologies and by exploiting the known biases in the distribution of repair outcomes. (**D**). Web-based tools facilitating the post-experimental sequencing data analysis. MH—microhomology.

2. Experimental Design—the Choice of a Genome Editing System

Naturally, the research goals determine the model of choice, the targeted genomic site and the course of the experiment. There are, however, considerations particularly important in the context of CRISPR-Cas research. The crucial decisions concern the choice of Cas nuclease, format of CRISPR-Cas components and delivery of the system to the cell. The choices are particularly determined by the aim of the experiment (knockin, knockout, base editing, need for transient or stable nuclease activity), as well as the model used and its susceptibility to different forms of transfection or transduction.

SpCas9 is the most widely used nuclease in genome engineering, but advancements in the identification of its orthologs have resulted in a repertoire of nucleases characterized by different sizes, PAM requirements and cleavage patterns (e.g., Cas9s of *Staphylococcus aureus* and *Streptococcus thermophilus* or Cas12a of *Francisella novicida*) [32]. The other branch of research has resulted in "rationally engineered" nucleases with even more improved specificity [33]. The other strategies that proved useful in certain applications are the use of nickase (Cas9n) and catalytically inactive version of Cas9 (dCas9) [24,25,34,35]. Most of the gRNA design tools integrate PAM requirements and cleavage profile of many nucleases offering a considerable degree of flexibility in aiming at the desired editing site.

The format of CRISPR-Cas components (DNA, mRNA, RNP complex) may impact the efficiency and specificity of the nuclease. Purified Cas9 protein and in vitro transcribed gRNA can be combined to form the RNP complex delivered subsequently by electroporation or lipofection. This format leads to immediate yet short-lasting nuclease activity, as opposed to a more stable and longer expression from

a plasmid. The peak of Cas9 activity delivered in the form of RNP complex occurs a few hours after transfection and then declines quickly as gRNA and Cas9 protein are degraded, leading to a significant increase in specificity [36,37]. A strategy based on a similar rationale utilizes in vitro transcribed Cas9 mRNA and gRNA [38,39].

Lipofection or calcium phosphate transfection are high-throughput and cost-effective delivery methods but may be not suitable in the case of difficult to transfect cells, such as primary or stem cells. Electroporation or viral transduction may be used instead but are more demanding in terms of labor and expense. Viral, specifically lentiviral, delivery allows a stable integration of genes into the genome. Adeno-associated viruses (AAVs) are preferable due to the relatively low immunogenicity but their packaging limit renders simultaneous delivery of sgRNA and SpCas9 gene problematic [40,41].

Another factor worth considering is that different cells exhibit a bias towards different DNA repair outcomes [3,6]. The phenomenon may be potentially levered to increase the chance of obtaining the desired type of edits.

Services such as Addgene or IDT offer not only plasmids, proteins and oligos necessary for CRISPR-Cas experiments but also repositories of protocols that the researcher may find useful at the experiment planning stage [42,43]. Springer Nature provides a browser of CRISPR-related methods papers, whereas Bio-protocol has released a curated collection of CRISPR-Cas-related protocols [44,45]. Both resources can be helpful in fulfilling more specific demands.

All the choices briefly outlined above need to be accounted for during the experiment planning and, specifically, guide RNA design. The tools described in the subsequent sections offer solutions appropriate for a given experimental approach. Nonetheless, the intrinsic uncertainty of in silico predictions, as well as variance in experimental conditions, make it necessary to empirically verify the efficiency and specificity of the Cas edits and to test more than one gRNA. The standard approach includes mismatch detection assays (T7E1 or SURVEYOR) or, preferentially, targeted NGS sequencing, but the repertoire of novel methods allowing evaluation of the editing efficiency is diverse [46–51]. The topic has been covered by extensive reviews of available methods [52–54].

3. Guide RNA Design

The problem of designing a specific and efficient gRNA may be avoided in certain cases by referring to existing libraries of validated gRNAs. These libraries offer assistance in choosing gRNAs with confirmed activity for specific applications; e.g., CRISPRlnc offers validated guides for lncRNAs, whereas CRISPRz is a database of validated CRISPR targets in zebrafish [55,56]. Additionally, Addgene delivers a list of prevalidated gRNA sequences for various species and applications, as well as genome-wide libraries of sgRNAs for gene knockout experiments [57,58].

However, these validated gRNAs are usually available only for frequently targeted *loci* such as GFP and for generic experiments. Additionally, to reproduce their activity, one needs to recreate the experimental condition, which is not necessarily possible or desirable. This limitation gives rise to a demand for custom gRNAs tailored for every particular purpose.

The main group of CRISPR-Cas computational tools is, therefore, dedicated to aid in gRNA design. At the moment of writing of this manuscript, the WeReview database indexes more than 60 tools in this category [59]. The programs usually score the possible guides according to the predicted cleavage efficiency (maximizing on-target activity) and specificity (minimizing off-target activity). The predictive models differ in the theory underlying the scoring methods and in generated results, forming a large and diverse category.

The observed variation in the activity of gRNAs has led to several attempts at identifying the relationship between the guide sequence characteristics and the efficiency of Cas cleavage. The problem was approached by conducting large-scale screening experiments analyzing the activity of hundreds or thousands of gRNAs at different *loci* by high-throughput methods [10,15–17,19,60–63]. These efforts resulted in sets of rules for the design of highly active gRNAs and predictive models that were subsequently implemented in public design tools (sgRNA Designer, SSC, CRISPRscan, sgRNA Scorer,

CRISPRater) [16,17,19,60,62]. These models differ in weighting particular features, such as nucleotides preferred at every position, thermodynamics of secondary structures, occurrence of the same nucleotide tracts or GC content.

Doench et al. tested a pool of 1841 sgRNAs aiming at all possible targets of a panel of six mouse and three human genes coding cell surface markers and tested quantitatively their ability to generate biallelic knockouts by fluorescence-activated cell sorting (FACS), thereby separating the most active sgRNAs [63]. They identified sequence features that predict sgRNA efficacy: guanine was strongly preferred whereas cytosine was strongly unfavorable at position 20 counting from the 5′ end, the opposite was true for the position 16, and thymine was unfavored in the 3′ end region of the sgRNA which is related to the termination signal of the Pol III transcription machinery used to express the sgRNAs. Additionally, sgRNAs with low or high GC content appeared to be less active. The authors used the obtained data to determine sequence feature weights for activity predictions and built a predictive model for sgRNA activity by training a logistic regression classifier to identify sgRNAs exhibiting the highest activity for each gene using sequence features.

Doench et al. continued their research and proposed improvements to their model [16]. They used screening data generated with human and mouse genome-wide sgRNA libraries designed according to the previously proposed rules. They also included in the model additional features such as position-independent nucleotide counts, the location of the sgRNA target site within the gene, and thermodynamic properties of the sgRNAs. The authors designated the new model as Rule Set 2 and demonstrated its improved performance versus Rule Set 1 across different data sets. The Rule Set 2 was implemented subsequently in many gRNA design tools, such as CHOPCHOP, CRISPOR, GPP sgRNA Designer, or Benchling.

A different approach was proposed by Moreno-Mateos et al. [19]. They analyzed the molecular features that influence sgRNA activity and stability in vivo by using 1280 sgRNAs targeting 128 genes in the zebrafish genome. They injected in vitro transcribed sgRNAs together with Cas9-encoding mRNA into zebrafish embryos at the one-cell stage. The use of in vitro transcribed sgRNAs allowed the authors to focus on the influence of sgRNA stability and its relation to the sequence composition of the sgRNA. They demonstrated that more stable sgRNAs exhibit guanine enrichment and adenine depletion and, predictably, that the sgRNA stability positively correlates with the efficiency of the system. The authors identified also sequence features of the most active sgRNAs. The region spanning nucleotides 1–14 exhibited guanine enrichment, and positions 15–18 were characterized by cytosine enrichment. Thymidines and adenines were generally unflavoured, with the exception of positions 9 and 10. The authors applied randomized logistic regression to select the main features that determine the sgRNA efficiency. Then, they trained a linear regression model to predict the activity of a given sgRNA and implement it into the CRISPRscan tool [19].

It has been demonstrated, however, that the predictive accuracy of the many developed models is variable, and there is no superior algorithm in terms of reliability across different datasets [64,65]. The most plausible explanation for the observed inconsistency is that different experimental settings are used to generate the training datasets for every predictive model. A particular distinction identified to be of importance is the method of gRNA expression and delivery. As described earlier, the delivery modality (plasmid and viral vs. mRNA or RNP complex) can impact the cleavage kinetics and, consequently, the efficiency of edits. The majority of experiments were conducted in cell cultures and utilized the U6 promoter for relatively long-term and strong gRNA expression from a plasmid, whereas the other group of studies used the T7 promoter for in vitro transcription, and then, the gRNAs were injected into zebrafish embryos resulting in a short editing window. Thus, the general advice is to use a predictive algorithm that is based on an expression system analogous to the planned experiment. Specifically, Haeussler et al. suggested using the Doench score to rank the activity of gRNAs expressed from plasmids and the Moreno-Mateos score for gRNAs transcribed in vitro [64].

Another source of inconsistency among predictive algorithms is based on different approaches to gRNA activity assessment. A subset of studies used an indirect measurement and trained their

models on data based on phenotypic changes, e.g., protein expression or drug resistance [16,63]. These models may be more appropriate in cases of functional studies but require reproduction of the original experimental conditions to achieve high accuracy. Another group of studies measured Cas9 activity directly by sequencing the targeted sites and by indel detection [19,60]. The applicability of the resulting models is therefore wider. Some tools integrate several models in their predictions (CRISPOR, CHOPCHOP, E-CRISP), allowing the user to evaluate different scores [64,66–68].

The problem of Cas9 off-target activity prediction was initially approached by using common sequence alignment algorithms such as BWA or Bowtie [68]. This approach offers, however, moderate usefulness due to the problem of falsely positive and falsely negative results. Algorithms miss targets with a high number of mismatches, leading to an omission of a proportion of potential off-target sites. On the other hand, they still deliver false positives, as many identified off-target sites are not actually cleaved. Tsai et al., demonstrated a considerable discrepancy between results predicted by the MIT CRISPR Design Tool (currently unavailable), E-CRISP, and the off-target sites empirically verified by GUIDE-seq [12]. Similar results were presented by Cameron et al., who compared predictions of Cas-OFFinder, CCTop and the off-target sites identified by Site-seq [69]. All these issues led, similarly to the problem of gRNA activity, to the development of more advanced scoring tools that aim to accurately identify the most likely off-target sites [10,16,61,70,71]. The algorithms were built on data from screening studies focused on the relationship between the observed Cas9 activity and the number and distribution of mismatches. The most widely applied algorithms are the MIT specificity score and the Cutting Frequency Determination (CFD) score [10,16].

Since all these tools were developed on the basis of different data-sets and differ in the way that they score particular factors, it is advisable to compare results generated by different algorithms. Certain tools implement more than one predictive algorithm, allowing the user to choose the most appropriate approach (CRISPRseek, CRISPRscan, CRISPR Multi-Targeter, CRISPOR).

The majority of available tools still use only the basic sequence features to predict off-target activity of a given gRNA. The involvement of more complex characteristics, such as genomic context, chromatin environment, DNA/RNA thermodynamics, and the occurrence of DNA or RNA bulges, is ambiguous and not fully understood. There are, however, attempts to incorporate these factors into predictive models, e.g., CROP-IT, DeepCRISPR, uCRISPR or CRISTA [72–75].

Among tools facilitating the design of efficacious and specific sgRNAs, we recommend CRISPOR. It integrates the Doench score, in our opinion one of the most robust efficiency prediction systems, along with many other algorithms, allowing a direct comparison of their results, making the prediction more reliable. It supports 15 nucleases and more than 150 genomes. CRISPOR provides sequences for in vitro expression or cloning of a given sgRNA using different systems as well as primers for amplification of on- and off-target sites. The tool integrates also the Bae et al., 2014 predictions of repair profile and the Chen et al., 2018 frameshift prediction [5,76]. Moreover, CRISPOR warns the user when a designed gRNA contains motifs that should be avoided as described by Graf et al. [77].

In Supplementary Table S1, we present a summary of the most popular tools for gRNA design.

4. Repair Outcome Predictions

The relatively new branch of tools utilizes the fact that DNA repair is, to a certain extent, a deterministic process. The tools allow us to predict Cas edit results by the analysis of microhomologies and by exploiting the biases in the repair outcome. High-fidelity prediction of DSB repair outcomes is one of crucial concerns in CRISPR-Cas experimental design. Depending on the purpose of genome editing different repair products are desired, such as frame shift indels used for gene knockout or, contrarily, in-frame deletions maintaining the transcriptional activity of the gene. Genome editing outcomes are determined by the DNA repair pathway, which is in turn influenced by many factors, including the cell cycle stage and repair-related proteins expressed in a given cell type [78]. The most common DSB repair mechanisms, canonical nonhomologous end joining (c-NHEJ) and alternative nonhomologous end joining (alt-NHEJ), including microhomology mediated end joining (MMEJ),

often lead to insertion or deletion at the cutting site [79]. It has been proven that the distribution of repair outcomes is not random but strongly reproducible within a cell population and depends mainly on the targeted sequence. Two dominant categories of mutations occurring in the effect of DSB repair has been revealed: NHEJ-dependent single base insertion (duplication of nucleotide in position -4 from PAM) and MH-mediated deletion. [4,6,80,81] These observations were adopted by the developers of bioinformatic tools dedicated to detailed characterization of repair outcomes and improve experimental design. A comparison of these tools is presented in Table 1 [3,7,82].

Table 1. Comparison of repair outcome prediction tools.

Compared Feature	FORECasT	SPROUT	inDelphi
Indel frequency	YES	YES	YES
Average indel length	NO	YES	NO
Sequences of predictions	YES	NO	YES
Alignment of predicted sequences	YES	NO	YES
Frequency count for every predicted sequence	YES	NO	YES
Frameshift/ in frame frequency	YES	NO	YES
MH strength of target site	NO	NO	YES
Level of homogeneity of predictions	NO	NO	YES
Distinction between MH and MH-less deletions	NO	NO	YES
Cell type enable to choose	NO	NO	YES (HEK293, HCT116, K562, mESC, U2OS)
Batch mode	Available in command line tool	NO	YES
Gene mode	NO	NO	YES (for human and mouse)
Shareable link to results	NO	NO	YES
Summary statistics for download	YES	NO	YES
References	[3]	[82]	[7]

Allen et al. and Shen et al. created online tools for Cas9 edit result prediction (FORECasT and InDelphi, respectively) based on similar experiments [3,7]. Both groups used artificial constructs containing a gRNA expressed under the human U6 promoter and target sequence complementary to the guide flanked by different nucleotide contexts. Dozens of these constructs were transfected into cell lines stably expressing Cas9, leading to edits of the provided target sites [3,7]. Leenay et al., used an endogenous model to develop an online tool named SPROUT. The authors targeted 1656 genomic *loci* in primary CD4+ T cells isolated from 18 human donors and used an RNP complex containing crRNA, tracrRNA and the Cas9 protein [82]. In each experiment, cut-site-proximal repair products were characterized after PCR amplification of target sites with their flanking regions and deep-sequencing, revealing a sequence-dependent mutation profile. Then, data from sequencing were used to train algorithms developed by the authors to precisely predict DSB repair outcomes, and subsequently, the accuracy of prediction was tested. Each algorithm exhibited high accuracy of repair profile prediction for a given cell type [82].

The algorithms developed by authors are available online. SPROUT requires only sequences of the gRNA and PAM (total of 23 nt) to output information regarding predicted results, including the fraction of total reads with insertion, insertion to deletion ratio, average insertion and deletion length and the inserted base pair. SPROUT does not show any information about sequences of outcomes [82]. FORECasT requests the target DNA sequence and a location of PAM to output sequences of the ten most likely results and percentage of in-frame mutations [3]. InDelphi provides many capabilities and information not presented in other tools. Aside from sequences of predicted outcomes with their

frequency, this tool presents information on the frameshift frequency and microhomology (MH) strength of a target site. Predicted deletions are described as microhomology-resulted or microhomology-less. Moreover, InDelphi allows choosing the cell type of interest, HEK293, HCT116, K562, mESC and U2OS, which is required to achieve a high accuracy of prediction. Furthermore, in batch mode, InDelphi allows quick searching for up to 80 gRNAs followed by a chosen PAM in the input sequence and comparing their precision, frameshift frequency and microhomology strength with the results of a single mode for each gRNA. In the gene mode InDelphi scans exonic sequences of selected human or mouse genes to provide hundreds of gRNAs with statistics similar to those in batch mode, and includes information about distances from the 5′ and 3′ ends of the targeted exon for each gRNA [7].

We compared predictions generated by ForeCast and InDelphi to empirical data obtained in genome editing experiments carried out at two distinct loci. The sites differ in nucleotide sequence architecture: the *TP53* (tumor protein P53) locus is microhomology-rich and contains homologous tracts in close proximity to the cleavage site (Figure 2A, marked in yellow) making MMEJ a highly likely mechanism of DNA repair. In the targeted region of *HBB* (hemoglobin β) longer microhomologies are relatively less abundant.
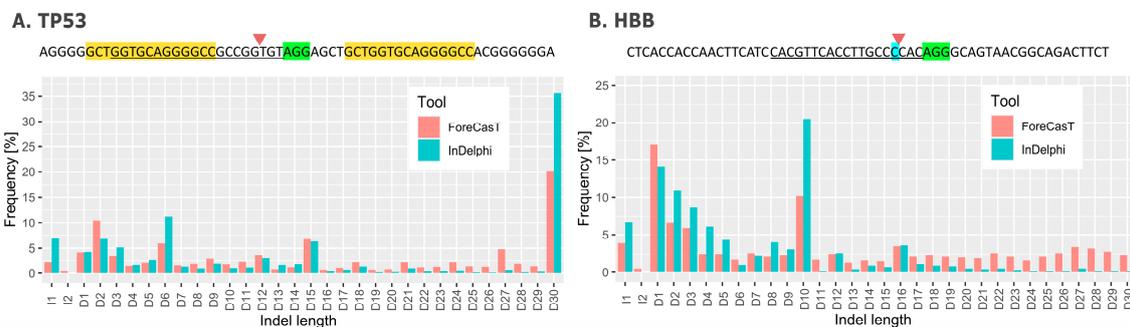

Figure 2. Editing outcome predictions generated by ForeCasT and InDelphi with a relevant sequence presented above. (**A**). Results obtained with gRNA targeting the human tumor protein P53 (*TP53*) gene; (**B**). Results obtained with gRNA targeting the human hemoglobin β (*HBB*) gene. Green rectangles— PAM sequences, yellow rectangles—tracts of microhomology, blue rectangle—duplicated nucleotide, red triangles—cutting sites.

In the case of *TP53*, both algorithms correctly predicted the occurrence of a relatively long, 30 bp deletion as the most probable editing result. Indeed, we observed a biallelic deletion, 30 and 50 bp long at this site in HEK293T cells [47]. The 50 bp deletion was not predicted since it exceeds the indel length capacity of both tools. The results confirm that MMEJ-based repair may be harnessed for highly predictable and precise genome editing. In the case of *HBB*, neither FOREcasT nor InDelphi was able to accurately assess the probability distribution across the range of indels. As reported by Cradick et al. introduction of a DSB resulted in an exceptionally high frequency of a one nucleotide insertion at the locus (43.2%) in HEK293T cells, whereas the tools assigned its relatively low probability (3.9% vs. 6.7%, ForeCasT and InDelphi, respectively) [83]. Nonetheless, both tools were able to predict which nucleotide is the most probable to be inserted (cytidine). The principles determining the preference of a nucleotide to be inserted appear to be relatively straightforward—the inserted base is usually homologous to the nucleotide at position −4 from PAM (Figure 2B, marked in blue) [3,4]. The high frequency of single nucleotide deletions predicted by both algorithms stems from the occurrence of CCCC motif directly at the cutting site. It has been reported that cytosine at the −5 and −4 positions relative to the cutting site results in a high probability of a single nucleotide excision [3,4]. However, the real frequency of this event turned out to be significantly lower (6.8%) [83].

The examples presented above point to the conclusion that, in spite of the considerable progress achieved in the area emerging from the intersection of CRISPR-Cas technology and DNA repair research, there is still room for improvement in terms of the ability to reliably predict DNA editing

outcomes. Longer microhomologies are a strong predictive factor of MMEJ engagement and the occurrence of predictable deletions. In cases when the editing precision is essential it may be reasonable to include the presence of microhomologies as a criterion for choosing a site to be edited. It narrows the targetable sequence range significantly, but the achieved accuracy may compensate for this restriction. The uncertainty intrinsic for in silico predictions notwithstanding, the idea of accurate prediction of DNA edits may prove exceptionally useful in the field of genome engineering.

5. Outcomes Analysis

Another group of algorithms is designed to facilitate the postexperiment (posterior) analysis of the CRISPR-Cas9 edition outcomes, including off-target assessment. Essentially, the postexperiment analysis algorithms enable quantification and visualization of CRISPR-Cas9 outcomes in tabular view with relevant graphs. These algorithms show the frequency of indels and repair types (NHEJ or HDR) simplifying data interpretation.

Some of tool such as Tracking of Indels by Decomposition (TIDE) are intended to analyze Sanger sequencing results of the targeted *locus*, providing a comprehensive profile of all insertions and deletions that occurred in a sample [49]. The redesigned version of this tool (TIDER—Tracking of Insertions, Deletions and Recombination events) also enables estimations of the efficacy of templated genome editing and assesses the frequency of HDR events [84]. The advantage of these approaches is that they require basic and easily accessible laboratory equipment and are relatively inexpensive. The other group of programs is concerned with NGS data analysis. Tools such as CRISPRpic, CRIS.py or CrispRVariants provide a command-line interface, which enables a detailed analysis of the results but may be difficult to approach for less experienced users [85–88]. Therefore, to address this issue, various web-based, command line-less and user-friendly programs for NGS analysis have been developed.

CRISPR-Genome Analyzer (CRISPR-GA) uses NGS data to quantify and characterize indels and homologous recombination events at the targeted sites [89]. Its pipeline consists of quality control and alignment of reads, which is subsequently used to compute the editing efficiency and frequency of HDR and NHEJ and to create charts with indel localizations and sizes.

Another tool available online is Cas-Analyzer, a JavaScript-based implementation for NGS data analysis [88]. This program provides results similar to those of CRISPR-GA but has one practical advantage: avoiding of a time-consuming step of data upload to a server. The Cas-Analyzer algorithm runs wholly on the side of a web browser, leading to a significant time savings. Moreover, it supports various nucleases used by scientists, including the paired nickases approach.

One additional program that facilitates the analysis of sequencing data is CRISPResso2, an updated version of the CRISPResso tool [90,91]. CRISPResso2, aside from performing edit outcome frequency assessment, enables allele-specific quantification of results by aligning reads to multiple reference sequences and computing statistics separately for each allele. Furthermore, CRISPResso2 may be very useful for base editing assays due to its detailed analysis of substitutions. The main disadvantage of this tool is the limit of data that can be uploaded for an analysis (up to 100 Mb in one FASTQ file). However, the command line version of CRISPResso2 containing additional tools and no data upload limits is available online. We compared the main features of the described tools in Table 2.

Table 2. Comparison of web-based postexperimental analysis tools.

Compared Feature	CRISPResso2	Cas-Analyzer	CRISPR-GA	TIDE/TIDER
Type of analysis	NGS	NGS	NGS	Sanger sequencing
File type	FASTQ	FASTQ	FASTQ	ABI
Output	-indel sizes and positions -HDR/NHEJ frequency -sequence alignment with reference -allele specific quantification	-indel sizes and positions -HDR/NHEJ frequency -sequence alignment with reference	-indel sizes and positions -HDR/NHEJ frequency	-indel sizes and positions -HDR/NHEJ frequency
Batch functionality	YES	NO	NO	NO
Base editing experiments	YES	NO	NO	NO
Template-mediated editing	YES	YES	YES	YES (TIDER)
Supported nucleases	Cas9, Cpf1	SpCas9, StCas9, NmCas9, SaCas9, CjCas9, AsCpf1/LbCpf1, paired nucleases: ZFNs, TALENs, Cas9 nickases, dCas9-FokI	Cas9	SpCas9, SaCas9, St1Cas9, NmCas9, AsCpf1, FnCpf1, LbCpf1
Need to upload dataset to a server	YES (up to 100Mb)	NO	YES	YES
Command line interface tool available	YES	NO	NO	NO
References	[90]	[88]	[89]	[49,92]

There are also tools dedicated to the analysis of NGS results from pooled screening experiments. Several tools developed for such purposes such as like MAGeCK and BAGEL utilize a command line interface, which may limit their accessibility for less experienced users [93,94]. Newly developed programs that offer web-based and user-friendly interfaces for the deconvolution of pooled screening data, such as CRISPRAnalyzeR, CRISPRcloud2 and PinAPL-Py, are addressed for researchers with zero programming background [95–97]. Each of these platforms requires uploading FASTQ files with NGS data and a gRNA library file (predefined gRNA libraries are also provided on websites) to start the workflow. The output of the mentioned tools includes information regarding quality control, detailed statistical analysis and visualization of the data. Moreover, CRISPRAnalyzeR and CRISPRcloud2 offer tutorial videos on their websites, simplifying the usage of these tools.

The methods described above can be used to analyze editing outcomes at on-target sites as well as at the off-target sites predicted in silico. This approach is, however, biased towards the predicted sites, and, usually, only the most probable cleavage sites are actually assayed. This, consequently, poses the risk of omitting the unpredicted or low-likely cleavages. The solution for this issue is to apply whole-genome sequencing (WGS) but the technique is expensive and unsuitable for analysis of many clones, additionally, it requires high coverage to detect low-frequency edits. All these problems resulted in the development of a distinct group of methods dedicated to cleavage detection across the entire genome without the requirement for WGS data. This group consists of techniques used to identify DSBs generated in living cells and in cell-free in vitro systems. The first type of methods includes unbiased identification of double-strand breaks evaluated by sequencing (GUIDE-seq), High-Throughput, Genome-wide translocation Sequencing (HTGTS), Integration-Deficient Lentiviral Vector (IDLV), direct in situ Breaks Labeling, Enrichment on Streptavidin, and next-generation Sequencing (BLESS) and Discovery of In Situ Cas Off-targets and VERification by sequencing (DISCOVER-seq) [12,98–101]. The second type consists of Selective enrichment and Identification of Tagged genomic DNA Ends by sequencing (SITE-Seq) and Circularization for In vitro Reporting of CLeavage Effects by sequencing (CIRCLE-seq) [69,102]. The latter group has the advantage of eliminating the necessity for cell culture and transfection of cells with components of the CRISPR-Cas system. In spite of different technical approaches, all these methods rely eventually on NGS sequencing. Some of these techniques are supplemented with dedicated computational pipelines facilitating the data analysis such as BLENDER (BLunt End finder) used for identification of off-target edits sites by ChIP-Seq in DISCOVER-seq method [101].

6. Discussion

Computational tools serve as valuable support for genome editing experiments. The significant progress in the development of scoring algorithms notwithstanding, the majority of predictive models relay on relatively basic sequence features. There is still room for improvement in terms of the understanding of the exact role of chromatin structure, nucleic acid thermodynamics or the occurrence of bulges in the process of Cas9 cleavage. As mentioned earlier, the issue of chromatin state and DNA accessibility is considered one of the key factors to be analyzed in the context of either the efficiency or specificity of CRISPR-Cas9 cleavage. Indeed, the association between Cas9 cleavage efficiency and chromatin accessibility has been demonstrated [103–109]. Avoidance of, e.g., strongly positioned nucleosomes may be beneficial in terms of efficacious edits and reliable assessment of the off-target risk.

Another important problem is naturally occurring genetic variation. The majority of gRNA design tools refer to the human reference genome and do not include variant information. Thus, SNPs and mutations in target sequences may have considerable effects on cleavage specificity either by disrupting the targeted sequence or by creating unexpected off-target sites [110]. These effects may, in turn, lead to unintended edits of essential genes, e.g., tumor suppressors. In fact, it has been demonstrated that genetic variation leads to a wide distribution of sites containing unexpected PAMs and absent PAMs across the genome [102,111,112]. Certain tools incorporate variation information into the process of gRNA design (CRISPOR, WGE, E-CRISP, CCTop). Moreover, there are tools dedicated to variant-specific CRISPR-Cas9 edits (SNP-CRISPR, AlleleAnalyzer) [113,114].

Generally, we suggest using tools that offer more than one scoring algorithm to assess the activity of gRNAs as accurately as possible. By this criterion CRISPOR, is the most advisable tool. Additionally, as we described earlier, the variety of additional features makes this tool particularly useful. Nevertheless, the activity and specificity of any custom gRNA need to be empirically validated due to the reported inconsistency of the models, as well as the risk of unexpected genetic variation or heterogeneity of cell populations. Another point to bear in mind is the fact that some models are not universally applicable. For example, the Moreno-Mateos score is best suited for experiments with in vitro expressed gRNAs.

At the moment of the manuscript writing, InDelphi is the most advanced predictive tool, providing the deepest analysis of results and several utilities that may be helpful in outcomes analysis. However, the data that the predictive algorithms are built on are limited to only certain cell lines, offering much less reliable estimation for cells with different genetic backgrounds. Cell line-specific DNA repair machinery disruption is a phenomenon that may lead to differential patterns of results obtained among different cell lines. Another issue worth consideration is the fact that accuracy of the predictive tools is moderate in the case of "difficult" sequences, such as long tracts of repeated sequences.

Of the tools aimed at post experiment results analysis, TIDE is the most straightforward and approachable solution; however, it has important limitations. Its accuracy is reduced in the case of low-frequency outcomes, and because it is a PCR-based method, it may not be well suited for the analysis of sequences containing tracts of tandem repeats. Additionally, the maximum size of indels that this tool detects is limited to 50 base pairs, making it not applicable in the analysis of larger rearrangements.

The outlined limitations may be overcome by applying tools based on NGS data analysis such as CRISPR-GA or CRISPResso. The downsides of these approaches are, however, tied to the inherent limitations of NGS techniques. NGS is a costly and time-consuming system and, in many cases, requires the use of an external service. Despite these issues, NGS is still a method of choice for a reliable analysis of CRISPR-Cas experimental results.

For a quick search of tools useful in the field of CRISPR-Cas experiments we recommend the WeReview: CRISPR Tools website [59]. It is a large, user-updated repository of computational tools related to CRISPR-Cas research. It offers an extensive list of services categorized by main characteristics, searching and sorting options, yet the presented features may be too basic for advanced users.

References

1. Jiang, F.; Doudna, J.A. CRISPR–Cas9 Structures and Mechanisms. *Annu. Rev. Biophys.* **2017**, *46*, 505–529. [CrossRef] [PubMed]
2. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J.A.; Charpentier, E. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **2012**, *337*, 816–821. [CrossRef] [PubMed]
3. Allen, F.; Crepaldi, L.; Alsinet, C.; Strong, A.J.; Kleshchevnikov, V.; De Angeli, P.; Páleníková, P.; Khodak, A.; Kiselev, V.; Kosicki, M.; et al. Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* **2019**, *37*, 64–72. [CrossRef] [PubMed]
4. Chakrabarti, A.M.; Henser-Brownhill, T.; Monserrat, J.; Poetsch, A.R.; Luscombe, N.M.; Scaffidi, P. Target-Specific Precision of CRISPR-Mediated Genome Editing. *Mol. Cell* **2019**, *73*, 699–713.e6. [CrossRef] [PubMed]
5. Chen, W.; McKenna, A.; Schreiber, J.; Haeussler, M.; Yin, Y.; Agarwal, V.; Noble, W.S.; Shendure, J. Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair. *Nucleic Acids Res.* **2019**, *47*, 7989–8003. [CrossRef] [PubMed]
6. Van Overbeek, M.; Capurso, D.; Carter, M.M.; Thompson, M.S.; Frias, E.; Russ, C.; Reece-Hoyes, J.S.; Nye, C.; Gradia, S.; Vidal, B.; et al. DNA Repair Profiling Reveals Nonrandom Outcomes at Cas9-Mediated Breaks. *Mol. Cell* **2016**, *63*, 633–646. [CrossRef]
7. Shen, M.W.; Arbab, M.; Hsu, J.Y.; Worstell, D.; Culbertson, S.J.; Krabbe, O.; Cassa, C.A.; Liu, D.R.; Gifford, D.K.; Sherwood, R.I. Predictable and precise template-free CRISPR editing of pathogenic variants. *Nature* **2018**, *563*, 646–651. [CrossRef]
8. Lin, Y.; Cradick, T.J.; Brown, M.T.; Deshmukh, H.; Ranjan, P.; Sarode, N.; Wile, B.M.; Vertino, P.M.; Stewart, F.J.; Bao, G. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res.* **2014**, *42*, 7473–7485. [CrossRef]
9. Fu, Y.; Foden, J.A.; Khayter, C.; Maeder, M.L.; Reyon, D.; Joung, J.K.; Sander, J.D. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **2013**, *31*, 822–826. [CrossRef]
10. Hsu, P.D.; Scott, D.A.; Weinstein, J.A.; Ran, F.A.; Konermann, S.; Agarwala, V.; Li, Y.; Fine, E.J.; Wu, X.; Shalem, O.; et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **2013**, *31*, 827–832. [CrossRef]
11. Pattanayak, V.; Lin, S.; Guilinger, J.P.; Ma, E.; Doudna, J.A.; Liu, D.R. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat. Biotechnol.* **2013**, *31*, 839–843. [CrossRef] [PubMed]
12. Tsai, S.Q.; Zheng, Z.; Nguyen, N.T.; Liebers, M.; Topkar, V.V.; Thapar, V.; Wyvekens, N.; Khayter, C.; Iafrate, A.J.; Le, L.P.; et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **2015**, *33*, 187–197. [CrossRef] [PubMed]
13. Cong, L.; Ran, F.A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P.D.; Wu, X.; Jiang, W.; Marraffini, L.A.; et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **2013**, *339*, 819–823. [CrossRef] [PubMed]
14. Mali, P.; Yang, L.; Esvelt, K.M.; Aach, J.; Guell, M.; DiCarlo, J.E.; Norville, J.E.; Church, G.M. RNA-Guided Human Genome Engineering via Cas9. *Science* **2013**, *339*, 823–826. [CrossRef] [PubMed]
15. Shalem, O.; Sanjana, N.E.; Hartenian, E.; Shi, X.; Scott, D.A.; Mikkelsen, T.S.; Heckl, D.; Ebert, B.L.; Root, D.E.; Doench, J.G.; et al. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **2014**, *343*, 84–87. [CrossRef]

16. 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]

17. Xu, H.; Xiao, T.; Chen, C.-H.; Li, W.; Meyer, C.A.; Wu, Q.; Wu, D.; Cong, L.; Zhang, F.; Liu, J.S.; et al. Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **2015**, *25*, 1147–1157. [CrossRef]

18. Mendoza, B.J.; Trinh, C.T. Enhanced guide-RNA design and targeting analysis for precise CRISPR genome editing of single and consortia of industrially relevant and non-model organisms. *Bioinformatics* **2018**, *34*, 16–23. [CrossRef]

19. Moreno-Mateos, M.A.; Vejnar, C.E.; Beaudoin, J.-D.; Fernandez, J.P.; Mis, E.K.; Khokha, M.K.; Giraldez, A.J. CRISPRscan: Designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods* **2015**, *12*, 982–988. [CrossRef]

20. Kleinstiver, B.P.; Pattanayak, V.; Prew, M.S.; Tsai, S.Q.; Nguyen, N.T.; Zheng, Z.; Joung, J.K. High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **2016**, *529*, 490–495. [CrossRef]

21. Vakulskas, C.A.; Dever, D.P.; Rettig, G.R.; Turk, R.; Jacobi, A.M.; Collingwood, M.A.; Bode, N.M.; McNeill, M.S.; Yan, S.; Camarena, J.; et al. A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells. *Nat. Med.* **2018**, *24*, 1216–1224. [CrossRef] [PubMed]

22. Slaymaker, I.M.; Gao, L.; Zetsche, B.; Scott, D.A.; Yan, W.X.; Zhang, F. Rationally engineered Cas9 nucleases with improved specificity. *Science* **2016**, *351*, 84–88. [CrossRef] [PubMed]

23. Ran, F.A.; Cong, L.; Yan, W.X.; Scott, D.A.; Gootenberg, J.S.; Kriz, A.J.; Zetsche, B.; Shalem, O.; Wu, X.; Makarova, K.S.; et al. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **2015**, *520*, 186–191. [CrossRef]

24. Ran, F.A.; Hsu, P.D.; Lin, C.-Y.; Gootenberg, J.S.; Konermann, S.; Trevino, A.; Scott, D.A.; Inoue, A.; Matoba, S.; Zhang, Y.; et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **2013**, *154*, 1380. [CrossRef] [PubMed]

25. Tsai, S.Q.; Wyvekens, N.; Khayter, C.; Foden, J.A.; Thapar, V.; Reyon, D.; Goodwin, M.J.; Aryee, M.J.; Joung, J.K. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* **2014**, *32*, 569–576. [CrossRef] [PubMed]

26. Chen, Y.; Liu, X.; Zhang, Y.; Wang, H.; Ying, H.; Liu, M.; Li, D.; Lui, K.O.; Ding, Q. A Self-restricted CRISPR System to Reduce Off-target Effects. *Mol. Ther.* **2016**, *24*, 1508–1510. [CrossRef]

27. Gao, Z.; Herrera-Carrillo, E.; Berkhout, B. A Single H1 Promoter Can Drive Both Guide RNA and Endonuclease Expression in the CRISPR-Cas9 System. *Mol. Ther. Nucleic Acids* **2019**, *14*, 32–40. [CrossRef]

28. Liu, K.I.; Ramli, M.N.B.; Woo, C.W.A.; Wang, Y.; Zhao, T.; Zhang, X.; Yim, G.R.D.; Chong, B.Y.; Gowher, A.; Chua, M.Z.H.; et al. A chemical-inducible CRISPR–Cas9 system for rapid control of genome editing. *Nat. Chem. Biol.* **2016**, *12*, 980–987. [CrossRef]

29. Cui, Y.; Xu, J.; Cheng, M.; Liao, X.; Peng, S. Review of CRISPR/Cas9 sgRNA Design Tools. *Interdiscip. Sci. Comput. Life Sci.* **2018**, *10*, 455–465. [CrossRef]

30. Wilson, L.O.W.; O'Brien, A.R.; Bauer, D.C. The Current State and Future of CRISPR-Cas9 gRNA Design Tools. *Front. Pharmacol.* **2018**, *9*. [CrossRef]

31. Brazelton, V.A.; Zarecor, S.; Wright, D.A.; Wang, Y.; Liu, J.; Chen, K.; Yang, B.; Lawrence-Dill, C.J. A quick guide to CRISPR sgRNA design tools. *GM Crop. Food* **2015**, *6*, 266–276. [CrossRef] [PubMed]

32. Fonfara, I.; Le Rhun, A.; Chylinski, K.; Makarova, K.S.; Lécrivain, A.-L.; Bzdrenga, J.; Koonin, E.V.; Charpentier, E. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.* **2014**, *42*, 2577–2590. [CrossRef] [PubMed]

33. Mueller, K.; Carlson-Stevermer, J.; Saha, K. Increasing the precision of gene editing in vitro, ex vivo, and in vivo. *Curr. Opin. Biomed. Eng.* **2018**, *7*, 83–90. [CrossRef] [PubMed]

34. Cho, S.W.; Kim, S.; Kim, Y.; Kweon, J.; Kim, H.S.; Bae, S.; Kim, J.-S. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **2014**, *24*, 132–141. [CrossRef] [PubMed]

35. Gaudelli, N.M.; Komor, A.C.; Rees, H.A.; Packer, M.S.; Badran, A.H.; Bryson, D.I.; Liu, D.R. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **2017**, *551*, 464–471. [CrossRef]

36. Kim, S.; Kim, D.; Cho, S.W.; Kim, J.; Kim, J.-S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **2014**, *24*, 1012–1019. [CrossRef]

37. Zuris, J.A.; Thompson, D.B.; Shu, Y.; Guilinger, J.P.; Bessen, J.L.; Hu, J.H.; Maeder, M.L.; Joung, J.K.; Chen, Z.-Y.; Liu, D.R. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat. Biotechnol.* **2015**, *33*, 73–80. [CrossRef]

38. Williams, D.J.; Puhl, H.L.; Ikeda, S.R. A Simple, Highly Efficient Method for Heterologous Expression in Mammalian Primary Neurons Using Cationic Lipid-mediated mRNA Transfection. *Front. Neurosci.* **2010**, *4*, 181. [CrossRef]

39. Lattanzi, A.; Meneghini, V.; Pavani, G.; Amor, F.; Ramadier, S.; Felix, T.; Antoniani, C.; Masson, C.; Alibeu, O.; Lee, C.; et al. Optimization of CRISPR/Cas9 Delivery to Human Hematopoietic Stem and Progenitor Cells for Therapeutic Genomic Rearrangements. *Mol. Ther.* **2019**, *27*, 137–150. [CrossRef]

40. Xu, X.; Wan, T.; Xin, H.; Li, D.; Pan, H.; Wu, J.; Ping, Y. Delivery of CRISPR/Cas9 for therapeutic genome editing. *J. Gene Med.* **2019**, *21*, e3107. [CrossRef]

41. Lino, C.A.; Harper, J.C.; Carney, J.P.; Timlin, J.A. Delivering CRISPR: A review of the challenges and approaches. *Drug Deliv.* **2018**, *25*, 1234–1257. [CrossRef] [PubMed]

42. User Guides & Protocols. Available online: https://eu.idtdna.com/pages/support/guides-and-protocols (accessed on 19 February 2020).

43. Addgene: CRISPR References and Information. Available online: https://www.addgene.org/crispr/reference/#protocols (accessed on 19 February 2020).

44. CRISPR Protocols and Methods | Springer Nature Experiments. Available online: https://experiments.springernature.com/techniques/crispr (accessed on 19 February 2020).

45. Protocols for the CRISPR/Cas Technology_Bio-protocol. Available online: https://bio-protocol.org/Special_Issue_info.aspx?siid=11 (accessed on 19 February 2020).

46. Vouillot, L.; Thélie, A.; Pollet, N. Comparison of T7E1 and Surveyor Mismatch Cleavage Assays to Detect Mutations Triggered by Engineered Nucleases. *G3: Genes Genomes Genet.* **2015**, *5*, 407–415. [CrossRef] [PubMed]

47. Dabrowska, M.; Czubak, K.; Juzwa, W.; Krzyzosiak, W.J.; Olejniczak, M.; Kozlowski, P. qEva-CRISPR: A method for quantitative evaluation of CRISPR/Cas-mediated genome editing in target and off-target sites. *Nucleic Acids Res.* **2018**, *46*, e101. [CrossRef] [PubMed]

48. Bell, C.C.; Magor, G.W.; Gillinder, K.R.; Perkins, A.C. A high-throughput screening strategy for detecting CRISPR-Cas9 induced mutations using next-generation sequencing. *BMC Genom.* **2014**, *15*, 1002. [CrossRef] [PubMed]

49. 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]

50. 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]

51. Findlay, S.D.; Vincent, K.M.; Berman, J.R.; Postovit, L.-M. A Digital PCR-Based Method for Efficient and Highly Specific Screening of Genome Edited Cells. *PLoS ONE* **2016**, *11*, e0153901. [CrossRef]

52. Lomov, N.A.; Viushkov, V.S.; Petrenko, A.P.; Syrkina, M.S.; Rubtsov, M.A. Methods of Evaluating the Efficiency of CRISPR/Cas Genome Editing. *Mol. Biol.* **2019**, *53*, 862–875. [CrossRef]

53. 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*, 1–8. [CrossRef]

54. Germini, D.; Tsfasman, T.; Zakharova, V.V.; Sjakste, N.; Lipinski, M.; Vassetzky, Y. A Comparison of Techniques to Evaluate the Effectiveness of Genome Editing. *Trends Biotechnol.* **2018**, *36*, 147–159. [CrossRef]

55. Chen, W.; Zhang, G.; Li, J.; Zhang, X.; Huang, S.; Xiang, S.; Hu, X.; Liu, C. CRISPRlnc: A manually curated database of validated sgRNAs for lncRNAs. *Nucleic Acids Res.* **2019**, *47*, D63–D68. [CrossRef] [PubMed]

56. Varshney, G.K.; Zhang, S.; Pei, W.; Adomako-Ankomah, A.; Fohtung, J.; Schaffer, K.; Carrington, B.; Maskeri, A.; Slevin, C.; Wolfsberg, T.; et al. CRISPRz: A database of zebrafish validated sgRNAs. *Nucleic Acids Res.* **2016**, *44*, D822–D826. [CrossRef] [PubMed]

57. Addgene: Validated gRNA Sequences. Available online: https://www.addgene.org/crispr/reference/grna-sequence/ (accessed on 19 February 2020).

58. Tzelepis, K.; Koike-Yusa, H.; De Braekeleer, E.; Li, Y.; Metzakopian, E.; Dovey, O.M.; Mupo, A.; Grinkevich, V.; Li, M.; Mazan, M.; et al. A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep.* **2016**, *17*, 1193–1205. [CrossRef] [PubMed]

59. Torres-Perez, R.; Garcia-Martin, J.A.; Montoliu, L.; Oliveros, J.C.; Pazos, F. WeReview: CRISPR Tools—Live Repository of Computational Tools for Assisting CRISPR/Cas Experiments. *Bioengineering* **2019**, *6*, 63. [CrossRef]

60. Chari, R.; Mali, P.; Moosburner, M.; Church, G.M. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat. Methods* **2015**, *12*, 823–826. [CrossRef]

61. Wong, N.; Liu, W.; Wang, X. WU-CRISPR: Characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol.* **2015**, *16*, 218. [CrossRef]

62. Labuhn, M.; Adams, F.F.; Ng, M.; Knoess, S.; Schambach, A.; Charpentier, E.M.; Schwarzer, A.; Mateo, J.L.; Klusmann, J.-H.; Heckl, D. Refined sgRNA efficacy prediction improves large- and small-scale CRISPR–Cas9 applications. *Nucleic Acids Res.* **2018**, *46*, 1375–1385. [CrossRef]

63. Doench, J.G.; Hartenian, E.; Graham, D.B.; Tothova, Z.; Hegde, M.; Smith, I.; Sullender, M.; Ebert, B.L.; Xavier, R.J.; Root, D.E. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nat. Biotechnol.* **2014**, *32*, 1262–1267. [CrossRef]

64. 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]

65. Lee, C.M.; Davis, T.H.; Bao, G. Examination of CRISPR/Cas9 design tools and the effect of target site accessibility on Cas9 activity. *Exp. Physiol.* **2018**, *103*, 456–460. [CrossRef]

66. Montague, T.G.; Cruz, J.M.; Gagnon, J.A.; Church, G.M.; Valen, E. CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **2014**, *42*, W401–W407. [CrossRef] [PubMed]

67. Labun, K.; Montague, T.G.; Gagnon, J.A.; Thyme, S.B.; Valen, E. CHOPCHOP v2: A web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res.* **2016**, *44*, W272–W276. [CrossRef] [PubMed]

68. Heigwer, F.; Kerr, G.; Boutros, M. E-CRISP: Fast CRISPR target site identification. *Nat. Methods* **2014**, *11*, 122–123. [CrossRef] [PubMed]

69. Cameron, P.; Fuller, C.K.; Donohoue, P.D.; Jones, B.N.; Thompson, M.S.; Carter, M.M.; Gradia, S.; Vidal, B.; Garner, E.; Slorach, E.M.; et al. Mapping the genomic landscape of CRISPR–Cas9 cleavage. *Nat. Methods* **2017**, *14*, 600–606. [CrossRef] [PubMed]

70. Bae, S.; Park, J.; Kim, J.-S. Cas-OFFinder: A fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **2014**, *30*, 1473–1475. [CrossRef]

71. Listgarten, J.; Weinstein, M.; Kleinstiver, B.P.; Sousa, A.A.; Joung, J.K.; Crawford, J.; Gao, K.; Hoang, L.; Elibol, M.; Doench, J.G.; et al. Prediction of off-target activities for the end-to-end design of CRISPR guide RNAs. *Nat. Biomed. Eng.* **2018**, *2*, 38–47. [CrossRef] [PubMed]

72. Abadi, S.; Yan, W.X.; Amar, D.; Mayrose, I. A machine learning approach for predicting CRISPR-Cas9 cleavage efficiencies and patterns underlying its mechanism of action. *PLoS Comput. Biol.* **2017**, *13*, e1005807. [CrossRef]

73. Singh, R.; Kuscu, C.; Quinlan, A.; Qi, Y.; Adli, M. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic Acids Res.* **2015**, *43*, e118. [CrossRef]

74. Chuai, G.; Ma, H.; Yan, J.; Chen, M.; Hong, N.; Xue, D.; Zhou, C.; Zhu, C.; Chen, K.; Duan, B.; et al. DeepCRISPR: Optimized CRISPR guide RNA design by deep learning. *Genome Biol.* **2018**, *19*, 80. [CrossRef]

75. Zhang, D.; Hurst, T.; Duan, D.; Chen, S.-J. Unified energetics analysis unravels SpCas9 cleavage activity for optimal gRNA design. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 8693–8698. [CrossRef]

76. Bae, S.; Kweon, J.; Kim, H.S.; Kim, J.-S. Microhomology-based choice of Cas9 nuclease target sites. *Nat. Methods* **2014**, *11*, 705–706. [CrossRef] [PubMed]

77. Graf, R.; Li, X.; Chu, V.T.; Rajewsky, K. sgRNA Sequence Motifs Blocking Efficient CRISPR/Cas9-Mediated Gene Editing. *Cell Rep.* **2019**, *26*, 1098–1103.e3. [CrossRef] [PubMed]

78. Shibata, A. Regulation of repair pathway choice at two-ended DNA double-strand breaks. *Mutat. Res. Fundam. Mol. Mech. Mutagenesis* **2017**, *803–805*, 51–55. [CrossRef] [PubMed]

79. Her, J.; Bunting, S.F. How cells ensure correct repair of DNA double-strand breaks. *J. Biol. Chem.* **2018**, *293*, 10502–10511. [CrossRef]

80. Taheri-Ghahfarokhi, A.; Taylor, B.J.M.; Nitsch, R.; Lundin, A.; Cavallo, A.-L.; Madeyski-Bengtson, K.; Karlsson, F.; Clausen, M.; Hicks, R.; Mayr, L.M.; et al. Decoding non-random mutational signatures at Cas9 targeted sites. *Nucleic Acids Res.* **2018**, *46*, 8417–8434. [CrossRef] [PubMed]

81. Lemos, B.R.; Kaplan, A.C.; Bae, J.E.; Ferrazzoli, A.E.; Kuo, J.; Anand, R.P.; Waterman, D.P.; Haber, J.E. CRISPR/Cas9 cleavages in budding yeast reveal templated insertions and strand-specific insertion/deletion profiles. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E2040–E2047. [CrossRef]

82. Leenay, R.T.; Aghazadeh, A.; Hiatt, J.; Tse, D.; Hultquist, J.F.; Krogan, N.; Wu, Z.; Marson, A.; May, A.P.; Zou, J. Systematic characterization of genome editing in primary T cells reveals proximal genomic insertions and enables machine learning prediction of CRISPR-Cas9 DNA repair outcomes. *bioRxiv* **2018**, 404947. [CrossRef]

83. Cradick, T.J.; Fine, E.J.; Antico, C.J.; Bao, G. CRISPR/Cas9 systems targeting β-globin and CCR5 genes have substantial off-target activity. *Nucleic Acids Res.* **2013**, *41*, 9584–9592. [CrossRef]

84. 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]

85. Lee, H.; Chang, H.Y.; Cho, S.W.; Ji, H.P. CRISPRpic: Fast and precise analysis for CRISPR-induced mutations via prefixed index counting. *Nar. Genom. Bioinform.* **2020**, *2*. [CrossRef]

86. Connelly, J.P.; Pruett-Miller, S.M. CRIS.py: A Versatile and High-throughput Analysis Program for CRISPR-based Genome Editing. *Sci. Rep.* **2019**, *9*, 1–8. [CrossRef] [PubMed]

87. Lindsay, H.; Burger, A.; Biyong, B.; Felker, A.; Hess, C.; Zaugg, J.; Chiavacci, E.; Anders, C.; Jinek, M.; Mosimann, C.; et al. CrispRVariants charts the mutation spectrum of genome engineering experiments. *Nat. Biotechnol.* **2016**, *34*, 701–702. [CrossRef] [PubMed]

88. 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]

89. Guell, M.; Yang, L.; Church, G.M. Genome editing assessment using CRISPR Genome Analyzer (CRISPR-GA). *Bioinformatics* **2014**, *30*, 2968–2970. [CrossRef] [PubMed]

90. Clement, K.; Rees, H.; Canver, M.C.; Gehrke, J.M.; Farouni, R.; Hsu, J.Y.; Cole, M.A.; Liu, D.R.; Joung, J.K.; Bauer, D.E.; et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **2019**, *37*, 224–226. [CrossRef]

91. Pinello, L.; Canver, M.C.; Hoban, M.D.; Orkin, S.H.; Kohn, D.B.; Bauer, D.E.; Yuan, G.-C. Analyzing CRISPR genome editing experiments with CRISPResso. *Nat. Biotechnol.* **2016**, *34*, 695–697. [CrossRef]

92. Brinkman, E.K.; Chen, T.; de Haas, M.; Holland, H.A.; Akhtar, W.; van Steensel, B. Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Mol. Cell* **2018**, *70*, 801–813.e6. [CrossRef]

93. Li, W.; Xu, H.; Xiao, T.; Cong, L.; Love, M.I.; Zhang, F.; Irizarry, R.A.; Liu, J.S.; Brown, M.; Liu, X.S. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **2014**, *15*, 554. [CrossRef]

94. Hart, T.; Moffat, J. BAGEL: A computational framework for identifying essential genes from pooled library screens. *BMC Bioinform.* **2016**, *17*, 164. [CrossRef]

95. Jeong, H.-H.; Kim, S.Y.; Rousseaux, M.W.C.; Zoghbi, H.Y.; Liu, Z. CRISPRCloud2: A cloud-based platform for deconvolving CRISPR screen data. *bioRxiv* **2018**. [CrossRef]

96. Spahn, P.N.; Bath, T.; Weiss, R.J.; Kim, J.; Esko, J.D.; Lewis, N.E.; Harismendy, O. PinAPL-Py: A comprehensive web-application for the analysis of CRISPR/Cas9 screens. *Sci. Rep.* **2017**, *7*, 15854. [CrossRef] [PubMed]

97. Winter, J.; Schwering, M.; Pelz, O.; Rauscher, B.; Zhan, T.; Heigwer, F.; Boutros, M. CRISPRAnalyzeR: Interactive analysis, annotation and documentation of pooled CRISPR screens. *bioRxiv* **2017**. [CrossRef]

98. Frock, R.L.; Hu, J.; Meyers, R.M.; Ho, Y.-J.; Kii, E.; Alt, F.W. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.* **2015**, *33*, 179–186. [CrossRef] [PubMed]

99. Wang, X.; Wang, Y.; Wu, X.; Wang, J.; Wang, Y.; Qiu, Z.; Chang, T.; Huang, H.; Lin, R.-J.; Yee, J.-K. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.* **2015**, *33*, 175–178. [CrossRef]

100. Crosetto, N.; Mitra, A.; Silva, M.J.; Bienko, M.; Dojer, N.; Wang, Q.; Karaca, E.; Chiarle, R.; Skrzypczak, M.; Ginalski, K.; et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat. Methods* **2013**, *10*, 361–365. [CrossRef] [PubMed]

101. Wienert, B.; Wyman, S.K.; Richardson, C.D.; Yeh, C.D.; Akcakaya, P.; Porritt, M.J.; Morlock, M.; Vu, J.T.; Kazane, K.R.; Watry, H.L.; et al. Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* **2019**, *364*, 286–289. [CrossRef]

102. Tsai, S.Q.; Nguyen, N.T.; Malagon-Lopez, J.; Topkar, V.V.; Aryee, M.J.; Joung, J.K. CIRCLE-seq: A highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. *Nat. Methods* **2017**, *14*, 607–614. [CrossRef]

103. Horlbeck, M.A.; Witkowsky, L.B.; Guglielmi, B.; Replogle, J.M.; Gilbert, L.A.; Villalta, J.E.; Torigoe, S.E.; Tjian, R.; Weissman, J.S. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *eLife* **2016**, *5*, e12677. [CrossRef]

104. Isaac, R.S.; Jiang, F.; Doudna, J.A.; Lim, W.A.; Narlikar, G.J.; Almeida, R. Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *eLife* **2016**, *5*, e13450. [CrossRef]

105. Yarrington, R.M.; Verma, S.; Schwartz, S.; Trautman, J.K.; Carroll, D. Nucleosomes inhibit target cleavage by CRISPR-Cas9 in vivo. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 9351–9358. [CrossRef]

106. Daer, R.M.; Cutts, J.P.; Brafman, D.A.; Haynes, K.A. The Impact of Chromatin Dynamics on Cas9-Mediated Genome Editing in Human Cells. *Acs. Synth. Biol.* **2017**, *6*, 428–438. [CrossRef] [PubMed]

107. Jensen, K.T.; Fløe, L.; Petersen, T.S.; Huang, J.; Xu, F.; Bolund, L.; Luo, Y.; Lin, L. Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Lett.* **2017**, *591*, 1892–1901. [CrossRef] [PubMed]

108. Uusi-Mäkelä, M.I.E.; Barker, H.R.; Bäuerlein, C.A.; Häkkinen, T.; Nykter, M.; Rämet, M. Chromatin accessibility is associated with CRISPR-Cas9 efficiency in the zebrafish (Danio rerio). *PLoS ONE* **2018**, *13*, e0196238. [CrossRef] [PubMed]

109. Chen, X.; Rinsma, M.; Janssen, J.M.; Liu, J.; Maggio, I.; Gonçalves, M.A.F.V. Probing the impact of chromatin conformation on genome editing tools. *Nucleic Acids Res.* **2016**, *44*, 6482–6492. [CrossRef] [PubMed]

110. Canver, M.C.; Joung, J.K.; Pinello, L. Impact of Genetic Variation on CRISPR-Cas Targeting. *Cris. J.* **2018**, *1*, 159–170. [CrossRef] [PubMed]

111. Lessard, S.; Francioli, L.; Alfoldi, J.; Tardif, J.-C.; Ellinor, P.T.; MacArthur, D.G.; Lettre, G.; Orkin, S.H.; Canver, M.C. Human genetic variation alters CRISPR-Cas9 on- and off-targeting specificity at therapeutically implicated loci. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E11257–E11266. [CrossRef]

112. Scott, D.A.; Zhang, F. Implications of human genetic variation in CRISPR-based therapeutic genome editing. *Nat. Med.* **2017**, *23*, 1095–1101. [CrossRef]

113. Chen, C.-L.; Rodiger, J.; Chung, V.; Viswanatha, R.; Mohr, S.E.; Hu, Y.; Perrimon, N. SNP-CRISPR: A Web Tool for SNP-Specific Genome Editing. *G3: Genes Genomes Genet.* **2020**, *10*, 489–494. [CrossRef]

114. Keough, K.C.; Lyalina, S.; Olvera, M.P.; Whalen, S.; Conklin, B.R.; Pollard, K.S. AlleleAnalyzer: A tool for personalized and allele-specific sgRNA design. *Genome Biol.* **2019**, *20*, 167. [CrossRef]