Synthetic Biology, 2026, 11(1), ysag006
https://doi.org/10.1093/synbio/ysag006
Advance Access Publication 23 April 2026
Research Article

Off-target detection of CRISPR-Cas9 nuclease *in vitro* with CROFT-Seq

Paulius Toliusis, Algirdas Grybauskas, Tomas Sinkunas, Tautvydas Karvelis, Giedrius Sasnauskas, Mindaugas Zaremba*

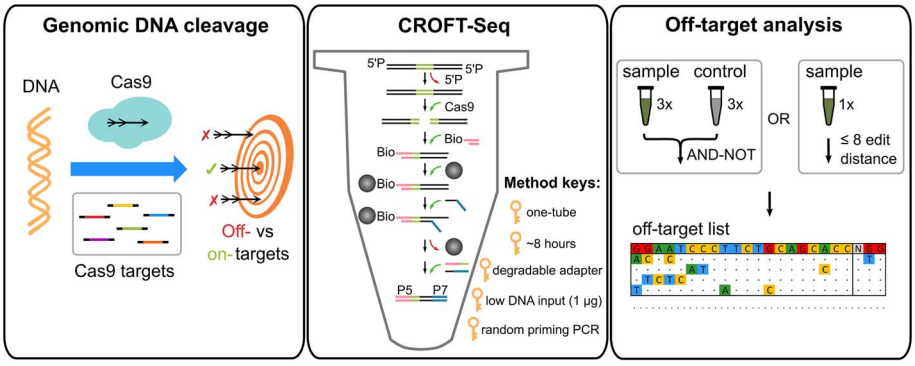
Institute of Biotechnology, Life Sciences Center, Vilnius University, Saulėtekio al. 7, LT-10257 Vilnius, Lithuania

*Corresponding author. Institute of Biotechnology, Life Sciences Center, Vilnius University, Saulėtekio al. 7, LT-10257 Vilnius, Lithuania. E-mail: mindaugas.zaremba@bti.vu.lt

Abstract

Programmable CRISPR-Cas9 nucleases have become invaluable tools for genome editing. However, off-target cleavage by these nucleases can lead to unintended changes in the edited genome. Detection of off-target sites is critical to make genome editing technology safe and predictable. Although current *in vitro* methods for off-target detection can identify these sites, they are time-consuming and complex. Here, we present CROFT-Seq (CRISPR nuclease off-target detection by sequencing), a sensitive, rapid, and affordable assay for the genome-wide detection of Cas9 off-target sites *in vitro*. CROFT-Seq performs comparably to the commonly used *in vitro* methods and serves as a valuable and efficient tool for the rapid assessment of genome-editing nuclease specificity. Notably, a high proportion of the top-ranked off-target sites identified by CROFT-Seq were validated in cells, highlighting its strong predictive performance.

Graphical Abstract



Keywords Cas9; *in vitro*; off-targets; CROFT-Seq

Introduction

Programmable genome editing nucleases, such as CRISPR-Cas9, hold great promise for therapeutic applications and have attracted widespread attention from the scientific community [1–3]. These enzymes generate double-strand breaks (DSBs) in genomic DNA that can be repaired in the cell by error-prone nonhomologous end joining (NHEJ), which can result in insertion/deletion mutations (indels), or precise homology-directed repair (HDR) [4, 5]. Nevertheless, unintentional cleavage can occur at various DNA sites similar to the target sequence, where even rare instances of cleavage may result in unwanted or potentially deleterious gene mutations and translocations [6].

The most widely used programmable nuclease for clinical use is Cas9 from *Streptococcus pyogenes* [7]. The Cas9 protein associates with guide RNA (gRNA) composed of either crRNA-tracrRNA or single-guide RNA and forms an active ribonucleoprotein (RNP) complex [8, 9]. This complex can be directed to any genomic site of interest based on complementarity between gRNA and target DNA. Recognition of the

Received: 25 March 2026. **Accepted:** 20 April 2026

target DNA also depends on a short protospacer adjacent motif (PAM) located downstream of the target site, which is recognized by the Cas9 protein. This leads to complementary hybridization of the gRNA to the 20-base pair (bp) target sequence [10]. Upon target binding, the HNH and RuvC nuclease domains of the Cas9 protein cleave both DNA strands, generating a double-strand break (DSB). However, the Cas9 complex can bind and cleave similar DNA target sites (off-targets) with nucleotide mismatches, affecting the precision of the programmable genome editing nucleases [11].

A number of genome-wide cell-based and cell-free methods have been developed to detect off-target cleavage sites. The most widely used cell culture-based assays, such as GUIDE-Seq [12], IDVL [13], LAM-HTGTS [14], BLESS/BLISS [15, 16], DISCOVER-Seq [17], and ChIP-Seq [18], can directly or indirectly detect Cas9-induced DSBs in the cell, making them valuable tools for preliminary evaluations prior to initiating effort-intensive *in vivo* genome editing experiments. However, these methods have several limitations, including low sensitivity, dependence on cell transfection efficiency, chromatin accessibility, cell type, and other factors [19]. Although the cell-free methods detect off-targets only *in vitro,* they detect a significantly higher number of off-target sites and allow precise control of RNP concentration, amount of purified DNA, and cleavage time. This capability enables efficient and rapid RNP prescreening to identify the most suitable RNP candidates for genome engineering applications. However, validation of off-target sites of the selected most suitable candidates is still required before their application *in vivo*. To date, several cell-free genome-wide off-target identification assays have been developed, including Digenome-seq [20], SITE-Seq [21], CIRCLE-Seq/CHANGE-Seq [22, 23], CLEAVE-Seq [24], AID-Seq [25], INDUCE-Seq [26], and RGEN-Seq [27]. While these techniques can identify a significantly greater number of potential off-target sites than cell-based approaches, they also have certain limitations. For example, Digenome-seq requires a large number (>400 million) of reads per sample due to sequencing of a high background of randomly sheared DNA [20]. SITE-Seq, CIRCLE-Seq, RGEN-Seq, AID-Seq, and CLEAVE-Seq are relatively sensitive but are prone to false-positive off-target detection (low off-target validation rate) [21, 22, 24, 25, 27]. There are also PCR-free methods, including RGEN-Seq and INDUCE-Seq [26]. Although these methods may provide a more accurate distribution of cleavage events due to the absence of PCR amplification, they lack the sensitivity of PCR-based methods and are associated with higher sequencing expenditure. Furthermore, all the methods discussed above are labour-intensive, time-consuming, and costly, and typically require DNA shearing and multiple DNA purification steps using spin columns or magnetic beads, resulting in significant DNA loss.

In this study, we present CROFT-Seq, an *in vitro* cell-free method designed to thoroughly detect genome-wide off-target sites induced by programmable genome editing nucleases, such as CRISPR-Cas9. We show here that CROFT-Seq exhibits sensitivity comparable to the widely used *in vitro* methods CIRCLE-Seq and SITE-Seq. Furthermore, CROFT-Seq outperforms existing *in vitro* off-target detection methods in terms of simplicity and hands-on time and provides a streamlined, one-tube method that is readily amenable to automation. Finally, we have shown that CROFT-Seq detects off-target sites at frequencies that closely match those observed in cells, making it an attractive method for benchmarking Cas9 proteins prior to their use for genome editing applications.

Materials and methods

Cell culture and transfection

HEK293T cells (ATCC CRL-3216) were cultured in Dulbecco's Modified Eagle Medium (DMEM) medium supplemented with 10% FBS (Sigma), 100 U/ml penicillin, and 100 μg/ml streptomycin (Thermo Fisher Scientific) at 37°C and 5% CO_2. The cells were maintained at confluency below 80%. Cells were plated in a 12-well plate with ∼250 000 cells per well in 500 μl culture medium 1 day before transfection. All transfections were performed with 3.6 μl TurboFect transfection reagent (Thermo Fisher Scientific) per 1.4 μg of plasmid encoding only Cas9 or Cas9 and gRNA in 180 μl of DMEM medium. The transfected cells were collected 3 days post-transfection. Genomic DNA was purified with GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific) and quantified by Qubit 4.0 fluorimeter (Thermo Fisher Scientific).

CROFT-Seq library preparation

CROFT-Seq experiments with *FANCF*, *VEGFA1*, *XRCC5*, *EMX1*, *RNF2*, *HBB*, *VEGFA2*, *VEGFA3,* and *CD151* targeting gRNAs (Supplementary Protocol) or without gRNAs were performed on human genomic DNA (Roche). Genomic DNA was treated with Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific). *In vitro* cleavage reactions were performed in a 40 μl reaction volume containing 100 nM SpCas9, 200 nM gRNA (Synthego), and 1 μg of human genomic DNA. Digested products were ligated to a biotinylated adapter and treated with Exonuclease I (Thermo Fisher Scientific). Ligated DNA products were immobilized on streptavidin-coated MyOne C1 magnetic beads (Thermo Fisher Scientific). The complementary DNA strand was removed with NaOH. The nonbiotinylated DNA strand was synthesized using a DNA oligonucleotide and T4 DNA polymerase (Thermo Fisher Scientific). DNA was then removed from the magnetic beads and amplified by PCR using Phusion Plus DNA polymerase (Thermo Fisher Scientific). The prepared libraries were quantified on Agilent 2100 Bioanalyzer (Agilent) and sequenced with 100 or 150 bp paired-end reads on an Illumina NextSeq 550 or 2000, or NovaSeq X Plus platform. The depth of coverage was ∼1–60 million reads per sample.

rhAmpSeq deep sequencing

First, off-target lists for each gRNA were identified by CROFT-Seq. Then, only those off-target sites, which were in the top 5% according to the read depth, were selected for validation in cells. For example, in the case of the VEGFA1 target site, the top 1 off-target read depth is 9438, so all off-target sites, whose read depth is above 5% of the top 1–ranked off-target site read value (above $9438 * 5\% = 471.9$) were selected for validation. A total of 190 off-target sites (including on-target sites) were selected for three target sites (*FANCF*, *VEGFA1,* and *XRCC5*) (Supplementary Tables S3 and S4). PCR rhAmpSeq primers were successfully designed for all 190 sites and ordered from Integrated DNA Technologies (IDT) (Supplementary Table S2). Following the manufacturer's protocol, rhAmpSeq PCR libraries were amplified from 25 ng of purified HEK293T genomic DNA duplicates obtained from edited HEK293T cells as described above using the rhAmpSeq CRISPR library kit (IDT). The resulting PCR products were pooled into different libraries corresponding to the different gRNAs used for transfection and purified using Ampure XP magnetic beads (Agencourt). The resulting rhAmpSeq libraries were quantified and analysed on an

Agilent 2100 Bioanalyzer (Agilent) and sequenced with 100 bp paired-end reads on a NextSeq 2000 instrument. The depth of coverage was ∼30 million reads per sample (∼300 000 reads per off-target).

The sequencing data were analysed using the manufacturer's recommended rhAmpSeq CRISPR analysis tool (https://eu.idtdna.com/pages/tools/rhampseq-crispr-analysis-tool). For each off-target site, indel frequencies (Δ indels, %) were calculated by subtracting two averaged control cells (edited with SpCas9 only) from two averaged transfected cells (edited with SpCas9-gRNA complex). The off-target site was counted as validated if the calculated Δ indel frequency was >0.1%.

CROFT-Seq data analysis pipeline

The analysis started with paired-end reads that were used as initial data. The removal of residual adapter sequences and low-quality reads was performed using AdapterRemoval [28], utilizing default parameters and appropriate adapter pairs specified in the method. If the sequencing was being performed with NovaSeq X Plus, which is known to produce polyG ends, the additional polyG removal process was applied with FASTP [29]. The cleaned reads were aligned to the genome using BWA-MEM [30] and sorted and indexed with SAMtools [31]. The alignment file generation and manipulation were executed using default settings. For each cleavage site search experiment, there were three target and three control samples. Initially, each of the six samples was analysed independently using the 'find-cleavage-patterns.jl' script. This script utilizes a 4 bp reading window sliding along the chromosomes to search for read start positions in the aligned R1 reads. If the number of read start positions within the reading window range is ≥20 (read depth change ≥20), the region is included for further analysis. From each included region, the cleavage position is selected based on the presence of reads with forward and reverse directions or, if absent, the position with the highest number of read start positions. If the reading window has multiple potential cleavage positions, then the position having bidirectional read start positions has the highest priority, even if there is another cleavage position with a greater count of monodirectional reads. Afterwards, cleavage positions with the greatest count of read start positions are prioritized. Subsequently, the target sequence is searched around the chosen cleavage position using the Needleman–Wunsch algorithm [32] (realized in the package 'BioJulia/BioAlignments.jl'). As the cleavage might occur in the sense or antisense direction, both variants are checked: from −19 bp to +5 bp forward and from −5 bp to +19 bp reverse complement sequences. Then, the best potential alignment position for the sequence is recorded in the BED format file. After all six samples were analysed individually (three target site replicates and three control replicates), the results were combined using the 'combine-cleavage-patterns.jl' script. The target and control samples were combined separately. This combining process was used to assess whether the identified cleavage sites were consistently found within their respective sample groups. The cleavage site was considered present if the exact cleavage position appeared in all three target samples (read depth change ≥20). Following the combinations, the cleavage sites, which were also found in all control samples (read depth change ≥20), were excluded from the target samples. The deduction process was performed using the same script as in the combination step. In parallel, to evaluate the plasticity and optimization capability of the CROFT-Seq bioinformatic analysis pipeline for off-target site detection, we applied an approach where the combination and deduction processes were omitted, as this workflow version uses only a single replicate of the target sample for generating the off-target list. Additional filtering was performed by removing potential off-target sites that have a sequence edit distance (mismatches and/or gaps) >8 compared to the target sequence. The final CROFT-Seq lists of potential on/off-target regions were exported in the BED format files.

Comparison of CROFT-Seq with other off-target detection methods

To compare CROFT-Seq with other *in vitro* and cell-based off-target detection methods, we obtained off-target sequences of different target sites from previously published datasets [12, 20–23, 25–27]. Start positions of the off-target sites in RGEN-Seq and SITE-Seq had to be shifted by −1 bp due to discrepancies in the chromosome coordinate system. The off-target lists of CIRCLE-Seq, DIGENOME-Seq, and GUIDE-Seq were originally mapped to the hg19 reference genome and had to be remapped to hg38 using the LiftOver tool [33]. DIGENOME-Seq off-target lists also lacked information about both strand directions and off-target sequence positions within the chromosomes. The information was retrieved by utilizing BLAT search [34]. Lastly, AID-seq and INDUCE-Seq lacked the information about the strand directions and they were detected with the simple usage of the pairwise sequence alignment comparison function found in BioPython's package [35]. All off-target sites detected by other methods and which were selected to use for comparisons with CROFT-Seq in this study are provided in Supplementary Table S5. In one set of comparisons where two or three methods are compared, we showed overlap of Venn diagrams at different ranges (top20, top50, and total lists). In another set of comparisons where more than three methods are compared, we showed the overlap of different methods in Upset plots. The DBSCAN clustering technique from the *scikit-learn* package [36] was used to accommodate possible cleavage position differences between the different methods, where 5 bp cleavage position variation was allowed.

To compare validation data from CROFT-Seq with those from SITE-Seq and CIRCLE-Seq, we obtained validation data for SITE-Seq and CIRCLE-Seq from the original studies. All of the validated off-target sites used for comparisons are provided in Supplementary Table S4. In the case of SITE-Seq and CIRCLE-Seq, all *in vitro*-detected off-target sites were used for the validation test. For CROFT-Seq, only *in vitro*-detected off-target sites with read depth exceeding 5% of the top1 off-target site were used. Then those off-target sites that met the criteria (Δ indels% > 0.1 – for CROFT-Seq and SITE-Seq, tag integration% > 0 – for CIRCLE-Seq) were considered as validated and used for comparisons.

Results

Development and optimization of CROFT-Seq

Here, we aimed to develop a robust *in vitro* off-target detection method that enhances multiple aspects of the off-target detection process while maintaining high sensitivity. We hypothesized that a one-pot method, without requiring physical or enzymatic DNA fragmentation, would be the most suitable assay for rapid and affordable genome-wide detection of off-target sites induced by CRISPR-Cas9 nucleases *in vitro* (Figure 1, Supplementary Figure S1A, and Supplementary Protocol). To achieve these objectives, we implemented and extensively optimized various reaction parameters. Before nuclease treatment, CROFT-Seq utilizes dephosphorylation (treatment with phosphatase)

to prevent randomly abundant DNA ends from adapter ligation, thereby increasing sensitivity [37]. Next, phosphatase-treated DNA is incubated with CRISPR-Cas9 nuclease, which produces cleavage products predominantly with blunt-end 5′ phosphate-containing termini [8, 9]. To verify cleavage activity, on-target cleavage is assessed using qPCR by comparing the amount of uncleaved DNA in the reaction treated with the SpCas9-gRNA complex to that in the negative-control reaction, which is treated with SpCas9 alone (Supplementary Figure S1A). The cleavage products are then ligated to a biotinylated adapter. To minimize DNA loss during the unligated adapter removal step, we designed a unique adapter that can be selectively removed by DNA exonuclease in the same reaction tube. The adapter-ligated DNA is then immobilized on streptavidin-coated magnetic beads to facilitate buffer exchange for subsequent reactions. Next, to reduce the amount of nonspecific DNA in the downstream steps, the complementary, nonbiotinylated DNA strand is removed by NaOH treatment [37]. The remaining immobilized ssDNA is then used as a template to synthesize new complementary, nonbiotinylated DNA strands. To avoid the need for DNA fragmentation, CROFT-seq utilizes T4 DNA polymerase along with a primer containing a 12-nucleotide random sequence at the 3′-end, generating short DNA fragments suitable for Illumina sequencing. The DNA is then released from streptavidin-coated magnetic beads and quantified by qPCR. Finally, DNA is amplified by PCR using oligonucleotides containing Illumina sequencing adapters, purified, and sequenced (Supplementary Figure S1A).

To identify cleavage sites, we developed a dedicated algorithm for the analysis of the sequencing data. Briefly, sequencing reads are trimmed to remove adapter sequences and then aligned to the reference genome. The aligned reads are then processed using the CROFT-Seq pipeline, which identifies read start positions using a 4 bp sliding-window approach. Candidate off-target site positions are accepted if they appear in all replicates and are filtered out if they are detected in control reactions (Figure 1C). This algorithm does not require filtering based on the total number of mismatches and/or gaps between on- and off-target sites. In parallel, we tested the bioinformatic pipeline for off-target analysis, which requires only a single replicate per target site. Consistent with other *in vitro* methods, this workflow applies filtering that permits up to eight mismatches and/or gaps between on- and off-target sites (Supplementary Figure S1B).

Off-target identification by CROFT-Seq

The CROFT-Seq method was tested on human genomic DNA treated with SpCas9 RNP complexes, targeting *FANCF, VEGFA1*, and *XRCC5* sites that have been previously used by other *in vitro* off-target detection methods. A total of 427, 385, and 263 off-target sites were identified for FANCF, VEGFA1, and XRCC5, respectively. Off-targets were ranked by the count of the mapped reads. As expected, most detected off-target sites showed sequence similarity to the on-target site (Supplementary Table S1). Furthermore, we generated sequence logos for off-target sites for the FANCF, VEGFA1, and XRCC5 guide RNAs, taking into account nucleotide frequency at each position (Figure 2A). For FANCF and VEGFA1 guides, mismatch/gap frequencies were lower in the PAM and seed sequences, consistent with prior mechanistic and structural studies, indicating that mismatches at the PAM and seed regions are less tolerated than mismatches at the PAM-distal end of off-targets [11, 38–40].

Next, we evaluated the reproducibility of CROFT-Seq by comparing read counts across technical replicates using gRNAs targeting the *FANCF*, *VEGFA1*, and *XRCC5* sites. The off-target site read counts for these gRNAs were highly reproducible between replicates (Pearson's correlation coefficient R^2 values exceeding 0.95, Spearman's rank correlation coefficient r_s values from 0.4 to 0.8, Supplementary Figure S2A–C, Supplementary Table S1). Additionally, we assessed the read count reproducibility between two independent CROFT-Seq experiments (each experiment averaged from three technical replicates) targeting the *FANCF* site, which also demonstrated high reproducibility and correlation (R^2 = 0.99, r_s = 0.823, Supplementary Figure S2D, Supplementary Table S1).

Testing if the read depth change threshold (10, 20, or 30 reads) for off-target site detection affects the variability in detected off-target sites by CROFT-Seq (Supplementary Figure S2E, Supplementary Table S1), we compared off-target sites of FANCF and VEGFA1. The results showed high overlap in top20 and top50 ranges, while the total off-target sites detected are highly influenced by the read depth. This indicates that the most abundant off-target sites (top20 and top50) are the most relevant when comparing different off-target site detection methods.

To test additional targets, we selected six target sites (EMX1, RNF2, HBB, CD151, VEGFA2, and VEGFA3) that have previously been examined using other *in vitro* off-target detection methods. Comparison of analysis approaches using three replicates with control reactions and using only a single replicate with an ≤8 edit distance filter on the FANCF, VEGFA1, and XRCC5 sites revealed that the top20 and top50 off-target lists overlapped by >85% across all target sites (Supplementary Figure S3). Due to sufficient sensitivity, the approach employing only a single replicate was used to generate off-target profiles for six additional target sites (Supplementary Table S1).

Validation of off-target sites identified by CROFT-Seq

We further tested whether the off-target sites identified by CROFT-Seq *in vitro* were also cleaved by SpCas9 in cells. HEK293T cells were transfected with plasmids encoding SpCas9 and gRNA targeting the *FANCF, VEGFA1,* and *XRCC5* sites. After 3 days post-transfection, the frequencies of insertion/deletion mutations (indels) at the respective sites were quantified using rhAmpSeq (Supplementary Table S2). All off-target sites for which read counts exceeded 5% of the number of read counts for the most abundant off-target (top1) were selected for validation. This included 38 off-target sites for *FANCF*, 52 for *VEGFA1,* and 100 for *XRCC5* (including on-target sites) detected by CROFT-Seq (Figure 2B and C and Supplementary Table S3). These sites were annotated as edited if the difference in indel frequency (Δ indels, %) between the sample treated with SpCas9 and gRNA (cleaved) and the sample treated with SpCas9 only (control) exceeded a threshold value of 0.1%. Validation experiments demonstrated that 10 of the 38 off-target sites tested for *FANCF*, 30 of the 52 off-target sites tested for *VEGFA1,* and 9 of the 100 off-target sites tested for *XRCC5* were successfully confirmed as edited in HEK293T cells (Figure 2B–D and Supplementary Table S4).

Comparison of CROFT-Seq with other *in vitro* off-target detection methods

To measure the performance of CROFT-Seq, we compared data for three different gRNAs, targeting *VEGFA1*, *FANCF,* and *XRCC5* sites in

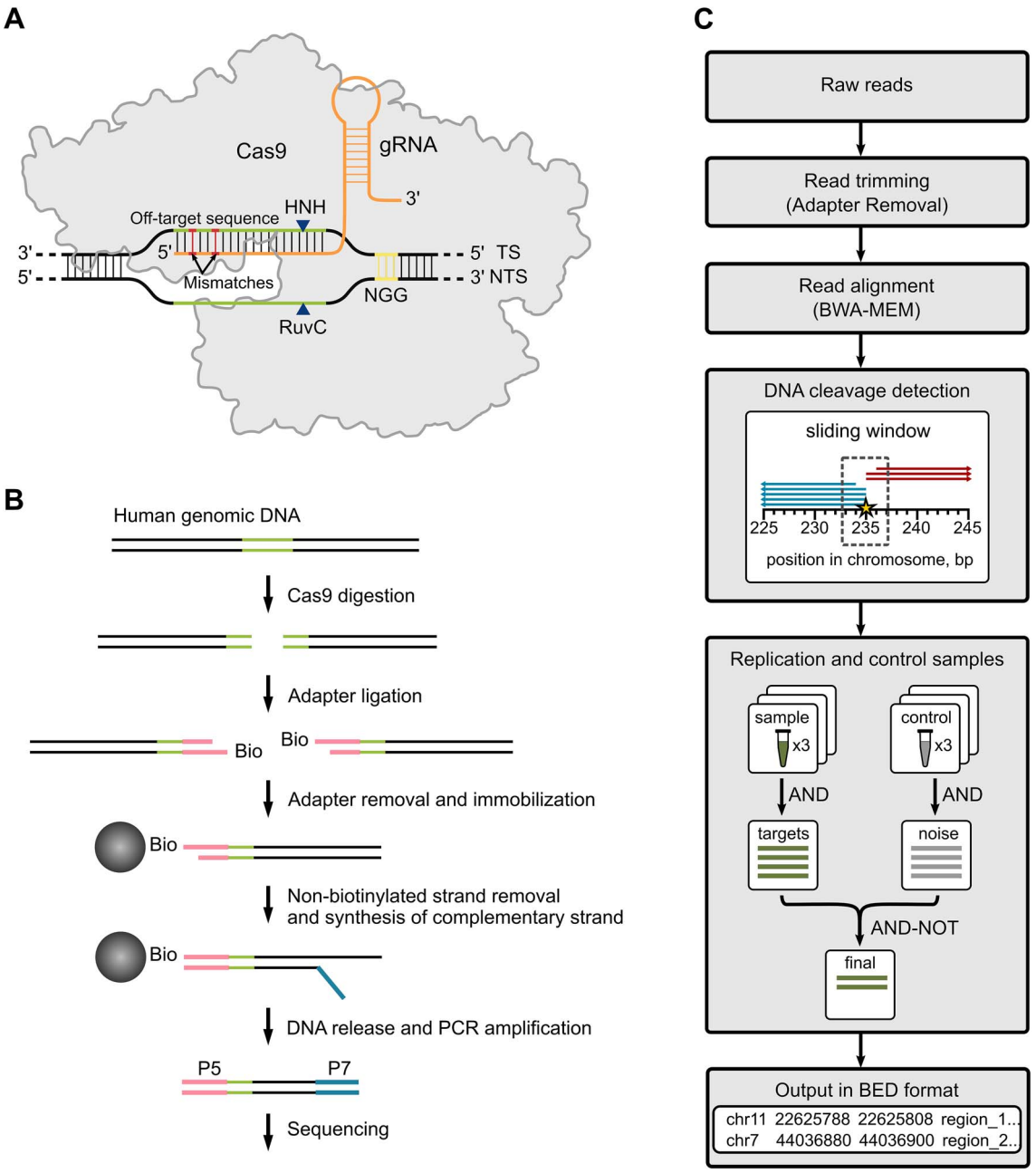


Figure 1 A schematic overview of CROFT-Seq. (A) A schematic illustration of *S. pyogenes* Cas9 with a gDNA bound to an off-target dsDNA containing mismatches proximal to an NGG PAM sequence. (B) Simplified schematic representation of the CROFT-Seq workflow. Human genomic DNA is treated with a phosphatase before digestion with the Cas9 nuclease. The resultant DNA ends are then selectively ligated to a biotinylated adapter. Excess of the adapter is then removed, and the ligated DNA is enriched with magnetic beads. The complementary, nonbiotinylated DNA strand is removed, and a new second DNA strand is synthesized. The resultant DNA is released from the beads and amplified by PCR for sequencing. (C) Standard bioinformatic analysis workflow of CROFT-Seq. Pair-end reads, sequenced and cleaned of residual adapter sequences, are first aligned to the reference genome. Aligned reads are then analysed with an off-target detection script that searches for steep read depth changes using a 4 bp reading window and prioritizes bidirectionality of the potential off-target-related reads and target sequence similarity. Only the off-target positions detected in all three sample replicates are analysed further. These positions are filtered out if they are found in all three control samples.

the human genome that had previously been tested by two of the most widely used *in vitro* off-target detection methods, SITE-Seq and CIRCLE-Seq [21, 22] (Figure 3A, Supplementary Tables S1 and S5).

For the *FANCF* and *VEGFA1* sites, the top20 and top50 off-target lists of all three CROFT-Seq, SITE-Seq, and CIRCLE-Seq methods overlapped by ∼50%. The CROFT-Seq and SITE-Seq top50 lists overlapped by approx. 40% of the same *XRCC5* target site (Supplementary Figure S4A, Supplementary Tables S1 and S5). These results indicate that all these methods perform with comparable accuracy in off-target detection.

In addition, we performed a more comprehensive comparison of CROFT-Seq with four *in vitro* off-target detection methods, CIRCLE-Seq, SITE-Seq, RGEN-Seq, and DIGENOME-Seq on *FANCF* and *VEGFA1* sites (Supplementary Figure S4C and D, Supplementary Tables S1 and S5).

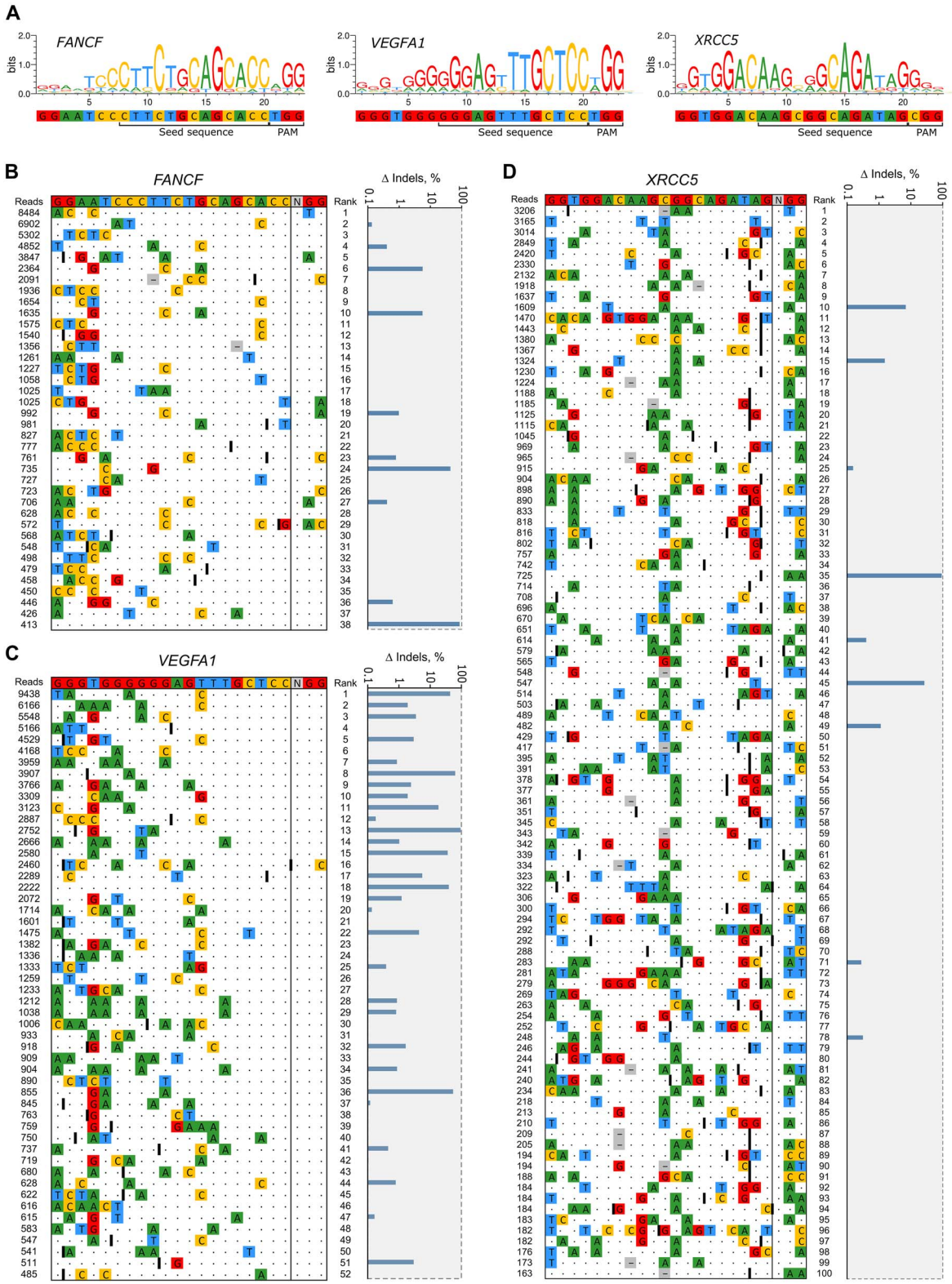


Figure 2 Validation of off-target sites detected by CROFT-Seq. (A) Sequence logos generated from FANCF, VEGFA1, and XRCC5 off-target sequences detected by CROFT-Seq. Validation results of off-target sites detected by CROFT-Seq for gDNA targeting *FANCF* (B), *VEGFA1* (C), and *XRCC5* (D). Only off-target sites with read counts exceeding 5% of the top-ranked off-target site (top1) that were selected for validation are shown. The on-target site is shown at the top of each off-target list; mismatched nucleotides are coloured, gRNA bulges are shown as grey squares with horizontal lines, and DNA bulges are shown as black lines (I). On the right is the logarithmic plot showing the changes in indel frequency observed by rhAmpSeq between nuclease-treated and control samples. Off-target sites are considered validated if the indel frequency change is >0.1%.

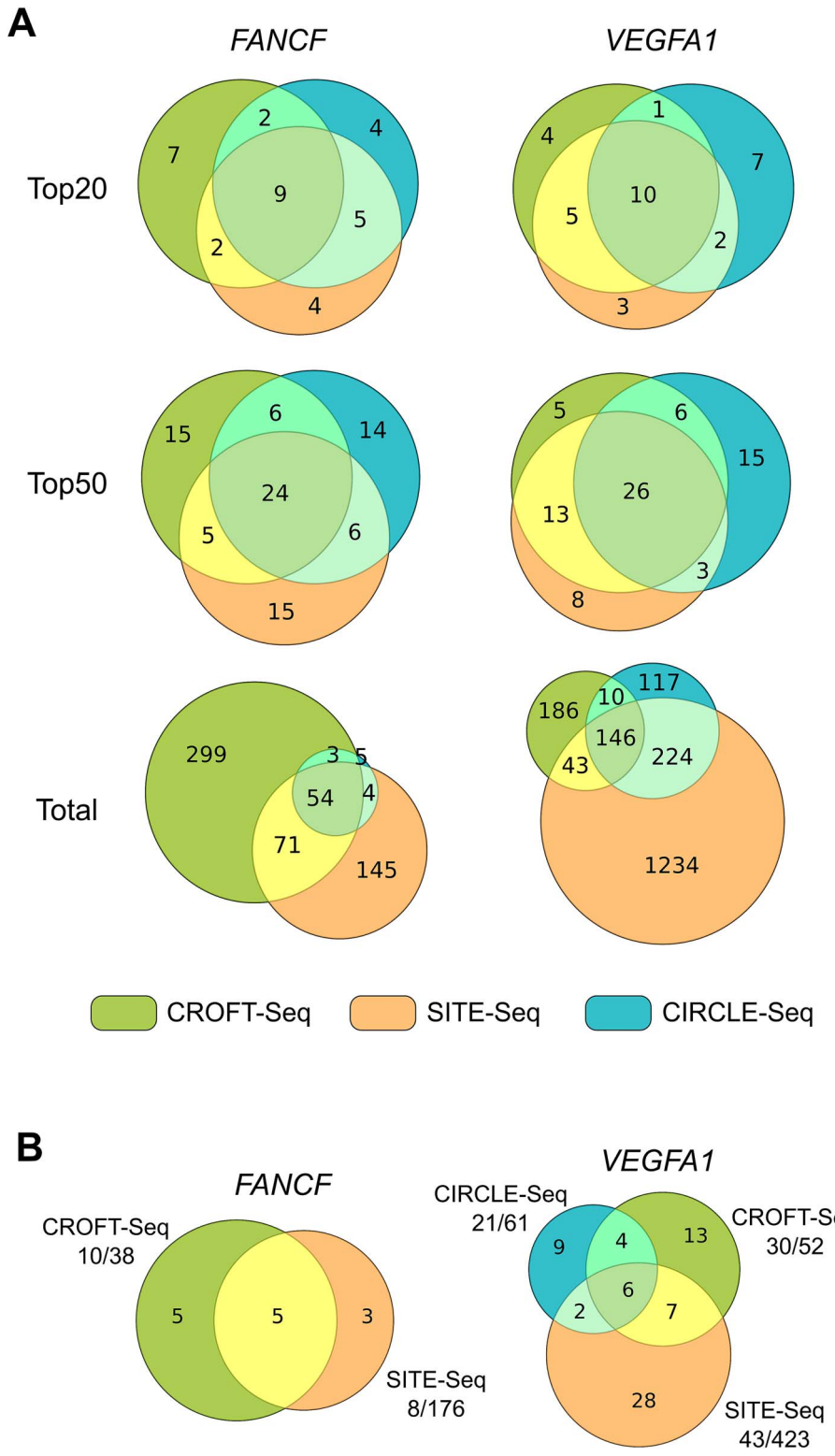


Figure 3 Comparison of CROFT-Seq with SITE-Seq and CIRCLE-Seq *in vitro* off-target detection methods. (A) Venn diagrams showing the overlap of detected off-target sites between CROFT-Seq, CIRCLE-Seq, and SITE-Seq. The used gRNA targeting *FANCF* or *VEGFA1* is shown at the top, and overlapping ranges (top20, top50, and the whole off-target list) used for comparison are shown on the left. (B) Venn diagrams showing overlap of in-cell validated off-target sites for *FANCF* and *VEGFA1* target sites between CROFT-Seq, SITE-Seq, and CIRCLE-Seq. The numbers shown after each off-target method name (e.g. 10/38) indicate the total number of off-target sites subjected to the validation test [38], of which ten [10] off-target sites showed a positive signal (Δ indels% $>$ 0.1 – for CROFT-Seq and SITE-Seq, tag integration% $>$ 0 – for CIRCLE-Seq).

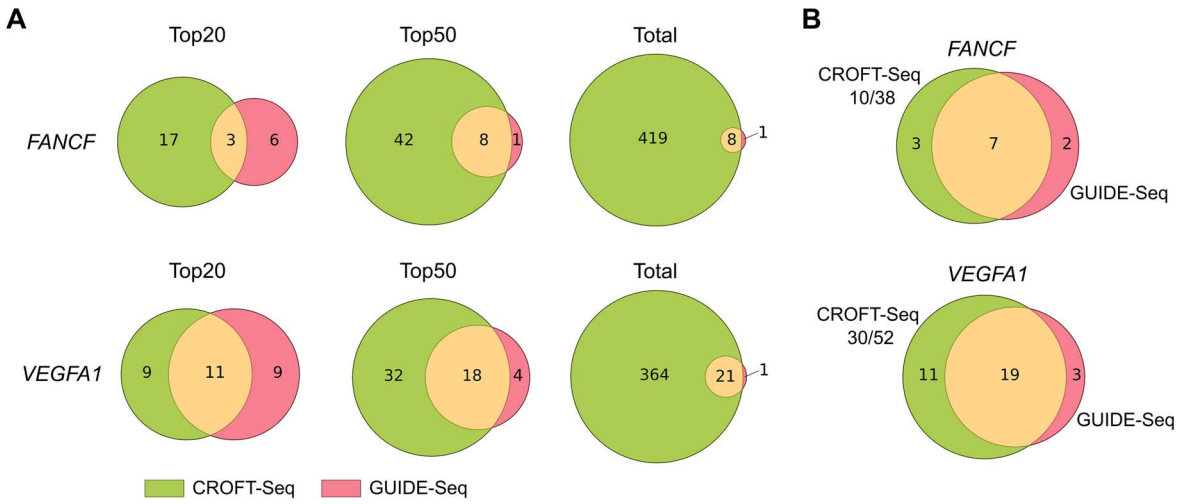


Figure 4 Comparison of CROFT-Seq with the *in vivo* off-target method GUIDE-Seq. (A) Venn diagrams showing the overlap at different ranges (top20, top50, and total off-target list) of detected off-target sites between CROFT-Seq and GUIDE-Seq using gRNA targeting the *FANCF* and *VEGFA1* sites. (B) Venn diagrams showing the overlap of validated off-target sites for the *FANCF* and *VEGFA1* target sites between CROFT-Seq and GUIDE-Seq. The numbers shown for CROFT-Seq (e.g. 10/38) indicate the total number of off-target sites subjected to validation [38], of which ten [10] off-target sites were considered as validated (Δ indels% $>$ 0.1).

In this analysis, CROFT-Seq showed one of the strongest intersections with other methods, underscoring its accuracy and robustness in off-target site detection.

Next, we compared CROFT-Seq with six *in vitro* off-target detection methods (CIRCLE-Seq, SITE-Seq, RGEN-Seq, DIGENOME-Seq, AID-Seq, and INDUCE-Seq) for six additional sites (*EMX1*, *RNF2*, *HBB*, *CD151*, *VEGFA2*, and *VEGFA3*) (Supplementary Figure S5, Supplementary Tables S1 and S5). Again, CROFT-Seq showed one of the strongest intersections with other methods, indicating its accuracy and robustness.

Finally, we sought to compare how the in-cell validated off-target sites of CROFT-Seq overlap with those of the SITE-Seq and CIRCLE-Seq methods (Figure 3B and Supplementary Figure S4B). The results show that the validated off-targets of CROFT-Seq overlap well with those of SITE-Seq and CIRCLE-Seq. Moreover, when compared to these methods, the CROFT-Seq showed a high yield of validated off-target sites, suggesting that CROFT-Seq is a sensitive and accurate predictor of off-target sites arising in cells.

Comparison of CROFT-Seq with the *in vivo* GUIDE-Seq method

We also compared *in vitro* detected and validated off-target sites of CROFT-Seq with the widely used *in vivo* off-target identification method, GUIDE-Seq (Figure 4A and B).

CROFT-Seq detects almost all off-target sites identified by GUIDE-Seq, and most of them are already detected at the top of the CROFT-Seq off-target list (Figure 4A, see top50). Careful inspection of the off-targets detected by GUIDE-Seq but undetected by CROFT-Seq revealed that one FANCF off-target was not detected by CROFT-Seq because the number of reads [15] at the cleavage site in one reaction replicate was below the set off-target detection threshold ($\geq$20 reads in each reaction replicate) (Supplementary Figure S6A). Another undetected *VEGFA1* off-target was excluded from the CROFT-Seq list by the bioinformatic pipeline due to not unambiguously mapped reads at the cleavage site (Supplementary Figure S6B). The validated CROFT-Seq off-targets overlapped quite well with the GUIDE-Seq off-targets (Figure 4B). Moreover, CROFT-Seq determined additional in-cell validated off-targets that were not detected by the GUIDE-Seq (Figure 4B). These results demonstrate the power of CROFT-Seq to accurately detect off-target sites occurring *in vivo*.

Discussion

Genome editing tools such as CRISPR-Cas9 are widely used in clinical trials and in the development of biotechnologically relevant products. While they can be precisely directed to specific genomic DNA sites, they are not always accurate and may lead to unintended genome edits [41–43]. Therefore, before selecting and using these tools *in vivo*, it is crucial to verify their efficiency and accuracy to ensure they cleave on-target with minimal off-target edits. *In vitro* methods for off-target detection are particularly well suited for this purpose.

Here, we present CROFT-Seq, our *in vitro* off-target detection method for the comprehensive evaluation of Cas9 cleavage sites in human genomic DNA. CROFT-Seq has several key improvements over other *in vitro* methods (Supplementary Table S6).

CROFT-Seq requires a relatively low amount of genomic DNA (1 μg), making it an effective tool for detecting nuclease-induced off-target sites in cells and tissues where the amount of isolated DNA is limited. In addition, we have significantly reduced both the cost per sample and hands-on time required to perform CROFT-Seq compared to other methods. This is achieved using individual enzymes, reagents, and buffers with known compositions rather than relatively expensive commercial kits with proprietary and undisclosed compositions used in other methods. Moreover, due to its unique reaction compositions and conditions, CROFT-Seq does not require multiple DNA library purification steps or DNA fragmentation, making it suitable for performing the whole protocol in a single reaction tube, thus allowing for simple automation.

Despite the simple and affordable design, CROFT-Seq performs comparably to the widely used *in vitro* off-target detection methods CIRCLE-Seq and SITE-Seq. In our comparison with SITE-Seq and CIRCLE-Seq, we observed that all methods detected ~50% of the most common (in the top20 and top50 ranges) *in vitro*-generated

off-target sites (Figure 3A). This indicates a comparable sensitivity between these methods. Moreover, CROFT-Seq detects more off-target sites for the FANCF target site but fewer for VEGFA1 than CIRCLE-Seq or SITE-Seq. However, comparing the total off-target site lists is less informative because the number of off-target sites detected by each method depends strongly on sequencing depth and the off-target site read threshold, and may vary substantially across methods. Thus, off-target sites at the bottom of the ranking are less reliable.

Furthermore, the CROFT-Seq data analysis algorithm does not necessarily require additional filtering steps for off-target site identification, such as setting specific mapping quality (MAPQ) scores for read alignment or applying mismatch count thresholds for off-target calling. Such filters may introduce bias and negatively impact the overall off-target detection. However, the mismatch and/or gap count threshold between on- and off-targets may be applied. As tested in the CROFT-Seq analysis pipeline, ≤8 edit distance filtering reduces the need for three replicates and control reactions per target site without affecting off-target detection accuracy. Lastly, our in-cell validation results indicate that CROFT-Seq is a reliable predictor of *in cellulo* off-target sites due to its high success rate in off-target validation (Figure 2).

Specific optimization of the CROFT-Seq assay for off-target site detection means that the on-target sites may not always rank at the top of the off-target list. Indeed, the FANCF, VEGFA1, and XRCC5 on-target sites were ranked at positions 40, 18, and 35 of the off-target lists, respectively. The ability to detect on-target sites is highly dependent on the concentration of RNP complex, which can influence their ranking [21]. Moreover, the RuvC nuclease domain of SpCas9 can trim the nontarget strand after cleavage [44], making the cleavage products less effectively ligatable to the blunt-ended adapter. Additionally, certain off-target sites might be cleaved more rapidly than on-target sites due to sequence context or the secondary structure of the target DNA [45, 46]. Consequently, these off-target sites are likely to be detected with higher read counts compared to on-target sites.

Next, we compared CROFT-Seq with the cell-based method GUIDE-Seq, a widely used approach for the direct detection of off-target sites in the edited cells. The results showed that the off-target sites detected by CROFT-Seq match well with those identified by the GUIDE-Seq method (Figure 4A). Furthermore, CROFT-Seq has detected additional validated off-target sites that were not identified by GUIDE-Seq (Figure 4B). However, a major limitation of cell-free methods is their tendency to produce a significant number of false-positive off-target sites, which are difficult to eliminate [47]. Addressing this challenge will require further optimization of experimental and data analysis workflows.

The sensitivity, affordability, time efficiency, and potential ease of automation make CROFT-Seq a powerful and attractive tool for high-throughput identification of off-target sites for genome editing nucleases. Looking ahead, CROFT-Seq could be further adapted for off-target detection of other emerging genome editing tools, including Cas12, TnpB, and base editors.

Acknowledgements

We thank Prof. V. Siksnys for discussions and suggestions. Also, we thank R. Zedaveinyte for her support with *in vivo* transfection experiments. We thank the EMBL GeneCore facility for performing high-throughput sequencing.

Author contributions

M.Z., G.S., T.K., and T.S. designed experiments to develop the CROFT-Seq method. P.T. performed all CROFT-Seq and cell-based validation experiments. A.G. wrote the CROFT-Seq bioinformatic analysis pipeline. P.T., A.G., T.K., G.S., and M.Z. contributed to developing the CROFT-Seq bioinformatic algorithm and sequencing data analysis. P.T. and M.Z. wrote the manuscript with the help of other co-authors.

Supplementary data

Supplementary data are available at *SYNBIO* online.

Conflict of interest

The authors declare competing financial interests. P.T., A.G., T.S., T.K., G.S., and M.Z. applied for a patent of CROFT-Seq.

Funding

Central Project Management Agency grant [01.2.2-CPVA-K-703-02-0010 to M.Z.]; Vilnius University Research Promotion grant (MSF-JM-05/2024 to P.T.); and Research Council of Lithuania (LMTLT) [S-MIP-23-131 to M.Z.].

Data availability

CROFT-Seq and rhAmpSeq sequencing raw data can be accessed at the NCBI BioProject database under accession number PRJNA1156337. Other supporting data of this study are available from the corresponding author upon request. CROFT-Seq open-source software is free and available at https://github.com/agrybauskas/croft-seq-analysis and https://doi.org/10.5281/zenodo.18791456. Pseudocodes for the cleavage detection algorithm and the combination of sample replicates and controls in CROFT-Seq workflow are presented in Supplementary Figures S7 and S8.

Material availability

Some of the material used in this study is available upon reasonable request and upon signing an MTA.

References

1. Hsu PD, Lander ES, Zhang F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 2014;**157**:1262–78. https://doi.org/10.1016/j.cell.2014.05.010
2. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014;**346**:1258096. https://doi.org/10.1126/science.1258096
3. Naldini L. Gene therapy returns to Centre stage. *Nature* 2015;**526**:351–60. https://doi.org/10.1038/nature15818
4. Xue C, Greene EC. DNA repair pathway choices in CRISPR-Cas9-mediated genome editing. *Trends Genet* 2021;**37**:639–56. https://doi.org/10.1016/j.tig.2021.02.008
5. Fu Y-W, Dai X-Y, Wang W-T *et al.* Dynamics and competition of CRISPR–Cas9 ribonucleoproteins and AAV donor-mediated NHEJ, MMEJ and HDR editing. *Nucleic Acids Res* 2021;**49**:969–85. https://doi.org/10.1093/nar/gkaa1251

6. Wienert B, Cromer MK. CRISPR nuclease off-target activity and mitigation strategies. *Front Genome Ed* 2022;**4**:1050507. https://doi.org/10.3389/fgeed.2022.1050507
7. Hryhorowicz M, Lipiński D, Zeyland J *et al.* CRISPR/Cas9 immune system as a tool for genome engineering. *Arch Immunol Ther Exp* 2017;**65**:233–40. https://doi.org/10.1007/s00005-016-0427-5
8. Gasiunas G, Barrangou R, Horvath P *et al.* Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* 2012;**109**:E2579–86. https://doi.org/10.1073/pnas.1208507109
9. Jinek M, Chylinski K, Fonfara I *et al.* A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;**337**:816–21. https://doi.org/10.1126/science.1225829
10. Nishimasu H, Ran FA, Hsu PD *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 2014;**156**:935–49. https://doi.org/10.1016/j.cell.2014.02.001
11. Hsu PD, Scott DA, Weinstein JA *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 2013;**31**:827–32. https://doi.org/10.1038/nbt.2647
12. Tsai SQ, Zheng Z, Nguyen NT *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* 2015;**33**:187–97. https://doi.org/10.1038/nbt.3117
13. Wang X, Wang Y, Wu X *et al.* Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat Biotechnol* 2015;**33**:175–8. https://doi.org/10.1038/nbt.3127
14. Frock RL, Hu J, Meyers RM *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol* 2015;**33**:179–86. https://doi.org/10.1038/nbt.3101
15. Crosetto N, Mitra A, Silva MJ *et al.* Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Methods* 2013;**10**:361–5. https://doi.org/10.1038/nmeth.2408
16. Yan WX, Mirzazadeh R, Garnerone S *et al.* BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat Commun* 2017;**8**:15058. https://doi.org/10.1038/ncomms15058
17. Wienert B, Wyman SK, Richardson CD *et al.* Unbiased detection of CRISPR off-targets in vivo using DISCOVER-Seq. *Science* 2019;**364**:286–9. https://doi.org/10.1126/science.aav9023
18. Iacovoni JS, Caron P, Lassadi I *et al.* High-resolution profiling of γH2AX around DNA double strand breaks in the mammalian genome. *EMBO J* 2010;**29**:1446–57. https://doi.org/10.1038/emboj.2010.38
19. Guo C, Ma X, Gao F *et al.* Off-target effects in CRISPR/Cas9 gene editing. *Front Bioeng Biotechnol* 2023;**11**:1143157. https://doi.org/10.3389/fbioe.2023.1143157
20. Kim D, Bae S, Park J *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods* 2015;**12**:237–43. https://doi.org/10.1038/nmeth.3284
21. Cameron P, Fuller CK, Donohoue PD *et al.* Mapping the genomic landscape of CRISPR–Cas9 cleavage. *Nat Methods* 2017;**14**:600–6. https://doi.org/10.1038/nmeth.4284
22. Tsai SQ, Nguyen NT, Malagon-Lopez J *et al.* CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR–Cas9 nuclease off-targets. *Nat Methods* 2017;**14**:607–14. https://doi.org/10.1038/nmeth.4278
23. Lazzarotto CR, Malinin NL, Li Y *et al.* CHANGE-seq reveals genetic and epigenetic effects on CRISPR–Cas9 genome-wide activity. *Nat Biotechnol* 2020;**38**:1317–27. https://doi.org/10.1038/s41587-020-0555-7
24. Young J, Zastrow-Hayes G, Deschamps S *et al.* CRISPR-Cas9 editing in maize: systematic evaluation of off-target activity and its relevance in crop improvement. *Sci Rep* 2019;**9**:6729. https://doi.org/10.1038/s41598-019-43141-6
25. Tian R, Cao C, He D *et al.* Massively parallel CRISPR off-target detection enables rapid off-target prediction model building. *Med* 2023;**4**:478–492.e6. https://doi.org/10.1016/j.medj.2023.05.005
26. Dobbs FM, Van Eijk P, Fellows MD *et al.* Precision digital mapping of endogenous and induced genomic DNA breaks by INDUCE-seq. *Nat Commun* 2022;**13**:3989. https://doi.org/10.1038/s41467-022-31702-9
27. Kuzin A, Redler B, Onuska J *et al.* RGEN-seq for highly sensitive amplification-free screen of off-target sites of gene editors. *Sci Rep* 2021;**11**:23600. https://doi.org/10.1038/s41598-021-03160-8
28. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes* 2016;**9**:88. https://doi.org/10.1186/s13104-016-1900-2
29. Chen S. Fastp 1.0: an ultra-fast all-round tool for FASTQ data quality control and preprocessing. *iMeta* 2025;**4**:e70078. https://doi.org/10.1002/imt2.70078
30. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 2013; arXiv:1303.3997.
31. Danecek P, Bonfield JK, Liddle J *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* 2021;**10**:giab008. https://doi.org/10.1093/gigascience/giab008
32. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 1970;**48**:443–53. https://doi.org/10.1016/0022-2836(70)90057-4
33. Hinrichs AS, Karolchik D, Baertsch R *et al.* The UCSC genome browser database: update 2006. *Nucleic Acids Res* 2006;**34**:D590–8. https://doi.org/10.1093/nar/gkj144
34. Kent WJ. BLAT —the BLAST -like alignment tool. *Genome Res* 2002;**12**:656–64.
35. Cock PJA, Antao T, Chang JT *et al.* Biopython: freely available python tools for computational molecular biology and bioinformatics. *Bioinformatics* 2009;**25**:1422–3. https://doi.org/10.1093/bioinformatics/btp163
36. Kramer O. Scikit-Learn. In: Kramer O (ed.), *Machine Learning for Evolution Strategies*. Studies in Big Data, Vol. **20**. Cham: Springer International Publishing, 2016, pp. 45–53. https://doi.org/10.1007/978-3-319-33383-0
37. Murgha YE, Rouillard J-M, Gulari E. Methods for the preparation of large quantities of complex single-stranded oligonucleotide libraries. *PLoS One* 2014;**9**:e94752. https://doi.org/10.1371/journal.pone.0094752
38. Fu Y, Foden JA, Khayter C *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 2013;**31**:822–6. https://doi.org/10.1038/nbt.2623
39. Pattanayak V, Lin S, Guilinger JP *et al.* High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* 2013;**31**:839–43. https://doi.org/10.1038/nbt.2673
40. Pacesa M, Lin C-H, Cléry A *et al.* Structural basis for Cas9 off-target activity. *Cell* 2022;**185**:4067–4081.e21. https://doi.org/10.1016/j.cell.2022.09.026
41. Hunt JMT, Samson CA, du Rand A *et al.* Unintended CRISPR-Cas9 editing outcomes: a review of the detection and prevalence of structural variants generated by gene-editing in human cells. *Hum Genet* 2023;**142**:705–20. https://doi.org/10.1007/s00439-023-02561-1
42. Lackner M, Helmbrecht N, Pääbo S *et al.* Detection of unintended on-target effects in CRISPR genome editing by DNA donors carrying diagnostic substitutions. *Nucleic Acids Res* 2023;**51**:e26–6. https://doi.org/10.1093/nar/gkac1254

43. Park SH, Cao M, Pan Y *et al.* Comprehensive analysis and accurate quantification of unintended large gene modifications induced by CRISPR-Cas9 gene editing. *Sci Adv* 2022;**8**:eabo7676. https://doi.org/10.1126/sciadv.abo7676
44. Stephenson AA, Raper AT, Suo Z. Bidirectional degradation of DNA cleavage products catalyzed by CRISPR/Cas9. *J Am Chem Soc* 2018;**140**:3743–50. https://doi.org/10.1021/jacs.7b13050
45. Lin Y, Cradick TJ, Brown MT *et al.* CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res* 2014;**42**:7473–85. https://doi.org/10.1093/nar/gku402
46. Boyle EA, Becker WR, Bai HB *et al.* Quantification of Cas9 binding and cleavage across diverse guide sequences maps landscapes of target engagement. *Sci Adv* 2021;**7**:eabe5496. https://doi.org/10.1126/sciadv.abe5496
47. Huang S, Huang X. A massively parallel approach for assessing CRISPR off-targets in vitro. *Cell Rep Methods* 2023;**3**:100561. https://doi.org/10.1016/j.crmeth.2023.100561

Synthetic Biology, 2026, **11**, ysag006
https://doi.org/10.1093/synbio/ysag006
Research Article