The CRISPR Journal
Volume 5, Number 1, 2022

DOI: 10.1089/crispr.2021.0113

The CRISPR Journal

RESEARCH ARTICLE

Inference of CRISPR Edits from Sanger Trace Data

David Conant,[†,*] Tim Hsiau,[†] Nicholas Rossi, Jennifer Oki, Travis Maures, Kelsey Waite, Joyce Yang, Sahil Joshi, Reed Kelso, Kevin Holden, Brittany L. Enzmann, and Rich Stoner

Abstract

Efficient and precise genome editing requires a fast, quantitative, and inexpensive assay to assess genotype following editing. Here, we present ICE (Inference of CRISPR Edits), which enables robust analysis of CRISPR edits using Sanger data. ICE proposes potential outcomes for editing with guide RNAs, and then determines which are supported by the data via regression. The ICE algorithm is robust and reproducible, and it can be used to analyze CRISPR experiments within days after transfection. We also confirm that ICE produces accurate estimates of editing outcomes across a variety of benchmarks, and within the context of other existing Sanger analysis tools. The ICE tool is free to use and open source, and offers several improvements over current analysis tools, such as batch analysis and support for a variety of editing conditions. It is available online at ice.synthego.com, and the source code is available at github.com/synthego-open/ice.

Introduction

CRISPR is a precise programmable tool used for genome editing, which involves a guide RNA (gRNA) that binds to a genomic locus and a Cas nuclease that creates a double-strand break (DSB) at the targeted site.[1,2] After a DSB occurs, non-homologous end joining (NHEJ) and/or microhomology-mediated end joining (MMEJ) can introduce insertions or deletions (indels) at the DSB.[3,4] Alternatively, if a DNA donor template is provided, knocking in a sequence of interest is possible through homology-directed repair (HDR).[5,6] While CRISPR offers precise control over the site of the DSB, both the overall efficiency of the edit (% non–wild type) and the particular genotypes resulting from a CRISPR experiment are highly variable and heterogeneous across cells within the same experiment.[7] This variability requires techniques to quantify the post-editing outcomes in order to extract meaningful results from CRISPR-based experiments.[8]

If the researcher cares only about the overall editing efficiency of a CRISPR experiment, techniques such as the T7 endonuclease assay [9] provide a fast and accessible estimate. However, these methods have limited precision and reproducibility, and do not provide insight into the specific genotypes in the edited population.[10] Next-generation sequencing (NGS) offers sequence-level resolution and gold-standard sensitivity,[10] but it is less widely available, has a long turnaround time, and comes at a high cost per sample in lower-throughput applications.

An alternative to NGS is the application of Sanger sequencing to CRISPR genotyping. Sanger sequencing is extremely fast and widely available, but the signal it produces is a convolution of all genotypes in the sample. Therefore, a computational tool is required in order to infer the individual sequences within the population. Previous tools for deconvolving individual contributions from ensemble Sanger data include base-calling methods,[11–15] compressed sensing,[16] and, most significantly, TIDE (Tracking of Insertions and DEletions)[17] and DECODR (DEconvolution of COmplex DNA Repair).[18] TIDE and DECODR are both effective tools at breaking down convoluted Sanger reads into component parts within a simple workflow, but both have shortcomings. First, TIDE requires that users independently sequence their donors for HDR edits, requiring additional time at the bench. In addition, TIDE is not open source, and many features, such as batch processing, require a license. DECODR improves on TIDE by not only supporting HDR edits without sequencing the donor but also by supporting genotyping edits produced from multiple guides, large deletions, and Cas12a edits.[18] However,

Synthego, Redwood City, California, USA.
[†]These authors contributed equally to this work.

**Address correspondence to: Travis Maures, PhD, Synthego, 3696 Haven Avenue, Suite A, Redwood City, CA 94063* Email: travis.maures@synthego.com

DECODR, like TIDE, is not open source and only supports batch analysis with edits that share the same guides and donors.

Here, we present an improved algorithm, called ICE (Inference of CRISPR Edits), which provides a unified front-end for low- and high-throughput genotyping as well as an open-source code base to serve the CRISPR editing community. ICE can be used to analyze knockout edits as well as donor-mediated editing. In this paper, we describe the ICE algorithm and its outputs, as well as validate it against two benchmarks: (1) a comparison with NGS for donor and donor-less editing, and (2) an analysis of single-nucleotide polymorphism (SNP) variants on known cell populations. Moreover, we review ICE within the context of TIDE and DECODR by contextualizing the performance of each tool on a large number of edits. The ease and flexibility of ICE enables experimenters to assess editing faster and cheaper, without compromising accuracy.

Methods

ICE algorithm

The ICE algorithm (Fig. 1) relies on two Sanger sequencing chromatogram trace files: one capturing the edited population, and one unedited control. Following the alignment of these sequences to the provided guide sequence, we generate a list of possible genotypes that might result from a DSB at the inferred cut site based on the known biological activity of Cas9. Because the generation of possible genotypes is anchored to our *a priori* understanding of Cas9, ICE may be less accurate in genotyping other nucleases' editing outcomes, including base editors. These possible genotypes encompass all frequent indels identified in the literature.[7] Finally, a regression algorithm is used to determine the linear combination of these possible genotypes that best explain the observed edited sequencing trace. The relative frequency of the resulting genotypes is inferred from the weights of the regression model. These steps are described in more detail below. This process shares many features of the TIDE and DECODR algorithms with added improvements to support more analysis cases with fewer demands on the user.

Step 1: The two trace files, as ab1 files, are aligned by finding a high-quality window of the control trace upstream of the cut site and trimming it to end at least 15 bp upstream of the cut site. This alignment window is defined as a region of the Sanger trace that has a windowed average with Phred quality scores of >30. The alignment window in the control is then aligned against the edited sample. By ignoring the poor-quality bases often found at the very beginning of a Sanger trace, we found that this alignment method is robust and scales well for reliably processing many ab1 files without manual tuning

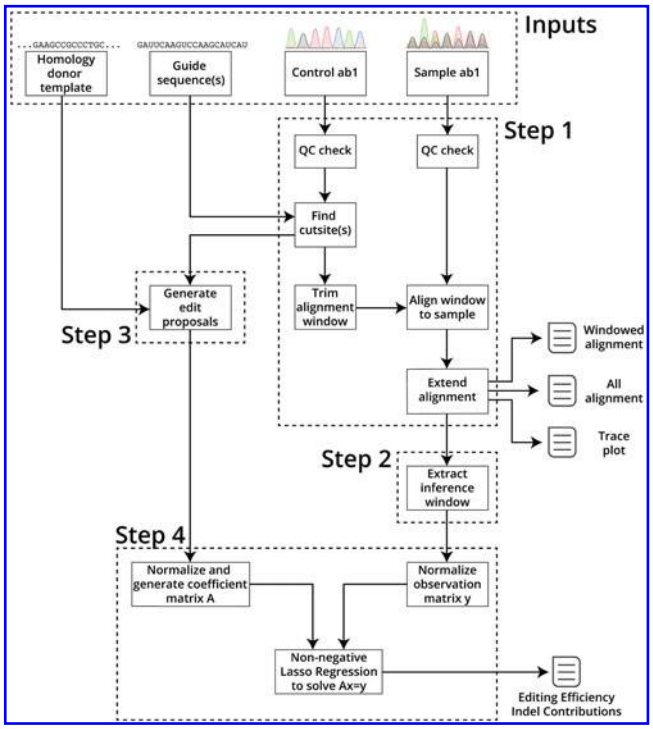

FIG. 1. Algorithm flow chart. The inputs to the algorithm (top) are the control ab1 file, the sample ab1 file, the guide sequence(s), and an optional DNA donor. The algorithm checks the data quality, generates edit proposals, and then runs a non-negative, L1-regularized regression to identify which edit proposal sequences are most likely present in the sample. The program then outputs the quality of the results, the percentage of the sample population that has been edited, and the identity of the edited sequences.

Step 2: Next, to locate the edited region of the sequence used for the regression, we define the inference window. The inference window starts 25 bp upstream of the cut site and extends up to 100 bp downstream of the cut site. The exact window length is calculated by the quality score of the control sample. The length of the inference window is limited due to the tendency of Sanger sequencing quality to decrease over the length of the read, which ultimately can diminish the precision of the regression.

Step 3: In order to support analysis of a variety of gene editing use cases, the edit proposal process can follow two paths:

(1) Knockouts/NHEJ: the algorithm uses a default indel range of deletions of up to 40 bp (20 bp from either

side of the cut site) and insertions of up to 14 bp to generate a list of potential edits (sequences and traces). For deletions, the associated trace data are deleted, while for insertions, a uniform distribution of 25% for each base is inserted. The trace data for other bases are copied from the wild type.

(2) Knock-ins/HDR: the algorithm first generates edit proposals as in (1) and then creates a HDR proposal by aligning the provided donor template with the control sequence and swapping any mismatched bases with that of the donor. This mimics the expected outcome of perfect donor integration. To achieve proper alignment, there must be at least 15 bp of target homology on both ends of the template.

Step 4: After the edit proposal stage, a regression is performed to infer the frequencies of each proposal sequence. In the regression, x is solved for in the equation $Ax = y$, where A is a matrix composed of the simulated traces and y is the edited sequencing trace. This regression finds a linear combination of the edit proposals that best explains the observed trace of the edited sample. We deviated from a standard regression solution in two ways. First, we imposed a non-negative constraint to the values of x, as negative prevalences of proposals are meaningless. Second, we added an L1 (Lasso) regularization penalty to the solution. Lasso regression mitigates overfitting to the noise in Sanger data via L1 regularization, producing more parsimonious results compared to unregularized solutions. Following the regression solution, the relative frequency of each edit proposal is extracted from the vector of weights of the regression (x). Frequencies of individual edits are rounded to the nearest whole percentage point to avoid overstating the confidence of the model's accuracy of contribution estimations. The correlation between the sum of regression-weighted contributions and the observed edited sequencing trace (r^2) measures the extent to which the edit proposals can explain the edited sequencing trace. A low r^2 therefore represents an experiment where ICE is unable to find a combination of indels that adequately explains the observed experimental trace, and the ICE results may not be reliable. This is typically caused by a nonspecific polymerase chain reaction (PCR) amplicon or poor Sanger sequencing quality.

Program outputs

After ICE has completed a run, it outputs a summary JSON and Microsoft Excel spreadsheet, as well as detailed reports on the sequencing quality, alignments, and results for each edit. The webtool additionally produces a variety of visualizations to help the user interpret their results (Fig. 2).

Source code

The source code for ICE is publicly available at github.com/synthego-open/ice, a docker container is on the docker hub at synthego/ice, and a publicly accessible webtool can be found at http://ice.synthego.com.

Using ICE

The user must first PCR amplify the genomic region (with >100 bp flanking both sides of the cut site) on both an edited sample, as well as an unedited control sample. Following Sanger sequencing on both samples, the user then provides the resulting ab1 files for edit and control samples and the guide target sequence used to generate the edit. If the edit was done with the presence of a donor, then the donor sequence can also be provided to estimate the prevalence of donor integration in the edited sample.

Following submission of the sequencing files and guide's target sequence, the algorithm will align the sequences, locate the cut site, populate possible repair genotypes, and infer the genotypes that best recapitulate the edited Sanger chromatogram trace (for more details, see Methods). The software outputs various files and visualizations that help the user check the quality of the editing reactions, verify the alignments, and interpret the results (Fig. 2).

CRISPR editing of cell cultures

Editing was performed with chemically modified synthetic guide RNAs (sgRNAs; Synthego, Redwood City, CA) at 96 unique genomic locations in HEK293 cells. The sgRNAs were complexed with wild-type SpCas9 (Aldevron, Fargo, ND) at a molar ratio of 9:1 (180 pmol sgRNA/20 pmol Cas9) to form ribonucleoproteins (RNPs) and transfected into cells using a Nucleofector™ 4D (Lonza, Basel, Switzerland). Transfected cells were recovered in normal growth medium, plated onto 96-well tissue culture plates, and incubated in humidified 37°C/5% CO_2. After 48 h, cells were lysed, and genomic DNA was extracted from the cells using QuickExtract™ DNA Extraction Solution (Lucigen, Middleton, WI) to each well of the plate in preparation for sequencing (see below).

To specifically test ICE's ability to estimate rates of HDR accurately, we evaluated an additional 15 sites in HEK293 cells using modified sgRNAs (Synthego) and single-stranded DNA (ssDNA) donor templates (Eurofins Genomics, Luxembourg). Each site was tested with three

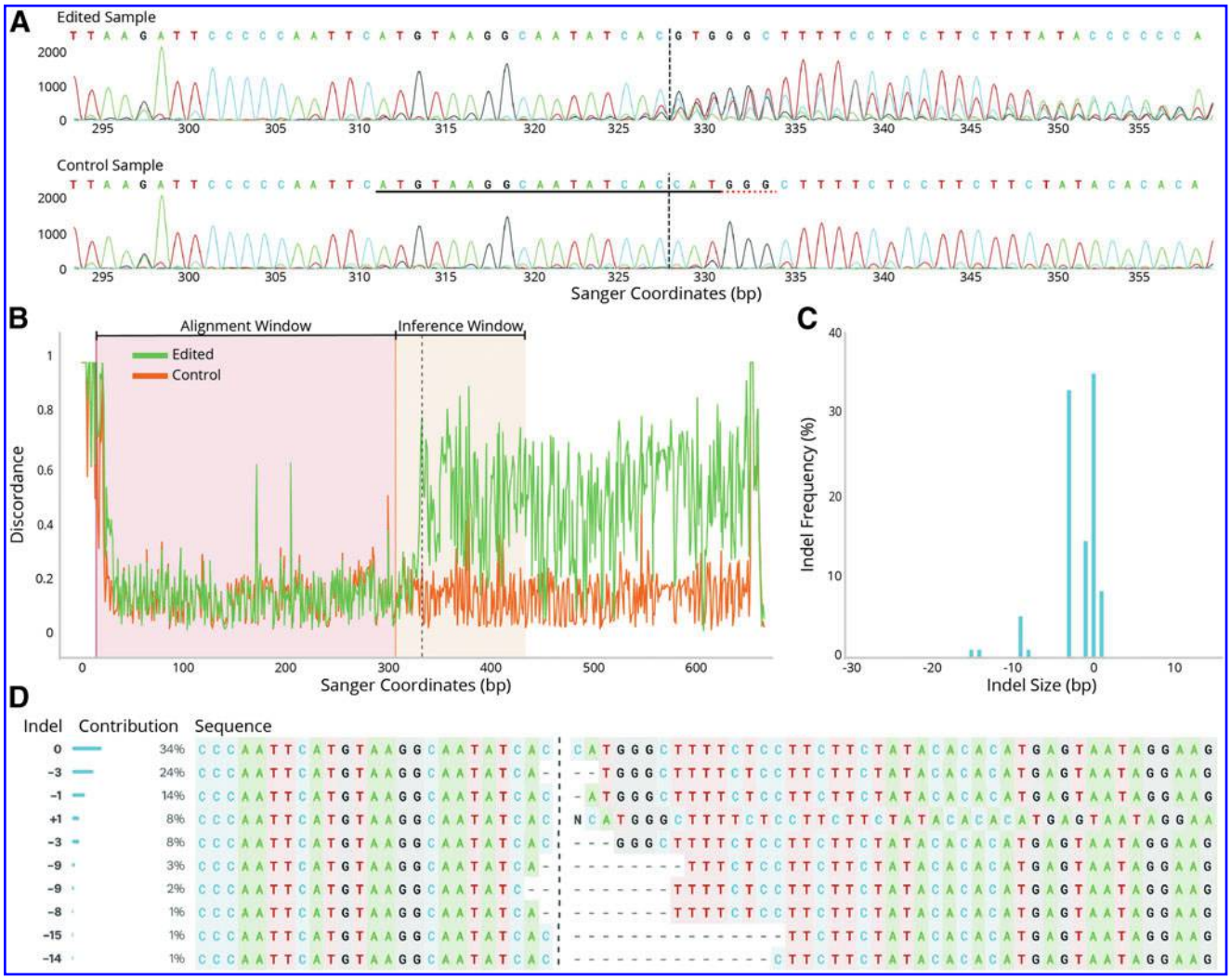

FIG. 2. An example of the outputs from the ICE (Inference of CRISPR Edits) software for a guide targeting the human gene *RYK*. **(A)** Trace file segments spanning the cut site from the control and the edited samples are generated for every analysis. The guide target sequence provided by the user is underlined in black, and the protospacer adjacent motif sequence is denoted by a dotted red underline in the control sample. Vertical dotted lines denote the expected cut site. **(B)** Discordance for the edited (green) and control (orange) trace files. The vertical dotted line marks the cut site. The alignment window marks the region of the traces with high Phred scores that is used to align the edited and control traces. The inference window marks the region of the traces around the cut site, which will be used to infer the change in sequence between the edited and control traces. Visualizing this way, we can see that discordance is a robust signal of trace irregularity that can approximate the location and prevalence of editing. **(C)** Insertion or deletion (indel) sizes along with their relative prevalence for this example, as calculated by ICE. **(D)** Exact sequence calls and their relative prevalence for this example.

different ssDNA templates designed to knock in sequences of varying length (+0 [SNP], +14 or +36 bp) with symmetric 40 bp homology arms (45 samples total). sgRNAs (90 pmol) were complexed with 10 pmol Cas9 and then transfected along with 30 pmol ssDNA. All other experimental details were the same as the donorless experiments.

Sanger sequencing and NGS

For all samples, PCR primers were designed to amplify a 500–800 bp genomic DNA segment containing the cut site. PCR was performed on lysed genomic samples using Amplitaq Gold 360 (Thermo Fisher Scientific, Waltham, MA), according to the manufacturer's instructions. Sanger sequencing was then performed through a

commercial vendor (Sequetech, Mountain View, CA) with one of the two primers used for amplification.

We additionally performed NGS on all single-guide samples. For NGS samples, a 200–300 bp segment containing the cut site was PCR amplified. The amplicons were purified, quantified using NanoDrop™, and sent to the Mass General Hospital (MGH, Boston, MA) DNA core facility for their CRISPR sequencing service. A summarization analysis was performed using the MGH NGS data pipeline, which reported the sequences and abundance of genotypes. Nine samples repeatedly failed Sanger sequencing and were dropped from subsequent analysis, leaving 92 donor-less samples and 40 donor-mediated samples.

Single nucleotide variant analysis

To assess ICE's ability to infer the prevalence of introduced single nucleotide variants (SNVs), we sequenced DNA mixtures that simulate a range of variant outcomes. We amplified the locus surrounding SNP rs2072579 from HEK293s (C/C) and the PGP1 iPSC line (G/G). Sanger sequencing confirmed the samples are homozygous and differ at only the SNP location. We then quantified the amplicons with a Fragment Analyzer (AATI, Ankeny, Iowa), mixed them at different molar ratios (5%, 10%, 20%, 40%, 60%, 80%, 90%, and 95% of PGP1 in the mixture), and sequenced the mixed samples. The resulting sequencing data were then analyzed with ICE, simulating an experiment in which the HEK293 cell line C is edited to a G at SNP rs2072579 with efficiencies ranging from 5% to 95%.

Comparison of ICE, TIDE, and DECODR

To compare ICE to TIDE and DECODR algorithms,[17,18] we also analyzed all knockout samples via the TIDE and DECODR websites using default parameters. The overall editing efficiency and specific indel frequencies were then compared between all three technologies. Moreover, HDR-mediated knock-in was compared between TIDER, ICE, and DECODR in a similar fashion. ICE and DECODR take the same experimental inputs, but TIDER requires an additional sequence file. For TIDER analysis, we used the same input control and sample Sanger file used in ICE, plus an additional reference sequence file containing the desired knock-in sequence. The reference Sanger sequence was generated using the protocol described in TIDER with the following modifications. Briefly, we generated three DNA fragments by PCR for Gibson assembly: one of the donor sequence, and two containing the genomic sequence between each of the primers used in the control PCR to the donor, plus 20 bp flanking the donor. All fragments were purified using AMPure XP bead cleanup kit (Beckman Coulter, Brea, CA), pooled, and assembled with the NEBuilder HiFi DNA Assembly Master Mix (New England Biolabs, Ipswich, MA) according to manufacturer's instructions. The annealed mix was used as input DNA for the final PCR with the same primer pair used in the control and sample PCR reactions. However, due to primer design restrictions around the mutation site, we were only able to generate a portion of the reference sequence amplicons for TIDER analysis.

Results

Validation of ICE Results

We tested the ICE algorithm against a variety of benchmarks in order to validate that it produces robust and accurate estimates of editing outcomes.

Comparison of ICE with NGS

In order to verify that ICE produces accurate estimates of indel frequency, we performed a head-to-head comparison between ICE and NGS across a variety of editing types. As illustrated in Figure 3, the indel frequencies reported by ICE (green bars) largely mirrored those determined by NGS (black bars). This is true across edit types, including NHEJ-mediated repair from single-guide edits (Fig. 3A) and HDR-mediated repair for insertions from a donor template (Fig. 3B). It should be noted that any PCR-based sequencing technology (Sanger or NGS) will preferentially introduce bias toward large deletions because they are faster to amplify. The severity of this bias will be a function of many experimental conditions [19]

While these results show anecdotally that there is a correspondence between ICE and NGS, it is necessary to perform more extensive comparisons to illustrate the correlation more rigorously and to eliminate the possibility of systematic bias in ICE. To generalize these results, we began by performing Sanger and NGS amplicon sequencing on 92 single-guide knockout edits. We found a high correlation between the ICE and NGS results for each indel size in all samples ($r^2 = 0.93$, Fig. 4A). Furthermore, we followed this with a separate large-scale donor-mediated editing experiment. After transfecting 40 samples with CRISPR components and a DNA donor containing a SNP, +14 or +36, the ICE and NGS analyses reported a mix of editing outcomes. Some outcomes were the expected SNPs and insertions resulting from HDR, whereas others were random indels resulting from NHEJ/MMEJ. There was a strong correlation between NGS and ICE indel frequencies for both HDR and NHEJ/MMEJ editing outcomes, indicating

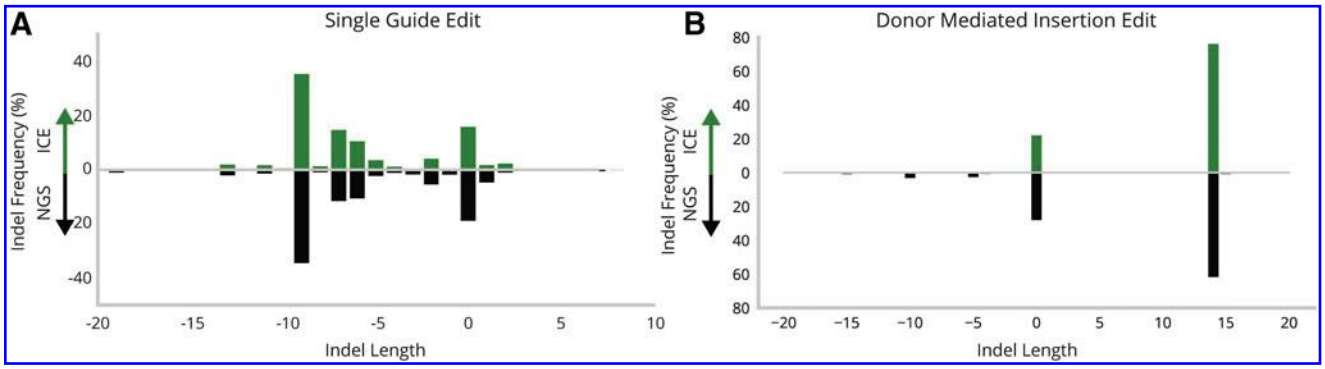

FIG. 3. Sample comparison of indel contributions across edit types comparing ICE and next-generation sequencing (NGS). The amplicon sequencing (black bars facing downward) and ICE results (green bars facing upward) for indel distributions for three samples. In each instance the indels predicted by ICE closely resemble those observed by NGS. **(A)** An example of a complex single-guide edit resulting in many genotypic outcomes. **(B)** An example donor-mediated insertion edit showing both wild-type and a +14 insertion.

that the frequency and identity of editing outcomes derived by ICE closely match those determined via NGS (Fig. 4B). ICE does tend to underestimate indel frequencies slightly compared to NGS due to the regularization step of the algorithm, although the effect size is small (median discrepancy across all indel frequencies = 1.04%). To interrogate further whether there are systematic biases between ICE and NGS results, we also analyzed whether the divergence of ICE and NGS indel frequency is related to indel length, and we found only minor effects (Supplementary Fig. S1).

Together, these results show that ICE and NGS produce very similar estimates of editing outcomes across various edits of differing types.

SNV analysis

To validate ICE's ability to measure edits resulting in SNVs, we sequenced controlled mixtures of DNA that simulate a range of variant outcomes. We amplified the locus surrounding SNP rs2072579 from HEK293s (which is homozygous C/C) and the PGP1 iPSC line (which is homozygous; G/G), quantified the amplicons with a Fragment Analyzer, mixed them at different molar ratios (5–95% of PGP1 in the mixture), and sequenced the mixed samples. The resulting sequencing data were then analyzed with ICE, simulating an experiment in which the HEK293 cell line C is edited to a G. ICE estimates of the percentage of SNV editing were strongly correlated with the percentages expected by the molar ratios (Fig. 5).

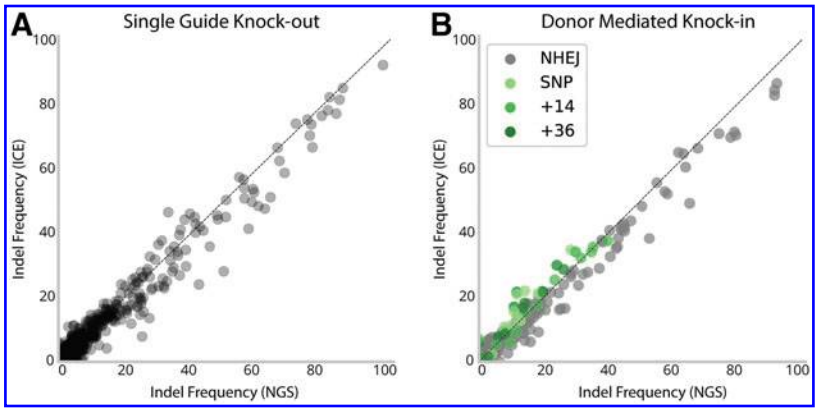

FIG. 4. Correlation of ICE and NGS genotyping results for single-guide knockout and donor-mediated knock-in experiments. **(A)** Pairwise comparisons of all indels resulting from ICE and NGS sequencing for 92 samples. The correlation between indel frequencies from ICE and NGS is $r^2 = 0.96$. **(B)** NGS results for homology-directed repair (HDR) experiments were compared to ICE results for 40 samples. Each data point represents an indel for each sample (HDR = green, NHEJ = gray). The overall correlation of NGS with ICE is $r^2 = 0.97$.

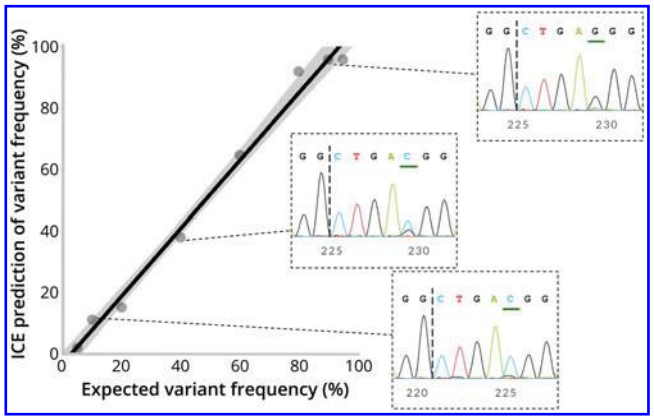

FIG. 5. ICE results are consistent with known molar ratios of single nucleotide variant populations. ICE variant results correlate with expected variant percentages ($r^2 = 0.99$). The traces on the right show the Sanger chromatogram with the C/G single-nucleotide polymorphism (SNP) underlined. The relative prevalence of the SNP is clearly visible in the chromatogram, and ICE correctly infers it.

Comparison of Sanger analysis tools

The figures above showcase the consistency of ICE as validated by NGS. However, ICE exists in a landscape with other Sanger analysis tools. To give insight into this competitive landscape, we repeated the above analysis using both TIDE and DECODR. First, we see that overall editing efficiency performance with single-guide edits versus NGS is consistent across algorithms (Supplementary Fig. S2). Next, at the level of the individual indel frequencies within these single-guide edits, we see that TIDE and DECODR are consistent with NGS, similar to ICE (Supplementary Fig. S3A and B). Summary statistics across single-guide experiments show that no one algorithm is consistently superior to another for this edit type for both mean-squared error (Supplementary Fig. S3C) and correlation (Supplementary Fig. S3D) with expected NGS indel frequencies. Within the single-guide edits, there is a slightly higher frequency of low-level indel contributions in NGS than the Sanger tools (Supplementary Fig. S4). ICE and DECODR have the least frequent low-level contributions due to the presence of regularization within their regression steps. Finally, we repeated a subset of our donor-mediated knock-ins from Figure 4A with the modified TIDE algorithm for including HDR outcomes, TIDER, and DECODR (Supplementary Fig. S5). DECODR produces similar performance to ICE versus NGS (Supplementary Fig. S5B). However, TIDER failed on more samples compared to ICE or DECODR (Supplementary Fig. S5C). Moreover, the samples that TIDER did return for successfully did not correlate with NGS as consistently as ICE or DECODR (Supplementary Fig. S5B).

Discussion

Here, we present our software tool, ICE, which uses Sanger sequencing data to quantify and identify genotypes following CRISPR editing. ICE can analyze editing outcomes for both donor (knock-in) and donor-less (knockout) experiments. While NGS has better sensitivity and quantitation, Sanger sequencing remains more widely accessible, faster, and cheaper for low-throughput applications. Moreover, we show a high correlation between the genotypes derived by ICE and NGS, suggesting that ICE can provide a reliable substitute in the majority of cases. We additionally compared the performance of other Sanger analysis tools (TIDE and DECODR) to NGS and ICE. While all tools broadly agreed in the case of knockout experiments, donor-mediated results did not line up as well. In particular, estimates from TIDER did not correlate with NGS results as well as ICE or DECODR. As the source code for TIDER is not publicly available, it is difficult to speculate on the cause of the poor correlation. Considering this is only a single data set, we are not inclined to believe that this poor correlation is systematically true of TIDER, but perhaps highlights a particular failure-mode for the algorithm. Further comparisons of these approaches to genotyping edited populations are necessary to provide clarity.

ICE performs competitively with other Sanger analysis tools in both knock-in and knockout applications. As tools such as DECODR, TIDE, and ICE remain in active development, it is clear the landscape of CRISPR Sanger analysis will continue to evolve. However, ICE continues to play a valuable role to the community as hundreds of thousands of samples have successfully been genotyped through the platform to date. ICE also serves the community by remaining completely free to use, with no features behind a paywall, and all source code transparently available.

While the ICE workflow is suitable for most CRISPR experiments, there are limitations to its capabilities. From the Sanger trace, ICE assumes at each position the peak signal for each base is linearly proportional to the molarity of the base. However, the peak height and phasing for a particular base in the Sanger trace is also a function of the local sequence context. This could result in sequences where Sanger signal ratios do not reflect the molar genotypic ratios of bases present at a given position. However, the high correlation between ICE and NGS indicates that this assumption does not affect ICE's ability to predict

insertions and deletions. We suspect that because an indel affects the signal for all bases downstream, the effect of peak signal variance cancels out over many bases.

While we have shown that ICE can be used to provide reasonable estimates of CRISPR editing outcomes using Sanger sequencing, the precision of any method that relies on Sanger sequencing will be limited by the noise that is inherent to Sanger. This can make interpreting the precise frequency of genotypes in a population particularly difficult, especially when the sequencing traces are of lower quality. If a high degree of precision about the frequency of a genotype in a population is required, then more sensitive measures such as NGS are better options.

Finally, ICE can only estimate the frequency of repair genotypes that it infers are likely, given the cut site (i.e., the proposal generation step). While the proposal generation methods in ICE have been shown to recapitulate those produced by NGS, this means that ICE is not a completely unbiased genotyping method, and it is possible that ICE may not recover all genotypes that exist in a population. Currently, we are developing an alternative to the core ICE algorithm that accommodates deletions without any of these proposal assumptions, enabling support for a variety of alternative Cas proteins, large deletion detection, and future gene-editing technologies yet to be deployed. Scientists within the CRISPR community are employing a growing arsenal of editing tools that produce a wide range of repair outcomes. As Sanger sequencing remains a core method for genotyping cells, ICE will continue development to accommodate these use cases.

Conclusion

ICE offers a robust and scalable method for analyzing CRISPR editing experiments. ICE can detect successful edits in just a few days after transfection and has already helped thousands of researchers with its public-facing Web site and transparent codebase. We found that ICE is able to offer results comparable to NGS but at a significant reduction in cost and time. The ICE workflow offers several advantages over the current state-of-the-art alternatives by providing a robust and reproducible way to analyze single-guide editing experiments. It also requires less work to analyze HDR experiments because it does not require a separate sequencing reaction for the donor. Because ICE reduces the labor, cost, and time associated with CRISPR experiments, analysis is no longer a limiting factor for precision genome editing.

Author Disclosure Statement

The authors of this publication are current or previous employees of, and hold stock options and/or equity in, Synthego Corporation.

Funding Information

No funding agencies were used for this article.

Supplementary Material

Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Figure S4
Supplementary Figure S5

References

1. van Overbeek M, Capurso D, Carter MM, et al. DNA repair profiling reveals nonrandom outcomes at Cas9-mediated breaks. *Mol Cell* 2016;63:633–646. DOI: 10.1016/j.molcel.2016.06.037.
2. Ran FA, Hsu PD, Wright J, et al. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 2013;8:2281–2308. DOI: 10.1038/nprot.2013.143.
3. Chen W, McKenna A, Schreiber J, et al. Massively parallel profiling and predictive modeling of the outcomes of CRISPR/Cas9-mediated double-strand break repair. *Nucleic Acids Res* 2019;47:7989–8003. DOI: 10.1093/nar/gkz487.
4. Mandal PK, Ferreira LM, Collins R, et al. Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. *Cell Stem Cell* 2014;15:643–652. DOI: 10.1016/j.stem.2014.10.004.
5. Lin S, Staahl BT, Alla RK, et al. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* 2014;3:e04766. DOI: 10.7554/eLife.04766.
6. Li K, Wang G, Andersen T, et al. Optimization of genome engineering approaches with the CRISPR/Cas9 system. *PLoS One* 2014;9:e105779. DOI 10.1371/journal.pone.0105779.
7. Chakrabarti AM, Henser-Brownhill T, Monserrat J, et al. Target-specific precision of CRISPR-mediated genome editing. *Mol Cell* 2019;73:699–713.e6. DOI: 10.1016/j.molcel.2018.11.031.
8. Clement K, Rees H, Canver MC, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat Biotechnol* 2019;37:224–226. DOI: 10.1038/s41587-019-0032-3.
9. Kim HJ, Lee HJ, Kim H, et al. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res* 2009;19:1279–1288. DOI: 10.1101/gr.089417.108.
10. Sentmanat MF, Peters ST, Florian CP, et al. A survey of validation strategies for CRISPR-Cas9 editing. *Sci Rep* 2018;8:888. DOI: 10.1038/s41598-018-19441-8.
11. Dehairs J, Talebi A, Cherifi Y, et al. CRISP-ID: decoding CRISPR mediated indels by Sanger sequencing. *Sci Rep* 2016;6:28973. DOI: 10.1038/srep28973.
12. Dmitriev DA, Rakitov RA. Decoding of superimposed traces produced by direct sequencing of heterozygous indels. *PLoS Comput Biol* 2008;4:e1000113. DOI: 10.1371/journal.pcbi.1000113.
13. Zhidkov I, Cohen R, Geifman N, et al. CHILD: a new tool for detecting low-abundance insertions and deletions in standard sequence traces. *Nucleic Acids Res* 2011;39:e47. DOI: 10.1093/nar/gkq1354.
14. Chang C-T, Tsai C-N, Tang CY, et al. Mixed sequence reader: a program for analyzing DNA sequences with heterozygous base calling. *ScientificWorldJournal* 2012;2012:365104. DOI: 10.1100/2012/365104.
15. Kluesner MG, Nedveck DA, Lahr WS, et al. EditR: a method to quantify base editing from Sanger sequencing. *CRISPR J* 2018;1:239–250. DOI: 10.1089/crispr.2018.0014.
16. Amir A, Zuk O. Bacterial community reconstruction using compressed sensing. *J Comput Biol* 2011;18:1723–1741. DOI: 10.1089/cmb.2011.0189.
17. Brinkman EK, Chen T, Amendola M, et al. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res* 2014;42:e168. DOI: 10.1093/nar/gku936.
18. Bloh K, Kanchana R, Bialk P, et al. Deconvolution of Complex DNA Repair (DECODR): establishing a novel deconvolution algorithm for comprehensive analysis of CRISPR-edited Sanger sequencing data. *CRISPR J* 2021;4:120–131. DOI: 10.1089/crispr.2020.0022.
19. Dabney J, Meyer M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* 2012;52:87–94. DOI: 10.2144/000113809.