# End-to-End Tools for Interrogation of CRISPR-Cas Associated Genotoxicity

Gavin Kurgan, Kyle Kinney, Ellen Schmaljohn, Morgan Sturgeon, He Zhang, Rolf Turk, Garrett Rettig, & Ashley Jacobi

Integrated DNA Technologies, Coralville, IA 52241, USA

Corresponding author: ajacobi@idtdna.com

*For research use only. Unless otherwise agreed to in writing, IDT does not intend these products to be used in clinical applications and does not warrant their fitness or suitability for any clinical diagnostic use. Purchaser is solely responsible for all decisions regarding the use of these products and any associated regulatory or legal obligations. RUO23-2612_001

## Overview – Tools to assess safety of CRISPR editing

- IDT rhAmp off-target nomination strategy results in similar total site counts as GUIDE-seq.
- IDT offers end-to-end on & off-target quantification using the rhAmpSeq™ CRISPR Analysis System, which includes a user interface and improved accuracy of indel calling.
- The development of an off-target classification tool (OTEasy) enables identification of <0.5% indels with high analytical specificity/sensitivity.
- A new translocation tool is capable of quantifying translocations using rhAmpSeq on & off-target panel data.
- IDT synthesis is demonstrated to be free of oligo contamination with very low limits of detection.

## IDT rhAmp Off-Target Nomination Performance

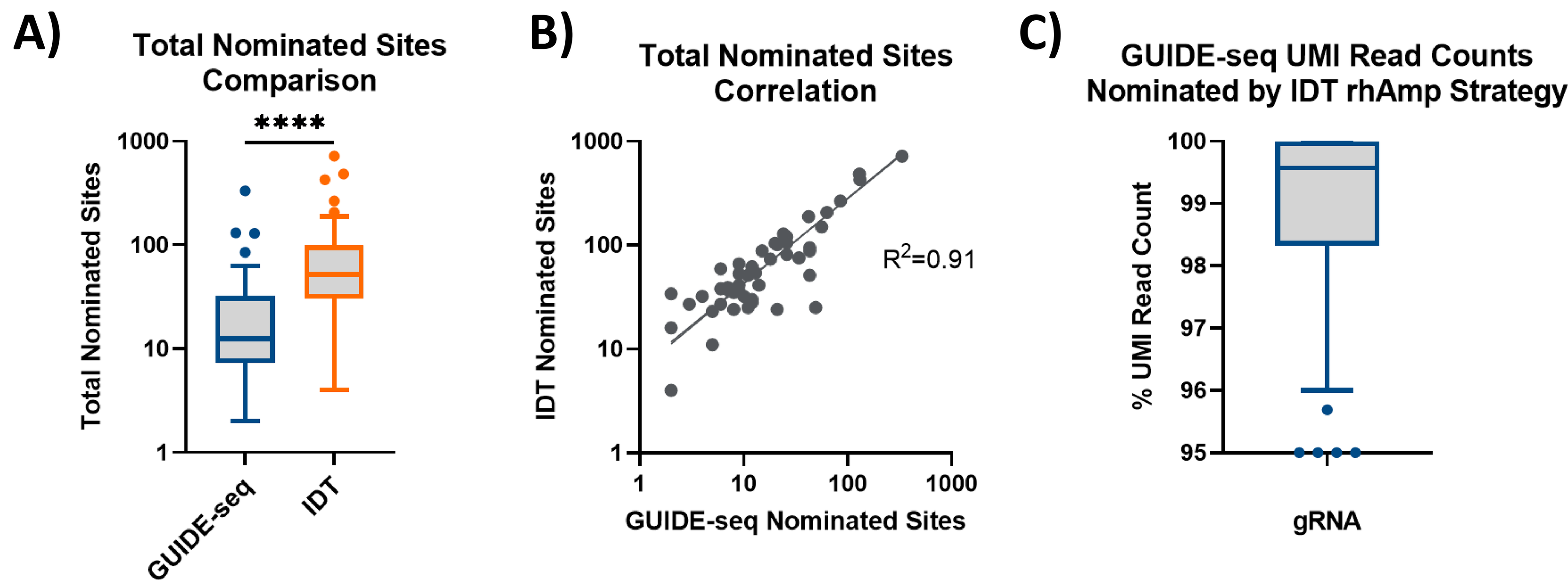
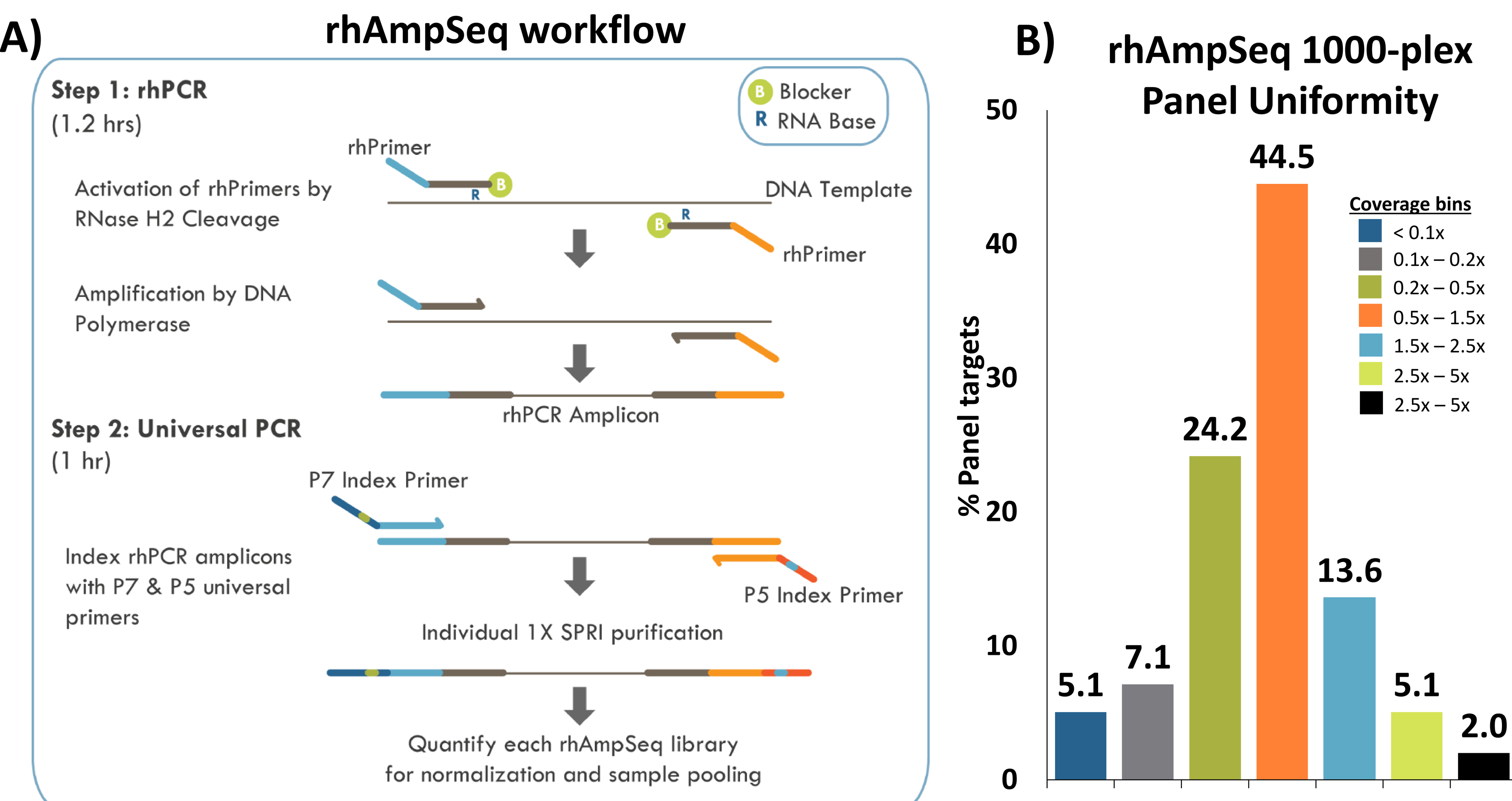



**Figure 1**. Comparison of IDT's rhAmp nomination strategy and the GUIDE-seq off-target analysis method. **A)** Tukey box plots showing the total nominated OTE sites for each indicated methodology. **B)** Scatterplots of total sites nominated by GUIDE-seq (x axis) and IDT's rhAmp nomination strategy (y axis). **C)** Tukey box plot of percent GUIDE-seq UMI read counts of sites nominated that were similarly nominated with IDT's rhAmp nomination strategy. All data is representative of 48 gRNAs (Targets: PDCD1, LAG3, CTLA4, NRP1, IL2RA, and TIGIT; 8 gRNAs per target). GUIDE-seq NGS libraries were processed through the GUIDE-seq analysis package. rhAmp NGS libraries were processed with IDT's proprietary OTE analysis pipeline. Statistical significance was determined using Wilcoxon matched-pairs signed rank test. Correlation was assessed using least squares regression. **A)** ****P < 0.0001.

## End-to-end on & off-target editing quantification using the rhAmpSeq™ CRISPR Analysis System




C)

rhAmpSeq Design (Primers)

rhAmpSeq (Library Prep)

NGS (Illumina MiSeq)

rhAmpSeq CRISPR Analysis Tool (Software)

Select or design rhAmpSeq panels

Prepare libraries

User-friendly data analysis

Custom rhAmpSeq Panels or rhAmpSeq Predesigned Panel

rhAmpSeq Library Kit rhAmpSeq Index Primers

**Figure 2. A)** Overview of the rhAmpSeq method for multiplexed amplicon sequencing. RNase H2 activates rhPCR primers flanking CRISPR-Cas9 cut sites by cleavage 5′ of the RNA base within the DNA:RNA duplex, removing a 3′ blocker. Activated primers are extended to generate target amplicons at validated and predicted Cas9 target sites. Universal PCR is used to incorporate sample indexes. **B)** rhAmpSeq panels are designed for uniformed amplification of up to 1,000 primer pairs in a single reaction. The mean (n=3) uniformity metrics of target amplification in unique 1000-plex panels is shown. Typical panel size for CRISPR-Cas9 gRNAs has ranged from approximately 40-400 targets per panel. **C)** The rhAmpSeq system (red text) accepts nominated on & off-target loci as input, designs singleplex or multiplex rhAmpSeq amplicon panels for library preparation and analyzes NGS amplicon sequencing data via rhAmpSeq CRISPR Analysis Tool (CRISPAltRations™).

## The rhAmpSeq CRISPR Analysis System has enhanced editing quantification accuracy

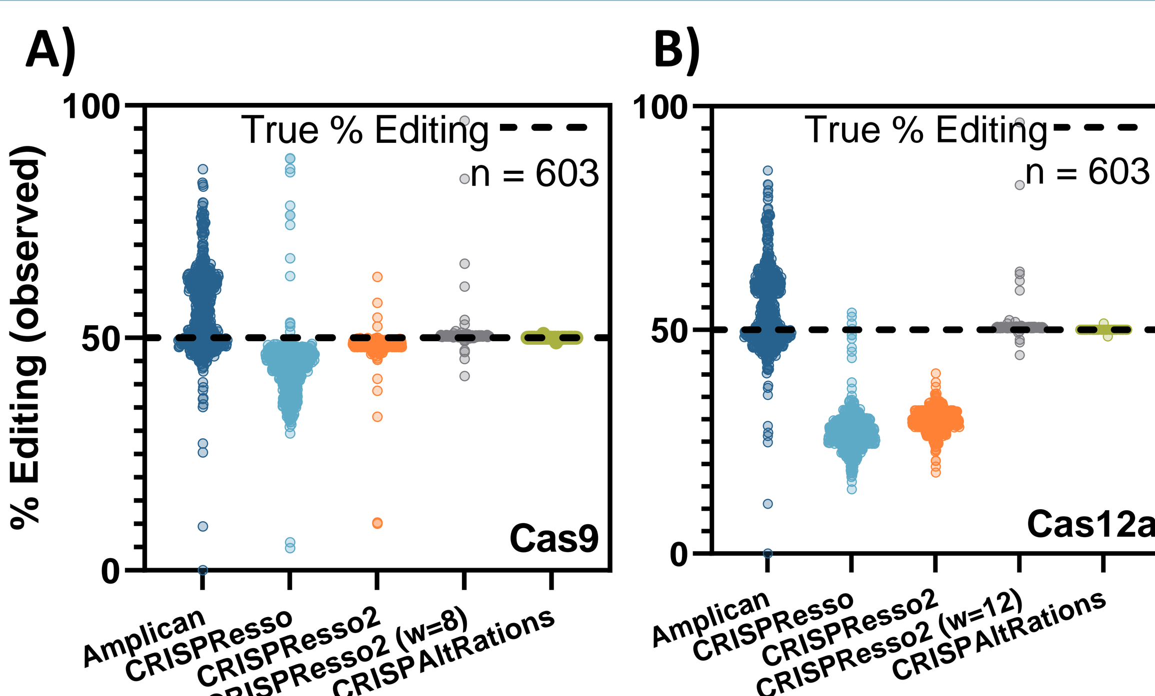
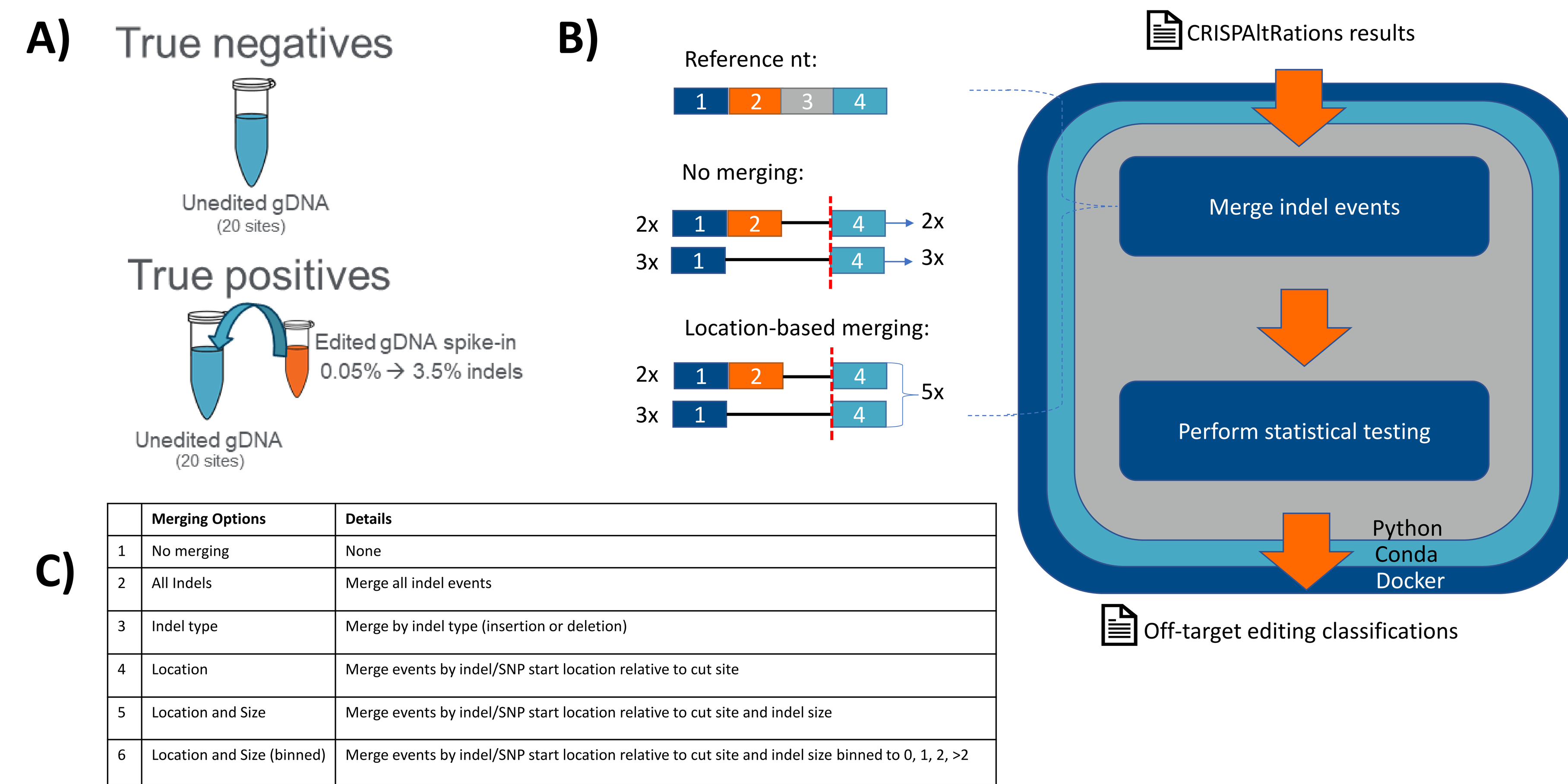



**Figure 3.** CRISPAltRations was developed by characterizing indel profiles from 516 unique guides and developing a Cas-specific alignment algorithm to enhance indel identification. Performance of synthetic data with known truth (n=603 sites) was utilized in comparing publicly available multiplex amplicon analysis tools against CRISPAltRations; indels represented are repaired DSBs caused by **A)** Cas9 and **B)** Cas12a. Synthetic loci (open circles) are denoted with ground truth of 50% editing (black dashed line); window size (w).

Mol Ther Methods Clin Dev. 2021 Apr 2;21:478-491. doi: 10.1016/j.omtm.2021.03.024. eCollection 2021 Jun 11.

## OTEasy tool produces a binary classification of CRISPR editing




| | Merging Options | Details |
|---|---|---|
| 1 | No merging | None |
| 2 | All Indels | Merge all indel events |
| 3 | Indel type | Merge by indel type (insertion or deletion) |
| 4 | Location | Merge events by indel/SNP start location relative to cut site |
| 5 | Location and Size | Merge events by indel/SNP start location relative to cut site and indel size |
| 6 | Location and Size (binned) | Merge events by indel/SNP start location relative to cut site and indel size binned to 0, 1, 2, >2 |

**Figure 4. A)** 20 loci within gene bodies were chosen in HAP1 cells (ctrl: <= 0.4% indels; treatment: 0.05 – 3.5% indels), transfected with high efficiency guides (~100% indels), and serial diluted into unedited HAP1 gDNA in triplicate to 0.05, 0.1, 0.2, 0.5, 1, 2, 3.5% indels, amplified using rhAmpSeq™ library prep, and analyzed using CRISPAltRations. **B)** OTEasy utilizes CRISPAltRations output files, **C)** merges identified indels based on a variety of strategies defined in input and performs statistical tests (hypergeometric for n=1, negative binomial model for n>2). The output is a csv containing binary classification of editing results (Edited, Unedited) and associated p-values. OTEasy is dockerized for easy portability and usage with any operating system. An example of location–based merging is depicted to show how this strategy merges deletions with the same location (relative to the cut site) for a total count of 5x, which are normally considered separate editing events with no merging.

## OTEasy off-target classification tool results in high analytical specificity and sensitivity

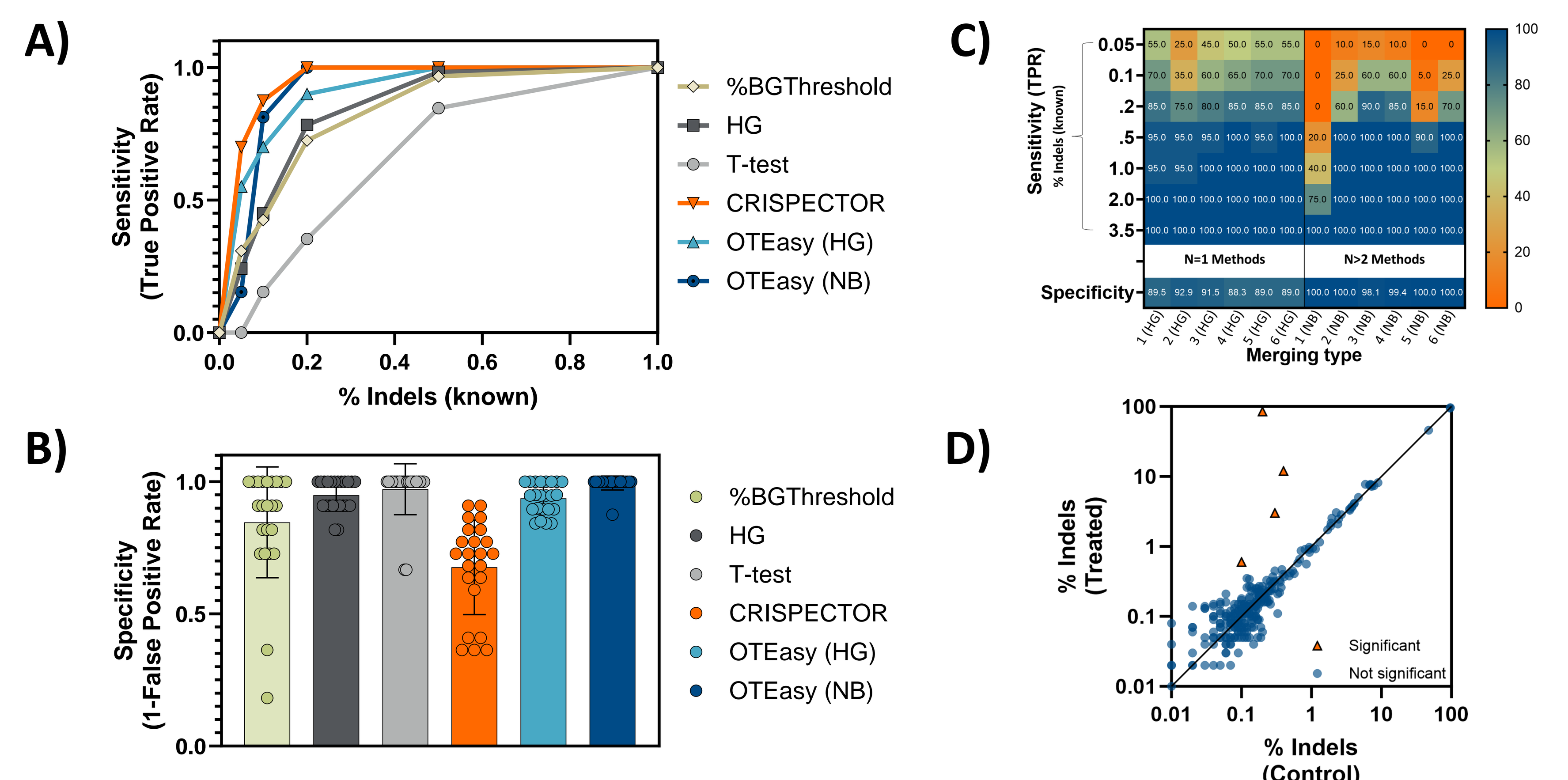
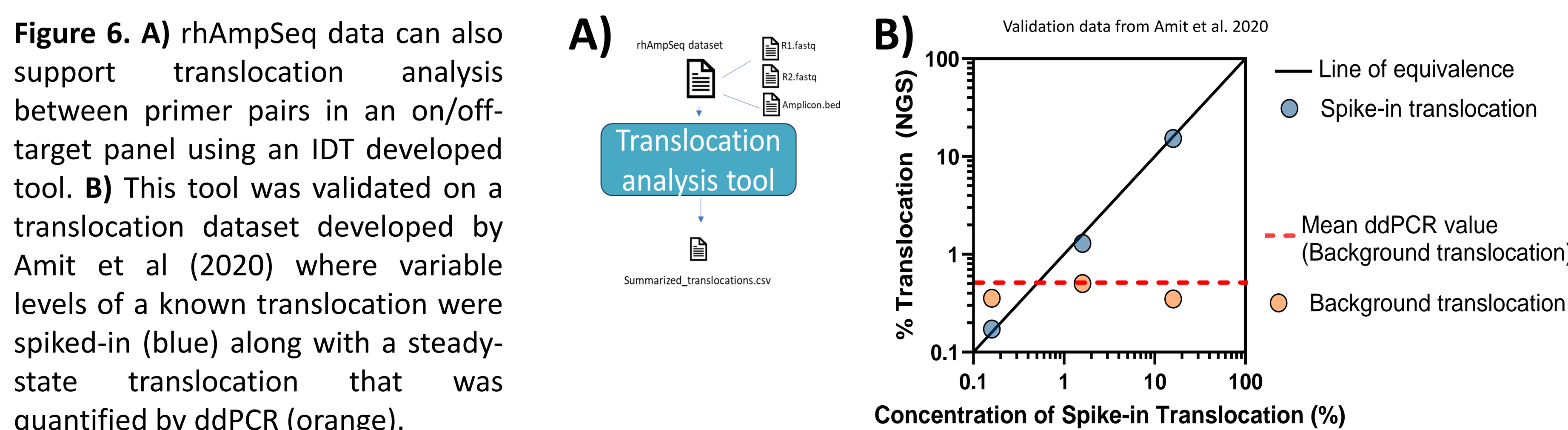



**Figure 5.** Implementation of OTEasy leads to increased analytical sensitivity and specificity of binary CRISPR editing classification. Performance of CRISPAltRations + OTEasy was compared to CRISPAltRations + CRISPECTOR using default parameters and for editing classification (data not shown). We compared the performance of the OTEasy location-based merging with hypergeometric (HG; p<0.05) and negative binomial (NB; p<0.05) statistical models against CRISPECTOR (>0% indel signal produced), a 75% relative change threshold (%BGThreshold), Hypergeometric (HG; p<0.05), T-test (p<0.05), and measured **A)** analytical sensitivity and **B)** analytical specificity. **C)** OTEasy merging methods (1-6; Description in Figure 4b) and statistical methods HG and NB results compared for analytical sensitivity (% true positive rate) and analytical specificity. **D)** Example results of off-target classification using statistical tests from OTEasy are shown.

## Detecting translocations at high accuracy with rhAmpSeq

**Figure 6. A)** rhAmpSeq data can also support translocation analysis between primer pairs in an on/off-target panel using an IDT developed tool. **B)** This tool was validated on a translocation dataset developed by Amit et al (2020) where variable levels of a known translocation were spiked-in (blue) along with a steady-state translocation that was quantified by ddPCR (orange).




## NHEJ DNA repair fingerprint method developed to measure gRNA contamination at sub 0.1% levels

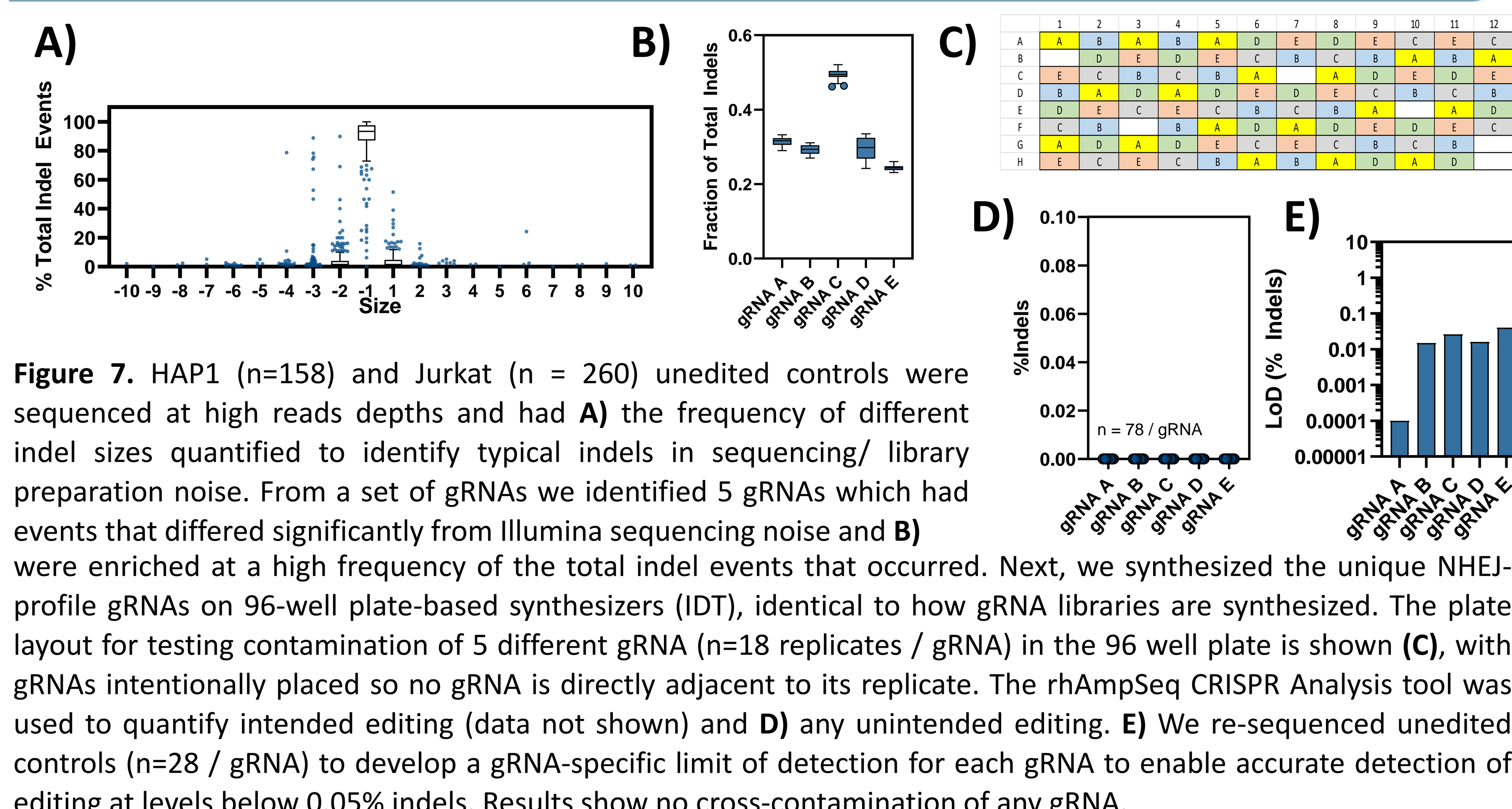



**Figure 7.** HAP1 (n=158) and Jurkat (n = 260) unedited controls were sequenced at high reads depths and had **A)** the frequency of different indel sizes quantified to identify typical indels in sequencing/ library preparation noise. From a set of gRNAs we identified 5 gRNAs which had events that differed significantly from Illumina sequencing noise and **B)** were enriched at a high frequency of the total indel events that occurred. Next, we synthesized the unique NHEJ-profile gRNAs on 96-well plate-based synthesizers (IDT), identical to how gRNA libraries are synthesized. The plate layout for testing contamination of 5 different gRNA (n=18 replicates / gRNA) in the 96 well plate is shown **(C)**, with gRNAs intentionally placed so no gRNA is directly adjacent to its replicate. The rhAmpSeq CRISPR Analysis tool was used to quantify intended editing (data not shown) and **D)** any unintended editing. **E)** We re-sequenced unedited controls (n=28 / gRNA) to develop a gRNA-specific limit of detection for each gRNA to enable accurate detection of editing at levels below 0.05% indels. Results show no cross-contamination of any gRNA.