UNCOVERseq Enables Sensitive and Controlled Gene Editing Off-Target Nomination Across CRISPR-Cas Modalities and Systems

Kyle Kinney[1], Kun Jia[2], He Zhang[1], Ellen Schmaljohn[1], Thomas Osborne[1], Bernice Thommandru[1], Karthik Murugan[1], Andrea Sanchez-Pena[1], Shengyao Chen[1], Roshani Codipilly[1], Morgan Sturgeon[1], Rolf Turk[1], Matthew McNeill[1], Mark Behlke[1], Ashley Jacobi[1], M. Kyle Cromer[2], Garrett Rettig[1], & Gavin Kurgan*[1]

[1]Integrated DNA Technologies, Coralville, IA 52241, USA
[2]University of California - San Francisco, San Francisco, CA, USA

*Corresponding author; gkurgan@idtdna.com

1138

Overview

- UNCOVERseq is a new off-target nomination methodology with improved, benchmarked performance confirming high true positive rates down to sub-0.05% editing frequencies
- Nominated off-targets in promiscuous proxy systems are sensitive lists for translational systems
- UNCOVERseq is more sensitive than GUIDE-seq and nominates real sites not caught with in vitro methods
- UNCOVERseq nominates relevant off-targets for both DSB and SSB methods confirmed in HSPCs
- IDT offers end-to-end on & off-target services using UNCOVERseq purpose-built for genotoxicity assessment

Want to learn more about IDT's Services?

Want to read the full pre-print manuscript?

UNCOVERseq as an improved in cellulo method

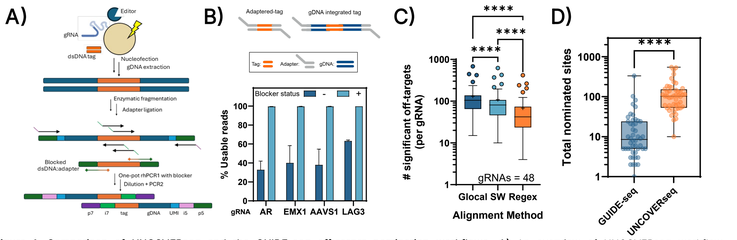


Figure 1. Comparison of UNCOVERseq and the GUIDE-seq off-target nomination workflows. A) An overview of UNCOVERSeq workflow demonstrates that cells with a genomically integrated dsDNA tag have gDNA extracted and amplified with rhPCR in a single reaction, with dsDNA tag:adapter byproducts being blocked by a targeted oligo before being sequenced and analyzed using our described workflow. **B)** A depiction is shown regarding what kind of events are targeted by the blocking oligo, with the usable reads (non-dsDNA:adapter reads) measured across nominating off-targets for 4 gRNAs in K562 (n = 3 per gRNA) with (light blue) or without (dark blue) the blocking oligo. **C)** Comparison of different alignment methods used in publicly available nomination packages was performed on UNCOVERSeq data to determine differences in levenshtein distance < 7 sites from a glocal alignment (Glocal), Smith-Waterman (SW; -100/-100 gap open/extension penalty as in original GUIDE-seq pipeline) or string-match method (Regex; same as current version "GUIDE-seq" pipeline). **D)** Tukey box plots show the end-to-end differences in total nominated OTE sites for the original GUIDE-seq method (wet lab protocol and alignment) compared with UNCOVERseq. Data is representative of 48 gRNAs (Targets: *PDCD1, LAG3, CTLA4, NRP1, IL2RA*, and *TIGIT*; 8 gRNAs per target). Statistical significance was determined for pairwise comparisons using Wilcoxon matched-pairs signed rank test and for multiple comparisons using a Friedman test with a post-hoc Dunn's test with Bonferroni correction. ****p < 0.0001.

Promiscuous systems are analytically sensitive and translational proxies for off-target nomination

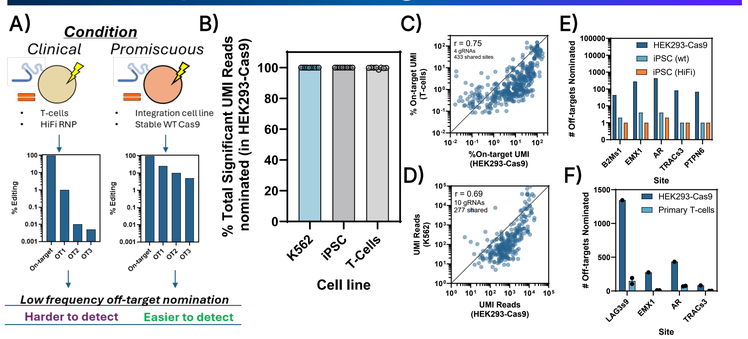


Figure 2. Analysis of performance of a promiscuous cell system and the ability to translate to other cellular systems. A) Example diagram demonstrating promiscuous nomination systems can help sensitively represent off-target lists by increasing the frequency that off-targets are detected. **B)** Comparison of total frequencies of nominated sites captured in a single replicate of HEK293-Cas9 vs wildtype Cas9 RNP transfection using three different cell types: K562 (10 gRNAs; biological duplicates per gRNA), iPSCs (5 gRNAs; one biological replicate per gRNA; wildtype and HiFi Cas9), and primary T-cells (4 gRNAs; 2 biological replicates per gRNA as two different donors). Total UMI-corrected reads, total number of nominated sites, and Spearman correlation are shown for shared nominated sites between HEK293-Cas9 and **C)** Primary T-cells, and **D)** K562. The total number of nominated sites between **E)** iPSCs and **F)** Primary T-cells are displayed in comparison to HEK293-Cas9

UNCOVERseq reproducibility and replication

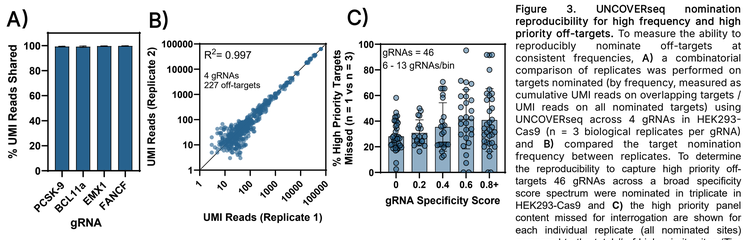
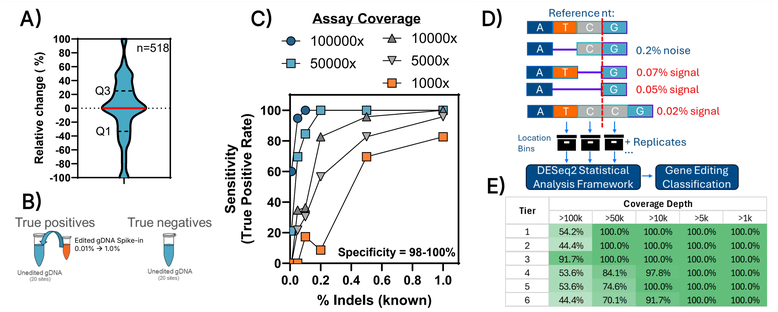


Figure 3. UNCOVERseq nomination reproducibility for high frequency and high priority off-targets. To measure the ability to reproducibly nominate off-targets at consistent frequencies, **A)** a combinatorial comparison of replicates was performed on targets nominated (by frequency, measured as cumulative UMI reads on overlapping targets / UMI reads on all nominated targets) using UNCOVERSeq across 4 gRNAs in HEK293-Cas9 (n = 3 biological replicates per gRNA) and **B)** compared the target nomination frequency between replicates. To determine the reproducibility to capture high priority off-targets 46 gRNAs across a broad specificity score spectrum were nominated in triplicate in HEK293-Cas9 and **C)** the high priority panel content missed for interrogation are shown for each individual replicate (all nominated sites) compared to the total # of high priority sites (Tier 1 to Tier 3) nominated as a biological triplicate.

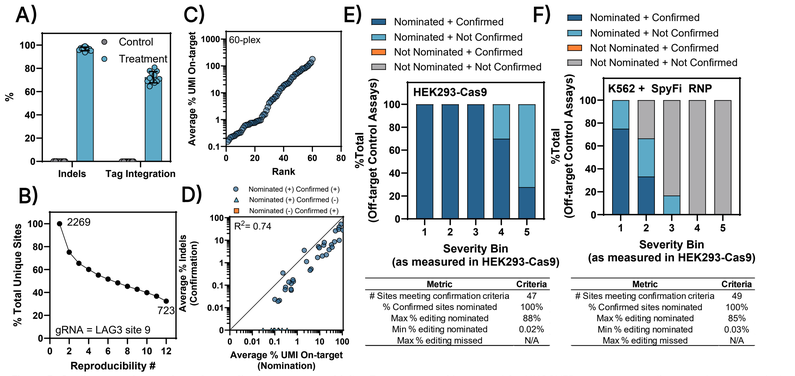
Improved classification algorithms for detection of indels (> 0.01%) and base editing (> 0.5%) activity



Tier	Coverage Depth >100k	>50k	>10k	>5k	>1k
1	54.2%	100.0%	100.0%	100.0%	100.0%
2	44.4%	100.0%	100.0%	100.0%	100.0%
3	91.7%	100.0%	100.0%	100.0%	100.0%
4	53.6%	84.1%	97.8%	100.0%	100.0%
5	53.6%	74.6%	100.0%	100.0%	100.0%
6	44.4%	70.1%	91.7%	100.0%	100.0%

Figure 4. Analysis algorithms reduce the impact of indel noise for high analytical sensitivity/specificity in genome editing classification. A) Relative change (%) of indels from the mean was quantified in unedited control samples for 23 unique genomic targets from Jurkat cells sequenced 22 independent times with high read depth (> 10,000 read pairs) to demonstrate the challenges in detecting low levels of indel editing using conventional statistical approaches. **B)** 20 loci within gene bodies were chosen in K562 cells (ctrl: <= 0.4% indels; treatment: 0.01 – 1.0% indels), transfected with high efficiency guides (~100% indels), and serial diluted into unedited K562 gDNA in triplicate to 0.01, 0.05, 0.1, 0.2, 0.5, 1% indels, amplified using rhAmpSeq™ library prep, sequenced at very high read depth, subsampled to various read depth/site, and analyzed using CRISPAltRations™ + OTEasy for binary editing classification against a known concentration of indels. **C)** Analytical sensitivity/specificity of the method; specificity represents the mean +/- 95% confidence interval of classifying controls as unedited (n = 506). **D)** OTEasy, an off-target classification method, utilizes CRISPAltRations™ output files and merges identified indels based on location relative to cut site and uses the Deseq2 statistical analysis framework for binary classification. **E)** Example of coverage depth tiered by risk used in IDT reporting.

Process controls enable reporting of FNR and sensitivity to reliably detect < 0.01% events



Metric	Criteria
# Sites meeting confirmation criteria	47
% Confirmed sites nominated	100%
Max % editing nominated	88%
Min % editing nominated	0.02%
Max % editing missed	N/A

Metric	Criteria
# Sites meeting confirmation criteria	49
% Confirmed sites nominated	100%
Max % editing nominated	85%
Min % editing nominated	0.03%
Max % editing missed	N/A

Figure 5. Quality control procedures for confirming assay sensitivity. To create a positive control for UNCOVERseq process quality control, we extensively characterized a promiscuous *LAG3* (site 9) gRNA in HEK293-Cas9 cells using 12 biological replicate transfections with paired controls (no gRNA). Following transfection, **A)** indel frequency and tag integration frequencies were measured at the *LAG3* on-target site via NGS and **B)** the frequency of unique sites relative to the cumulative total reaching each reproducibility frequency is shown. **C)** Twelve sites were selected per frequency bin (Bin 5= < 0.49%; Bin 4 = 0.50% – 0.99%; Bin 3=1 - 9.9%; Bin 2 = 10 - 49.9%; Bin 1 = > 50% UMI reads relative to the on-target) for routine targeted sequencing with average nomination frequency shown. To test application of this, paired confirmation **D)** quantification and **E)** classification was performed in a highly promiscuous condition (HEK293-Cas9 cells; n = 3 biological replicates) and **F)** high specificity condition (K562 nucleofected SpyFi RNP; n = 3 biological replicates) and the status of each of the 60 measured sites meeting coverage criteria (> 1,000x) recorded.

Optimal NGS Conditions for Peak Performance

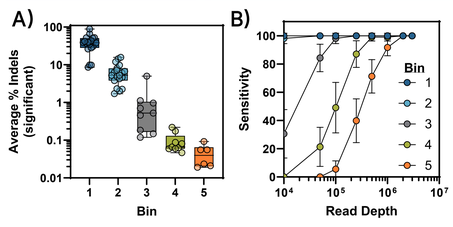


Figure 6. Definition of read depth requirements for maximum sensitivity. We extensively characterized the read depth requirements for the previous promiscuous *LAG3* (site 9) gRNA in HEK293-Cas9 cells using 12 biological replicate transfections with paired controls (no gRNA) by performing downsampling. **A)** Confirmed indel frequencies per bin are shown for confirmation data. **B)** Downsampling of the HEK293-Cas9 *LAG3* nomination samples (n = 12) was performed from 3 million reads down to 10,000 reads and sensitivity per severity bin was calculated for recovering all 60 *LAG3* positive control confirmation loci. We use this to recommend that all UNCOVERseq experiments should allocate a minimum of 500,000 reads per experiment, with ideally > 2 million reads per sample to ensure maximum sensitivity across observable frequencies.

Gavin Kurgan
Bioinformatics Applications Manager
Integrated DNA Technologies
gkurgan@idtdna.com

UNCOVERseq outperforms other methodologies

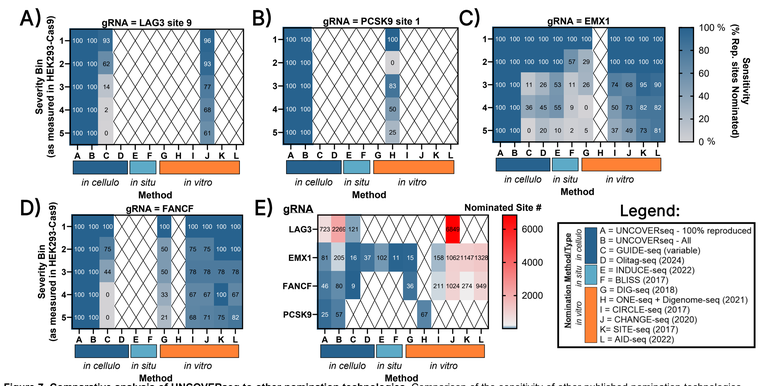


Figure 7. Comparative analysis of UNCOVERseq to other nomination technologies. Comparison of the sensitivity of other published nomination technologies to nominate 100% reproducible UNCOVERseq off-target sites segmented according to our previous nomination frequency bins (Severity Bins) using **A)** *LAG3* site 9 (n = 12 nomination replicates) **B)** *PCSK9* site 1 (n = 6 nomination replicates) **C)** *EMX1* (n = 6 nomination replicates) and **D)** *FANCF* (n = 6 nomination replicates). **E)** Total number of off-targets nominated per technology per gRNA. Legend displays the corresponding method to each column in the heatmaps and includes the number of fully reproduced UNCOVERseq sites (A) and all UNCOVERseq sites even without replicates (B). This demonstrates that UNCOVERseq largely improves on existing cell-based nomination methods, and no existing nomination method sensitively covers 100% of UNCOVERseq highly reproducible nominated sites, which likely represent true positives.

Interrogation of broad editor modalities in HSPCs

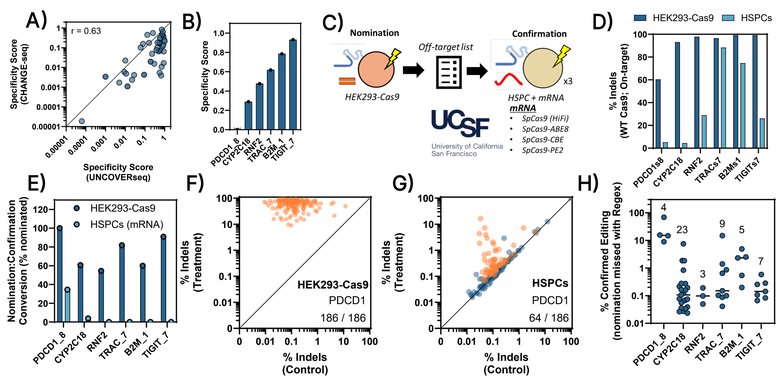


Figure 8. Interrogating DSB nomination translation capabilities in novel editors in translational systems. A) A total of 192 gRNAs were interrogated in HEK293-Cas9 cells with UNCOVERseq and specificity scores were compared to those derived from CHANGE-seq. **B)** 6 gRNAs were selected that could support all base editor modalities across a dynamic range. **C)** Experimental design interrogating off-targets effects in HSPCs with mRNA-delivered editors (HiFi Cas9, ABE8, CBE, PE2). **D)** On-target editing efficiencies of the gRNAs in the two systems HEK293-Cas9 cells and HSPCs and **E)** the frequency at which nominated off-targets are confirmed to have NHEJ editing (WT S.p Cas9 in HEK293-Cas9 cells and HiFi Cas9 in HSPCs) in the systems demonstrates the high-specificity of the method, but that most off-targets do not exist in therapeutically relevant conditions (HSPCs + HiFi Cas9 mRNA). An example of this demonstrates the frequency of significant (p <0.05) off-targets in **F)** HEK293-Cas9 cells as compared to **G)** HSPCs with mRNA delivery. **H)** Multiple off-targets were confirmed throughout this in all gRNAs and conditions that were not successfully nominated using the previously benchmarked Regex alignment method.

UNCOVERseq nominates relevant off-targets for both DSB and SSB editor modalities

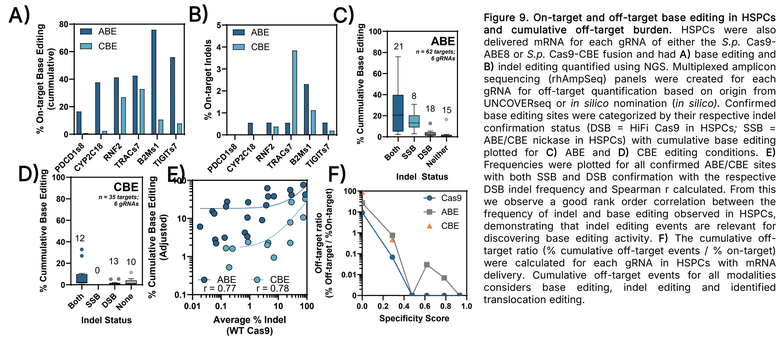


Figure 9. On-target and off-target base editing in HSPCs and cumulative off-target burden. HSPCs were also delivered mRNA for each gRNA of either the *S.p.* Cas9-ABE8 or *S.p.* Cas9-CBE fusion and had **A)** base editing and **B)** indel editing quantified using NGS. Multiplexed amplicon sequencing (rhAmpSeq) panels were created for each gRNA for off-target quantification based on origin from UNCOVERseq or *in silico* nomination (*in silico*). Confirmed base editing sites were categorized by their respective indel confirmation status (DSB = HiFi Cas9 in HSPCs; SSB = ABE/CBE nickase in HSPCs) with cumulative base editing plotted for **C)** ABE and **D)** CBE editing conditions. **E)** Frequencies were plotted for all confirmed ABE/CBE sites with both SSB and DSB confirmation with the respective DSB indel frequency and Spearman r calculated. From this we observe a good rank order correlation between the frequency of indel and base editing observed in HSPCs, demonstrating that indel editing events are relevant for discovering base editing activity. **F)** The cumulative off-target ratio (% cumulative off-target events / % on-target) were calculated for each gRNA in HSPCs with mRNA delivery. Cumulative off-target events for all modalities considers base editing, indel editing and identified translocation editing.