[1]Division of Cancer & Genetics
School of Medicine,
Cardiff University, UK
reedsh1@cardiff.ac.uk

INDUCE-seq: Ensuring the safe development of cell & gene therapies by gene editing

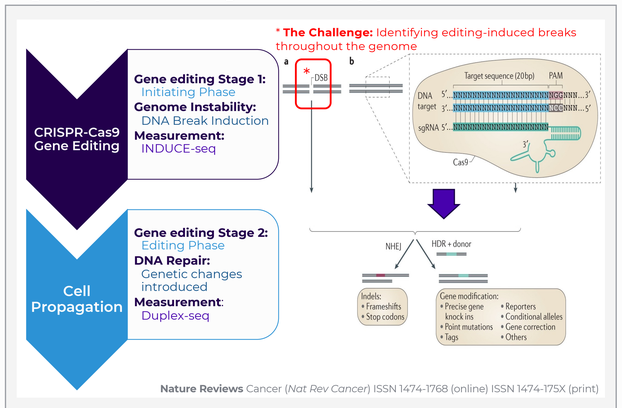
Simon Reed[1,2], Patrick van Eijk[1,2] & Felix Dobbs[2]

Background |

Advances in genome editing are facilitating the development of novel cell and gene therapies. These innovations enable precise modifications to existing genes and the creation of synthetic genes, particularly beneficial for generating immune-compatible cells for allogeneic therapy, providing readily available treatment options.

Despite the potential of gene editing in disease treatment, ensuring its safe and effective application demands the development of new tools. Risks associated with genome editing include the potential for malignant transformation of cultured cells, attributed to mutations acquired during cell culture, as well as the induction of genomic instability, especially in cells with pre-existing DNA repair defects. Moreover, the tools used for gene editing can inadvertently cause DNA breaks at unintended sites in the genome, leading to off-target effects. Such off-target editing poses risks such as the activation of proto-oncogenes or disruption of tumor suppressors, potentially driving carcinogenesis. Additionally, off-target mutagenesis may generate neoantigens, triggering autoimmune reactions or other forms of cellular dysfunction. Identifying and addressing off-target gene editing is crucial in mitigating these risks.

Figure 1 | Gene-editing is a two-step process



Nature Reviews Cancer (*Nat Rev Cancer*) ISSN 1474-1768 (online) ISSN 1474-175X (print)

Step 1 introduces some form of damage in the DNA targeted to a specific location (e.g. a DSB induced by CRISPR-Cas9). Step 2, gene editing occurs as a result of endogenous DNA repair mechanism in the cell, typically utilising one of two major DNA repair pathways (Non-Homologous End Joining (NHEJ) and Homology-Direct Repair (HDR)). The outcome of the repair reaction results in a mutation at the target-site, modifying the gene of interest.

INDUCE-seq™ |

More precise methods are needed for testing off-target gene-editing during all phases of therapeutic development, including treatment follow-up. At present, standardised assays to assess the safety of gene editing-based therapies are lacking[1]. INDUCE-seq was developed to address this[2].

INDUCE-seq is a scalable platform technology for mapping and characterizing DNA breaks. It leverages a novel PCR-free methodology for *in situ* break capture and sequencing by NGS, revealing the breaks induced by any nuclease-based genome editing system with high precision.

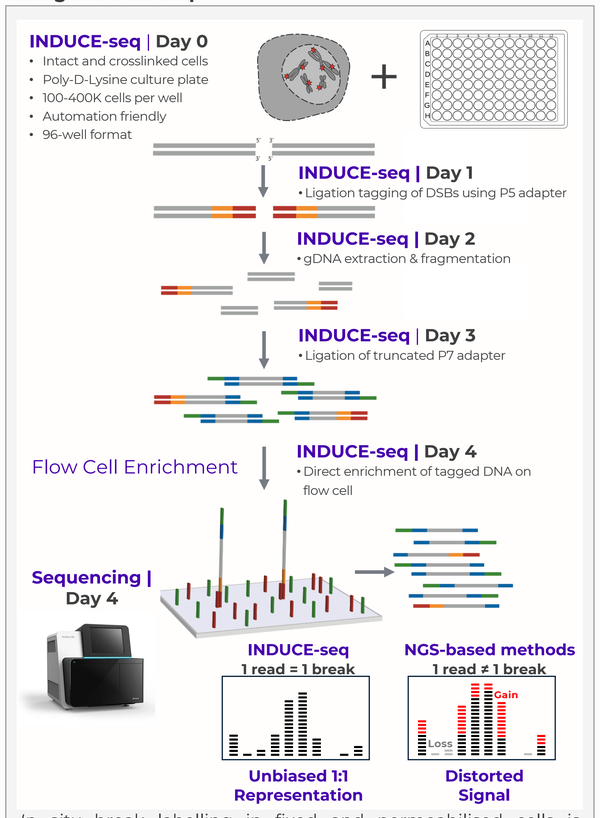
INDUCE-seq is the first unbiased cell-based solution that is free from PCR induced biases that distort measurements, has broad compatibility with a wide range of therapeutically relevant cells, and applicable to any nuclease-based gene editing system. INDUCE-seq provides data-driven and actionable insights to accelerate research & development, pre-clinical and clinical stages gene editing programs.

INDUCE-SEQ PERFORMANCE METRICS

INDUCE-seq's ability to differentiate between induced vs endogenous recurrent breaks was calculated using a standard induced break dataset.

Metrics		
Sensitivity	*How well we can call editing-induced breaks in a background of endogenous ones*	~94%
Specificity	*Our ability to call endogenous breaks from induced ones at low recurrence*	~85%
Precision	*Our ability to accurately call induced breaks at high recurrency*	~96%
FNR	*Occur at a recurrency at which mutations occur at background levels*	~6.5%
FPR	*What fraction of our off-target list of recurrent breaks are false positives*	~15%

Figure 2 | Detecting editing-induced breaks using INDUCE-seq

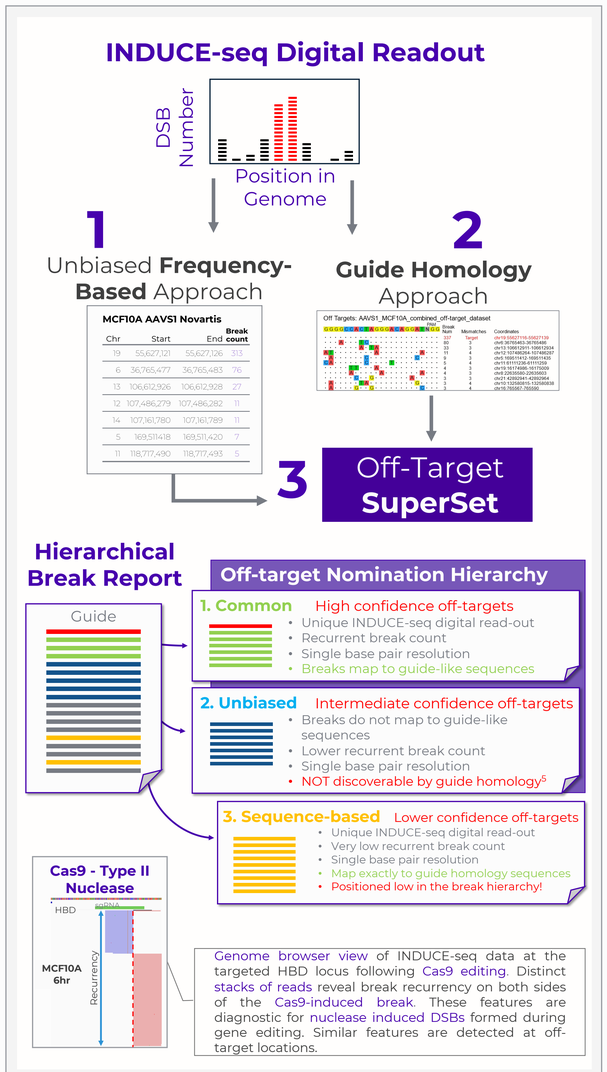


In situ break labelling in fixed and permeabilised cells is performed by ligating a full-length, chemically modified P5 sequencing adapter to end-prepared DSBs. Genomic DNA is then extracted, fragmented, end-prepared and ligated using a chemically modified half-functional P7 adapter. Resulting DNA libraries contain a mixture of functional DSB-labelled fragments (P5:P7) and non-functional genomic DNA fragments (P7:P7). Subsequent sequencing of INDUCE-seq libraries enriches for DSB-labelled fragments and eliminates all other non-functional DNA. As the INDUCE-seq library preparation is PCR-free, each sequencing read obtained is equivalent to a single labelled DSB-end from a cell[2].

HESI Cross-Validation Study |

Here we describe a pilot study to evaluate the method involving a consortium of partners under the umbrella of HESI CT-TRACS committee[3]. CRISPR-Cas9-based gene editing of five well-studied genetic targets was conducted by two independent industry partners (AZ & Novartis), using two different cell types. On- and off-target gene editing was assessed by measuring breaks in the genome using INDUCE-seq. The genetic changes at these locations were subsequently measured using error-corrected sequencing using Duplex-seq[4] which allows for sensitive detection of mutations.

Figure 3 | Identifying on & off-target editing-induced breaks with INDUCE-seq



Top Panel | Using the unique digital read-out of the INDUCE-seq DSB detection method, we developed an off-target nomination process that combines simple break recurrency with guide sequence homology to generate a superset of potential off-targets. **Middle Panel |** A hierarchical break report is generated enabling a data-driven analysis of off-target likelihood, as shown in three classes (i) common, (ii) unbiased & (iii) sequence-based. **Bottom Panel |** Genome browser view demonstrating break recurrency at HBD gene-edit locus.

Figure 4 | The experimental plan undertaken by the CT-TRACS consortium

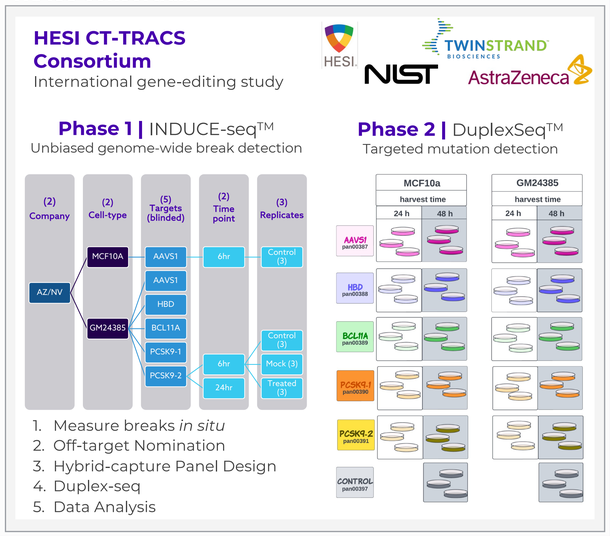


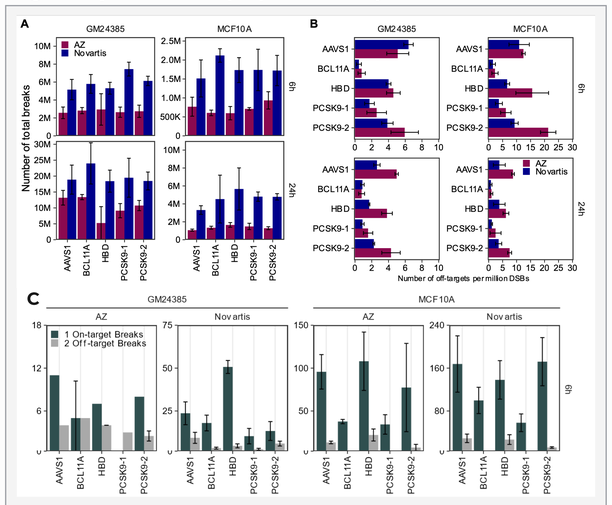
Figure 5 | Mutation analysis at on- and off-target sites by DuplexSeq

AAVS1 Off-target Report: Mutation & Break Data Summary

	Off-target name	chr	start	end	Break Score GM24385	Break Score MCF10A	GM24385 24h	GM24385 48h	GM24385 Control	MCF10A 24h	MCF10A 48h	MCF10A Control
	01_AAVS1_common	19	55,627,121	55,627,127	-	673	59.7%	72.55%	0.09%	88.76%	83.45%	0.04%
Common	02_AAVS1_common	6	36,765,477	36,765,483	40	156	12.5%	9.86%	0.00%	32.27%	32%	0.00%
Common	03_AAVS1_common	13	106,612,923	106,612,929	11	62	3.1%	5.49%	0.00%	19.83%	20.90%	-
Common	04_AAVS1_common	12	107,486,278	107,486,282	16	24	1.7%	2.98%	-	2.91%	3.74%	0.00%
Common	05_AAVS1_common	5	169,511,417	169,511,421	17	23	2.0%	4.74%	0.55%	5.34%	6.10%	0.01%
Common	06_AAVS1_common	11	61,111,251	61,111,254	16	8	0.4%	0.71%	0.00%	1.63%	3.25%	-
Common	07_AAVS1_common	10	132,580,828	132,580,833	12	3	0.1%	0.13%	-	0.29%	0.42%	-
Common	08_AAVS1_common	8	144,885,132	144,885,136	10	4	0.06%	0.13%	-	1.17%	0.48%	0.00%
Common	09_AAVS1_common	8	22,635,586	22,635,589	6	5	1.23%	1.40%	0.00%	2.24%	1.81%	-
Common	12_AAVS1_common	15	67,137,278	67,137,283	8	1	0.01%	0.00%	-	0.02%	0.04%	-
Common	14_AAVS1_common	11	65,297,767	65,297,771	8	0	0.01%	0.02%	-	0.04%	0.03%	-
Common	15_AAVS1_common	10	121,198,694	121,198,697	4	3	0.09%	0.16%	0.00%	0.26%	0.29%	-
Common	17_AAVS1_common	6	53,200,090	53,200,095	6	0	0.06%	0.10%	0.01%	0.03%	0.03%	-
Common	18_AAVS1_common	15	90,476,691	90,476,696	4	1	0.02%	0.03%	0.01%	0.07%	0.07%	-
Sequniq	01_AAVS1_sequniq	21	42,892,941	42,892,964	6	5	0.28%	0.54%	0.00%	0.65%	1.38%	-
Sequniq	03_AAVS1_sequniq	21	29,719,114	29,719,137	6	3	0.01%	0.01%	0.00%	0.07%	0.09%	-
Sequniq	04_AAVS1_sequniq	22	22,479,063	22,479,086	7	2	0.01%	0.00%	0.00%	-	0.03%	2.78%
Sequniq	06_AAVS1_sequniq	10	115,704,562	115,704,585	4	3	0.01%	0.02%	0.00%	0.19%	0.17%	-
Sequniq	07_AAVS1_sequniq	1	49,442,700	49,442,723	7	0	0.00%	0.01%	0.07%	0.02%	0.02%	0.06%
Sequniq	08_AAVS1_sequniq	13	100,280,739	100,280,762	4	2	-	0.00%	-	0.01%	0.01%	-
Sequniq	09_AAVS1_sequniq	16	31,484,902	31,484,925	3	3	0.00%	0.01%	-	0.01%	0.01%	-
Sequniq	10_AAVS1_sequniq	6	24,124,892	24,124,915	3	3	0.00%	0.01%	0.00%	0.01%	0.03%	0.00%
Sequniq	11_AAVS1_sequniq	15	89,175,079	89,175,102	2	3	0.02%	0.05%	2.04%	0.08%	0.09%	1.76%
Sequniq	12_AAVS1_sequniq	2	205,396,126	205,396,149	4	1	0.21%	0.38%	0.00%	0.86%	1.24%	-
Sequniq	14_AAVS1_sequniq	12	108,581,675	108,581,698	4	0	0.01%	0.02%	-	0.07%	0.10%	-
Sequniq	16_AAVS1_sequniq	15	45,827,892	45,827,915	3	1	0.01%	0.02%	-	0.04%	0.07%	-
Sequniq	17_AAVS1_sequniq	17	77,121,114	77,121,137	4	0	0.00%	0.00%	0.00%	-	0.01%	-
Sequniq	10_AAVS1_unbiasedsequniq	12	49,463,976	49,463,983	5	8	-	0.01%	0.00%	0.00%	0.01%	0.00%
Sequniq	15_AAVS1_unbiasedsequniq	1	162,106,766	162,106,769	11	0	0.01%	0.00%	0.00%	0.01%	0.00%	-
Sequniq	16_AAVS1_unbiasedsequniq	2	70,475,938	70,475,941	11	0	-	0.01%	0.01%	0.00%	0.01%	0.00%
Sequniq	17_AAVS1_unbiasedsequniq	2	159,047,922	159,047,923	11	0	-	0.00%	0.00%	0.00%	-	0.00%
Sequniq	18_AAVS1_unbiasedsequniq	5	68,530,719	68,530,726	-	-	0.00%	0.00%	-	-	0.01%	-

Mutations at the high-confidence class of off-target are shown here, demonstrating a strong proportional relationship between break count and the accumulation of mutations. Importantly, mutation rates in the control sample indicate the background level of mutation.

Figure 6 | INDUCE-seq reproducibility



INDUCE-seq measures DSBs in an unbiased way. An inventory of the high-level break numbers are presented here. **A) Total Breaks:** Breaks measured by INDUCE-seq are reproducible between triplicates. GM24385 and MCF10A cells have different total amounts of DSBs that accumulate over time post gene-editing. **B) Number of Off-targets:** The number of sequence-based off-targets detected by INDUCE-seq is consistent between replicates. The majority of CRISPR-Cas9 activity is registered at 6 hours after transfection. Off-targets discovery is consistent between the two experiments. **C) On- & Off-target Break relationship:** Detailed analysis of the on- and off-target DSBs demonstrates a highly reproducible break detection across the 5 gene-editing experiments. All data shown plotted as the mean for each triplicate with error bars representing the standard deviation.

Conclusions |

- INDUCE-seq reproducibly **detects on-target editing between the two laboratories and cell types.**
- INDUCE-seq **reproducibly identifies off-targets between laboratories.**
- INDUCE-seq **detects both common and unique off-targets in different cell types.**
- **DSB recurrency at both on- and off-targets correlates with mutational frequency.**

References

1. Vitelli et al. Recent Advancements in DNA Damage-Transcription Crosstalk and High-Resolution Mapping of DNA Breaks, Annu Rev Genomics Hum Genet. 2017
2. Dobbs et al. Precision digital mapping of endogenous and induced genomic DNA breaks by INDUCE-seq, Nature Communications, 2022
3. HESI CT-TRACS, https://hesiglobal.org/cell-therapy-tracking-circulation-safety-ct-tracs/
4. Kennedy et al. Detecting ultralow-frequency mutations by Duplex Sequencing, Nature Protocols, 2014
5. Newton et al. Negative DNA supercoiling induces genome-wide Cas9 off-target activity, Mol Cell, 2023

© 2024 Broken String Biosciences – All rights reserved

https://www.linkedin.com/company/broken-string-biosciences-ltd
https://twitter.com/brokenstringbio
https://www.brokenstringbio.com / info@brokenstringbio.com

Contact us:
[2] Unit AB3-04, Level 3
BioData Innovation Centre
Wellcome Genome Campus
Hinxton, Cambridge, CB10 1DR, UK