Improved Prediction and Activation of Homology-Directed Repair in Translational Ex-vivo Systems

Gavin Kurgan[1], Sridhar Selvaraj[2], Jessica Woodley[1], Ludwig Schmiderer[2], Steve Glenn[1], Javier Gomez Vargas[1], Morgan Sturgeon[1], Bernice Thommandru[1], Karthik Murugan[1], Cheryl Blackmer[1], Kirby Wallace[3], Garrett Rettig[1], Chris Vakulskas[1], Matthew Porteus[2], and Ashley Jacobi[1]

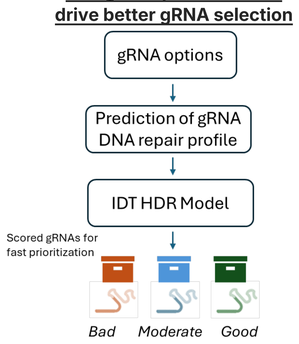
[1]Integrated DNA Technologies, Coralville, IA, USA
[2]Stanford University, Stanford, CA, USA
[3]Kamau Therapeutics, South San Francisco, CA, USA

* Corresponding author: gkurgan@idtdna.com

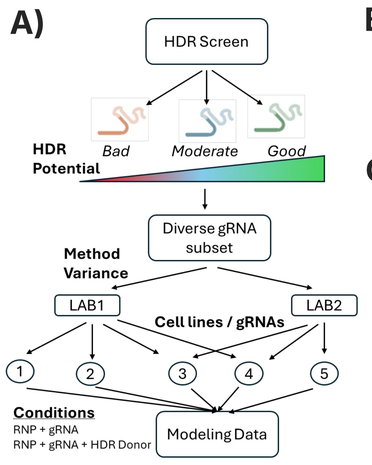
Overview

- HDR potential can be predicted apriori using *in silico* derived DNA repair predictions
- HDR predictions can generalize across clinical / non-clinical cell lines, and donor oligo types
- HDR prediction tool selects top 60% of HDR gRNAs with 97% precision / 66% recall
- Positively predicted sites show improved MOI and DNA repair kinetics performing HDR compared to negative prediction
- Next-generation HDR Enhancing Peptide (HEP) improves HDR without impacting off-targets in translational context

Using HDR predictions to drive better gRNA selection

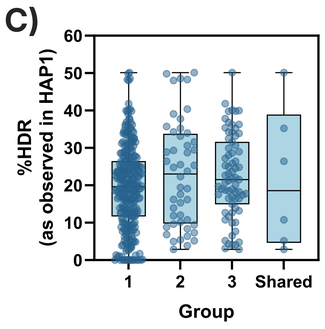


Defining a broad target set across cell lines



B)

Group	Cell lines	Lab	# gRNAs
1	HAP1	IDT	277
2	K562, Primary T, iPSC	IDT	48
3	HSPC, Primary T, iPSC	Porteus	72



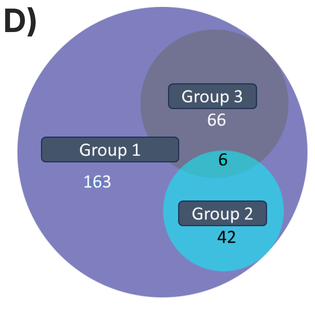


Figure 1. Experimental Design. A) First, we defined a workflow by which we would first screen 277 different gRNAs in a single cell line (HAP1) and measure the % homology directed repair (HDR) of the site when electroporating *S.p.* Cas9, the gRNA of interest, and either +/- a commonly designed ssODN oligo (+6 bp insertion, 40 bp homology arms on either end at the DSB site) to measure the native NHEJ repair profile in addition to its HDR potential. Following this, we subset gRNAs based on the goal of achieving a smaller set of gRNAs satisfying the dynamic range, and split these between two labs which would use their own optimized methods, using both common and unique gRNAs / cell lines to generate a sufficiently large dataset to attempt to make generalized conclusions regarding HDR independent of cell line. **B)** the different generalized data groups by lab / cell line / gRNA are displayed as well as the **C)** the HDR potential of these sites in HAP1 and **D)** the overlap of selected gRNAs per group.

HDR rank-order potential (Spearman r) is Broadly Conserved Across Cell lines

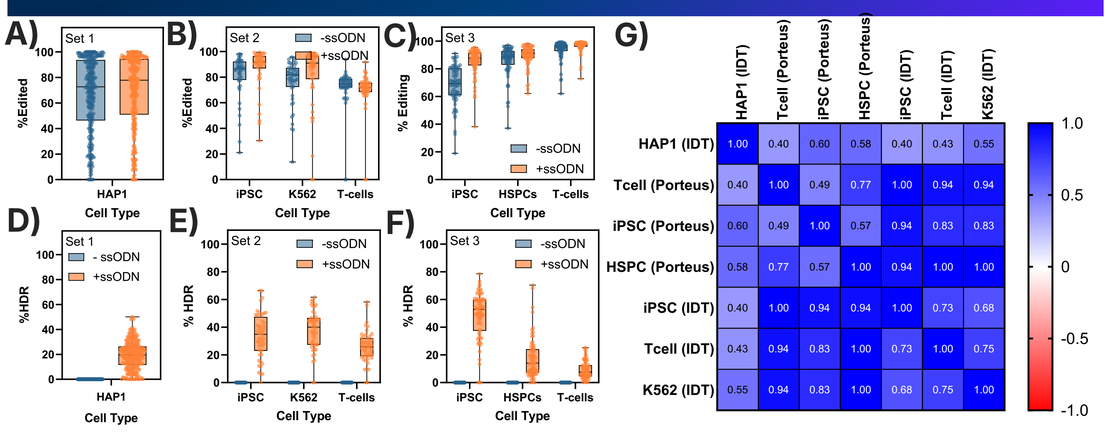


Figure 2. HDR Potential across cell lines. Following editing in the respective groups, the **A-C)** indel editing and **D-F)** %HDR was measured in the absence (blue) and presence (orange) of the commonly designed HDR donor (ssODN; +6 bp INS; 40 bp homology arms; insertion at the canonical DSB site). Following this, **G)** the Spearman correlation coefficient was measured for all gRNAs common between each respective cell line / lab. This demonstrates that general HDR rank order potential is relatively conserved between cell lines, across a number of the gRNAs and conditions tested.

Donor-type independent conservation observed

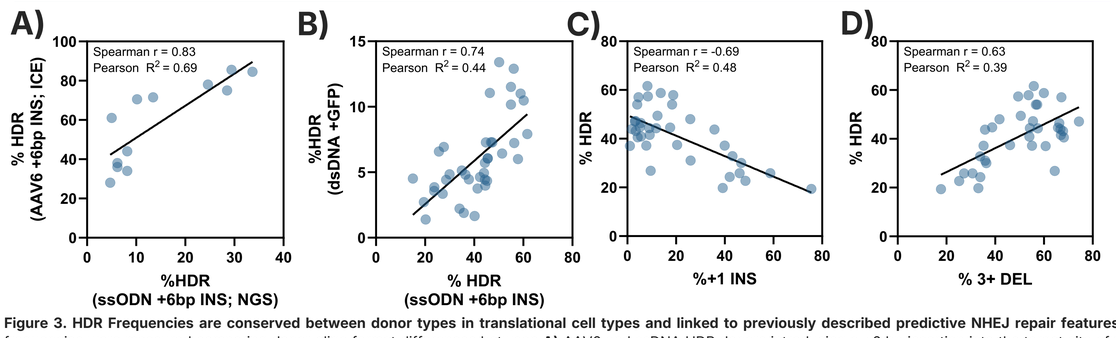


Figure 3. HDR Frequencies are conserved between donor types in translational cell types and linked to previously described predictive NHEJ repair features. HDR frequencies were measured comparing donor oligo format differences between **A)** AAV6 and ssDNA HDR donors introducing a +6 bp insertion into the target site of CD34⁺ HSPCs (n = 12 sites) and **B)** dsDNA and ssDNA HDR donors introducing a larger INS (~1kb) and shorter INS (+6 bp INS) using 200 bp and 40 bp homology arms in K562, respectively. Comparison of HDR to previously studied predictive features of HDR **C)** +1 INS frequency and **D)** 3+ DEL frequency show expected correlations to HDR in K562.

In silico and empirically derived features generally correlate on HDR predictive features

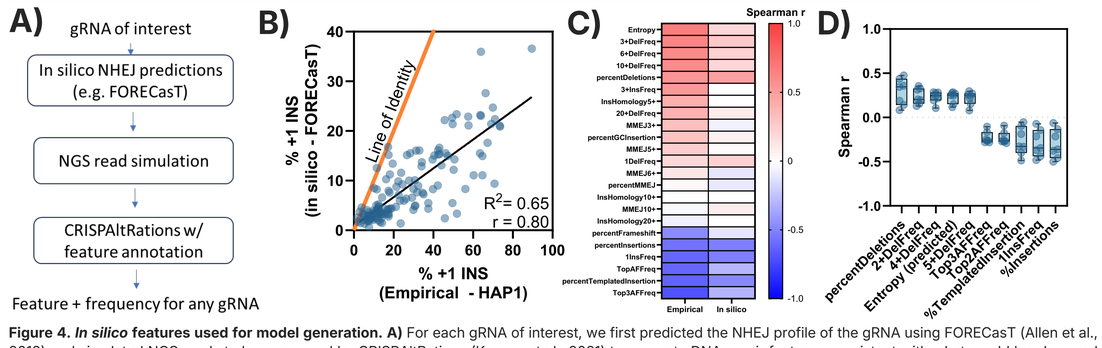


Figure 4. *In silico* features used for model generation. A) For each gRNA of interest, we first predicted the NHEJ profile of the gRNA using FORECasT (Allen et al., 2018) and simulated NGS reads to be processed by CRISPAltRations (Kurgan et al., 2021) to generate DNA repair features consistent with what would be observed empirically. **B)** Comparison of empirical and *in silico* +1 INS frequencies for Group 1 gRNAs tested in the HAP1 cell line show that well described features like +1 INS frequencies maintain a strong rank order correlation with empirical findings. **C**) Comparison of a broader set of features demonstrates additional features that maintain predictive rank-order correlations with HDR and maintain consistent relationships between empirical and *in silico* data in HAP1. **D)** A broader comparison of all cell lines / groups presented more diverse findings and the top median

Model predictions significantly improve selection of high HDR potential gRNAs across cell lines / conditions

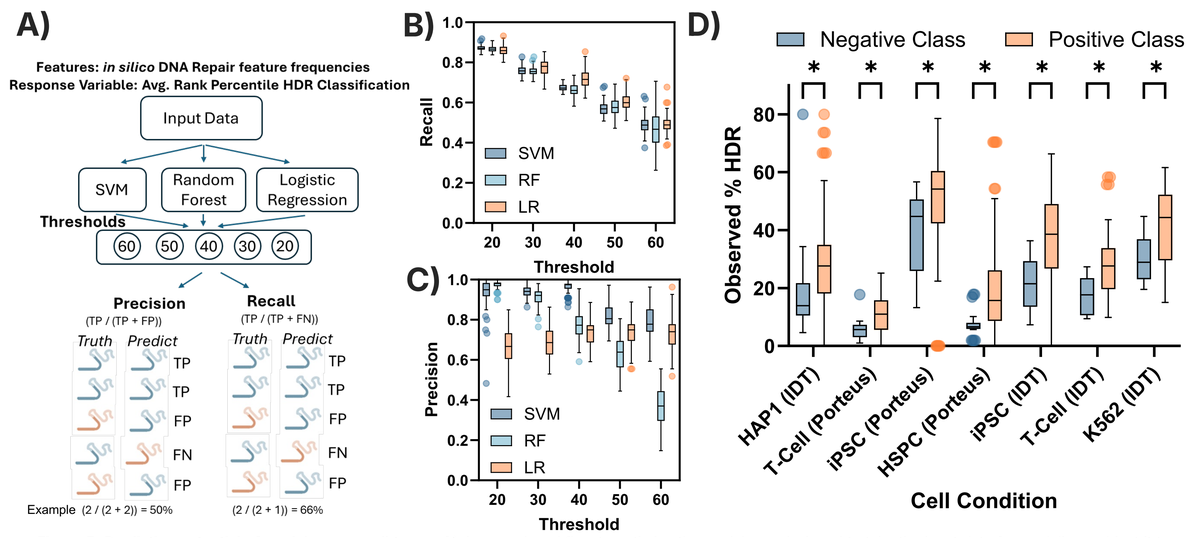


Figure 5. Predictive potential of model across cell types. A) A general workflow to optimize the model was designed to take the input data for modeling, subject it to three different model types (SVM, Random Forest, Logistic Regression, attempting to perform binary classification of the gRNA's predicted repair profile using avg percentile rank of HDR between cell lines as a response variable with classification performed at five different thresholds (60%, 50%, 40%, 30%, 20%) over 100 bootstrapped model iterations calculating both precision and recall. **B)** Recall and **C)** precision for all 100 bootstrapped iterations per model per threshold are shown. Upon selecting an SVM performing classification at the 40% threshold as "Moderate" HDR potential gRNAs and an SVM at the 60% threshold as "High" HDR potential gRNAs and **D)** we compared the performance of the model at the "Moderate" threshold on observed HDR frequencies for a held out set of test gRNA data during 10 bootstrap.

High HDR potential sites have more favorable MOI; kinetics are better too in the absence of enhancers

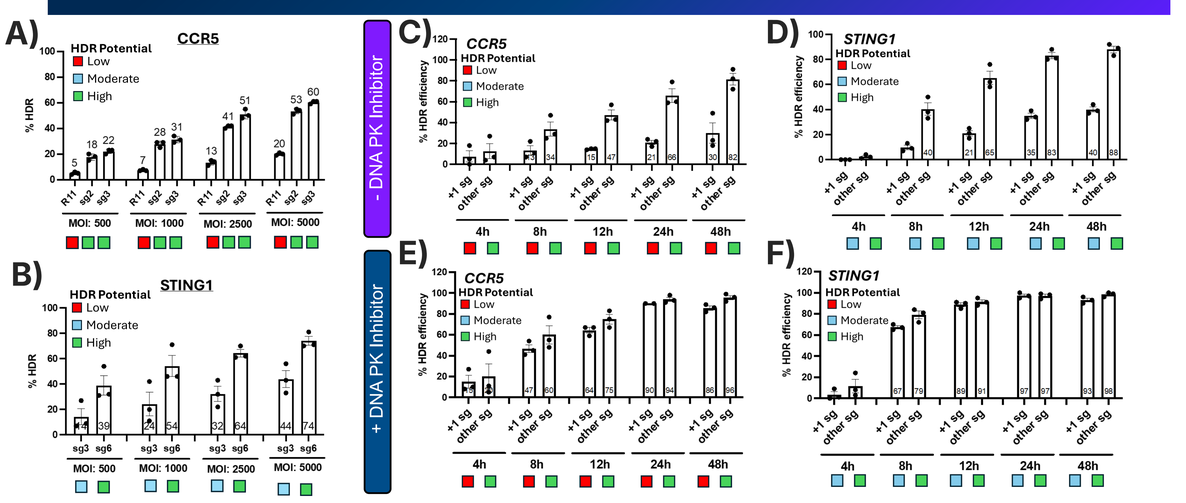


Figure 6. Predictive potential corresponds to lower MOI and better HDR kinetics. Using AAV6 as an HDR donor in CD34+ HSPCs, multiple gRNAs with HDR predictive potential labeled were tested at a variable multiplicity of infection (MOI) for the **A)** CCR5 and **B)** STING1 locus and % HDR measured. To investigate if lower HDR predictive potential corresponds to differences in HDR kinetics, we tested two gRNAs with different predictive potential in the absence (top row) and presence (bottom row) of a DNA PK inhibitor (AZD7648) in CD34+ HSPCs (n = 3) at the **C-D)** CCR5 and **E-F)** STING1 locus measuring HDR over variable intervals across 48h.

Next-generation HDR Enhancing protein (HEP) increases HDR without impacting off-target editing

A) V2, but not Ubv-A, inhibits tag integration by NHEJ

B) V2, but not Ubv-A, changes indel and translocation profiles

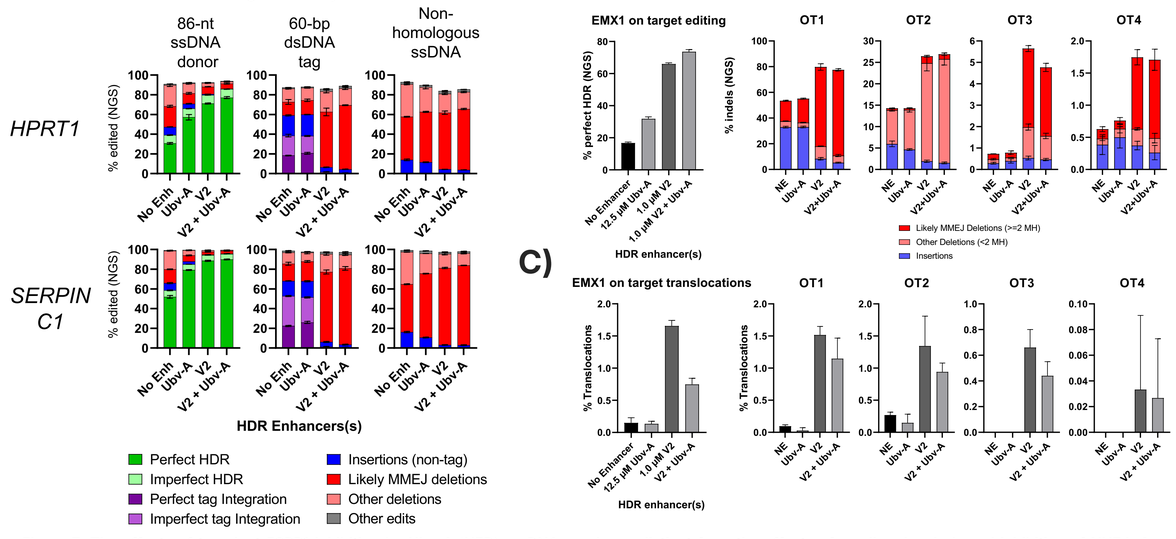


Figure 7. The effects of transient 53BP1 inhibition by Ubv-A (HEP) on DNA repair are distinct from the effects of small molecule-based inhibition of NHEJ. A. Breakdown of repair events at *HPRT1* and *SERPINC1* in HEK293 cells treated with Cas9 RNP, 0 or 25 µM tagless Ubv-A, and either 2 µM ssDNA donor (86 bp with 40 bp homology arms), 2 µM dsDNA (60 bp) tag without homology to the target site, or 2 µM ssDNA Alt-R Cas9 Electroporation Enhancer (IDT). Cells were then incubated either with or without 1 µM Alt-R enhancer V2 NHEJ inhibitor. Error bars indicate SD, n=3. **B.** Breakdown of repair events at *EMX1* on- and off-target sites in HEK293 cells treated with Cas9 RNP targeting EMX1, ssDNA donor, and 0 or 12.5 µM Ubv-A and incubated with 0 or 1 µM V2 for 24 hours. Editing measured using rhAmpSeq™ using a pool of NGS primers targeting the EMX1 on-target site and a set of known off-target sites. The graphs show the rate of perfect HDR at the on-target site and the distribution of repair events at the most frequently edited off-target sites with likely MMEJ repair events being deletions with microhomology lengths of at least 2 bp. **C.** Percent translocations to the on and off target sites, as calculated by summing all reads using the forward amplicon primer for each site that paired with a reverse primer for a different site in the primer pool for translocations events that reached a 0.05% FDR, divided by all reads that included the forward primer.

HDR enhancing protein has improved translational editing

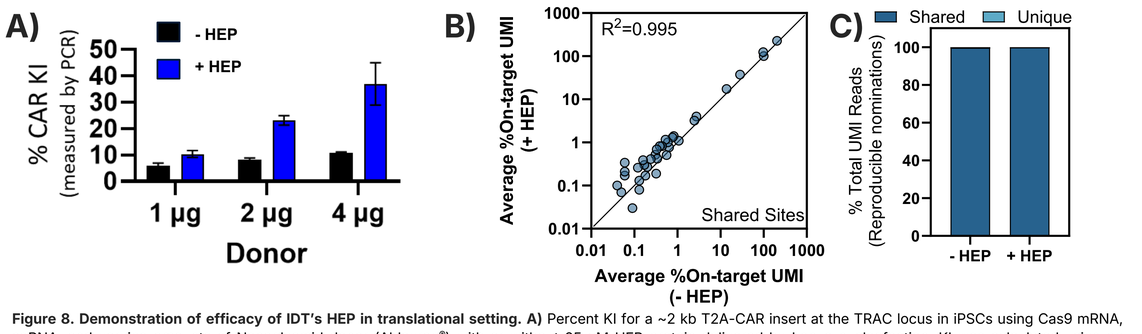


Figure 8. Demonstration of efficacy of IDT's HEP in translational setting. A) Percent KI for a ~2 kb T2A-CAR insert at the TRAC locus in iPSCs using Cas9 mRNA, sgRNA, and varying amounts of Nanoplasmid donor (Aldevron®) with or without 25 µM HEP protein delivered by Lonza nucleofection. KI was calculated using an Agilent Fragment Analyzer (n=3). Off-targets were nominated in HEK293-Cas9 using IDT's modified GUIDE-seq method treated with (- HEP) and without HEP (+ HEP). **B)** Scatterplot of average % On-target UMI counts of shared nomination sites and **C)** sum of average % On-target UMI reads for both shared and unique sites in the treatments conclude that frequencies of nomination are conserved and off-target profiles are not significantly altered by HEP