

Computational Systems Biology Deep Learning in the Life Sciences

6.802 6.874 20.390 20.490 HST.506

David Gifford

Lecture 19

April 16, 2020

Predicting genome editing outcomes with machine
learning methods



<http://mit6874.github.io>

Poll Warm Up:

Where are you located today?
How do you prefer to receive
lectures?

Predicting the outcomes of genome editing

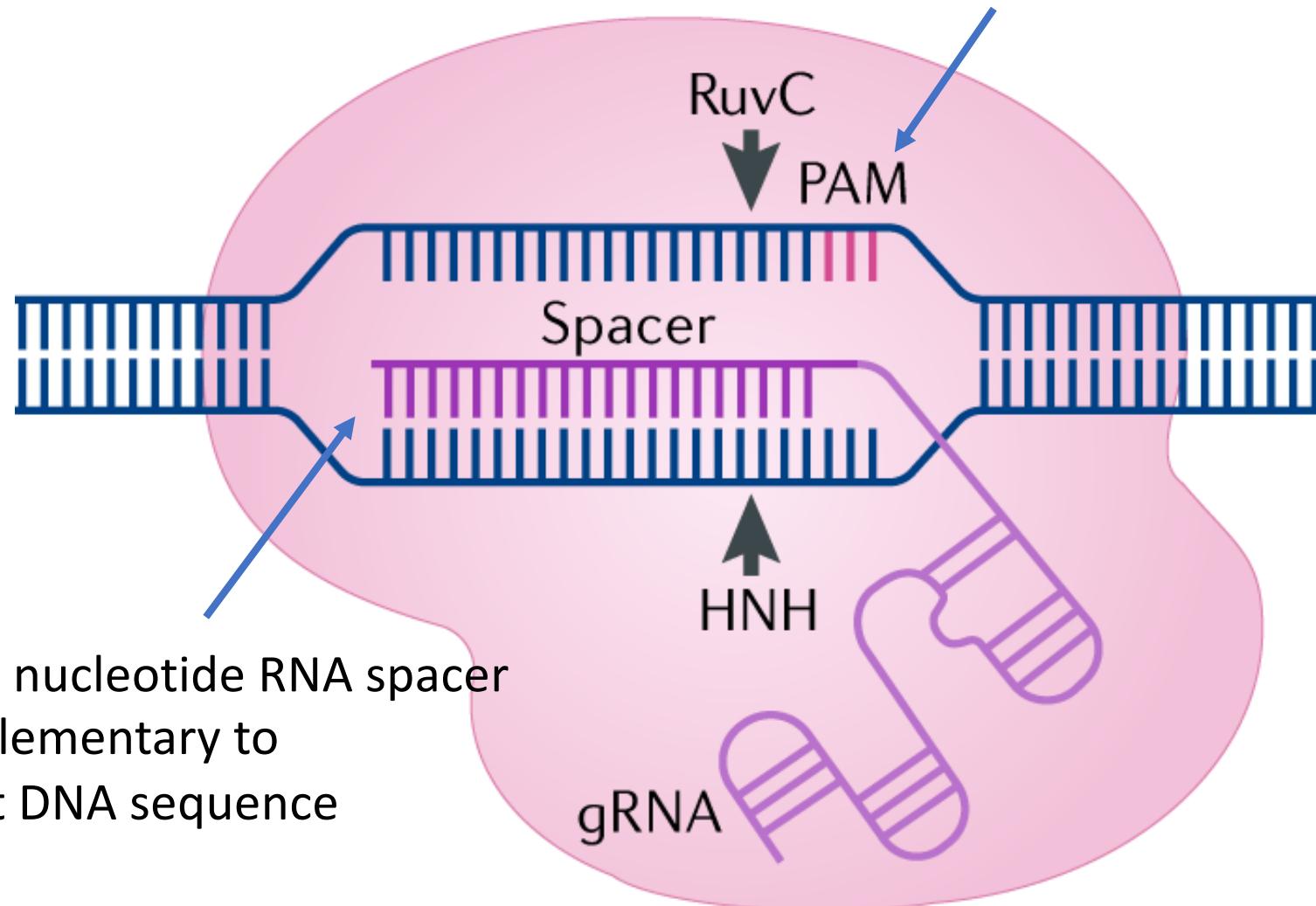
- CRISPR (cas9) genome editing in detail
- Assays to detect off target cutting
- Machine learning models to predict off target cutting
- Discovering the necessary genome for Tdgf1
- Machine learning models of on target cutting
- The limitations of base editing

CRISPR (clustered regularly interspaced short palindromic repeats) editing mechanics

Cas9 nuclease engaged in cutting

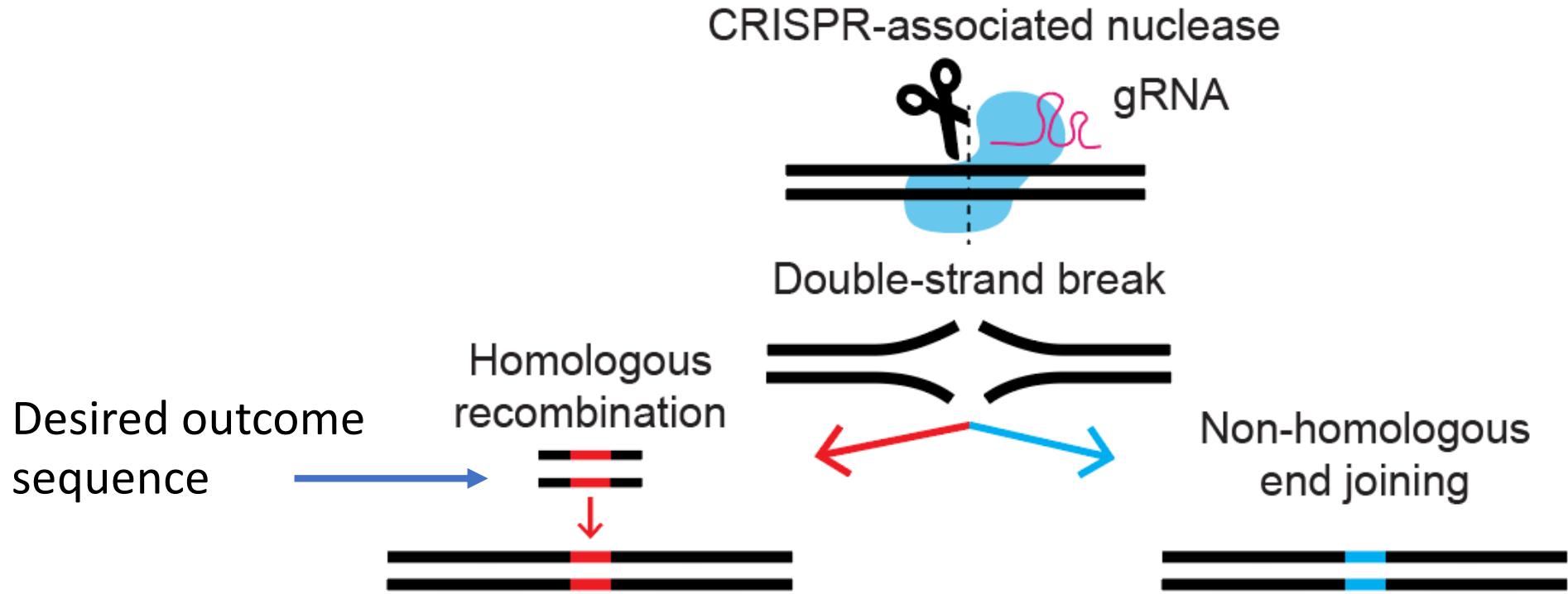
Cas9 nuclease

Required PAM sequence limits available cut sites (NGG Cas9)



17-24 nucleotide RNA spacer
complementary to
target DNA sequence

Genome cuts resolve in two ways



CRISPR is relevant as a therapeutic tool

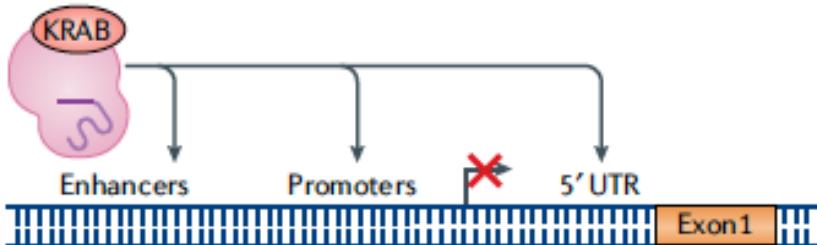
TABLE 1: Application of CRISPR as a therapeutic tool for common monogenetic disorders of humans.

Disease	Manipulated gene	SgRNA target	Cell type	Species	In vivo/ in vitro	Delivery	Outcome	Ref
Cystic fibrosis	CFTR locus F508 del	CFTR exon 11 or intron 11	Cultured intestinal stem cells from organoids isolated from CF patients	Human	<i>In vitro</i>	Cas9, sgRNA plasmid transfection	Successful and rescued CFTR protein	[28]
	F508	CFTR exon 10.	Induced pluripotent stem cells (iPSCs)	Human	<i>In vitro</i>	PiggyBac transposase nucleofection	Normal CFTR expression on differentiation	[29]
Sickle cell anemia	β -Globin	Exon 1 of the human β -globin gene	Blood-derived CD34 ⁺ cells CD34 ⁺ stem and progenitor cells from SCA patients	Human	<i>In vitro</i>	Lenti-viral vector	Successful 18% gene modification in <i>in vitro</i> cells. Wild type H b cells observed 30% HDR in CD34 ⁺ with 80% of them being viable and producing β -globin	[30]
	β -Globin		CD34 ⁺ cells	Human	<i>In vitro</i>	Nucleofection	Successful rectification achieved	[31]
	Glu6Val mutant gene		Stem and progenitor cells	Human	<i>In vitro</i>	Adeno-associated viral vector	Successful rectification achieved	[32]
Thalassemia	IV52-645	Gene targeting intron 2 mutation site near IVS2-654 C > T mutation	Thalassemia patient-derived iPSCs	Human	<i>In vitro</i>	PiggyBac transposon donor	Mutation corrected in the relevant site	[33]
	HBB gene CD 41/42 mutation		β -Thal iPSCs	Human	<i>In vitro</i>	Lenti-CRISPR V2 vector.	Cells have exhibited normal karyotype and have retained full pluripotency	[34]
	HBB mutation	TLTT deletion between 41 st and 42 nd amino acid producing Hbb gene	Native iPSCs obtained from urinary cells of β -thalassemic patient	Human	<i>In vitro</i>	pX330 vector	Exhibited normal karyotype and retained pluripotency	[35]
	HBB mutations	Exon 2	iPSCs from thalassemic patients	Human	<i>In vitro</i>	PiggyBac transposon vector	Full pluripotency or normal karyotypes and no off-target effects	[36]
	HBB mutations	2 nd intron of the Hbb gene.	Induced pluripotent stem cells (iPSCs)	Human	<i>In vitro</i>	B003 plasmid vector transfection	Normal cell types with no off-target effects	[37]
Huntington's chorea disorder	HTT gene	Promoter region, transcription start site, and expanded CAG mutation	Primary fibroblast cells	Human	<i>In vitro</i>	CRISPR/Cas9 vector transfection	Rectification of mutation successful	[38]
	Mutant HTT locus	HTT exon 1 deletion	Fibroblasts and <i>in vivo</i> HD transgenic mice	Human and mouse	<i>In vitro</i>	Plasmid pX330	Rectification of mutant HTT locus	[39]
	mHTT	CAG repeats in exon 1 of the human HTT gene	HEK 239 cell line	Mouse	<i>In vivo</i>	Adeno-associated virus vector	Reduction in expression of mHTT in the striatum of model mice	[40]

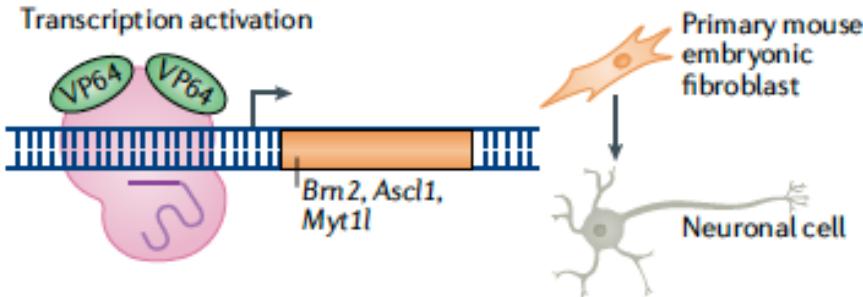
CRISPR derivatives can implement many functions

a Gene regulators

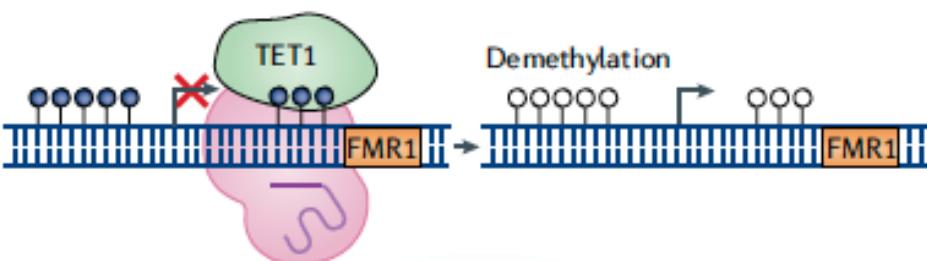
Transcription repression



Transcription activation

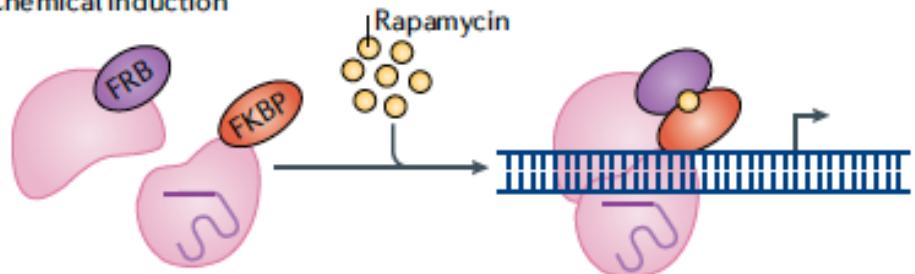


Epigenome editing

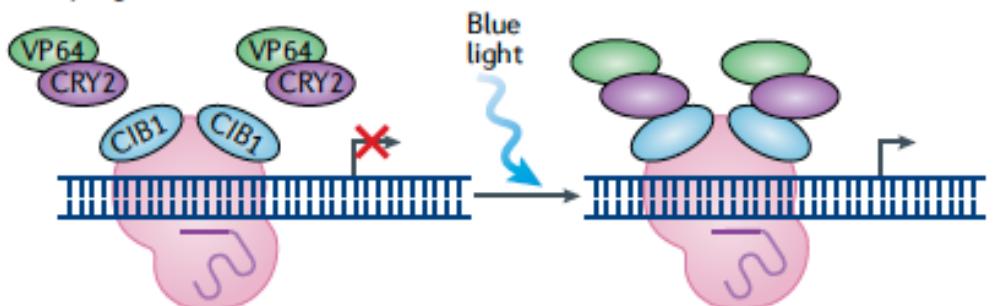


b Dynamic control

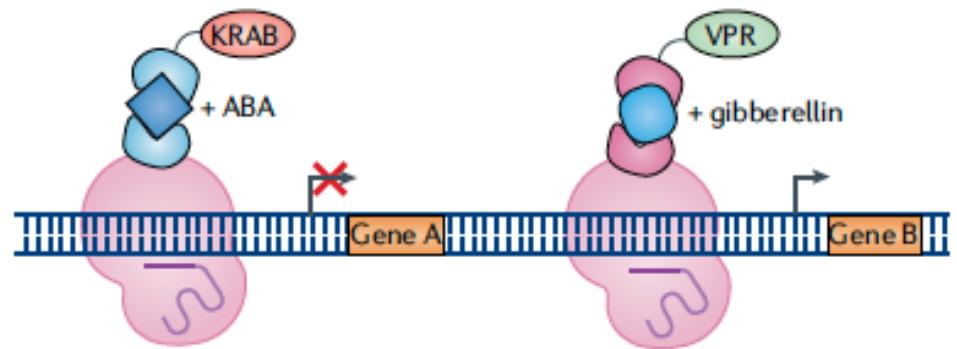
Chemical Induction



Optogenetics



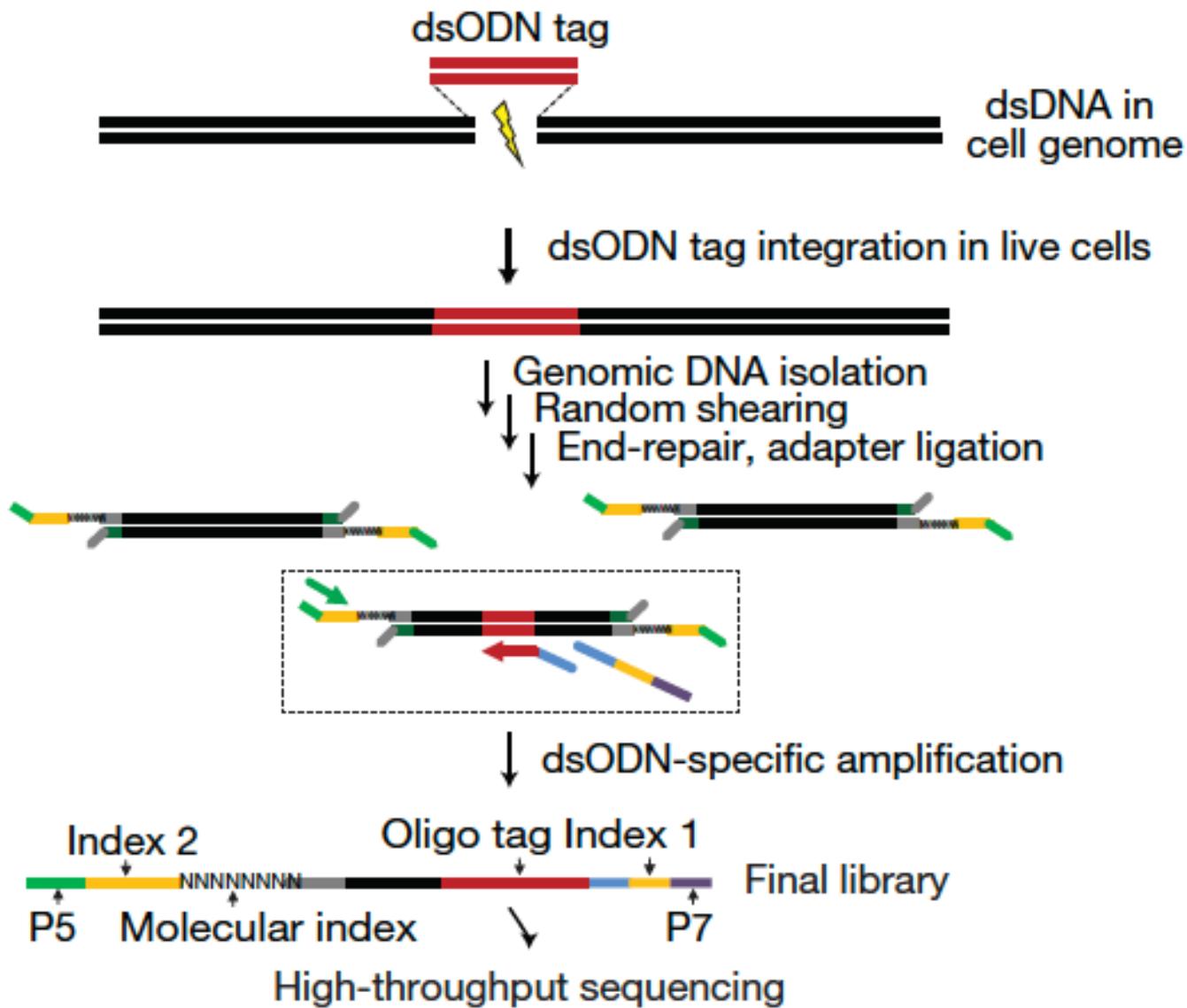
+ ABA



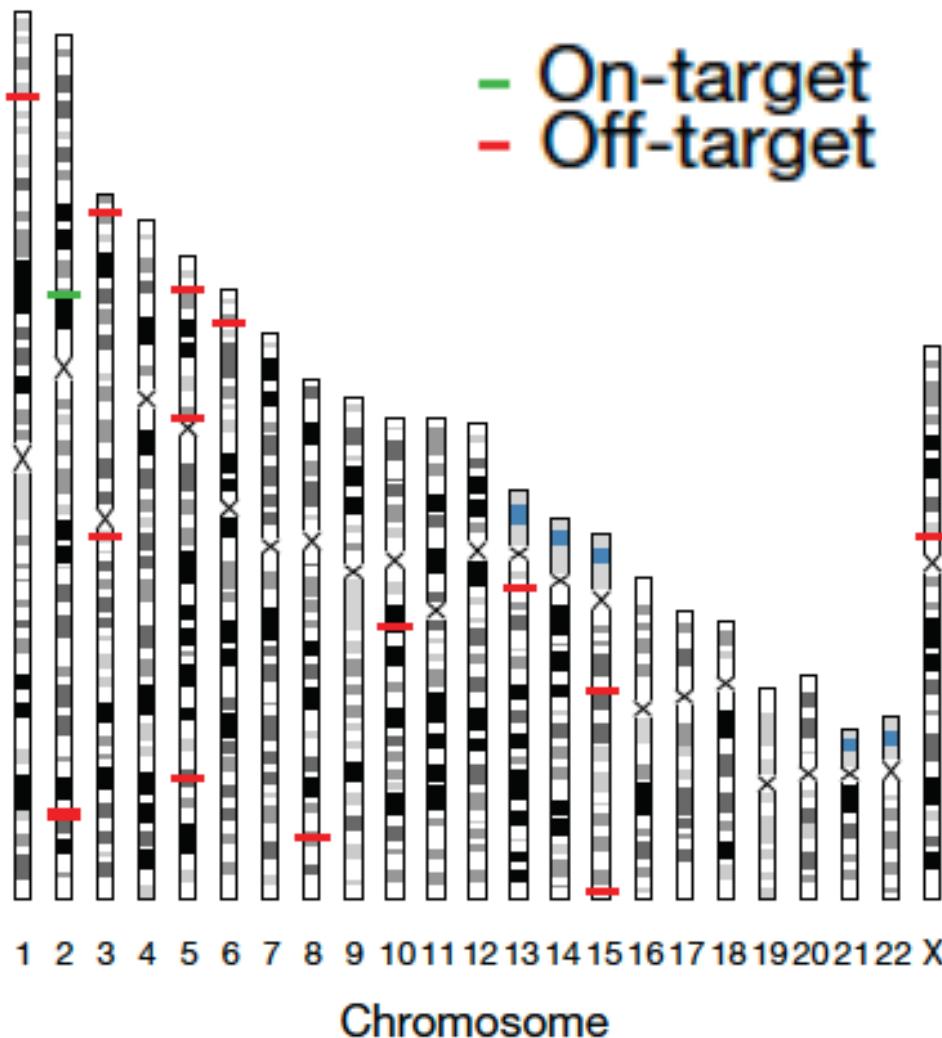
How well does CRISPR find its way to a specific site?

Characterizing off target effects with a genome wide assay

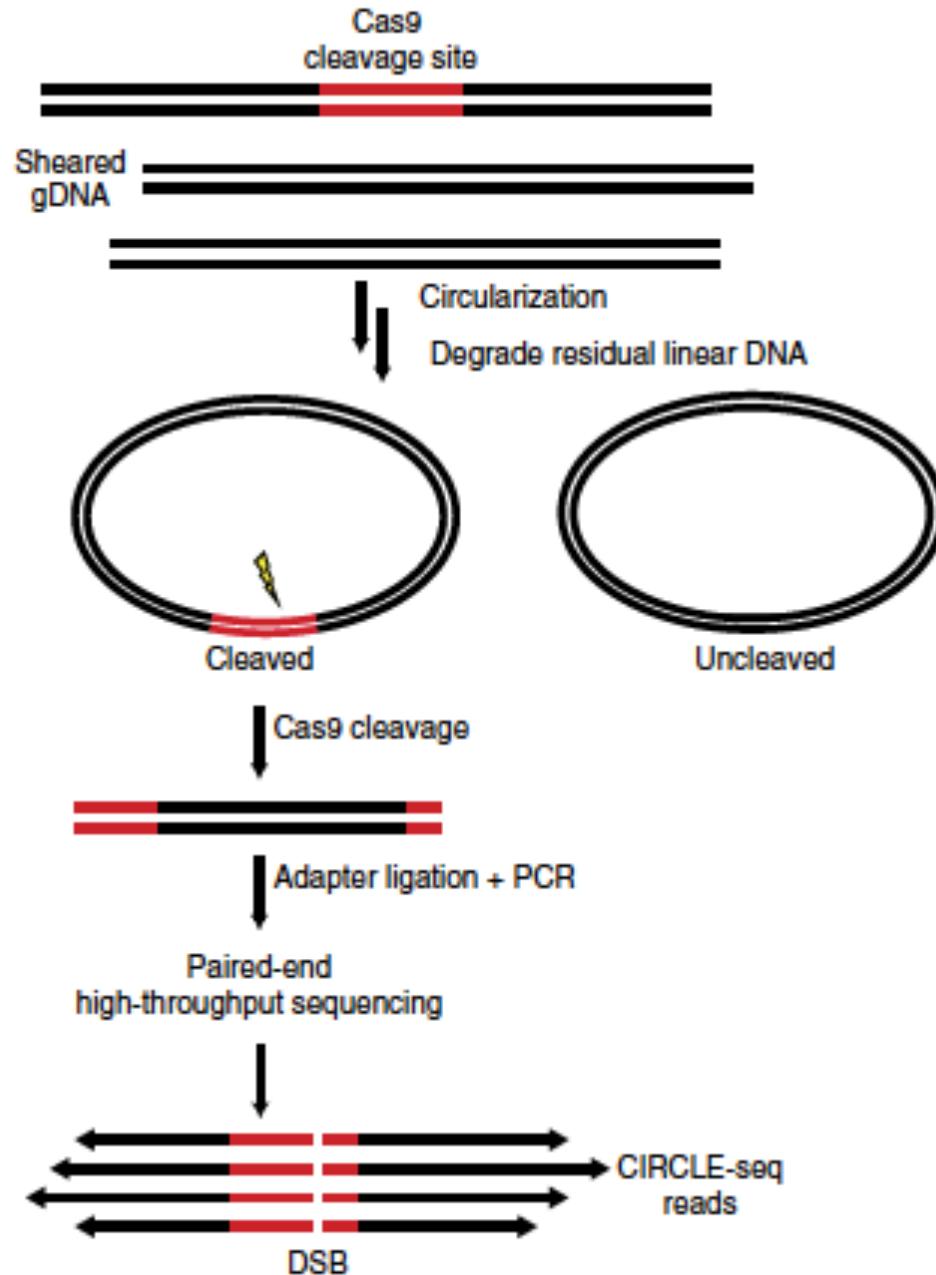
GUIDE-seq incorporates a 34-bp phosphothiorated double stranded DNA oligo (dsODN) into cut sites



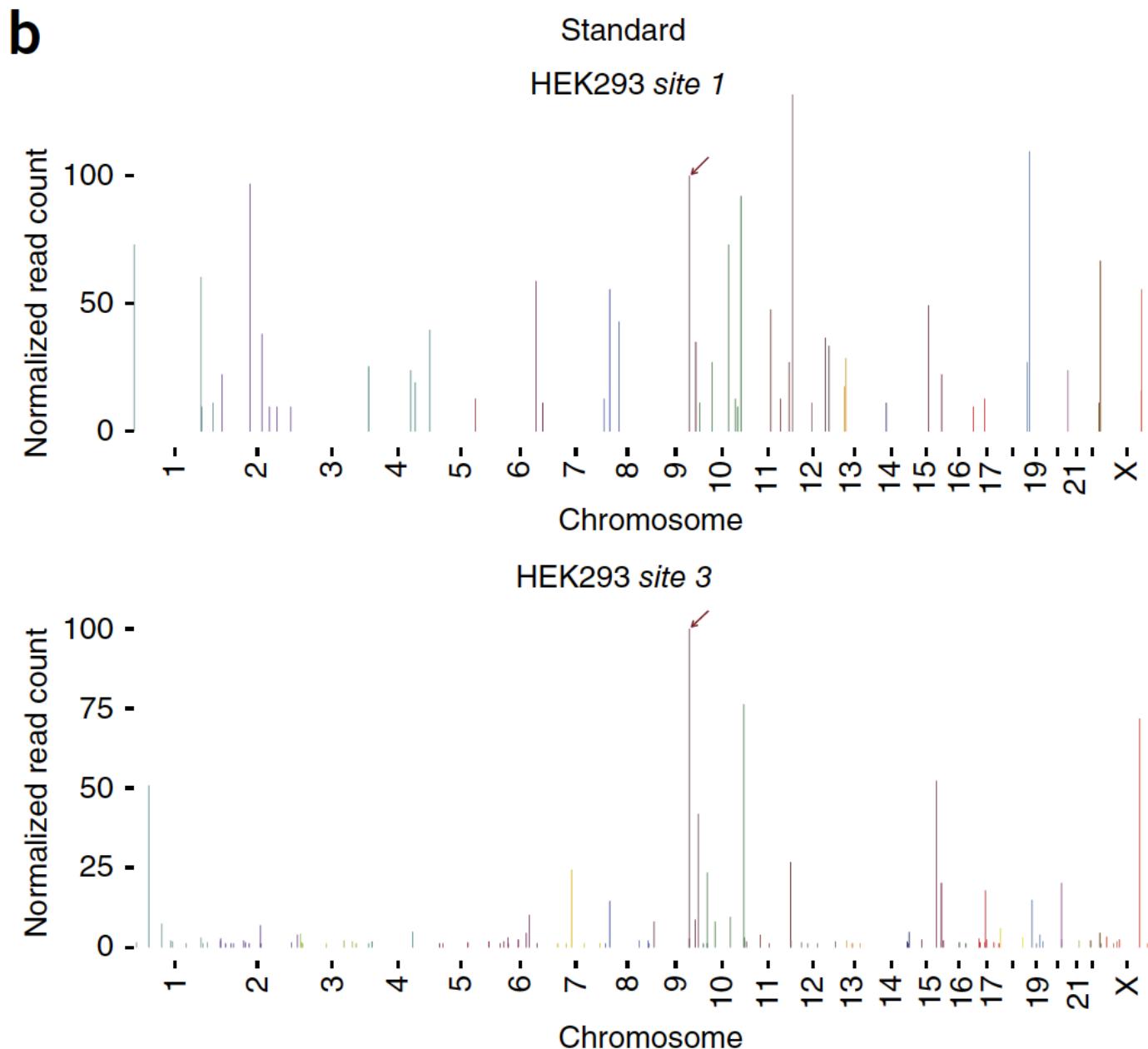
GUIDE-seq identifies off target cuts



CIRCLE-seq reveals CRISPR cut sites genome wide



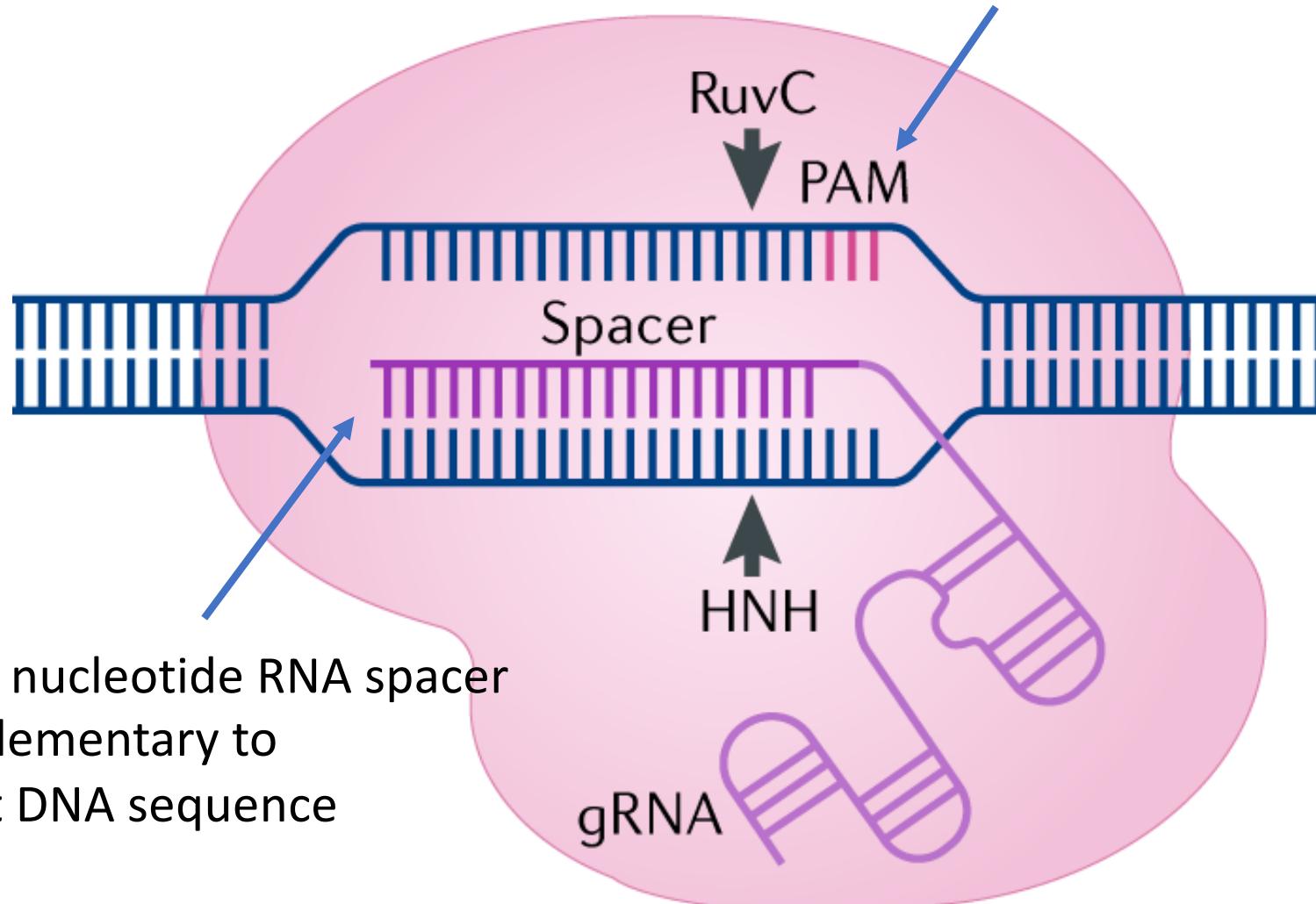
CIRCLE-seq reveals CRISPR cut sites genome wide (arrow is intended cut site)



How can we predict off target activity of a CRISPR based enzyme?

Recall our biological model
What features should we use to predict off target effects?

Cas9 nuclease



Example CRISPR features for off target prediction

- CROP-IT grades gRNA sequences dividing 23bp sequence into three regions with different weights, penalty scores for consecutive mismatched sites
- CCTOP and MIT score considers positions and counts of mismatches
- CFD (cutting frequency determination) emulates a large number of single base, deletion, and insertions in the gRNA and scores these with reference to validated gRNAs in a cellular assay

A	G	C	.	.	C	G	G
---	---	---	---	---	---	---	---

Target DNA

1	0	0	.	.	0	0	0
0	1	0	.	.	0	1	1
0	0	1	.	.	1	0	0
0	0	0	.	.	0	0	0

1	0	0	.	.	0	0	0
1	1	0	.	.	0	1	1
0	0	1	.	.	1	0	0
0	0	1	.	.	0	0	0

One-hot
CodeOR
operation
on
Mismatches

0	0	0	.	.	0	0	0
1	1	0	.	.	0	1	1
0	0	0	.	.	1	0	0
0	0	1	.	.	0	0	0

G	G	T	.	.	C	G	G
---	---	---	---	---	---	---	---

Complementary
guide RNA

C	C	A	.	.	G	C	C
---	---	---	---	---	---	---	---

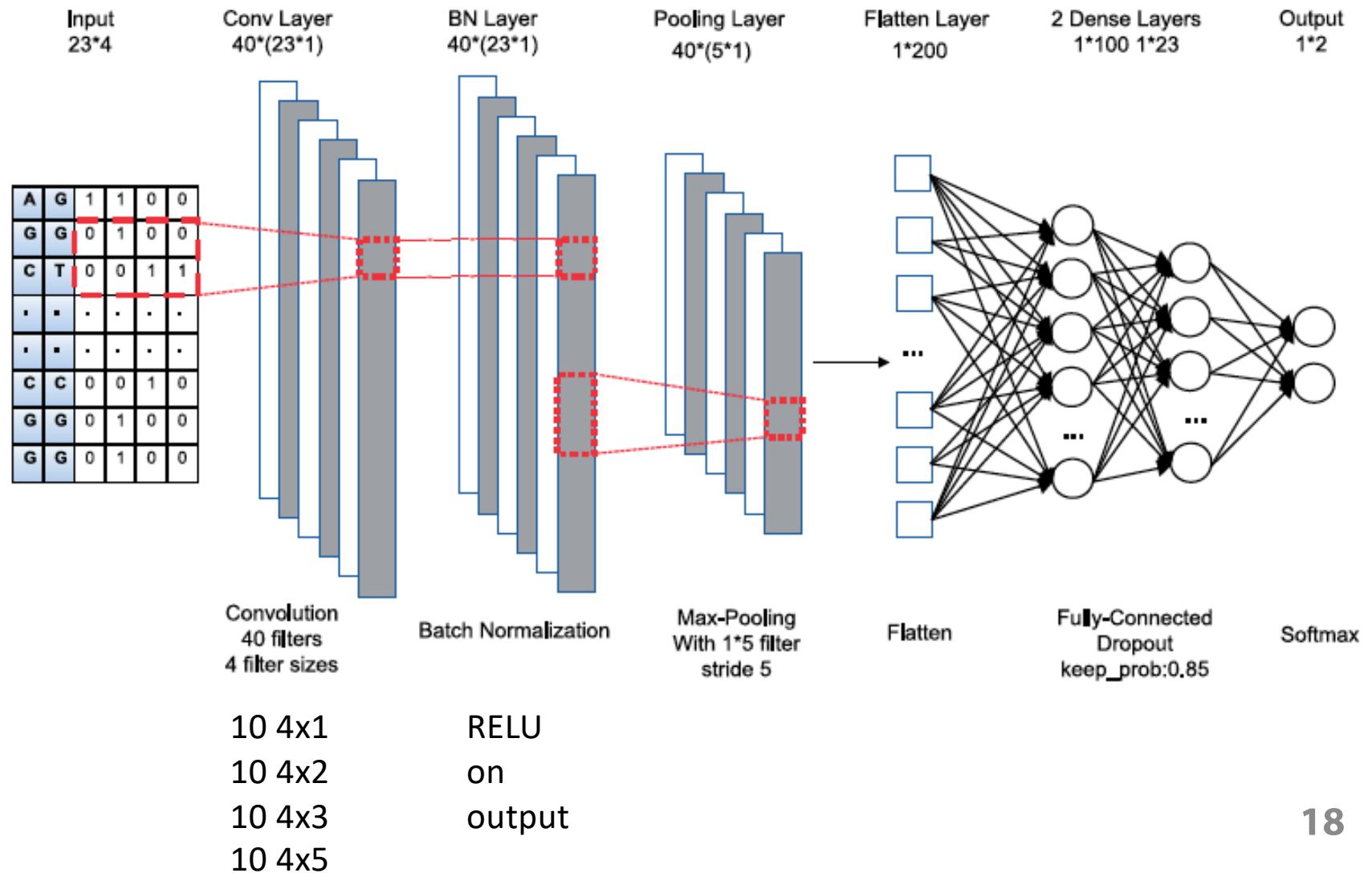
guide RNA

Off-target predictions in CRISPR-Cas9 gene editing using deep learning

Jiecong Lin and Ka-Chun Wong*

doi: 10.1093/bioinformatics/bty554

A deep neural network for classifying CRISPR recognition of a genomic site



Performance of different architectures on a 5x cross-validation on CRISPOR dataset

Table 1. The naming convention code and brief description of the variants of CNN and FNN models compared in this work

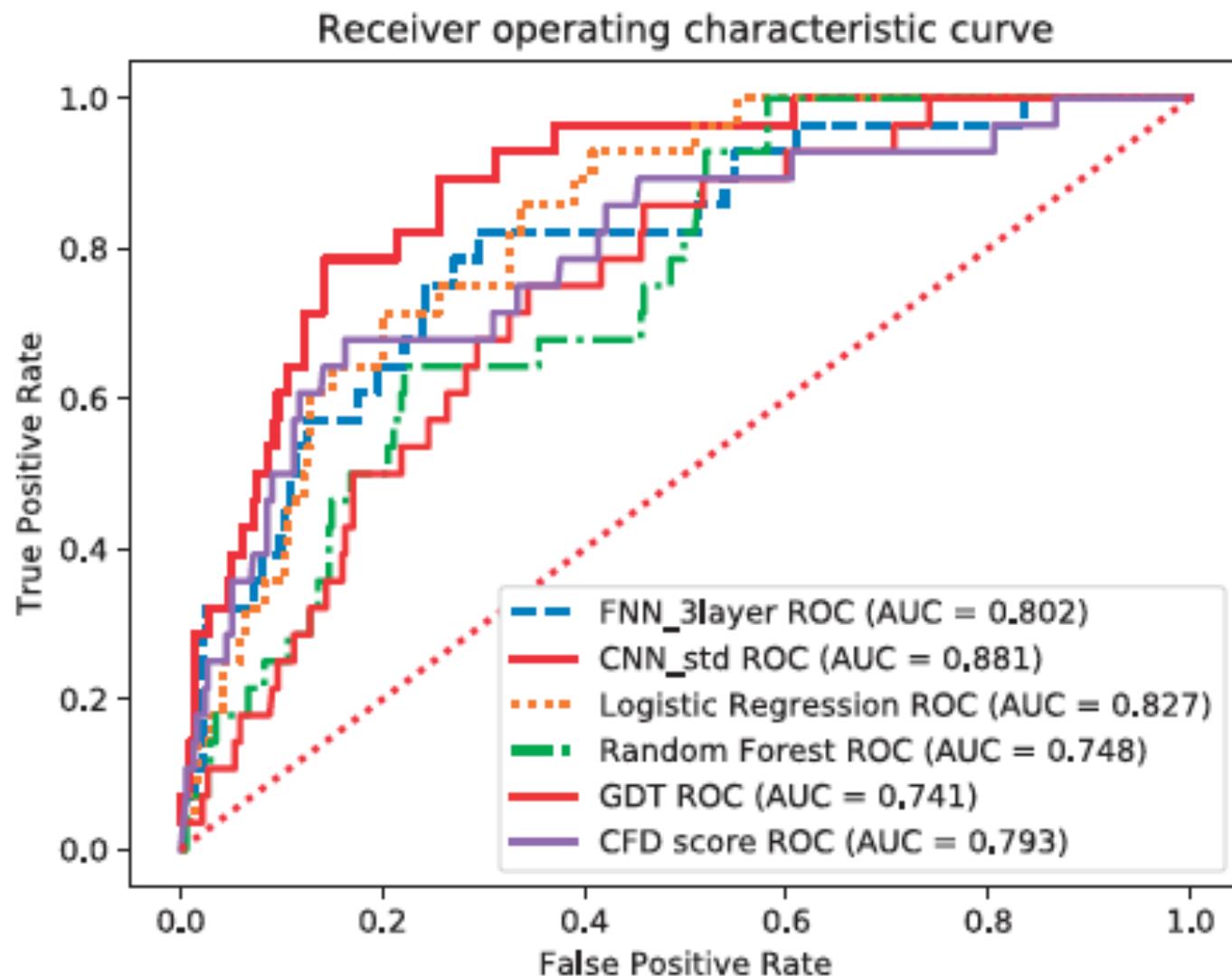
Model	Code	Architecture
FNN	FNN_2layer	Using 2 hidden layers with 250×40 neurons
FNN	FNN_3layer	Using 3 hidden layers with $50 \times 20 \times 10$ neurons
FNN	FNN_4layer	Using 4 hidden layers with $25 \times 10 \times 10 \times 4$ neurons
CNN	CNN_std	The basic structure as depicted in Figure 2
CNN	CNN_nbn	Without using BN layer
CNN	CNN_nd	Without drop-out layer
CNN	CNN_np	Without using max-pooling layer
CNN	CNN_pwin3	Using max-pooling layer of window size 3
CNN	CNN_pwin7	Using max-pooling layer of window size 7

Note: The descriptions of all CNN models are relative to CNN_std model depicted in Figure 2.

Table 2. Performance comparisons for different architectures under stratified 5-fold cross-validation on CRISPOR dataset

Model	Min_AUC	Max_AUC	Mean_AUC	Var_AUC
FNN_2layer	0.852	0.891	0.842	0.010
FNN_3layer	0.963	0.977	0.970	0.005
FNN_4layer	0.951	0.960	0.954	0.009
CNN_std	0.954	0.983	0.972	0.010
CNN_nbn	0.929	0.973	0.954	0.022
CNN_nd	0.953	0.974	0.969	0.013
CNN_np	0.720	0.981	0.899	0.093
CNN_pool_win3	0.632	0.979	0.903	0.137
CNN_pool_win7	0.943	0.983	0.967	0.015

Train on CRISPOR dataset Test on GUIDE-seq



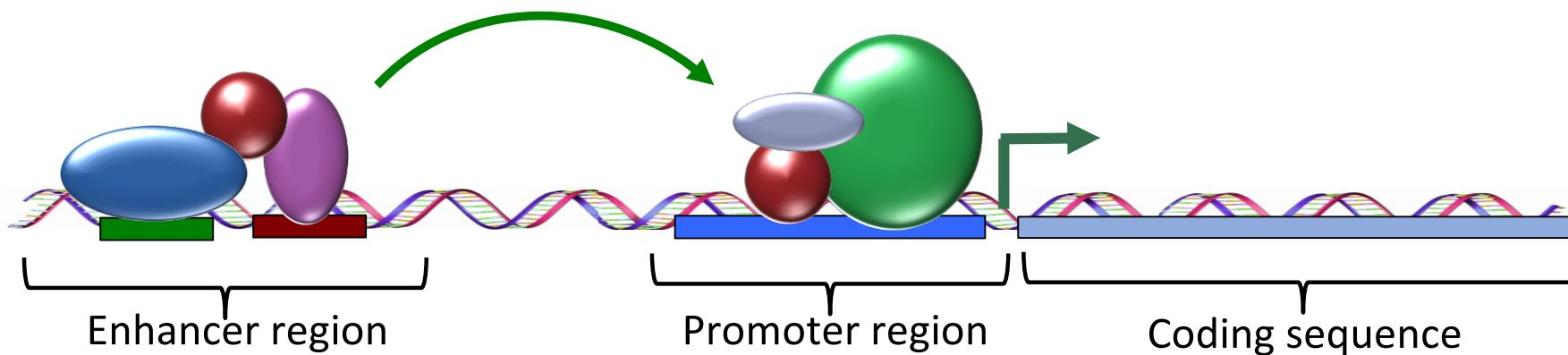
What bases are necessary for genome function?

Genome editing allow us to
change genome sequence
and observe the function of
each base in a selected
cellular context

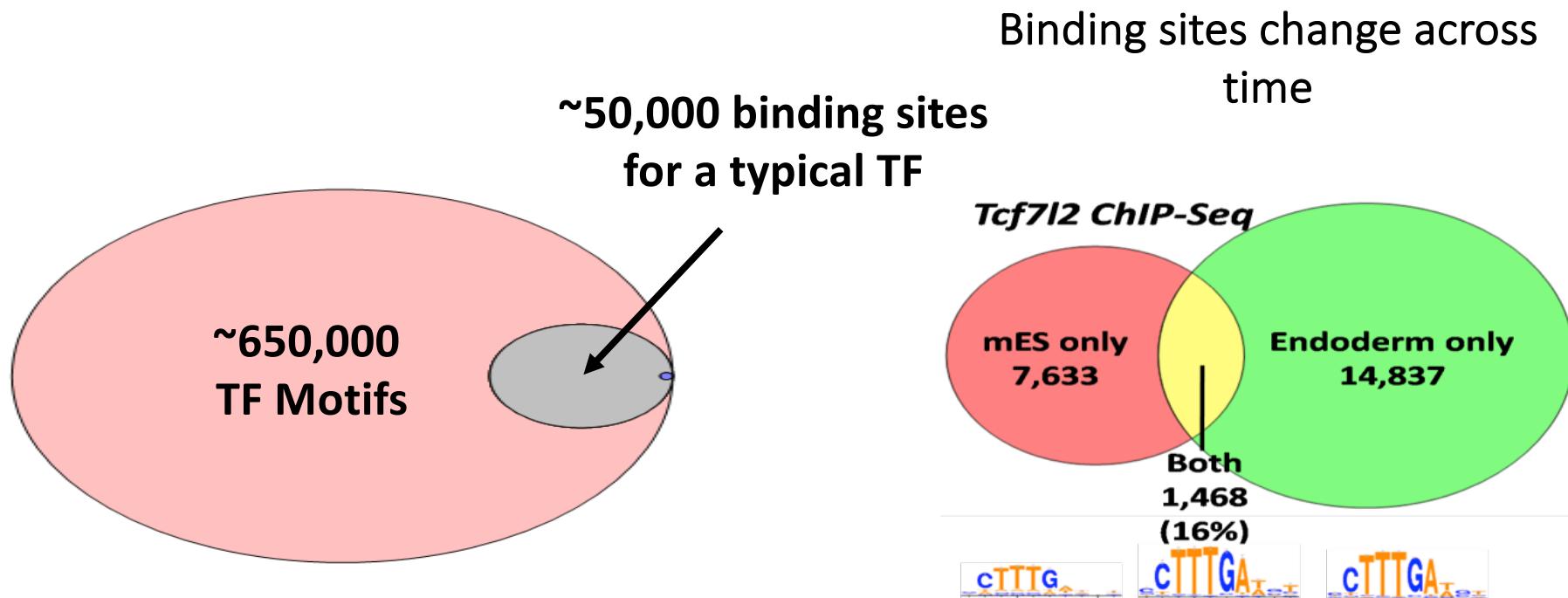
Genome editing allow us
to change genome
sequence and **observe the**
function of each base in a
selected cellular context

Genome editing allow us
to change genome
sequence and observe the
function of each base in a
selected cellular context

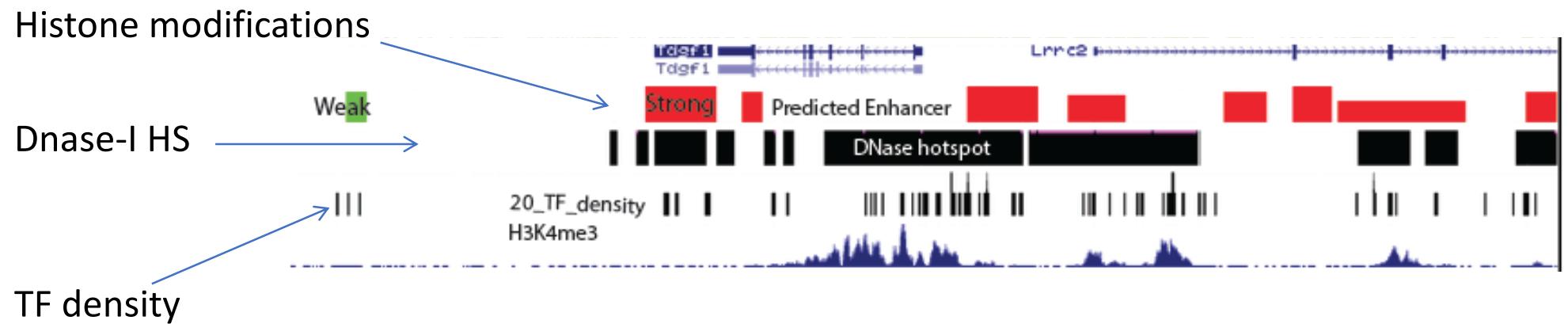
What are the key genomic elements that are necessary for gene expression?



Necessary elements will depend upon cell type –
the binding sites used by a given factor can depend
upon cell type (here Tcf7l2)

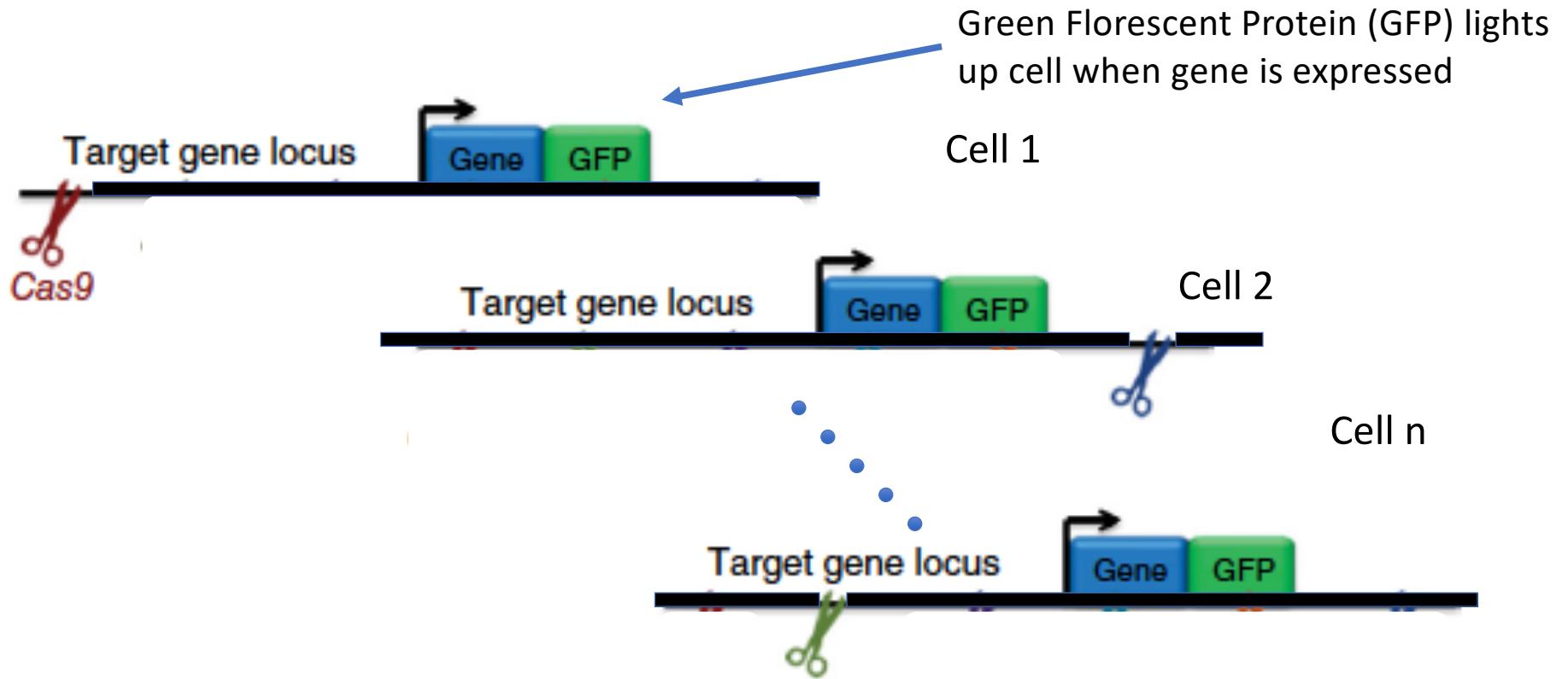


An annotation of potential Tdgf1 cis-regulation



Predictions of regulatory function based on indirect epigenomic measurements

Idea – break parts of the genome to see what is essential for gene expression



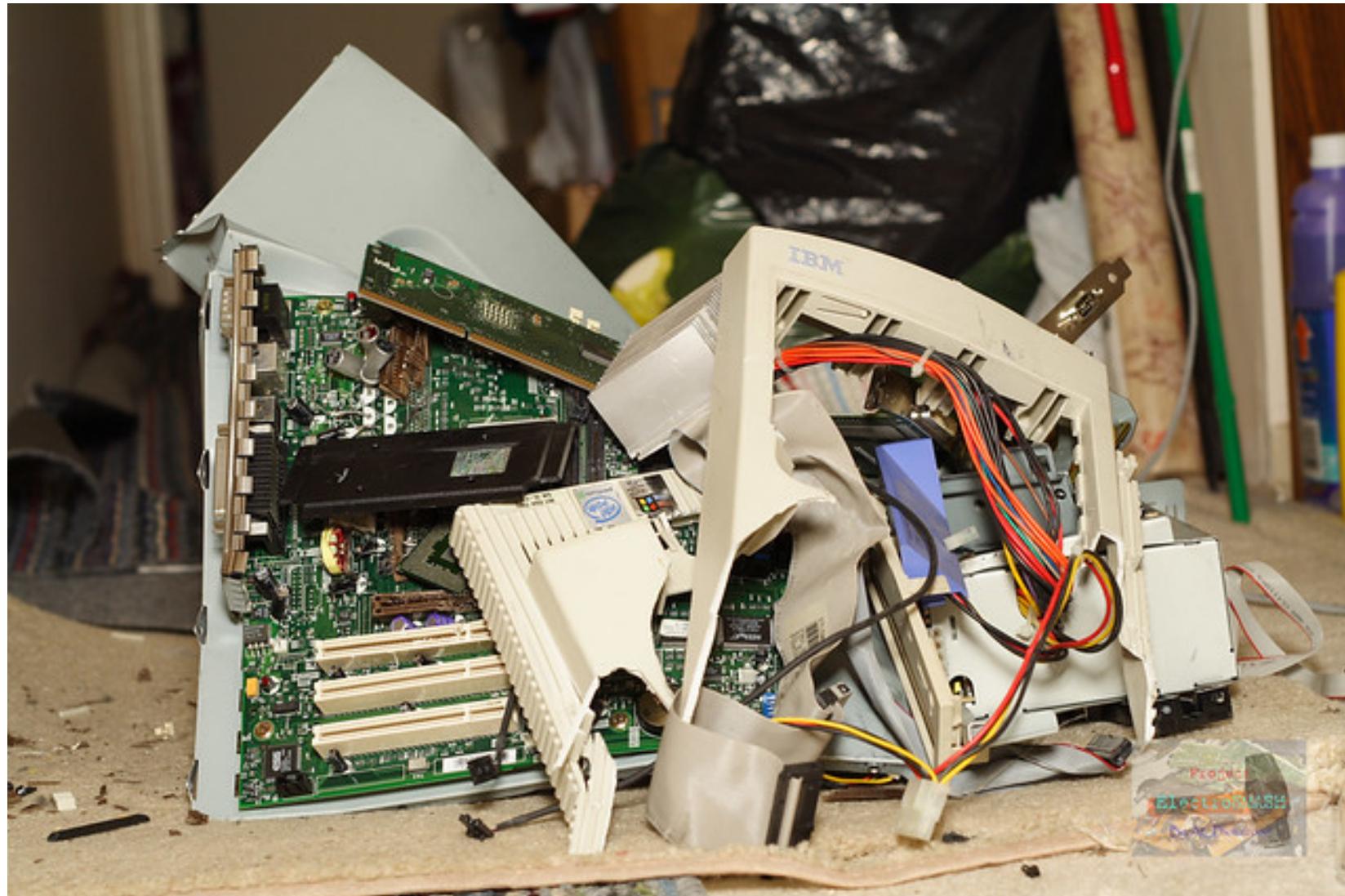
- Native context measurement
- High-throughput
- Directly observe expression of target gene via GFP
- Controlled delivery of only 1 gRNA per cell

Idea – determine parts of your computer
necessary for Zoom by breaking parts

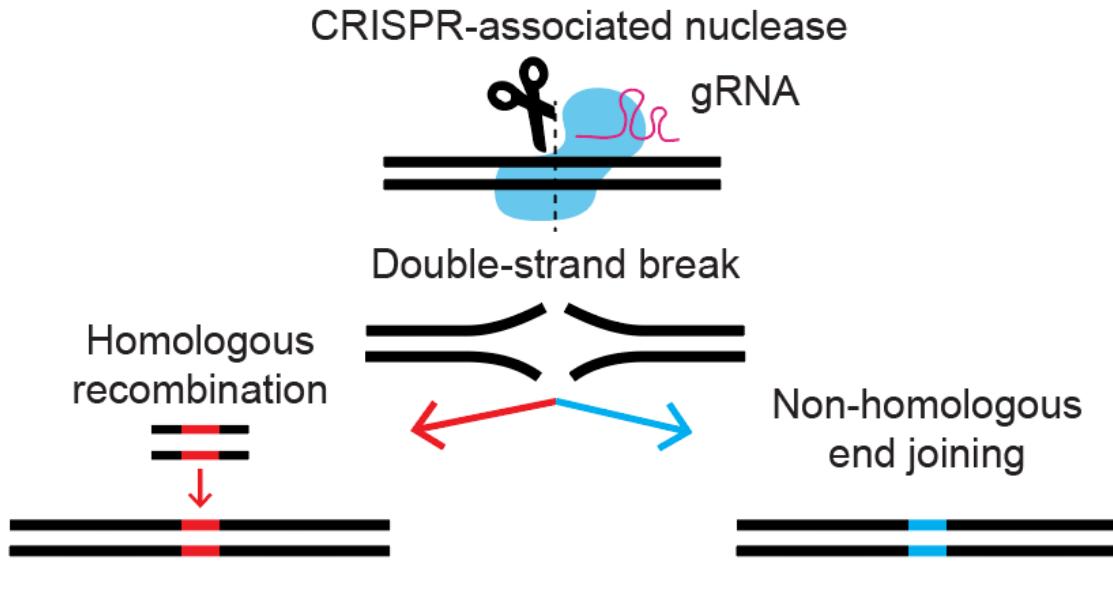
Idea – determine parts of your computer necessary for Zoom by breaking parts



Refinement – need fine grain resolution on what we break



We can use CRISPR genome editing to make localized genome alterations that are addressed by a guide RNA (gRNA)

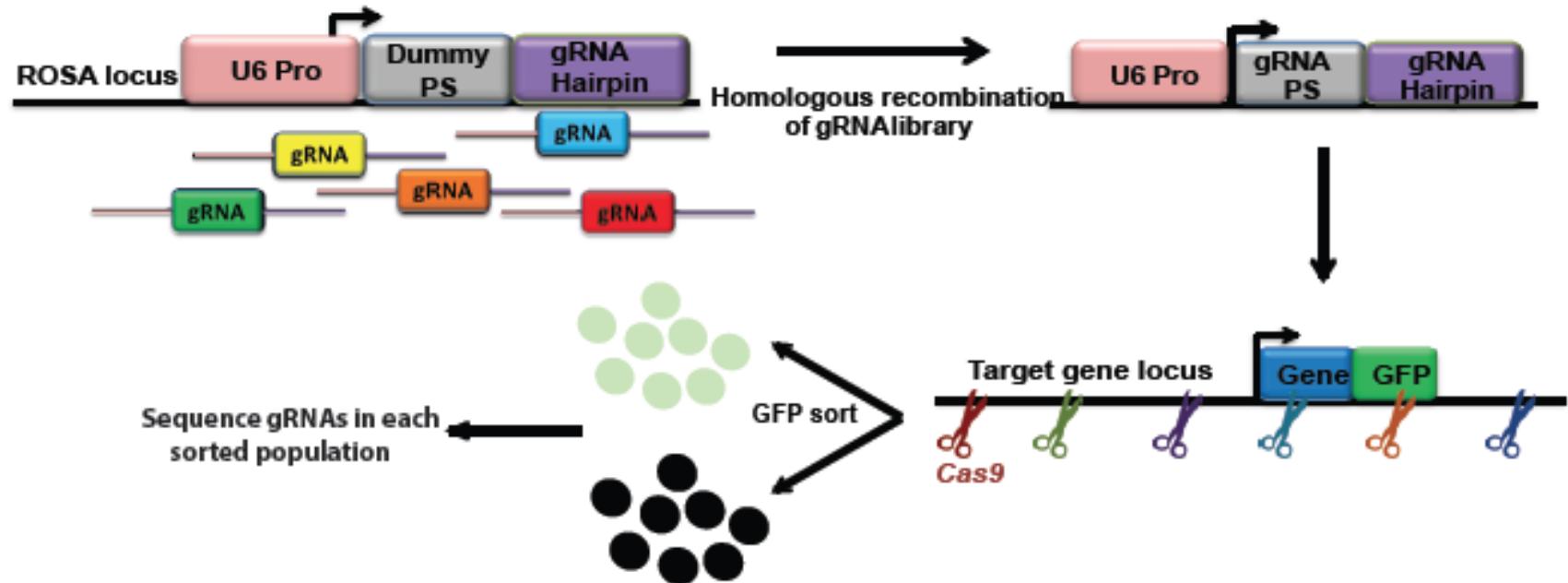


- often inefficient
- designable
- undesirable byproducts
- efficient
- random indels
- highly heterogeneous
- impractical beyond gene disruption
- “genome vandalism”

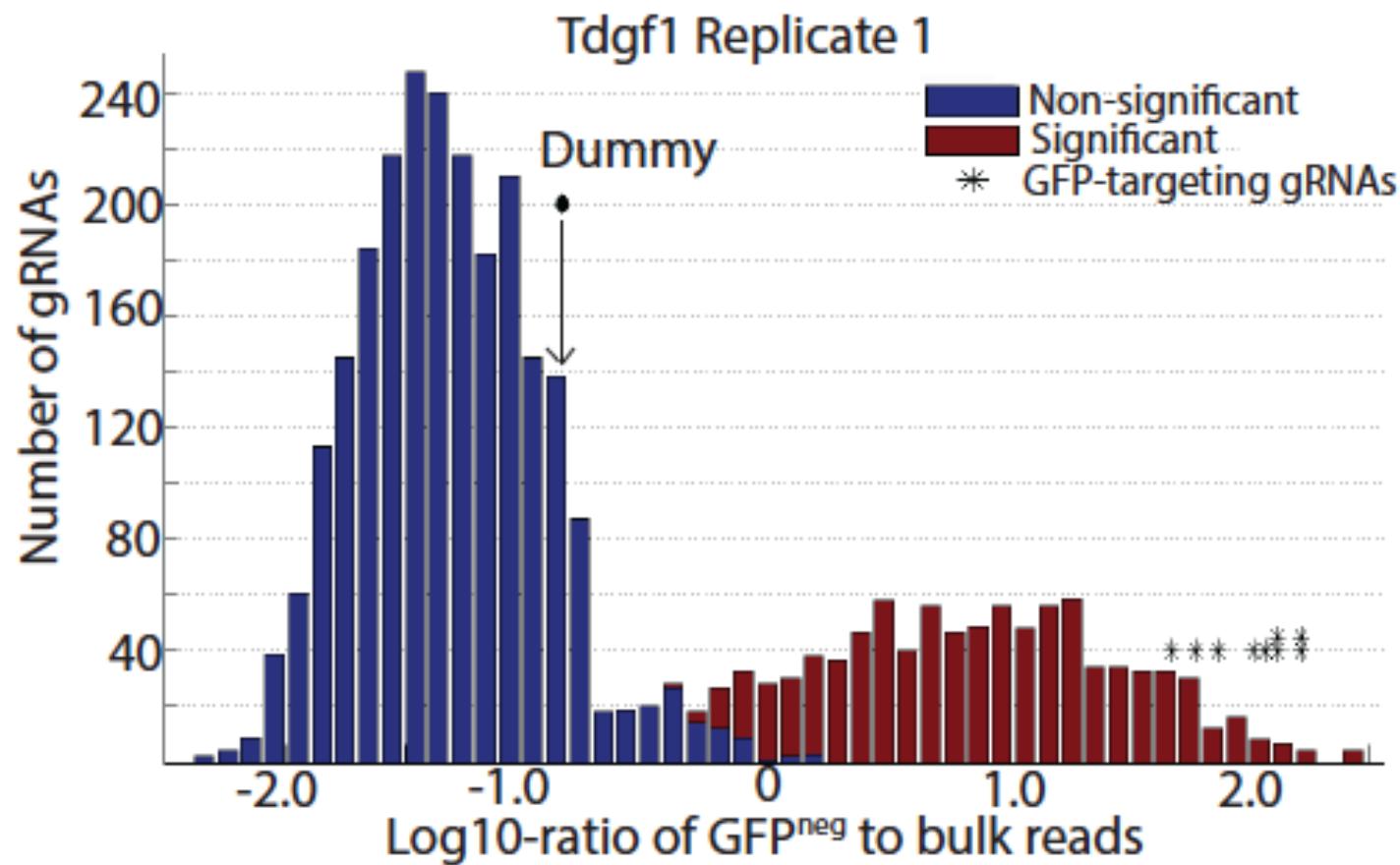
	reference
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 2986
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 2388
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 11060
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1961
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1956
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 465
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 281
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 3426
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 7266
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 26791
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1711
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1910
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 3673
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 972
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 646
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 4192
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1278
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 435
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 7388
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 2650
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 234
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 4588
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 878
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 745
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 933
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 5200
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 754
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1920
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1682
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 935
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 5347
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 3129
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1852
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1254
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 492
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 28310
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 28216
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 116
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 4485
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 2305
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1896
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1533
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 377
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 194
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 175
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 188
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 426
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 357
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 4572
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 484
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1136
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 488
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1929
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 458
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 3700
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 610
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 131
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 254
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 218
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 320
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 194
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 1354
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 308
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 720
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 351
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 178
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 652
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 3194
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 328
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCTTATAGGGGCCTAAATGT 149

Multiplexed Editing Regulatory Assay (MERA) experimental flow:

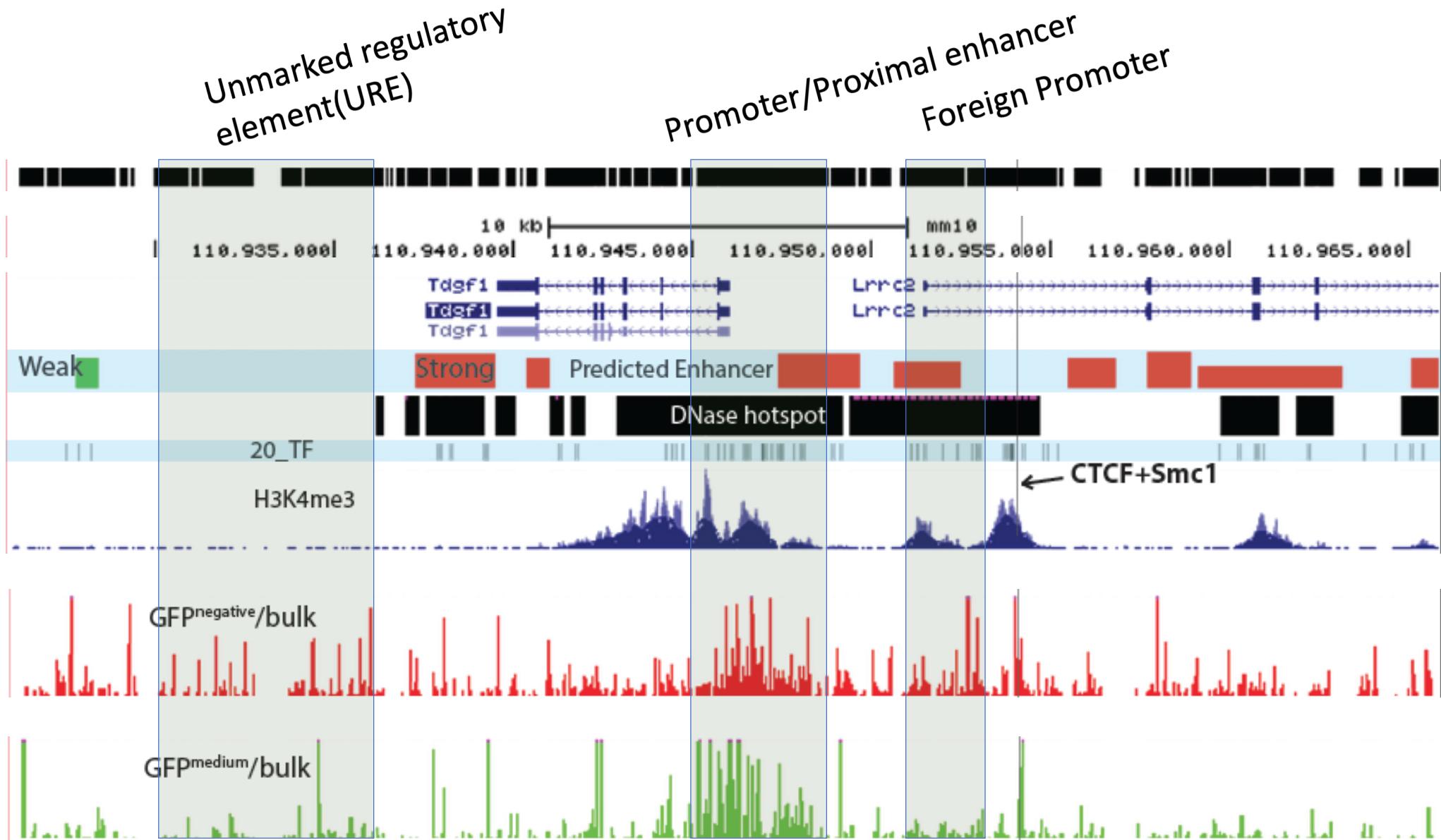
1. Put one gRNA in each cell that targets a location of interest
2. Use CRISPR to ablate the respective location in each cell
3. Sort cells by expression of GFP
4. Sequence gRNAs in each population to determine what locations are necessary, what locations are not necessary for GFP expression



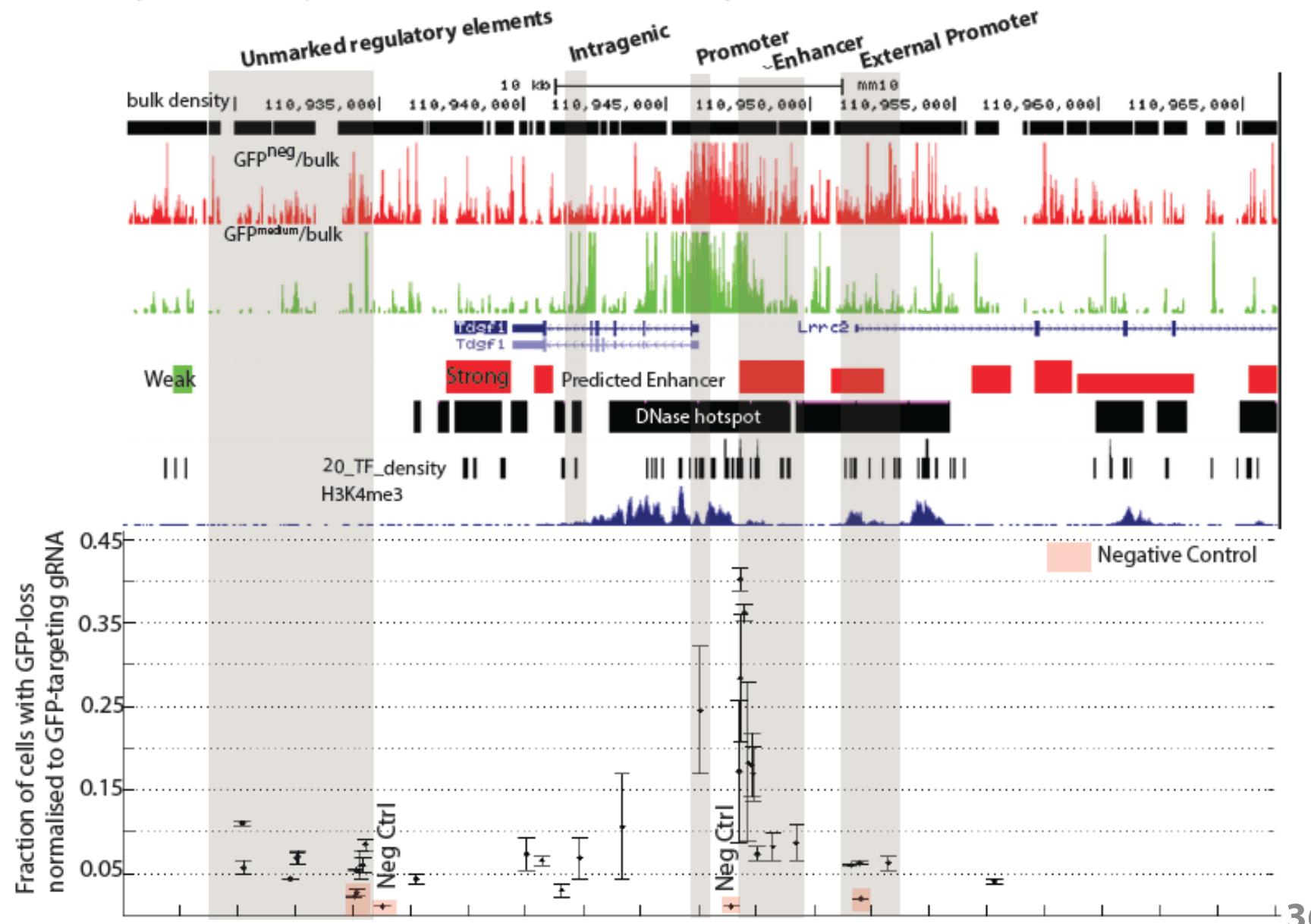
Distribution of the \log_{10} ratio of GFP^{neg} to bulk reads for all integrated gRNAs for Tdgf1



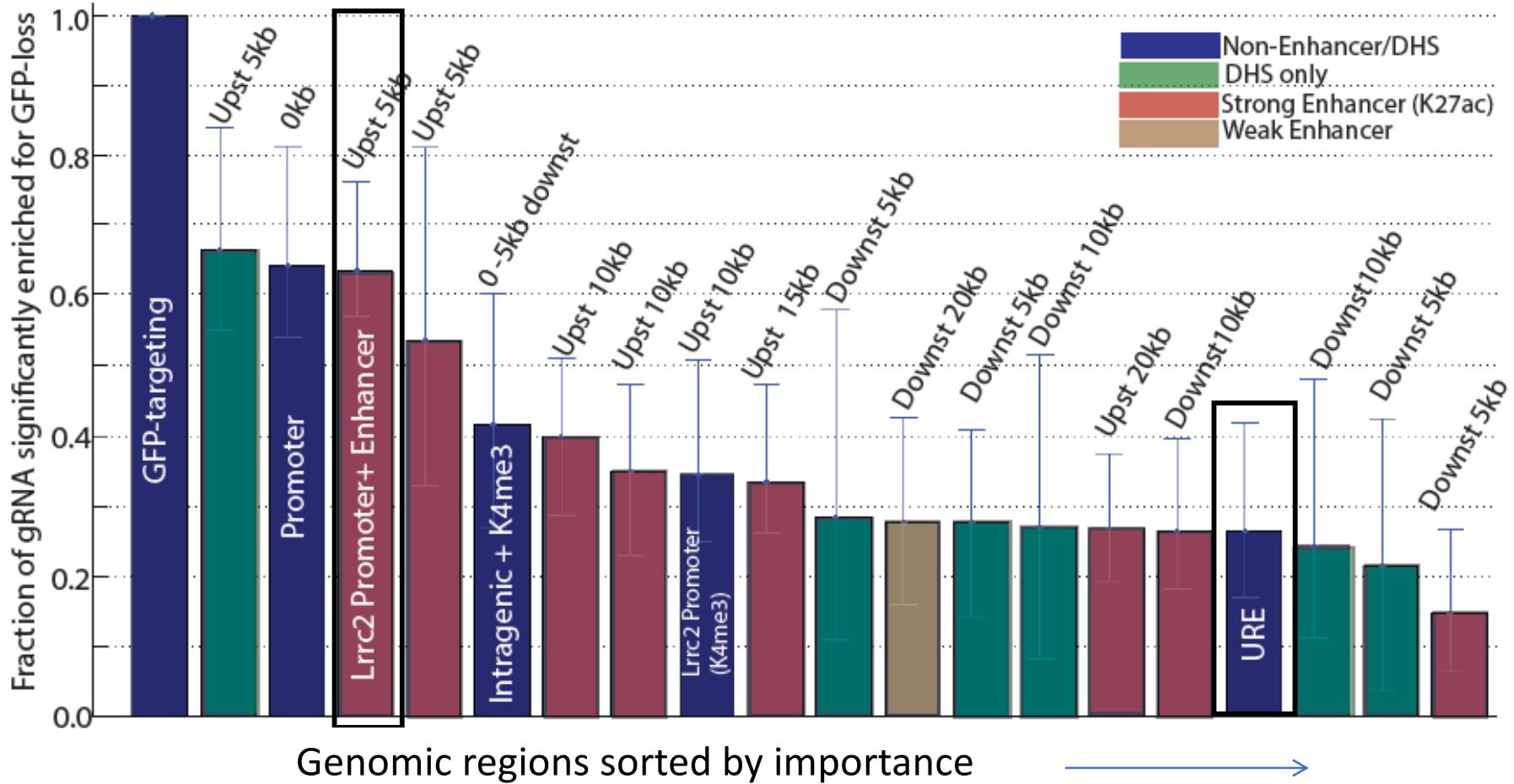
MERA enables systematic identification of required cis-regulatory elements for *Tdgf1*



Testing of individual gRNAs supports required cis-regulatory elements for Tdgf1

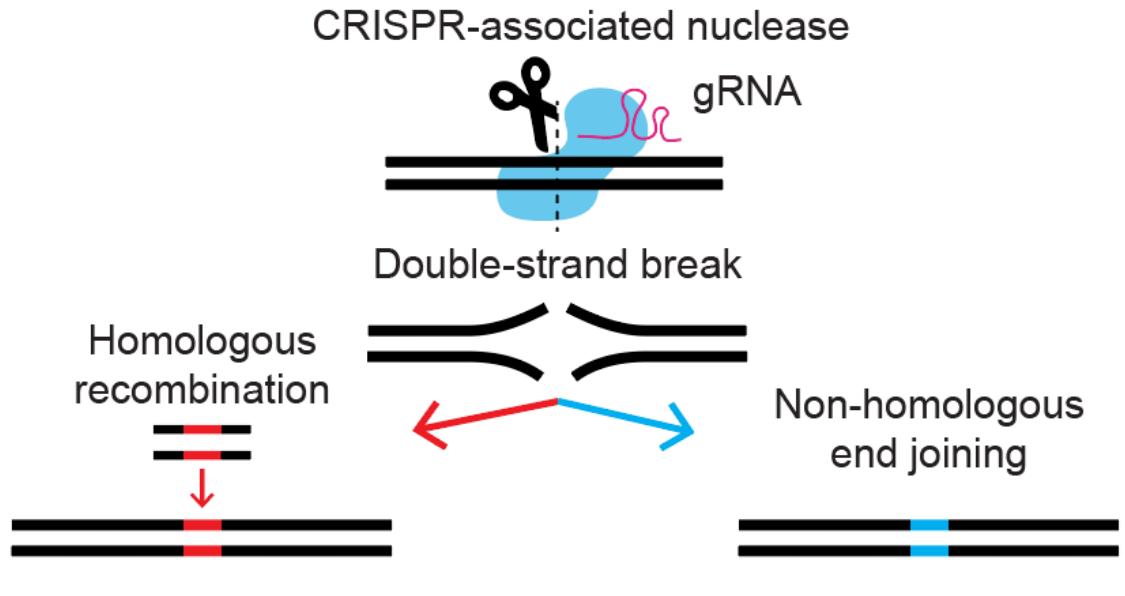


Necessary genome goes beyond known annotations (Tdgf1)



How can we predict the
genotypes of on target CRISPR
cuts?

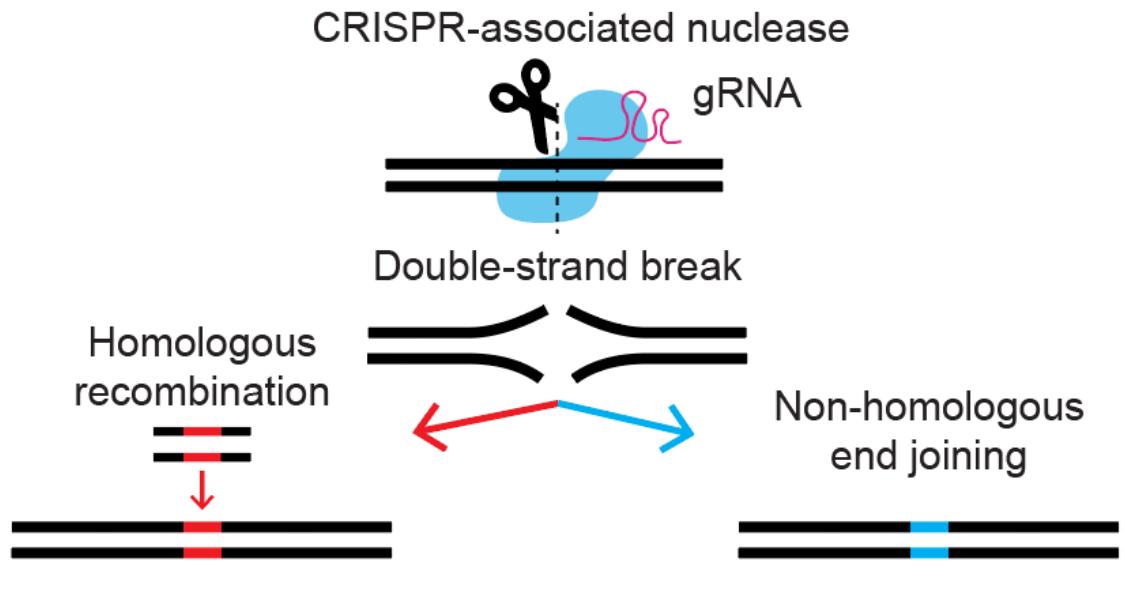
The state of CRISPR genome editing



- often inefficient
- designable
- undesirable byproducts
- efficient
- random indels
- highly heterogeneous
- impractical beyond gene disruption
- “genome vandalism”

	reference
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCGTTATAGGTGGCCTAAATGT 2986
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCGGCCGTTATAGGTGGCCTAAATGT 2388
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCGTTATAGGTGGCCTAAATGT 11060
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TCGGCCGTTATAGGTGGCCTAAATGT 1961
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCGTTATAGGTGGCCTAAATGT 1956
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CA-CGGCCGTTATAGGTGGCCTAAATGT 465
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGTACATCGGCCGTTATAGGTGGCCTAAATGT 281
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAGGTGGCCTAAATGT 3426
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAGGTGGCCTAAATGT 7266
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCGGCCGTTATAGGTGGCCTAAATGT 26791
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAGGTGGCCTAAATGT 1711
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCGGCCGTTATAGGTGGCCTAAATGT 1910
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAGGTGGCCTAAATGT 3673
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAGGTGGCCTAAATGT 972
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAGGTGGCCTAAATGT 646
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TACATCGGCCGTTATAGGTGGCCTAAATGT 4192
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAGGTGGCCTAAATGT 1278
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGTTATAAGGTGGCCTAAATGT 435
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GCCGTTATAAGGTGGCCTAAATGT 7388
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGGTGGCCTAAATGT 2650
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGGTGGCCTAAATGT 234
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGGTGGCCTAAATGT 4588
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GCCTGTTATAAGGTGGCCTAAATGT 878
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAGGTGGCCTAAATGT 745
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGGTGGCCTAAATGT 933
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TATAGGTGGCCTAAATGT 5200
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGGTGGCCTAAATGT 754
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGGTTGGCCTAAATGT 1920
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGTTATAAGGTGGCCTAAATGT 1682
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGGTGGCCTAAATGT 935
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TATAAGGTGGCCTAAATGT 5347
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGCTGTTATAAGGTGGCCTAAATGT 3129
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATATCGGCCGTTATAAGGTGGCCTAAATGT 1852
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TAGGTGGCCTAAATGT 1254
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT 492
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGGTGGCCTAAATGT 28310
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GTTAGGTGGCCTAAATGT 28216
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GTTAGGTGGCCTAAATGT 116
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGGTGGCCTAAATGT 4485
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT 2305
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAGGTGGCCTAAATGT 1896
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGCTGGCCTAAATGT 1533
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CCATAATGT 377
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT 194
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGGTGGCCTAAATGT 175
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT 188
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TATAGGT 426
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT 357
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGGTGGCCTAAATGT 4572
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TGT 1136
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CCTAAATGT 484
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GT 488
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TGGCCTAAATGT 1929
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGGTGGCCTAAATGT 458
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TGGCCTAAATGT 3700
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT 610
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGGTGGCCTAAATGT 131
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGGTGGCCTAAATGT 254
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAATGT 218
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CCTAAATGT 320
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CTAAATGT 458
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAAGGTGGCCTAAATGT 194
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAAGGTGGCCTAAATGT 354
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AAATGT 720
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AAATGT 351
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT 53056
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GT 652
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CCGTTATAAGGTGGCCTAAATGT 3194
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GT 328
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATGT 149

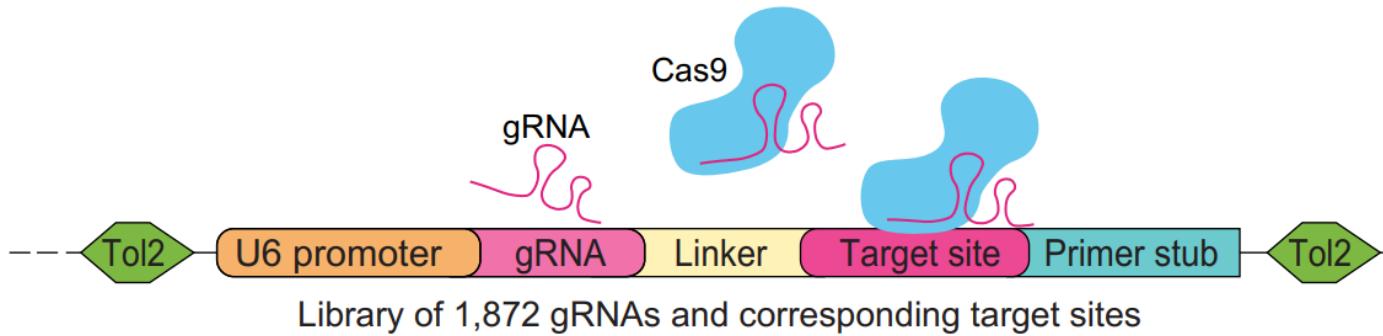
The state of CRISPR genome editing



- often inefficient
- designable
- predictable byproducts
- efficient
- predictable indels
- can be homogeneous
- practical: repair of pathogenic alleles to wild-type
- “genome art”

CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCGTTATAGGTGGCCTAAATGT	reference
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCGGCCGTTATAGGTGGCCTAAATGT	2986
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCGTTATAGGTGGCCTAAATGT	2388
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TCGGCCGTTATAGGTGGCCTAAATGT	11060
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCGTTATAGGTGGCCTAAATGT	1961
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CA-CGGCCGTTATAGGTGGCCTAAATGT	1956
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGATCGGCCGTTATAGGTGGCCTAAATGT	465
CGTCATAGATTTGGATATGGGCCATGTAGTA	-G-GCGCGTTATAGGTGGCCTAAATGT	281
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAGGTGGCCTAAATGT	3426
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCGGCCGTTATAGGTGGCCTAAATGT	7266
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCGGCCGTTATAGGTGGCCTAAATGT	26791
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCGGCCGTTATAGGTGGCCTAAATGT	17111
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCGGCCGTTATAGGTGGCCTAAATGT	1910
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCGGCCGTTATAGGTGGCCTAAATGT	3673
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGGCGTTATAGGTGGCCTAAATGT	972
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAGGTGGCCTAAATGT	646
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TATAGGTGGCCTAAATGT	4192
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAGGTGGCCTAAATGT	1278
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGTATAAGGTGGCCTAAATGT	435
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GCGCTTATAGGTGGCCTAAATGT	7388
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGGCGTTATAGGTGGCCTAAATGT	2650
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAGGTGGCCTAAATGT	234
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGGCGTTATAGGTGGCCTAAATGT	4588
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GCGCTTATAGGTGGCCTAAATGT	878
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAGGTGGCCTAAATGT	745
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGGTGGCCTAAATGT	933
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAGGTGGCCTAAATGT	5200
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCGGCCGTTATAGGTGGCCTAAATGT	754
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGGTGGCCTAAATGT	1920
CGTCATAGATTTGGATATGGGCCAA	-CGGTATAAGGTGGCCTAAATGT	1682
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAGGTGGCCTAAATGT	935
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TATAGGTGGCCTAAATGT	5347
CGTCATAGATTTGGATATGGGCC	-CGGTATAAGGTGGCCTAAATGT	3129
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCGGCCGTTATAGGTGGCCTAAATGT	1852
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TAGGTGGCCTAAATGT	1254
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	492
CGTCATAGATTTGGATATGGGCC	-CGGGCGTTATAGGTGGCCTAAATGT	28318
CGTCATAGATTTGGATATGGGCC	-GTTAGGTGGCCTAAATGT	28216
CGTCATAGATTTGGATATGGGCC	-GTTAGGTGGCCTAAATGT	116
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGGCGTTATAGGTGGCCTAAATGT	4485
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	2305
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAGGTGGCCTAAATGT	1896
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGCCTAAATGT	1533
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CCATAATGT	377
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	194
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGGTGGCCTAAATGT	175
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	188
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TATAGGT	426
CGTCATAGATTTGGATATGGGCC	-GGTGGCCTAAATGT	357
CGTCATAGATTTGGATATGGGCC	-AGGTGGCCTAAATGT	4572
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TG	1136
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CCTAAATGT	484
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GT	488
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TGGCCTAAATGT	1929
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGGTGGCCTAAATGT	458
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TGGCCTAAATGT	3700
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	610
CGTCATAGATTTGGATATGGGCC	-CGGGCGTTATAGGTGGCCTAAATGT	131
CGTCATAGATTTGGATATGGGCC	-CGGGCGTTATAGGTGGCCTAAATGT	254
CGTCATAGATTTGGATATGGGCC	-AATGT	218
CGTCATAGATTTGGATATGGGCC	-CCTAAATGT	320
CGTCATAGATTTGGATATGGGCC	-ATGT	458
CGTCATAGATTTGGATATGGGCC	-CTAAATGT	194
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAGGTGGCCTAAATGT	1354
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	308
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AATGT	720
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AATGT	351
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	53056
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GT	178
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CCGTTATAGGTGGCCTAAATGT	3194
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GT	652
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATGT	149

High-throughput genome-integrated assay of Cas9-mediated DNA repair

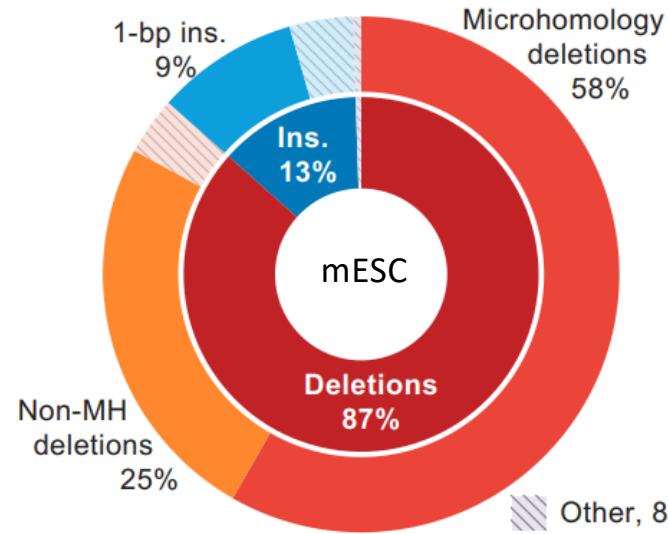


- 96 target sites in largest previous study
- Designed 1,872 target sites (55-bp) based on the human genome
- Observed 1,262 unique genotypes / target site

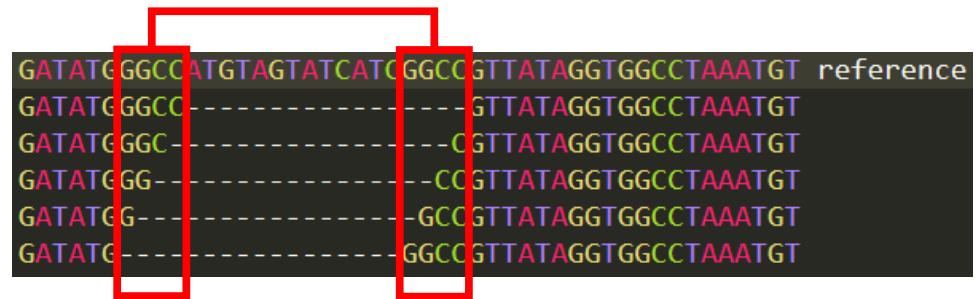
Target sequence	Count	Reference
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT reference
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 2986
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 2388
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 11069
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 1961
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 1956
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 465
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 281
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 3426
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 7266
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 26791
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 17111
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 1910
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 3673
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 972
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 646
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 4192
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 1278
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 435
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 7388
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 2650
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 234
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 4588
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 878
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 745
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 933
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 5200
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 754
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 1928
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 1682
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 5347
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 3129
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 1852
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 1254
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 492
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 28318
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 28216
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 116
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 4485
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 2305
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 1896
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 1533
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 377
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 194
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 175
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 188
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 181
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 426
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 4572
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 1136
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 484
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 480
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 1929
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 610
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 131
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 254
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 218
CATCGATAGATTGGATATGGGCCATGTAGTA	1	CATCGGCGGTATAAGGGGCCCTAAATGT 320

Cas9 editing without a homology template is predictable, can be precise, and can be practical for disease correction.

Cas9 primarily causes microhomology deletions in genome-integrated and endogenous settings

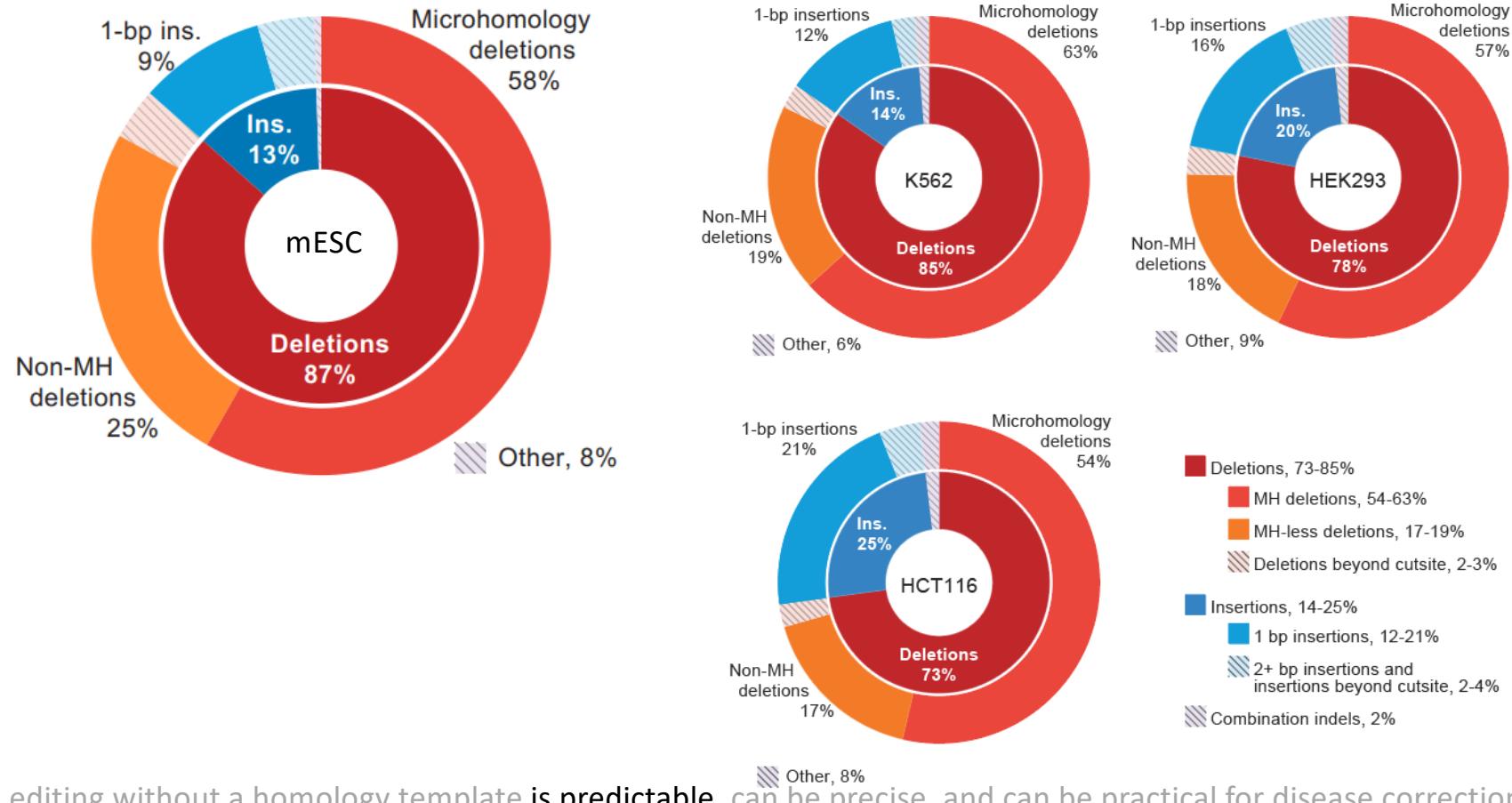


A microhomology deletion is a deletion with multiple equal-scoring alignments



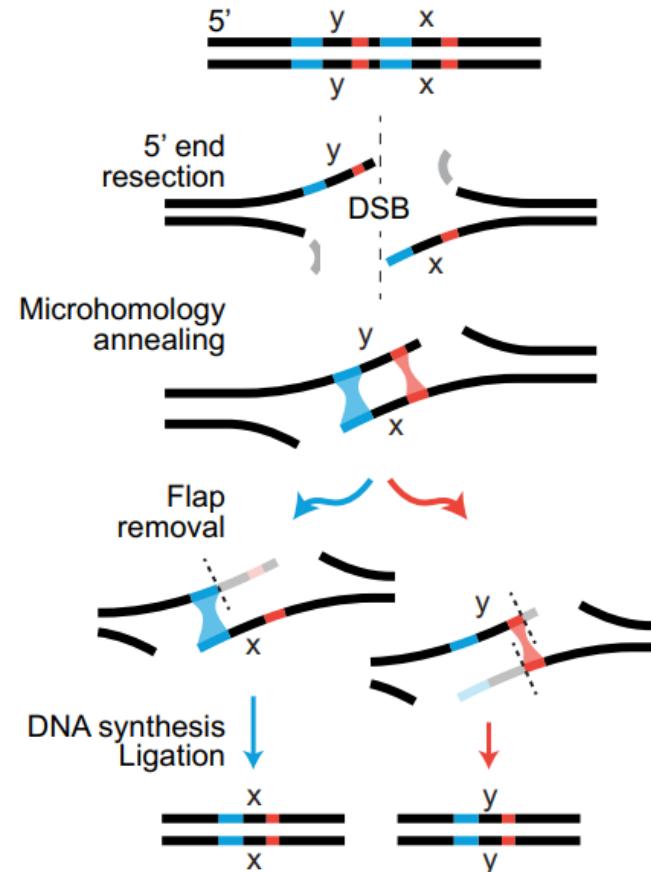
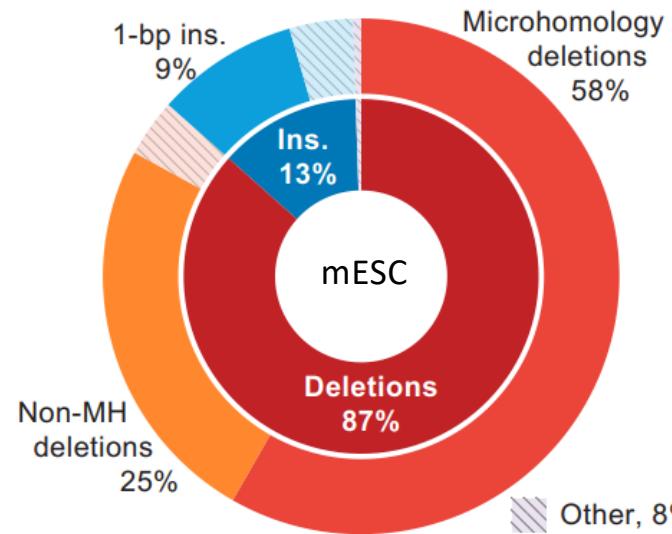
Cas9 editing without a homology template is predictable, can be precise, and can be practical for disease correction.

Cas9 primarily causes microhomology deletions in genome-integrated and endogenous settings



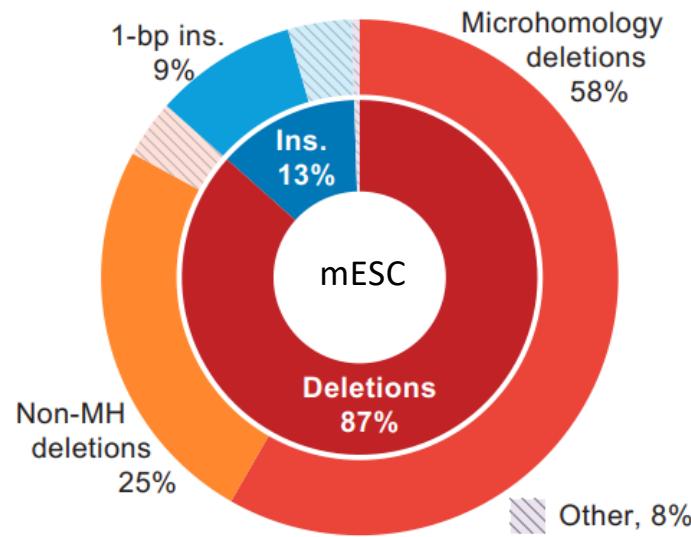
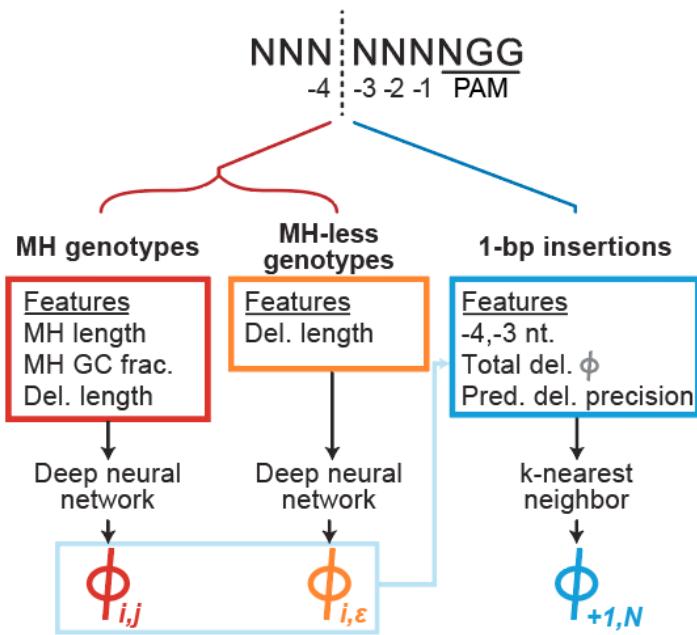
Cas9 editing without a homology template is predictable, can be precise, and can be practical for disease correction.

Majority of repair products arise from microhomology-mediated end-joining (MMEJ)

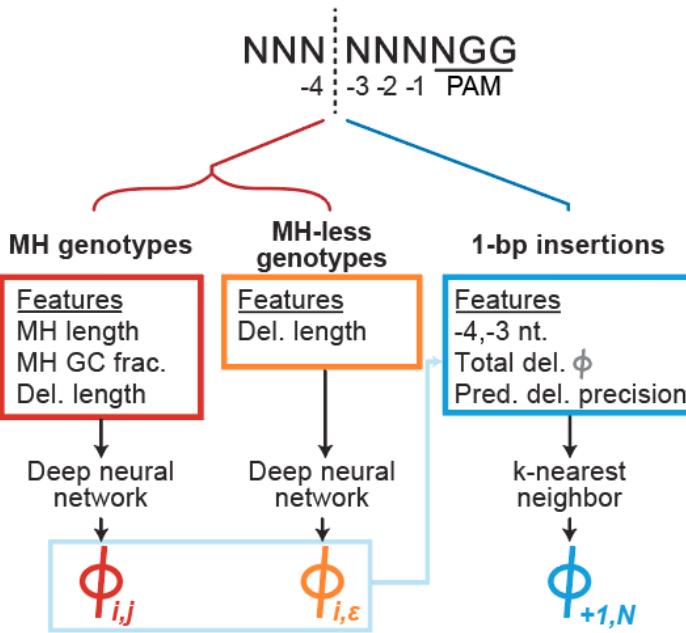


Cas9 editing without a homology template is predictable, can be precise, and can be practical for disease correction.

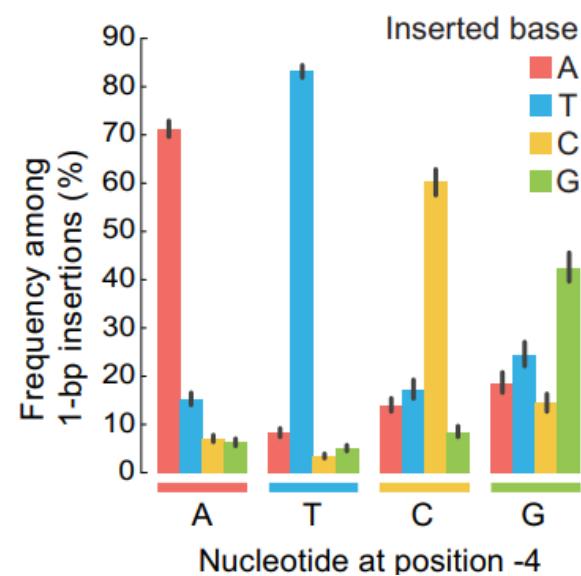
inDelphi predicts 90% of repair products from 3 major repair classes



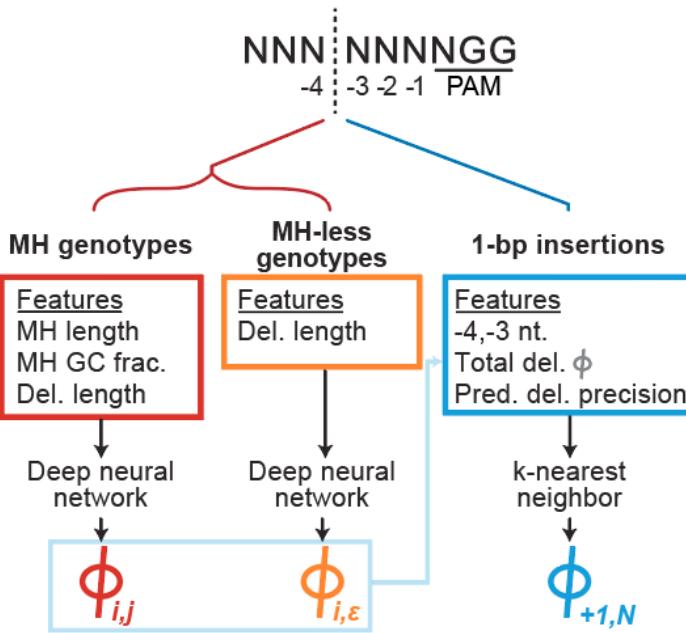
Cas9 editing without a homology template is predictable, can be precise, and can be practical for disease correction.



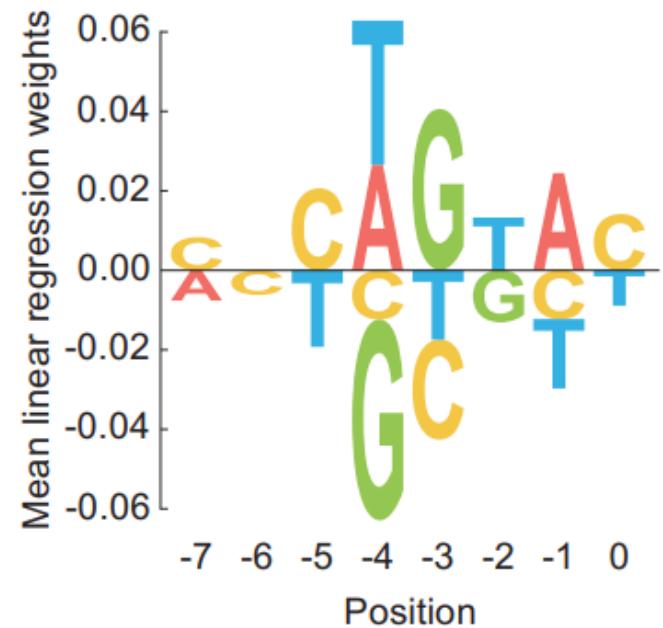
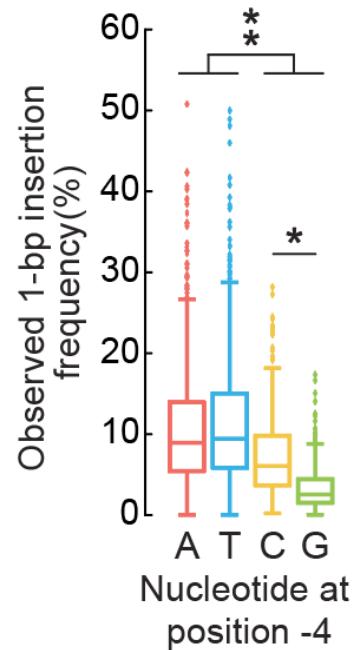
1-bp insertions copy the adjacent nucleotide



Cas9 editing without a homology template is predictable, can be precise, and can be practical for disease correction.

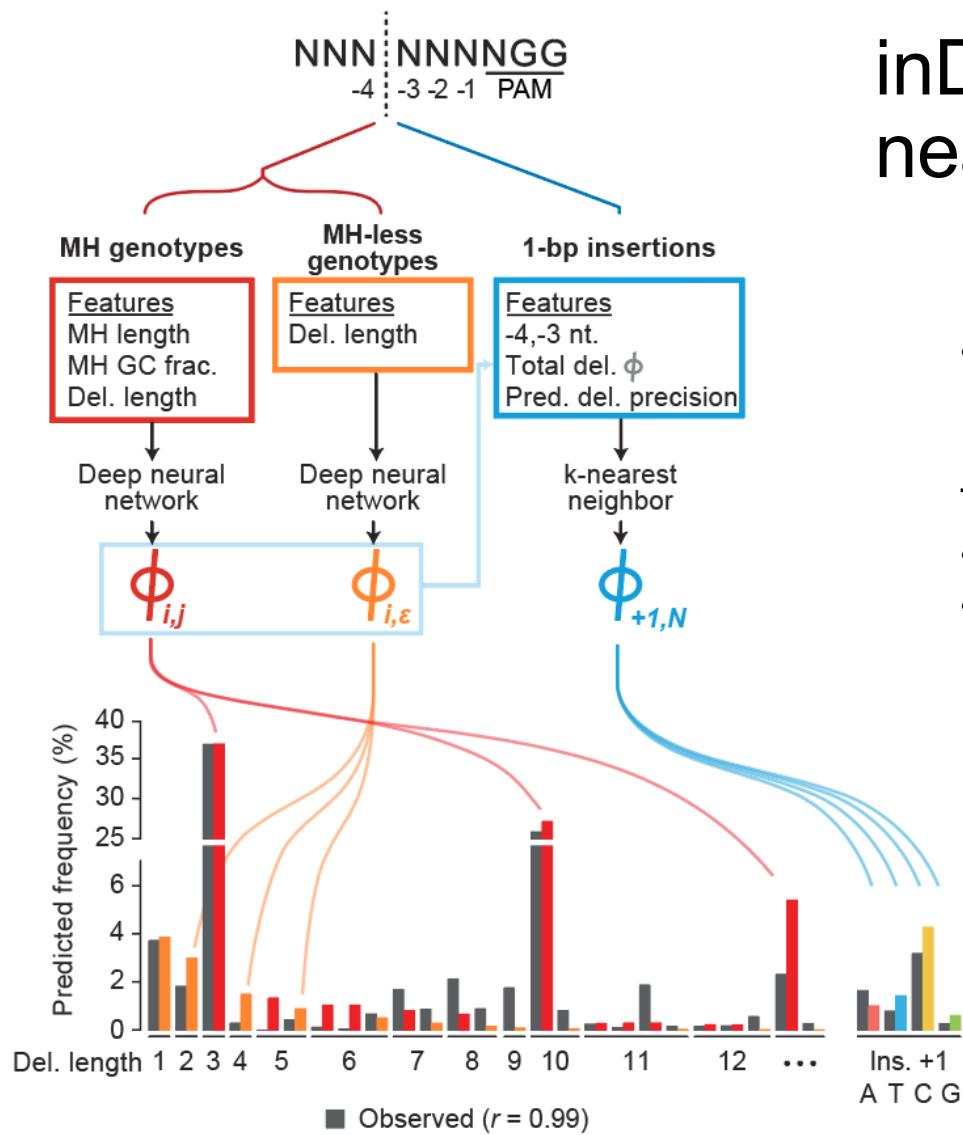


1-bp insertion frequency depends on local sequence context



Cas9 editing without a homology template is predictable, can be precise, and can be practical for disease correction.

inDelphi accurately predicts nearly all repair outcomes

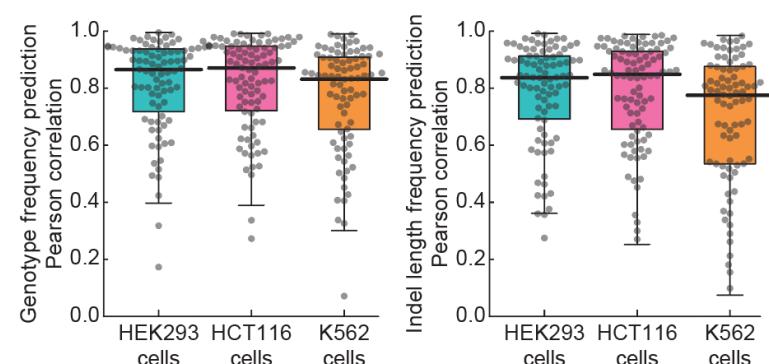


Input: Sequence, cutsite

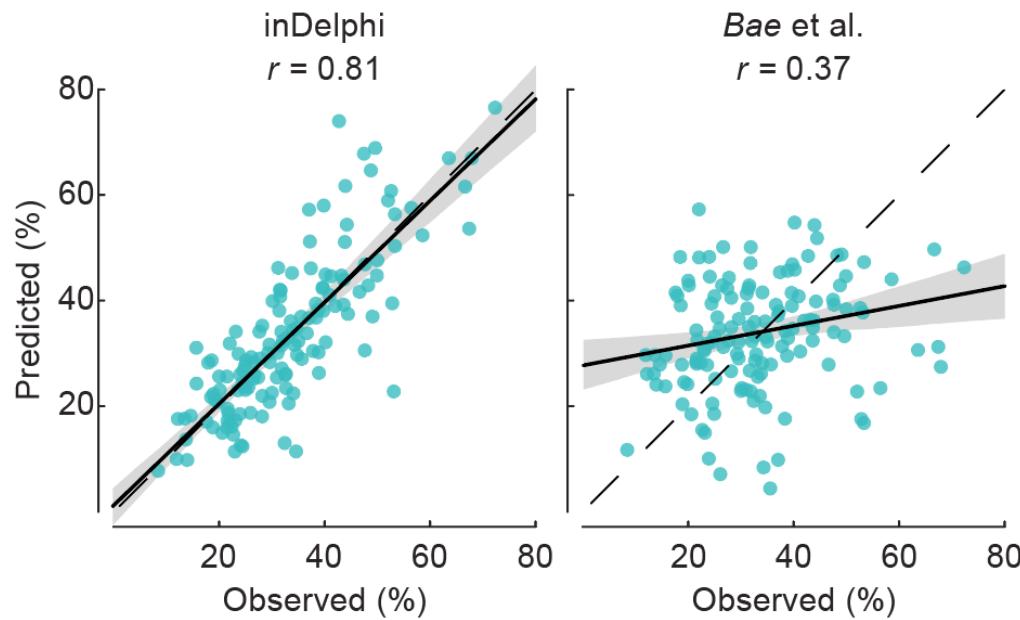
- Predicts 90% of observed repair outcomes
 - 70% at single-base resolution

Training & testing on held-out cell-types

- Median $r = 0.87$ on genotype prediction
- Median $r = 0.84$ on indel length prediction

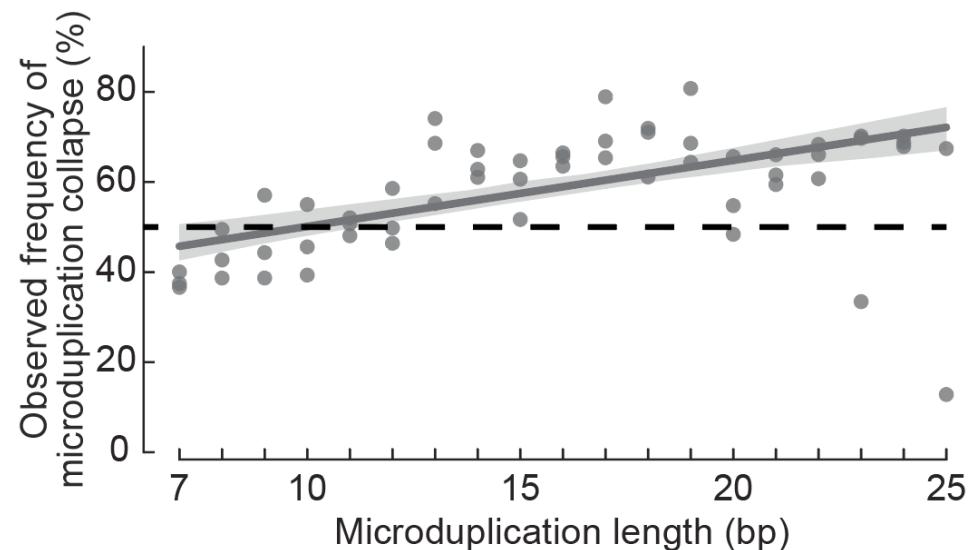
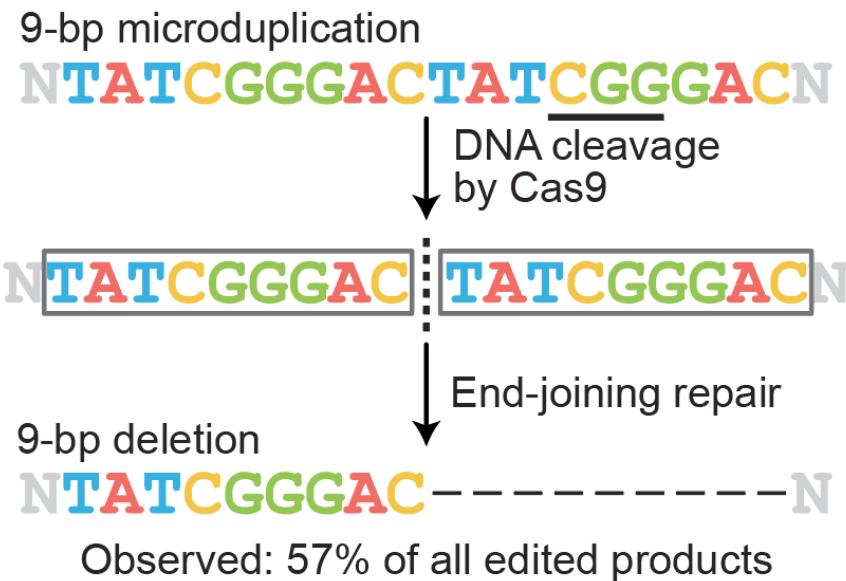


inDelphi accurately predicts frameshifts



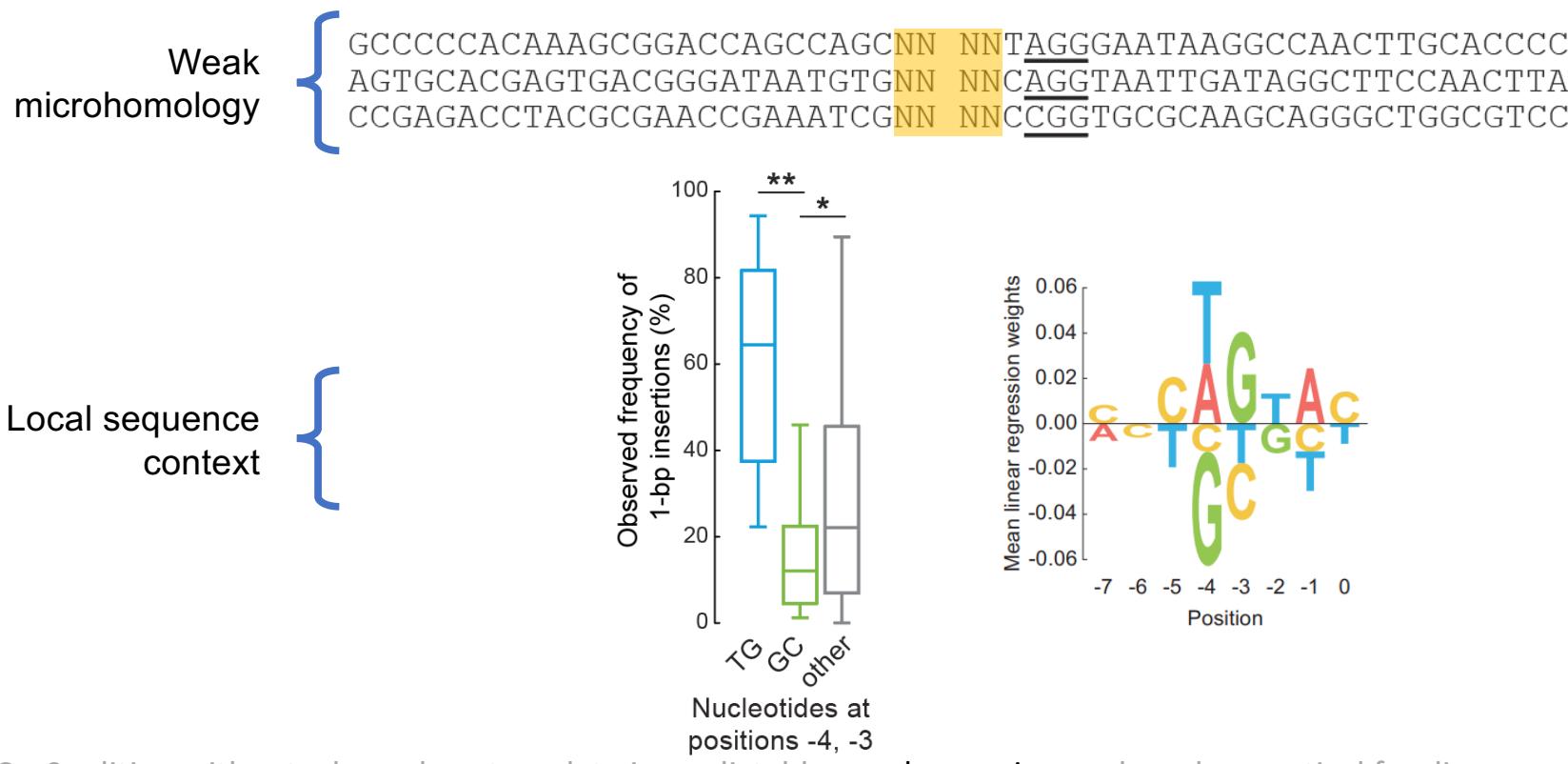
Cas9 editing without a homology template **is predictable**, can be precise, and can be practical for disease correction.

Target sites yielding a single deletion repair genotype >50% of the time



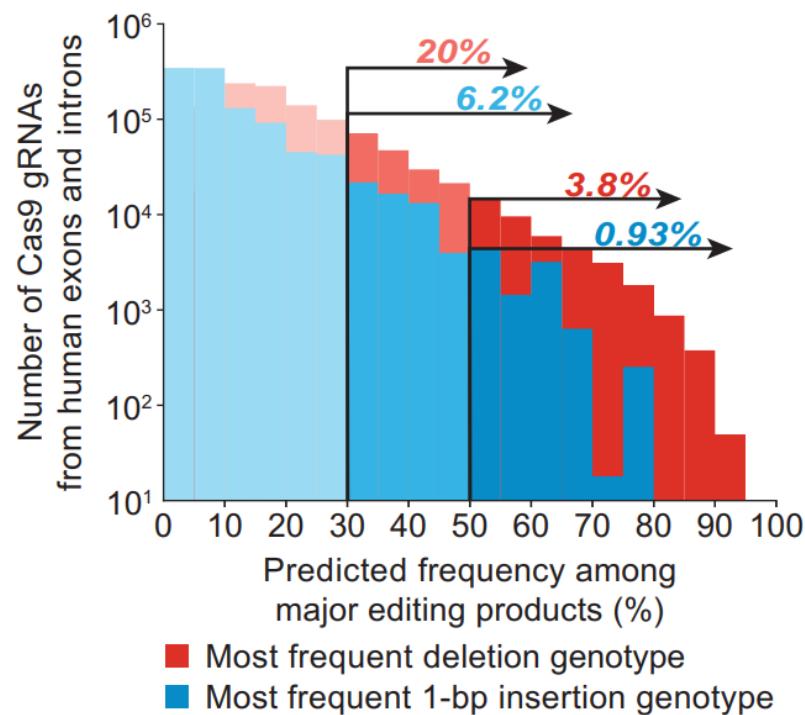
Cas9 editing without a homology template is predictable, can be precise, and can be practical for disease correction.

Target sites yielding a single insertion repair genotype >50% of the time



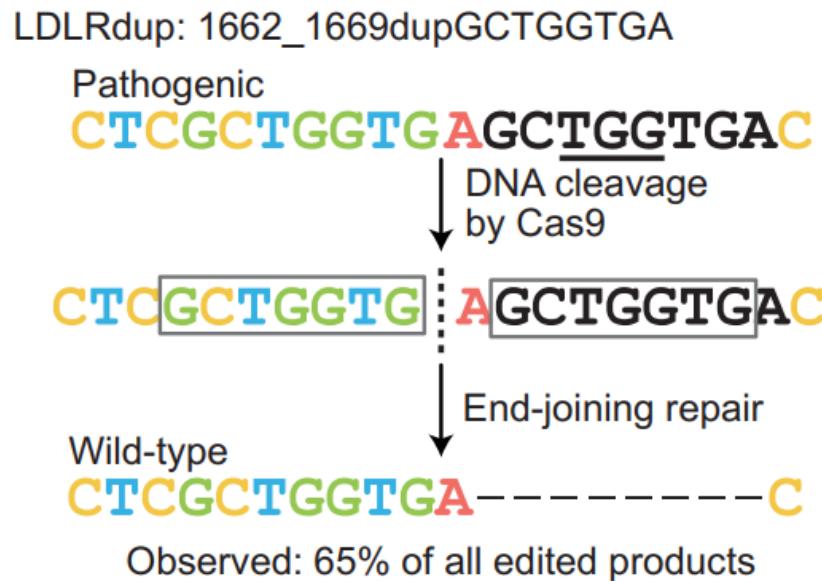
How can we use DNA cuts to restore function?

inDelphi predicts that 5% of gRNAs yield a single repair genotype the majority of the time



Cas9 editing without a homology template is predictable, can be precise, and can be practical for disease correction.

Pathogenic microduplications are efficiently repairable to wild-type with simple Cas9 cutting



Cas9 editing without a homology template is predictable, can be precise, and can be practical for disease correction.

inDelphi identified 183 pathogenic alleles corrected to wild-type at >50% frequency ($r = 0.64$)

23,018

- Clinvar and HGMD

1,592

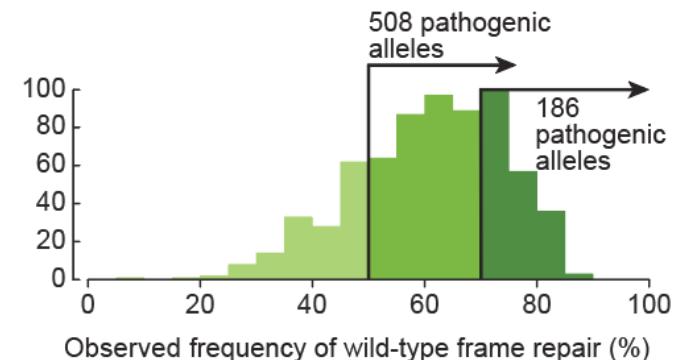
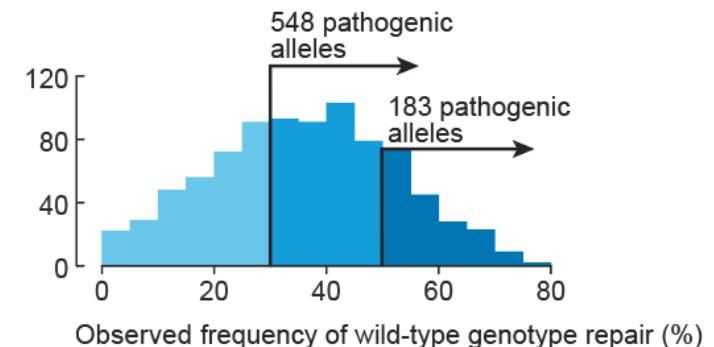
- Identified and designed with highest repair %

865

- Candidates for wild-type repair after quality filtering

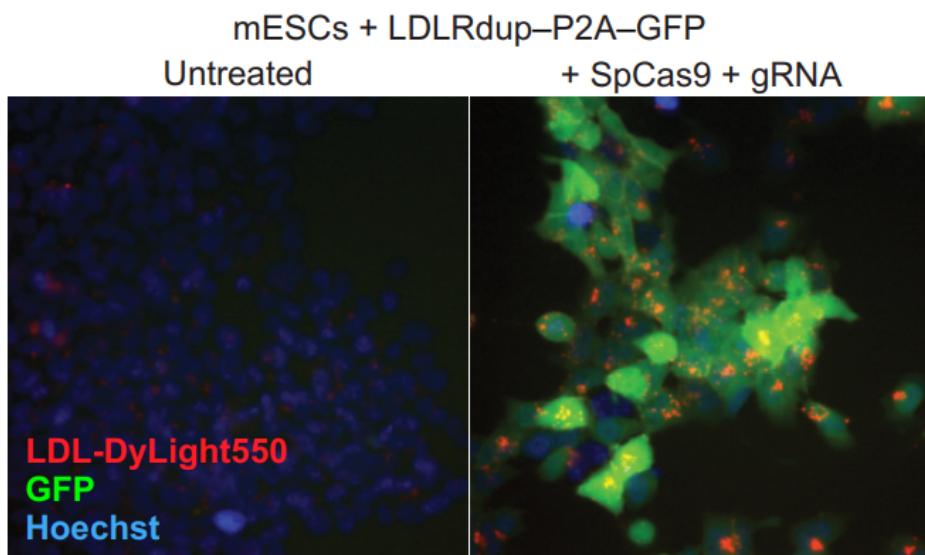
682

- Candidates for wild-type frame repair after quality filtering



Cas9 editing without a homology template is predictable, can be precise, and can be practical for disease correction.

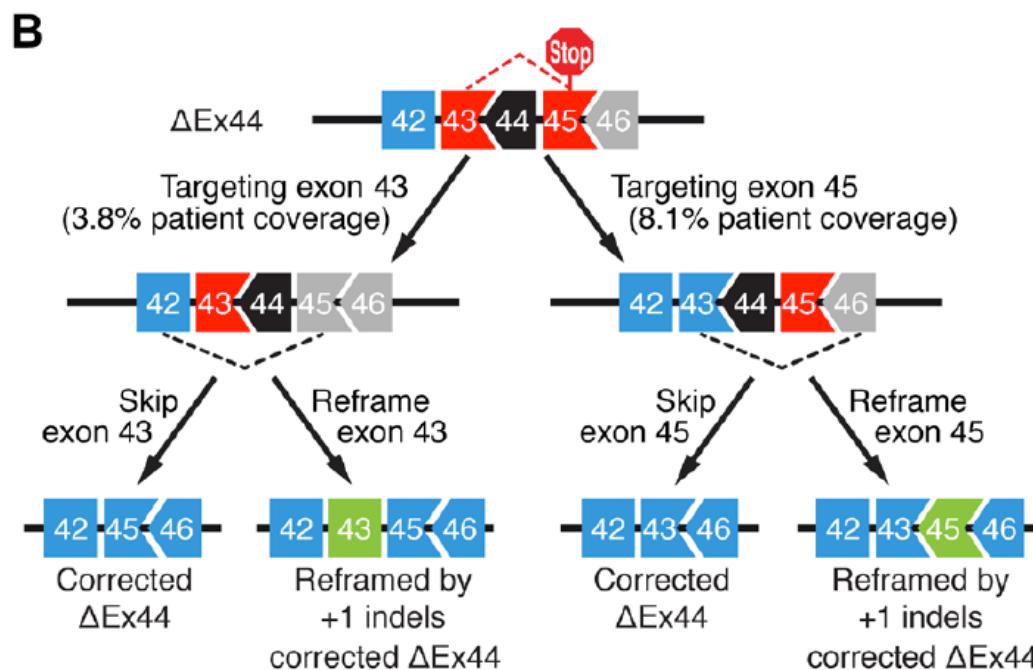
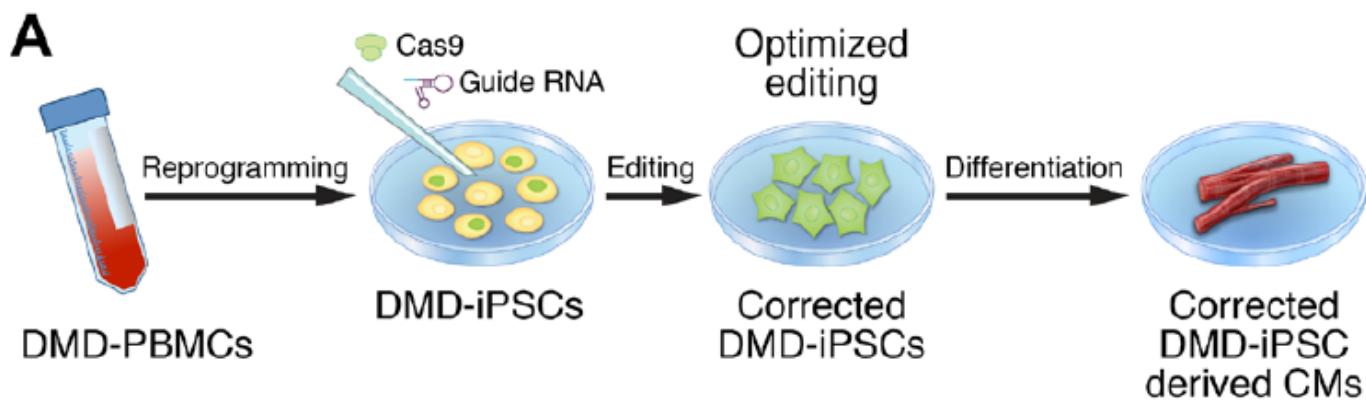
Efficient repair of pathogenic alleles to wild-type with template-free Cas9-nuclease treatment



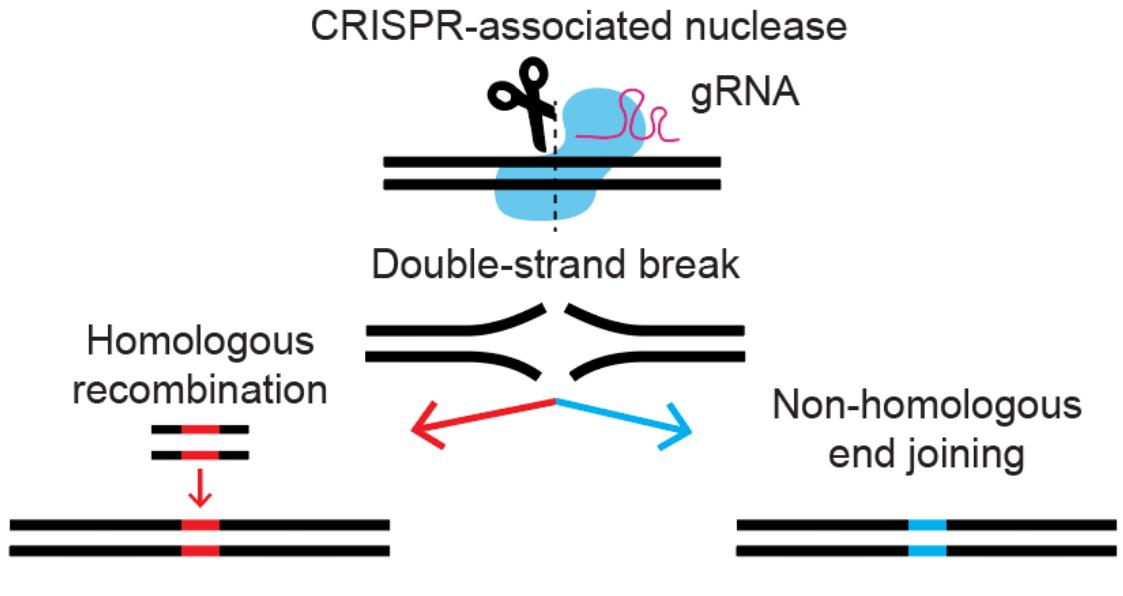
- Primary patient-derived fibroblasts
Human and mouse cell lines
- SpCas9 and SaCas9
- HPS1 71%
LDLR 77%
PORCN 48%
GAA 68%
GLB1 42%

Cas9 editing without a homology template is predictable, can be precise, and can be practical for disease correction.

Correcting Duchenne muscular dystrophy exon 44 deletion with CRISPR



The state of CRISPR genome editing



- often inefficient
 - designable
 - predictable byproducts
- efficient
 - predictable indels
 - can be homogeneous
 - practical: repair of pathogenic alleles to wild-type
 - “genome art”

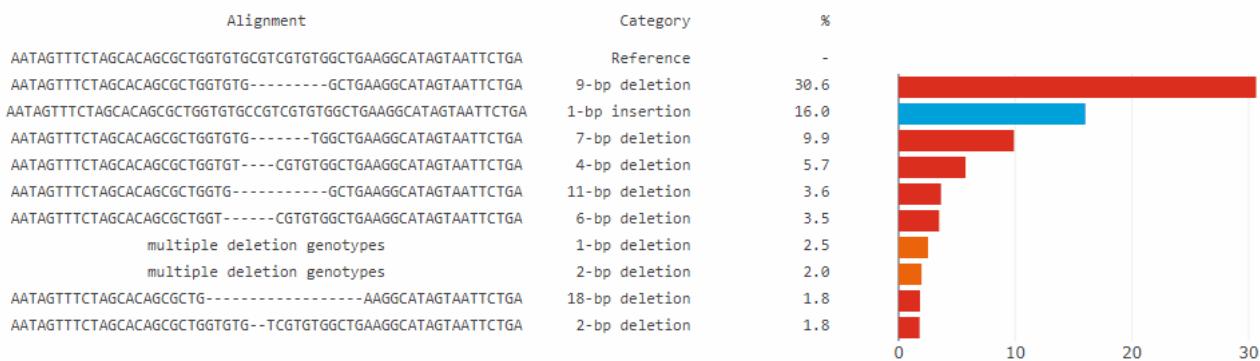
CGTCATAGATTTGGATATGGGCCATGTAGTA	CATCGGCCGTTATAAGTGGCCTAAATGT	reference
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCGGCCGTTATAAGTGGCCTAAATGT	2986
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGTGGCCTAAATGT	2388
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGTGGCCTAAATGT	11060
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TCGGCCGTTATAAGTGGCCTAAATGT	1961
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGTGGCCTAAATGT	1956
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CA-CGGCCGTTATAAGTGGCCTAAATGT	465
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGTATCGGCCGTTATAAGTGGCCTAAATGT	465
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGTGGCCTAAATGT	281
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGTGGCCTAAATGT	3426
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGTGGCCTAAATGT	7266
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCGGCCGTTATAAGTGGCCTAAATGT	26791
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGTGGCCTAAATGT	17111
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGTGGCCTAAATGT	1910
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGTGGCCTAAATGT	3673
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGTGGCCTAAATGT	972
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGTGGCCTAAATGT	646
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TACATCGGCCGTTATAAGTGGCCTAAATGT	4192
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGTGGCCTAAATGT	1278
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGTTATAAGTGGCCTAAATGT	435
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GCCGTTATAAGTGGCCTAAATGT	7388
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGTGGCCTAAATGT	2650
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGTGGCCTAAATGT	234
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGTGGCCTAAATGT	4588
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GCCTATAAGTGGCCTAAATGT	878
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAAGTGGCCTAAATGT	745
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGTGGCCTAAATGT	933
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAGTGGCCTAAATGT	5200
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGTGGCCTAAATGT	754
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGGGATGGCCTAAATGT	1920
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGTTATAAGTGGCCTAAATGT	1682
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CATCGGCCGTTATAAGTGGCCTAAATGT	935
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TATAAGTGGCCTAAATGT	5347
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGCTATAAGTGGCCTAAATGT	3129
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATCATCGGCCGTTATAAGTGGCCTAAATGT	1852
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TAGTGGCCTAAATGT	1254
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	492
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGTGGCCTAAATGT	28318
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GTTAGTGGCCTAAATGT	28216
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GTTAGTGGCCTAAATGT	116
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGTGGCCTAAATGT	4485
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	2305
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAAGTGGCCTAAATGT	1896
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGCTGGCCTAAATGT	1533
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CCATAATGT	377
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	194
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGGGGGCCTAAATGT	175
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	188
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GTTAGTGGCCTAAATGT	426
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	357
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGGGGGCCTAAATGT	4572
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TGT	1136
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CCTAAATGT	484
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GT	488
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TGGCCTAAATGT	1929
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AGGGGGCCTAAATGT	458
CGTCATAGATTTGGATATGGGCCATGTAGTA	-TGGCCTAAATGT	3700
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	610
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGTGGCCTAAATGT	131
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CGGCCGTTATAAGTGGCCTAAATGT	254
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AATAGT	218
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CCTAAATGT	320
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CTAAATGT	458
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAAGTGGCCTAAATGT	194
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATAAGTGGCCTAAATGT	1354
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AAATGT	308
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AAATGT	720
CGTCATAGATTTGGATATGGGCCATGTAGTA	-AAATGT	351
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GGTGGCCTAAATGT	53056
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GT	178
CGTCATAGATTTGGATATGGGCCATGTAGTA	-CCGTTATAAGTGGCCTAAATGT	3194
CGTCATAGATTTGGATATGGGCCATGTAGTA	-GT	652
CGTCATAGATTTGGATATGGGCCATGTAGTA	-ATGT	149

inDelphi

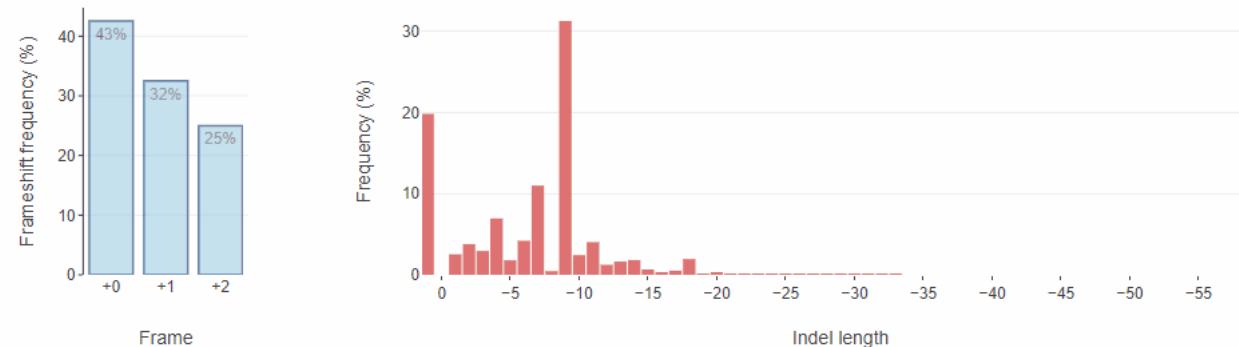
AATAGTTCTAGCACAGCGCTGGTGTGCGTCGTGGCTGAAGGCATAGTAATTCTGA GTCGTGTGGCTGAAGGCCATAGTAATTCTGA

◀ DSB ▶

Summary of predictions: Top 10 frequent events



Indel length predictions

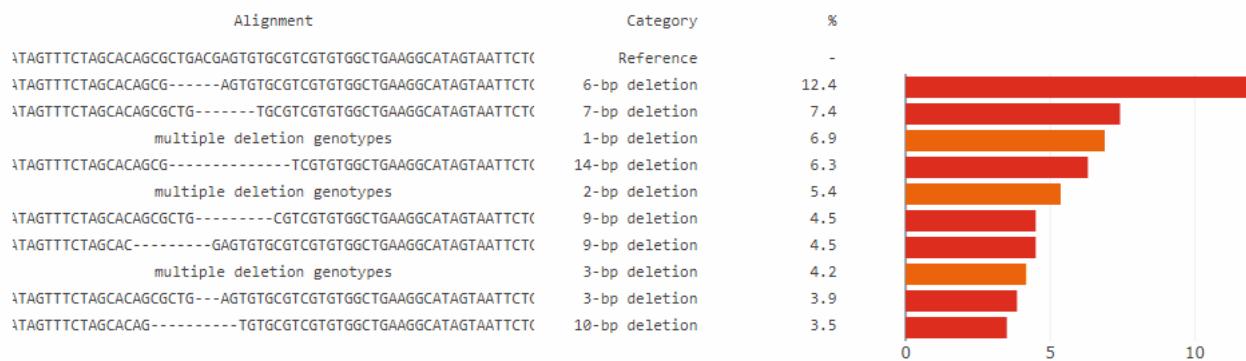


inDelphi

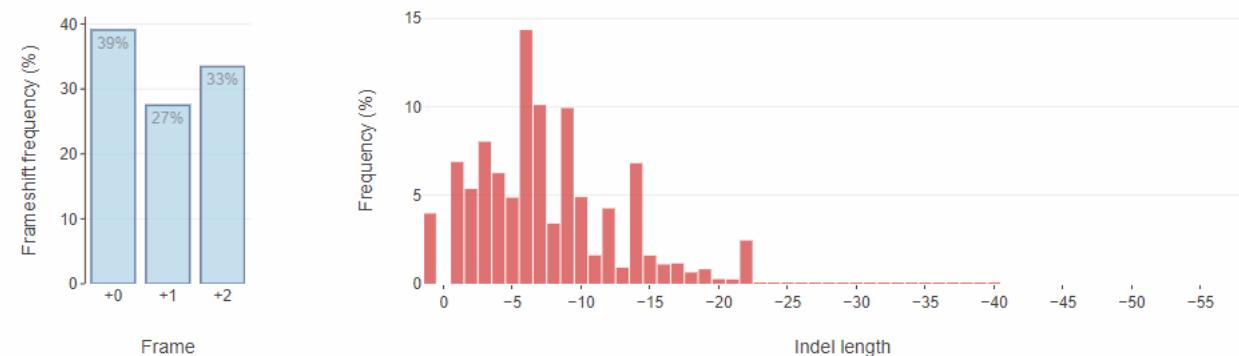
AATAGTTCTAGCACAGCGCTGACGAGTGTGCGTCGTGTGGCTGAAGGCATAGTAATTCTC ACGAGTGCGCTCGTGTGGCTGAAGGCATA

◀ DSB ▶

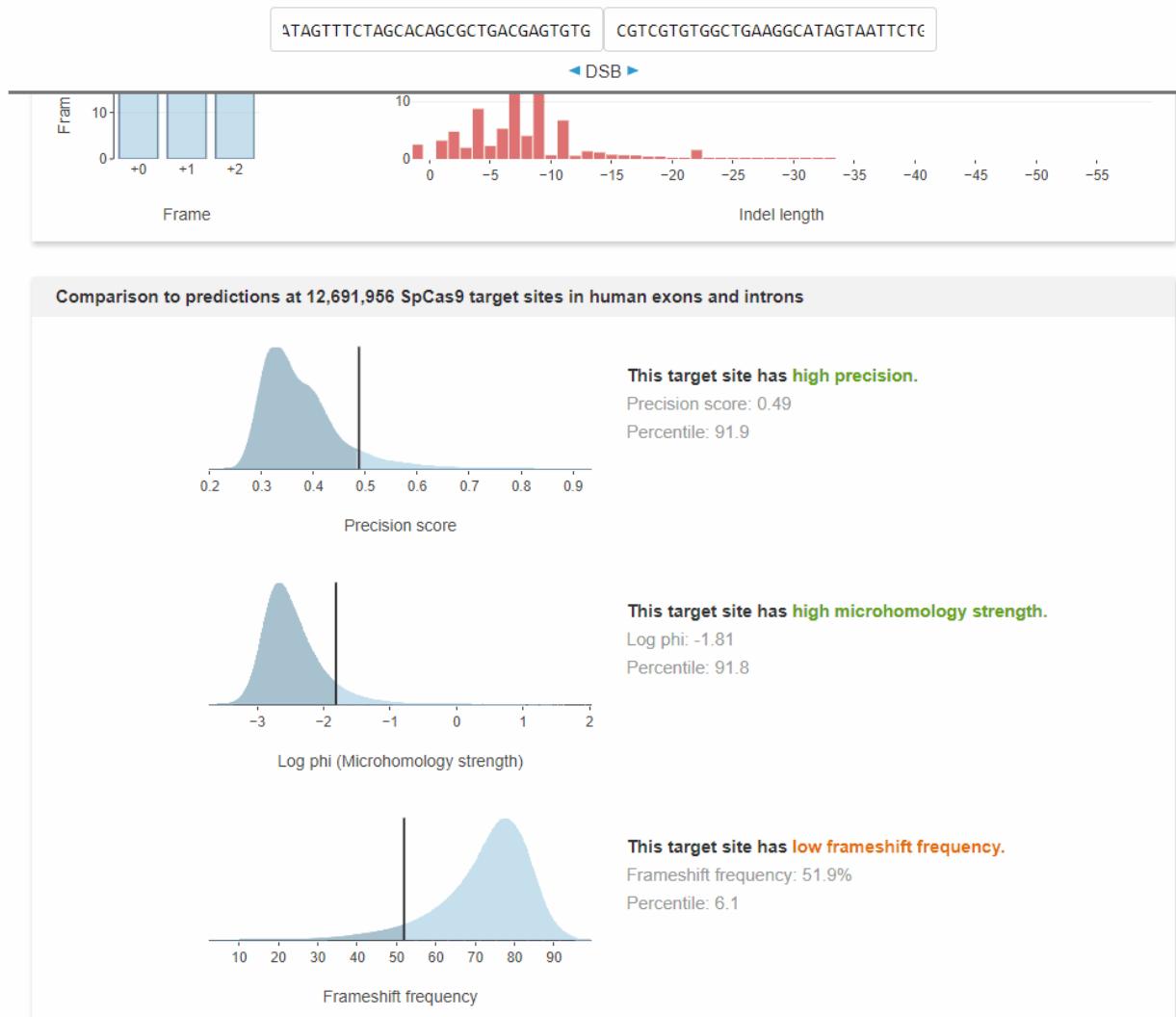
Summary of predictions: Top 10 frequent events



Indel length predictions

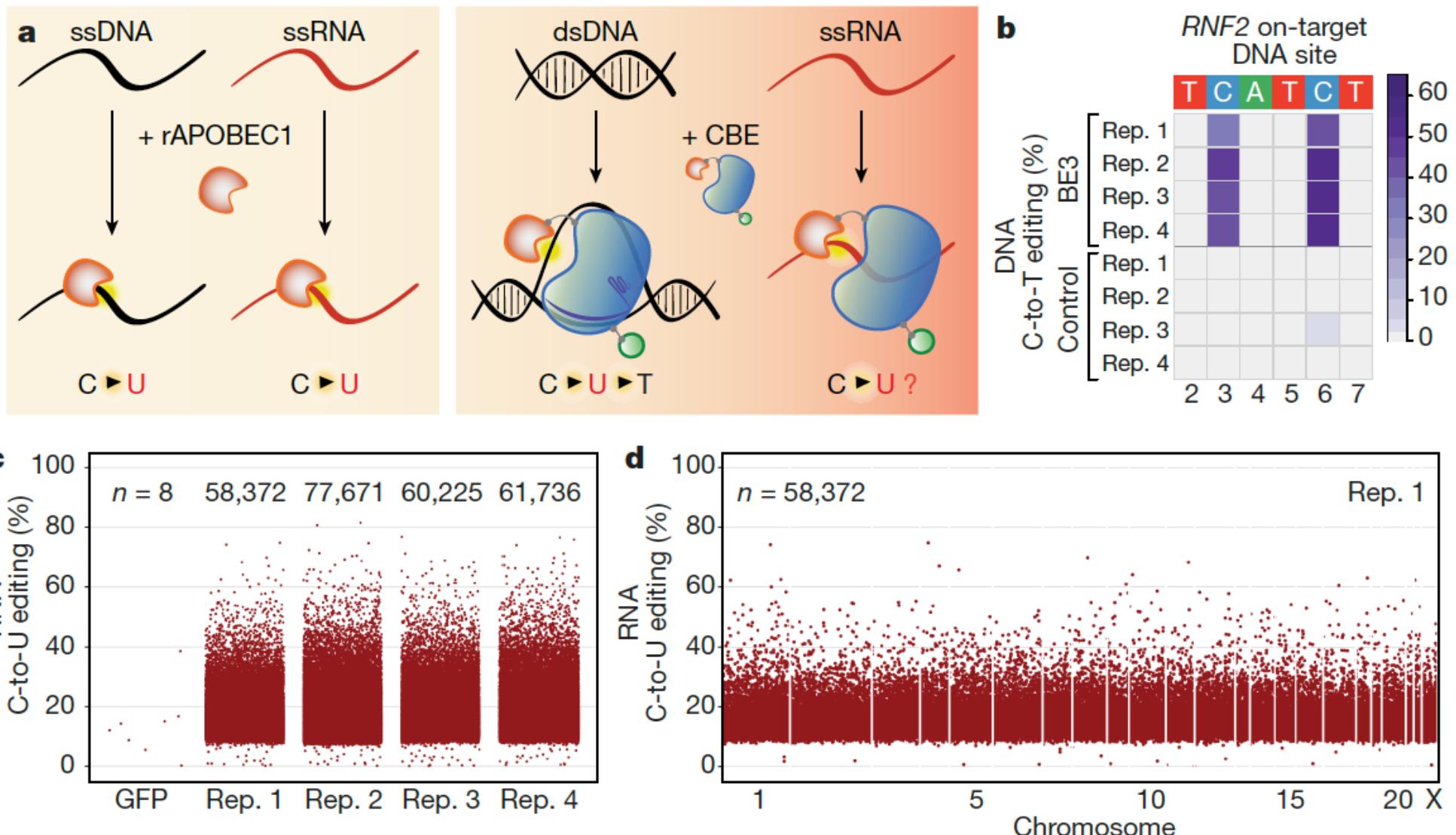


inDelphi

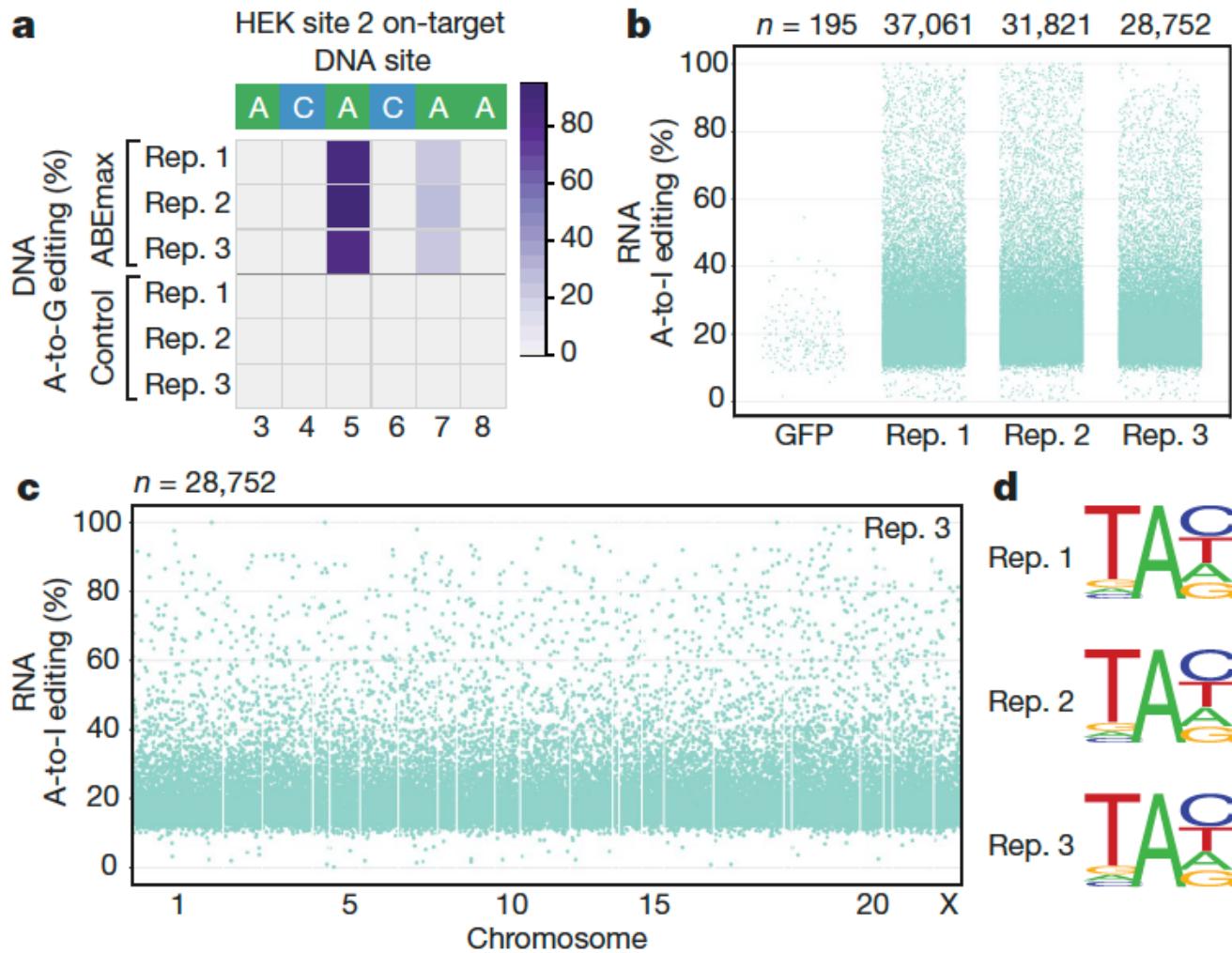


The limitations of base editing

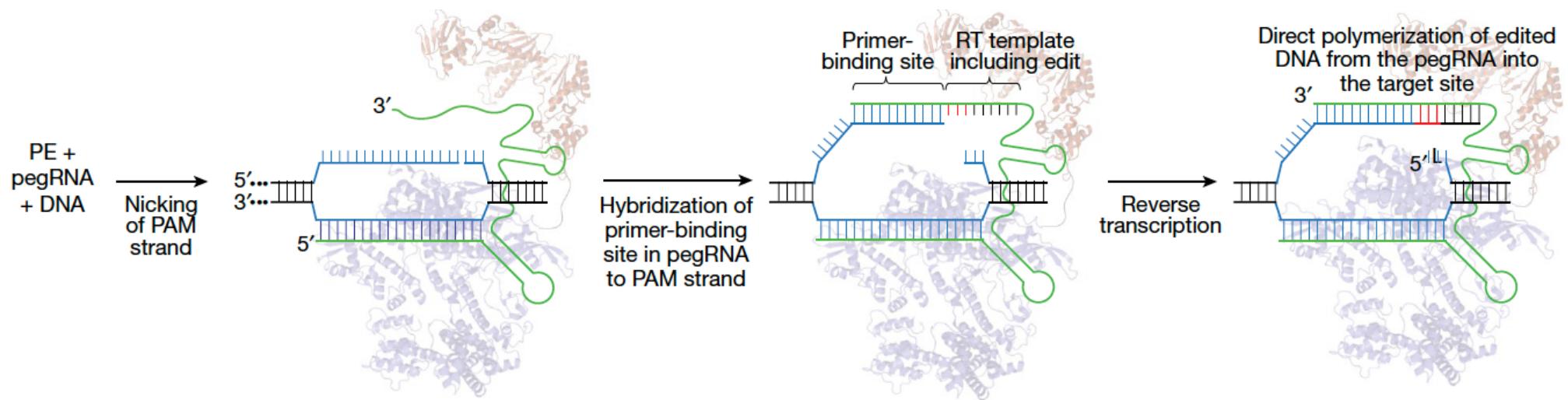
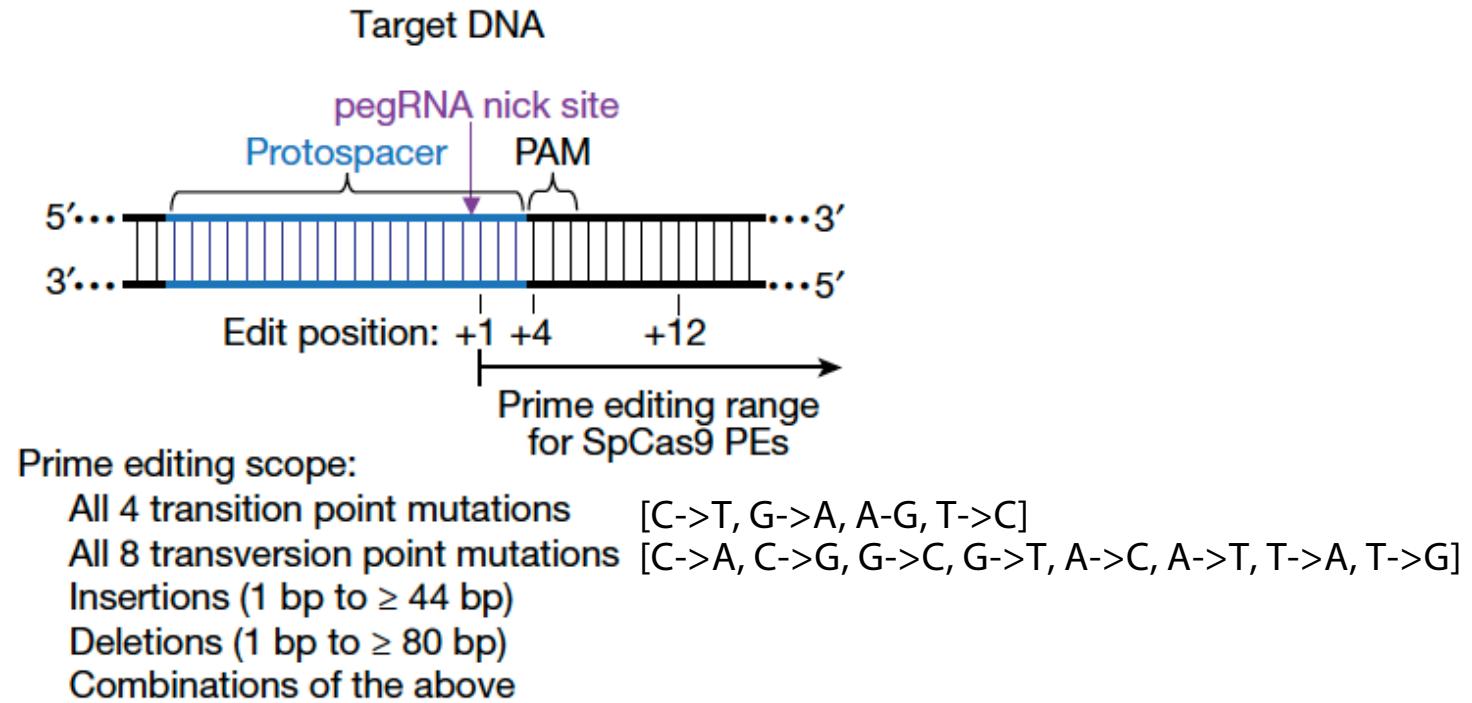
Cytosine base editors (CBEs) have issues with untargeted edits



Adenine base editors (ABEs) have issues with untargeted edits



Prime editing is guided by pegRNAs – Lower off target rate, but not zero



FIN - Thank You