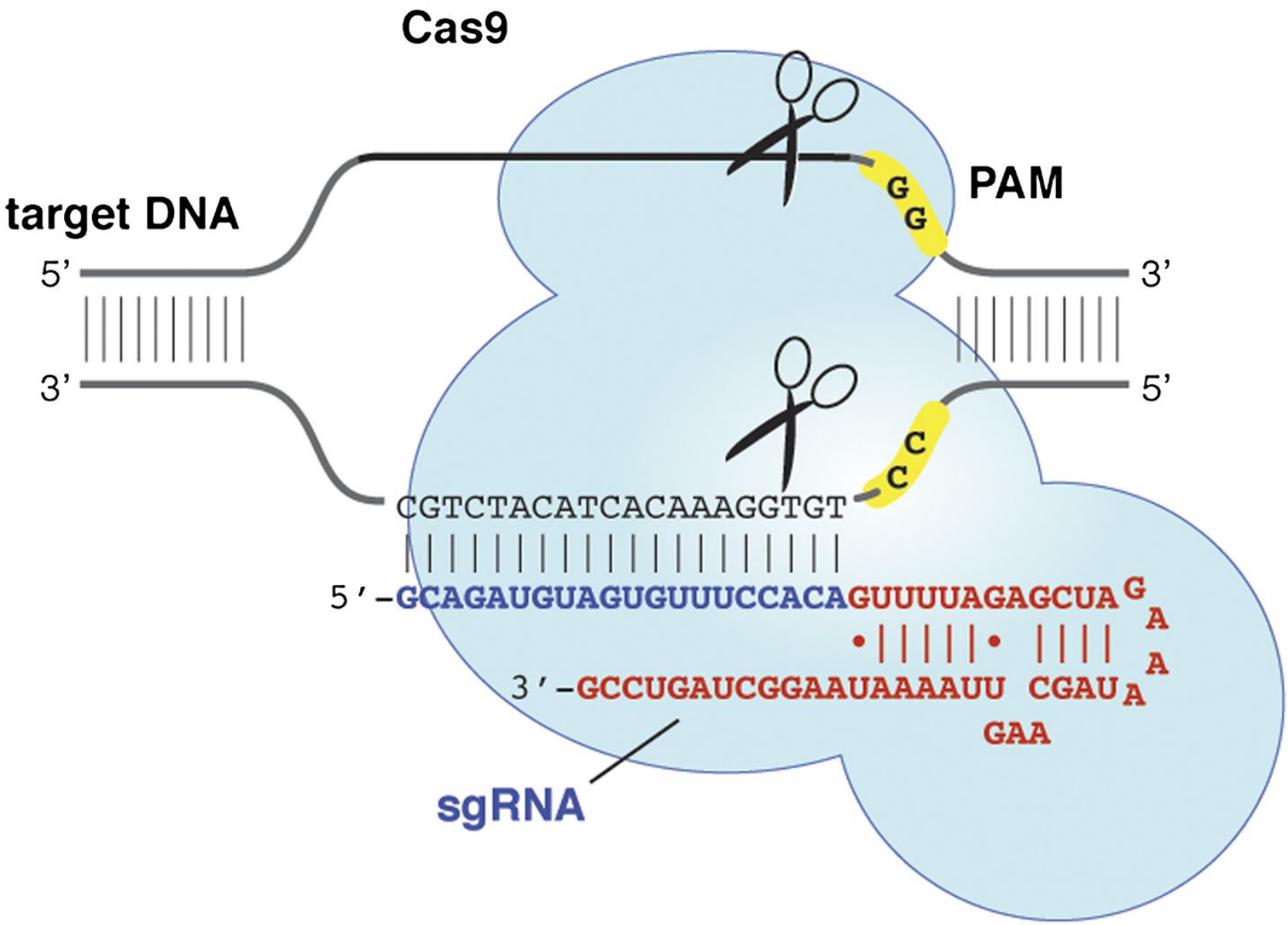
CRISPR Design Considerations

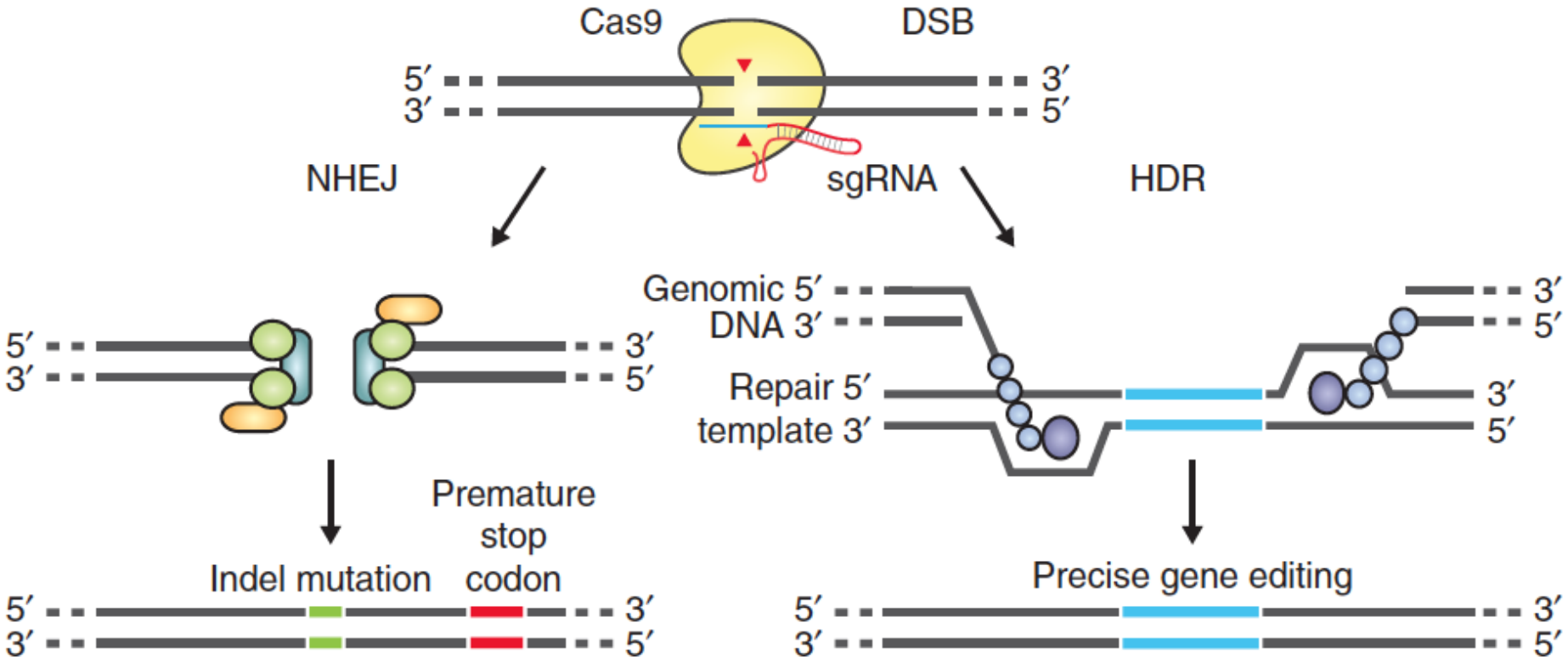
Shifra Ben-Dor

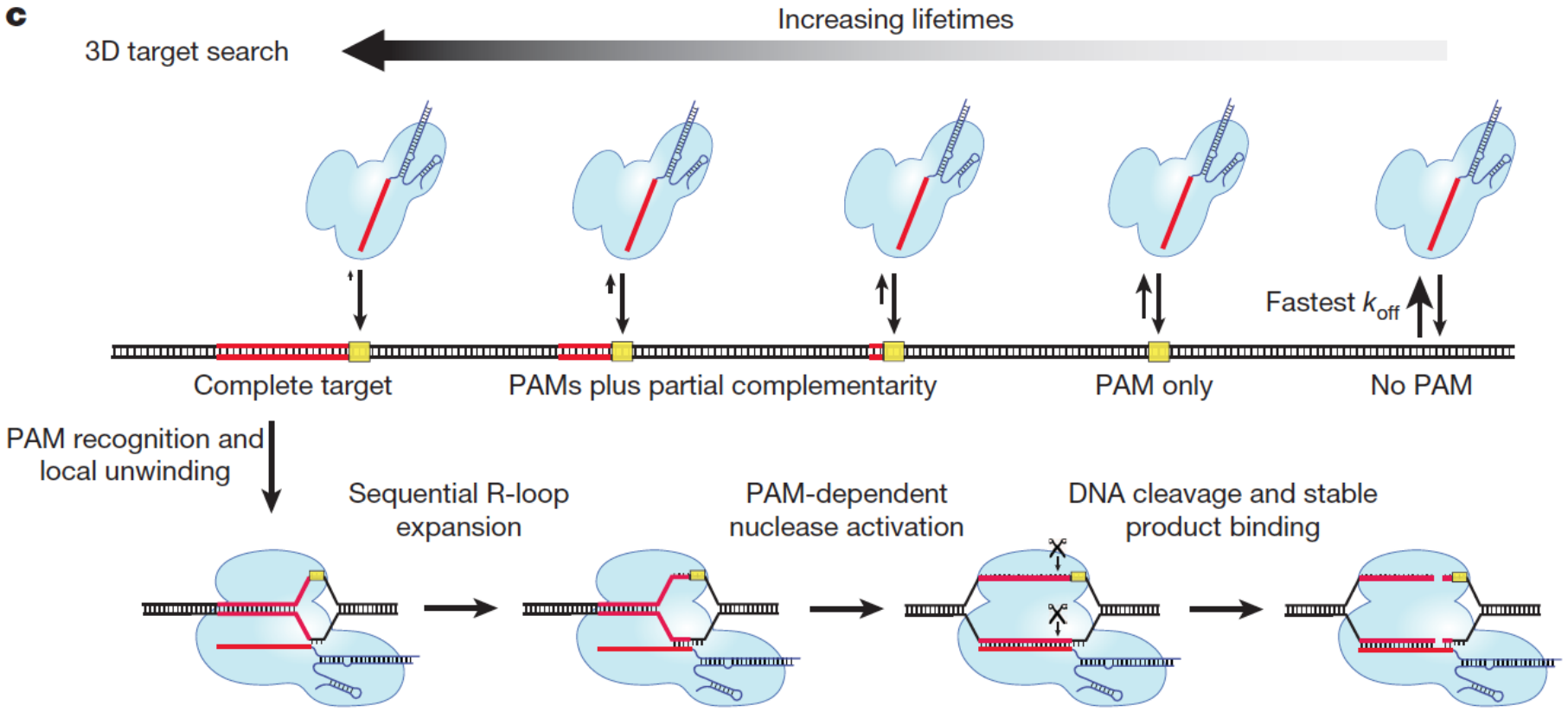
Bioinformatics Unit, Biological Services

April-May 2016



Jinek et al, eLife 2013;2:e00471


Cas9
DSB
5′
3′
3′
5′
NHEJ
sgRNA
HDR
5′
3′
3′
5′
Genomic 5′
DNA 3′
Repair 5′
template 3′
3′
5′
3′
5′
Premature
stop
codon
Indel mutation
Precise gene editing
5′
3′
3′
5′
5′
3′
3′
5′


c
Increasing lifetimes
3D target search
Fastest k_{off}
Complete target
PAMs plus partial complementarity
PAM only
No PAM
PAM recognition and local unwinding
Sequential R-loop expansion
PAM-dependent nuclease activation
DNA cleavage and stable product binding

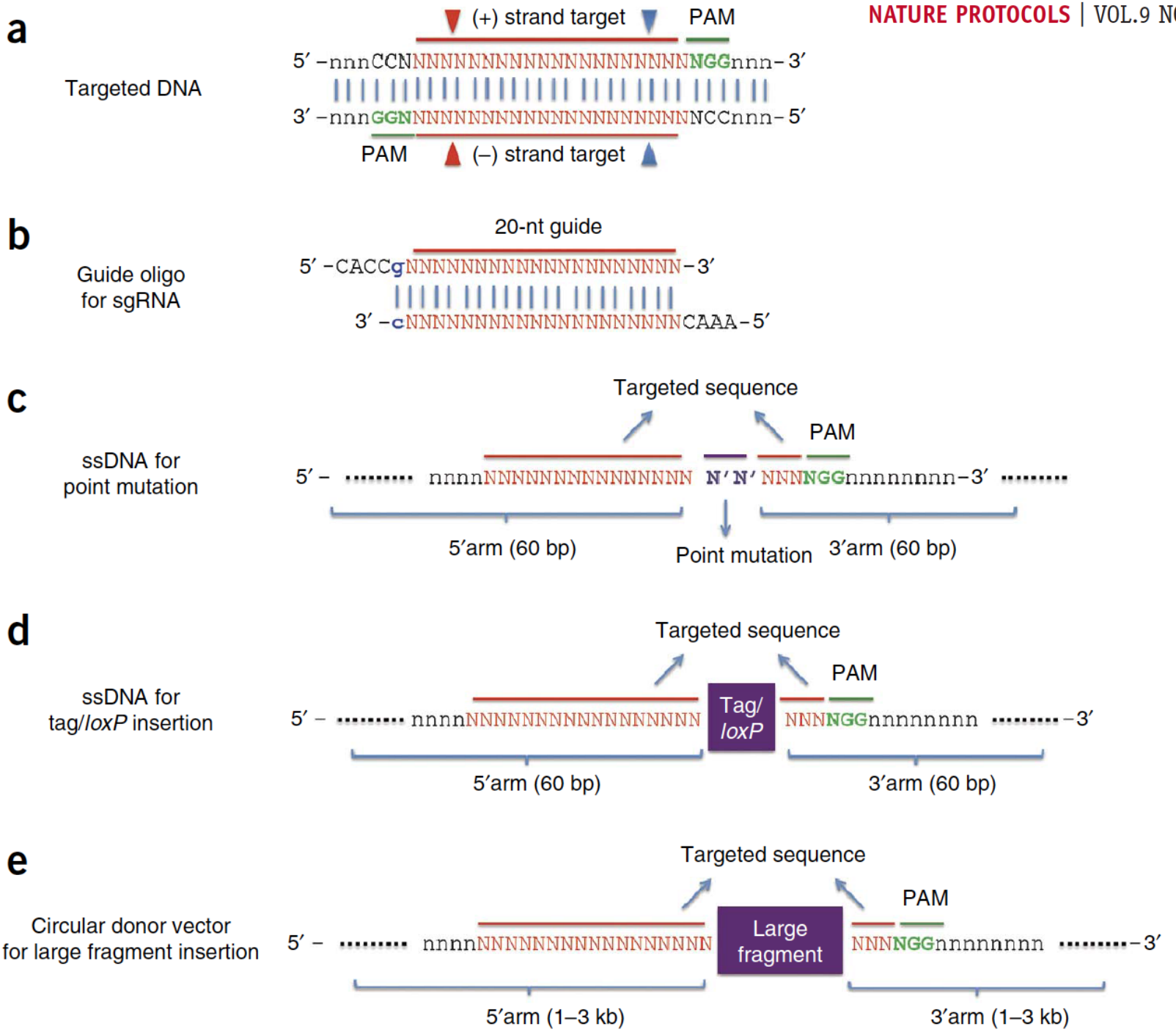
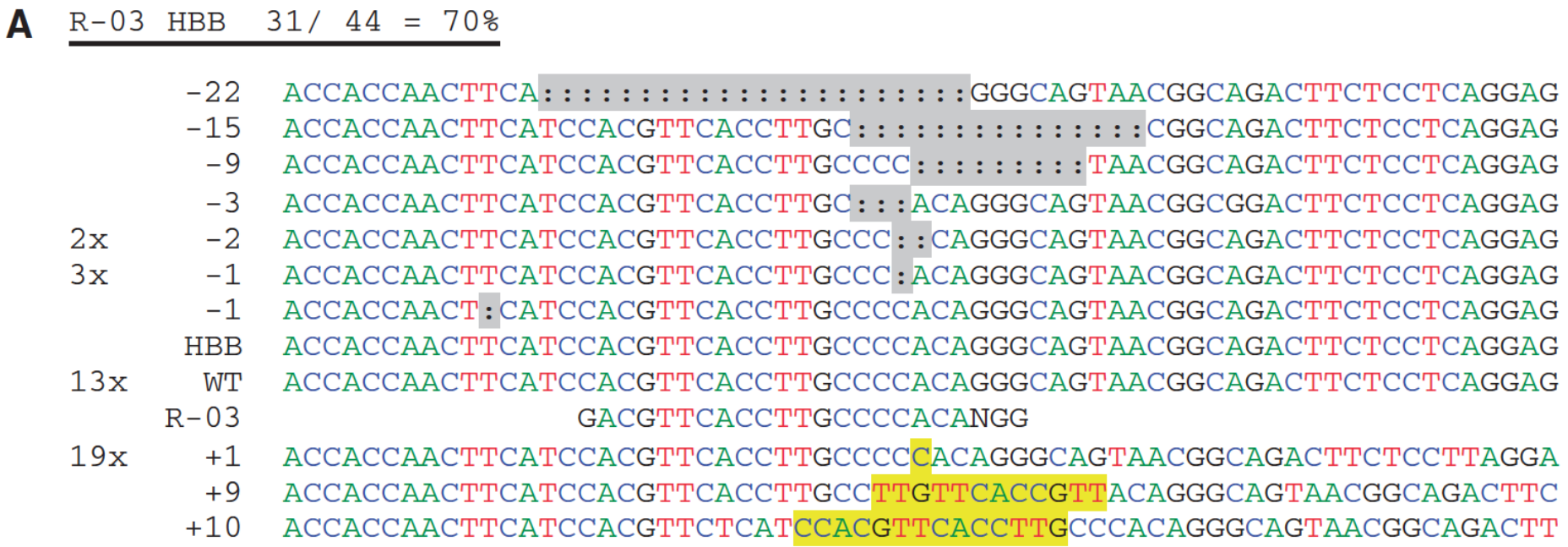

a
Targeted DNA
(+) strand target
PAM
5′ –nnnCCNNNNNNNNNNNNNNNNNNNNGGnnn–3′
3′ –nnnGGNNNNNNNNNNNNNNNNNNNNCCnnn–5′
PAM
(–) strand target
b
Guide oligo for sgRNA
20-nt guide
5′ –CACCgNNNNNNNNNNNNNNNNNNNN–3′
3′ –cNNNNNNNNNNNNNNNNNNNNCAAA–5′
c
ssDNA for point mutation
Targeted sequence
PAM
5′ – nnnnNNNNNNNNNNNNNNNN N′N′ NNNNGGnnnnnnnnn–3′
5′arm (60 bp)
Point mutation
3′arm (60 bp)
d
ssDNA for tag/loxP insertion
Targeted sequence
PAM
5′ – nnnnNNNNNNNNNNNNNNNN
Tag/ loxP
NNNNGGnnnnnnnnn –3′
5′arm (60 bp)
3′arm (60 bp)
e
Circular donor vector for large fragment insertion
Targeted sequence
PAM
5′ – nnnnNNNNNNNNNNNNNNNN
Large fragment
NNNNGGnnnnnnnnn –3′
5′arm (1–3 kb)
3′arm (1–3 kb)

TABLE 1 | Genetically modified mice generated by CRISPR/Cas-mediated genome engineering[13,14].

Type of gene modification	Components	Repair	Targeting efficiency (%)	Application
Indels	Cas9 and sgRNA	NHEJ	80–90	Gene disruption by frame shift
Point mutation	Cas9, sgRNA and ssDNA	HDR	50–80	Precise gene modification
Small tag insertion	Cas9, sgRNA and ssDNA	HDR	30–50	Gene labeling
Conditional allele generation	Cas9, sgRNA and ssDNA	HDR	10–20	Conditional gene knockout
Large fragment deletion	Cas9, sgRNA	NHEJ	30[a]	Gene disruption of specific domain
Large fragment insertion	Cas9, sgRNA and plasmid DNA	HR	10–20	Gene reporter or gene expression

[a]The efficiency of a 700-bp deletion is ~30%. The larger the fragment to be deleted, the lower the efficiency.



Cradick et al *Nucleic Acids Research, 2013, Vol. 41, No. 20*

Outline

- Define your question
- Know your gene
- Positive selection (on target efficiency)
- Negative selection (off target probability)
- Repair Oligo Design
- Screening

Outline

- Define your question
- Know your gene
- Positive selection (on target efficiency)
- Negative selection (off target probability)
- Repair Oligo Design
- Screening

Define your question

- Why are you doing CRISPR?
  - Gene knock-out
  - Point mutation
  - Deletion
  - Insertion (tag, lox sites)
- What?
  - NHEJ or HDR?
  - Cleave or nick (One strand or two)?

Define your question

- When?
  - Conditional or constitutive?
- Where?
  - Cell line or whole animal?
  - Tissue specific?

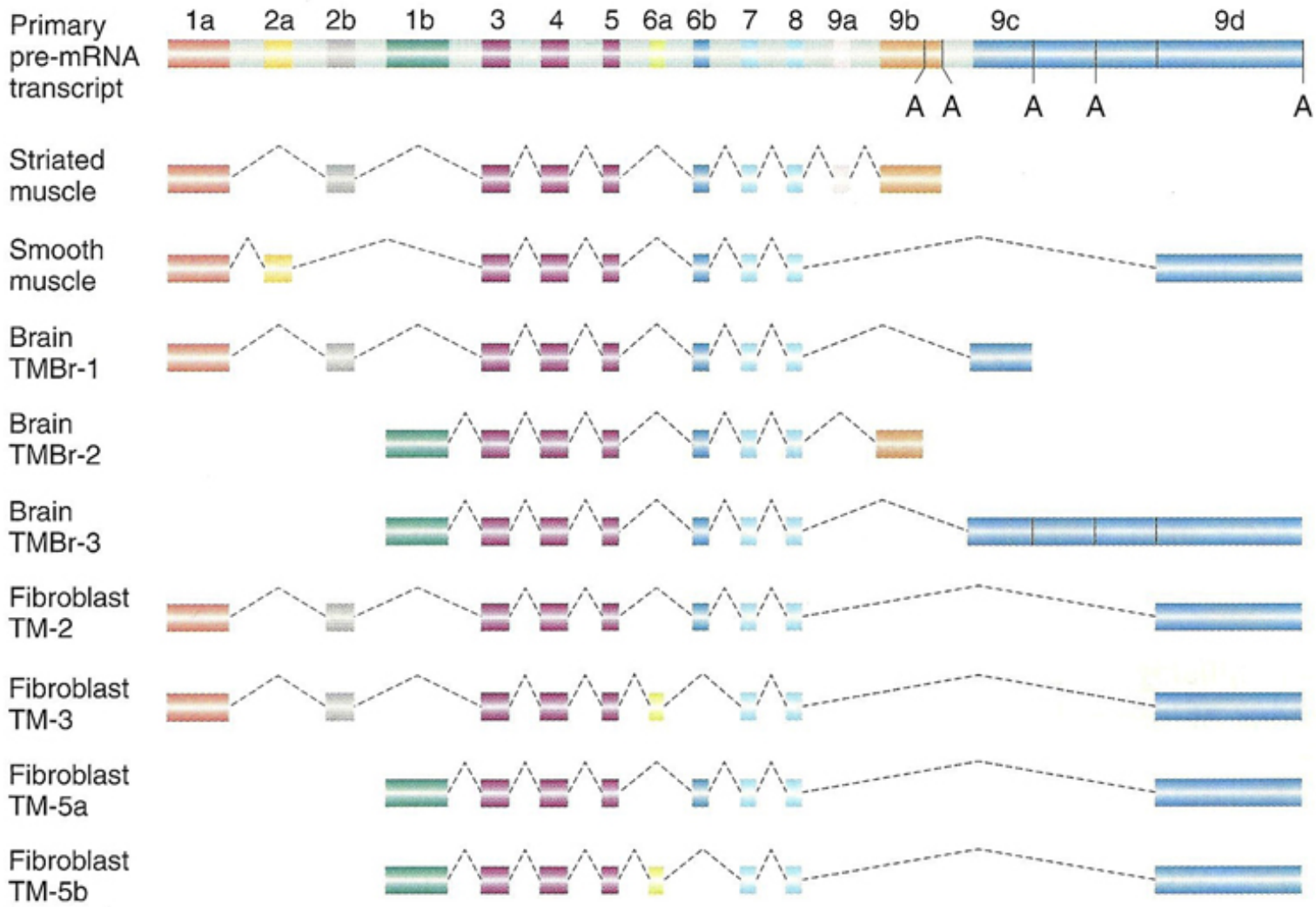
Outline

- Define your question
- Know your gene
- Positive selection (on target efficiency)
- Negative selection (off target probability)
- Repair Oligo Design
- Screening

Know your gene

- Splice Variants
  - Are there consistent exons?
  - Do they all have the same start?
  - Are the exons in frame?
- How far into the gene can you go?
- Does your gene overlap with other genes?

Alternative splicing in tropomyosin



Taken from: BIOL 202 Genetics taught by Dr. Paul Szauter at the University of New Mexico

Know your gene

- Splice Variants
  - Are there consistent exons?
  - Do they all have the same start?
  - Are the exons in frame?
- How far into the gene can you go?
- Does your gene overlap with other genes?

Your Favorite Protein

Know your gene

- Splice Variants
  - Are there consistent exons?
  - Do they all have the same start?
  - Are the exons in frame?
- How far into the gene can you go?
- Does your gene overlap with other genes?

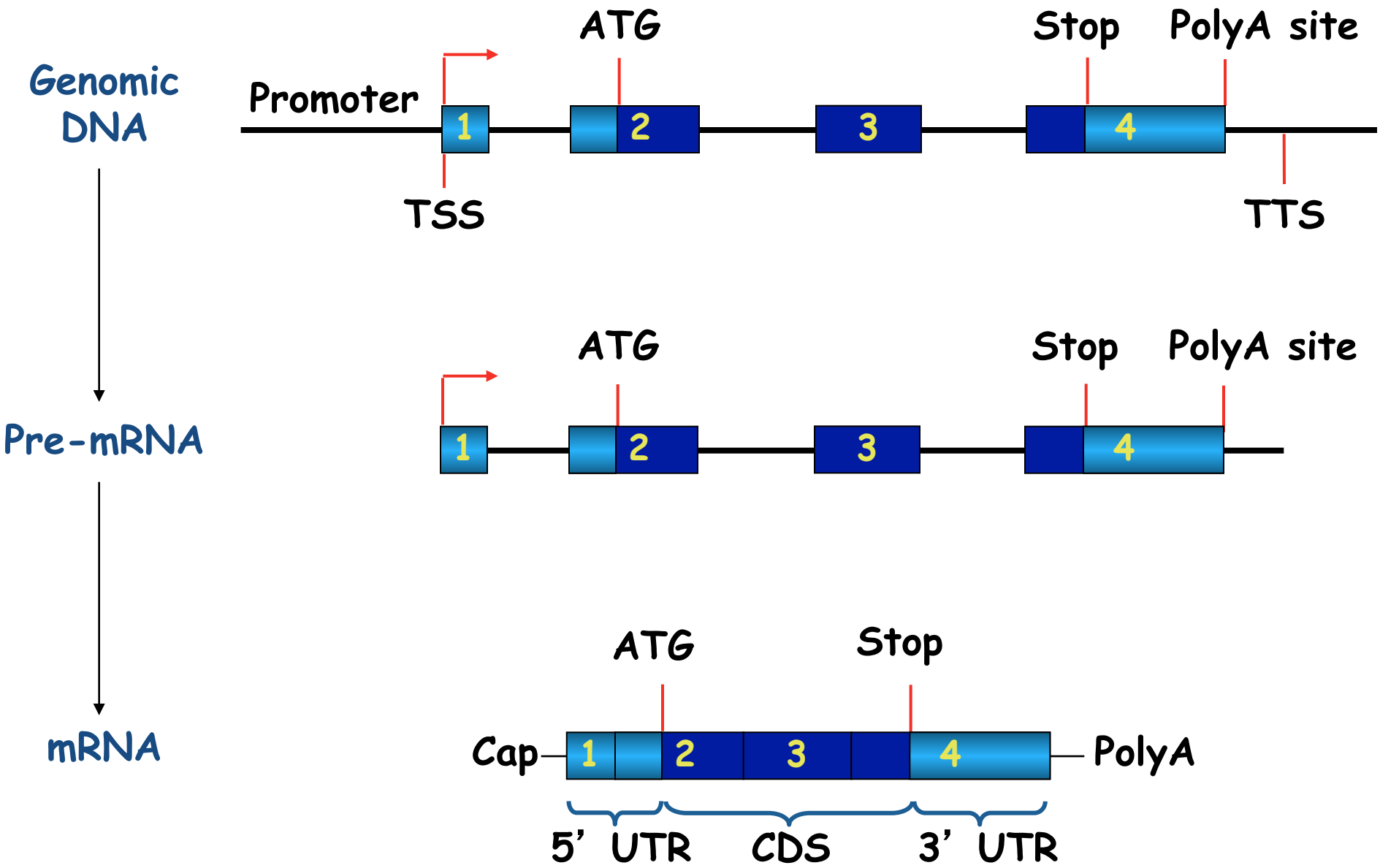
Other considerations

Residual protein

- If a whole exon was removed
- If an alternate ATG can be activated
- If the stop codon isn't introduced in the right place

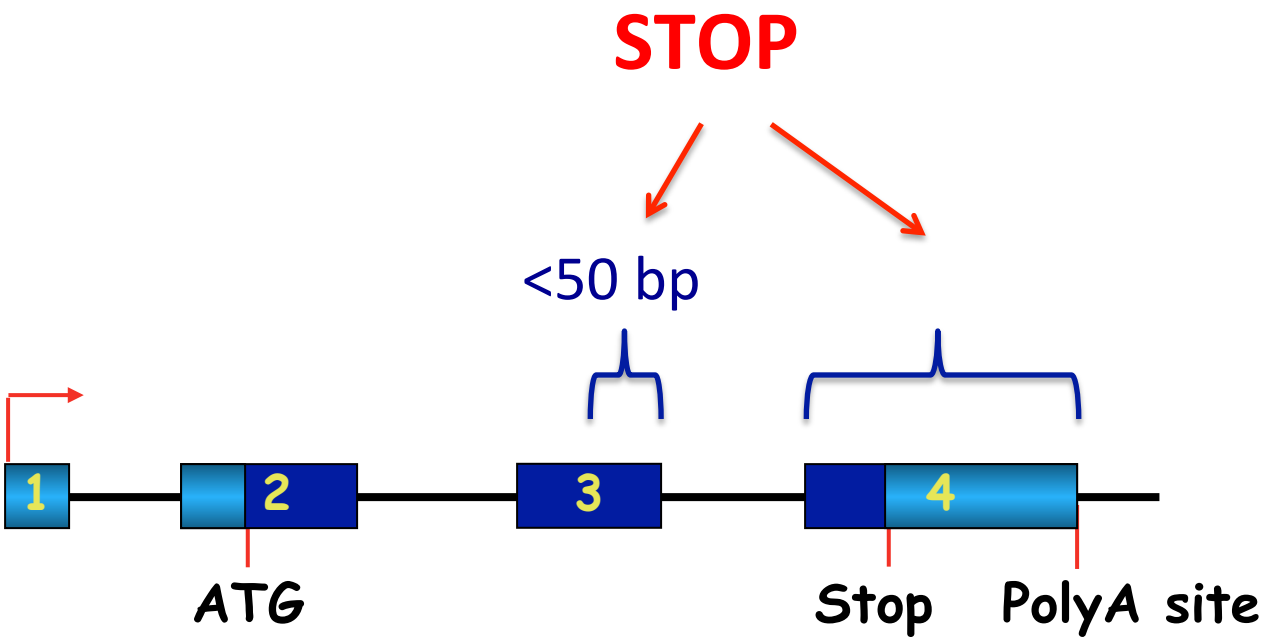
NMD

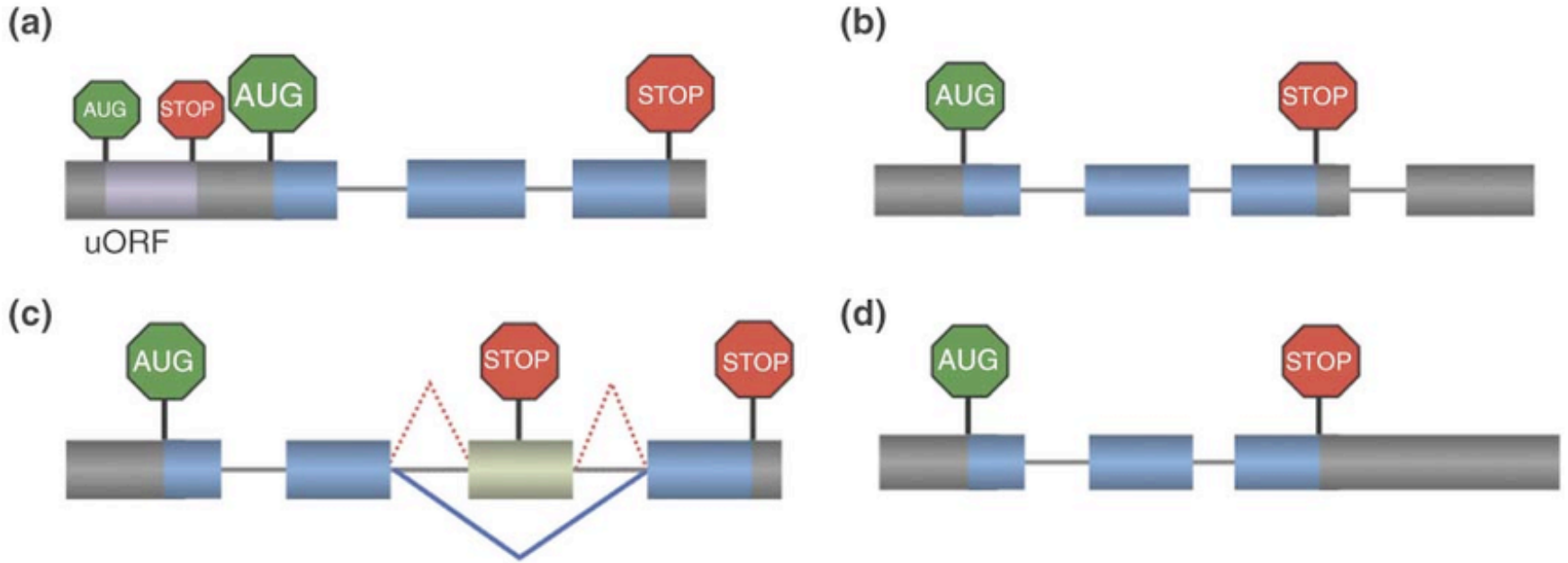
- Nonsense mediated decay



Modified from Zhang MQ Nat Rev Genet. 2002 Sep;3(9):698-709.

Stop codon must be in last exon, or within 50 bp of last junction





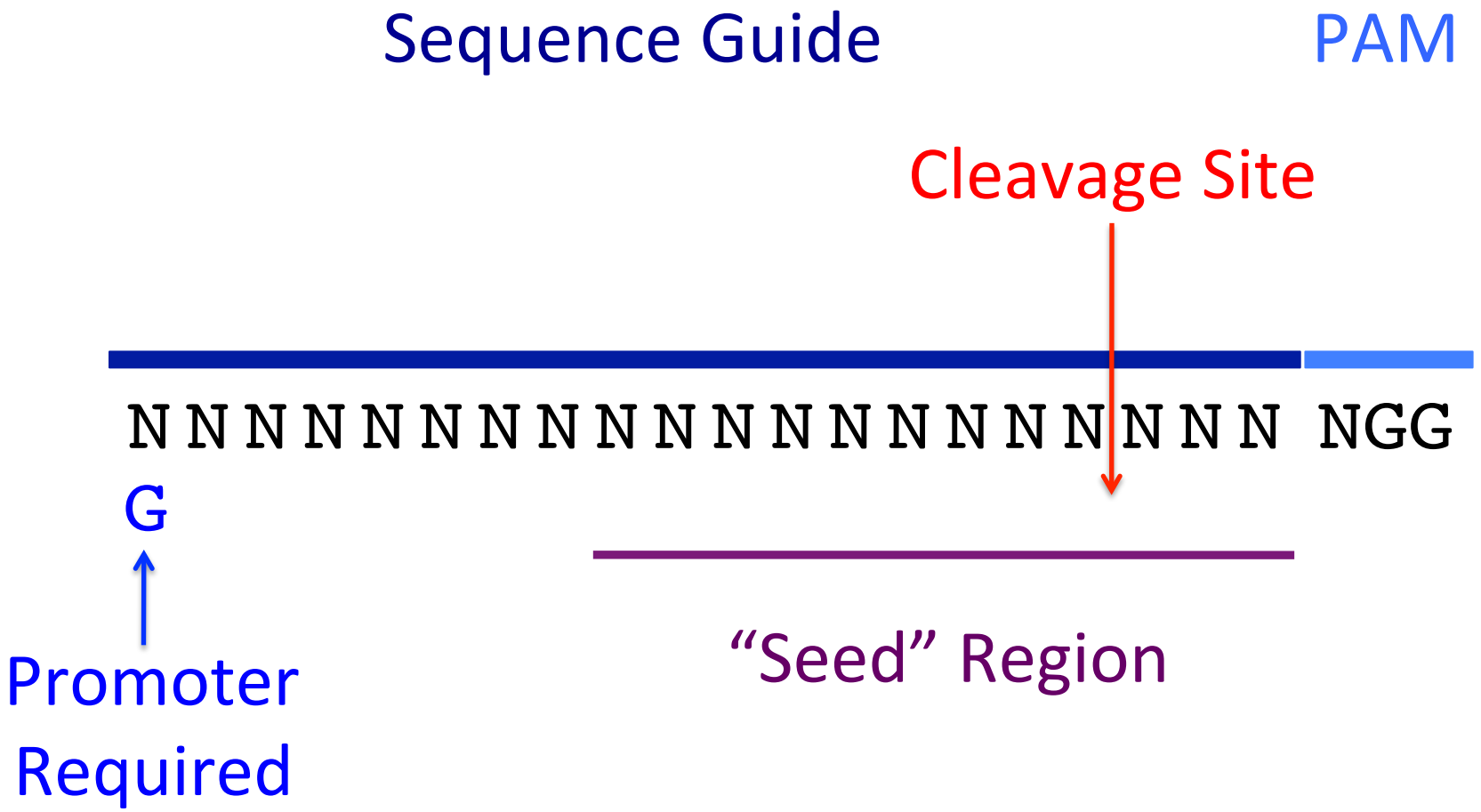
Regulation of nonsense-mediated mRNA decay. Huang L, Wilkinson MF. (2012)

Genomic Considerations

- Family members
- Pseudogenes
- Conditional Knockouts
  - Control regions (Promoter, Enhancer)
  - Overlapping genes

Guide Design

Starting Point



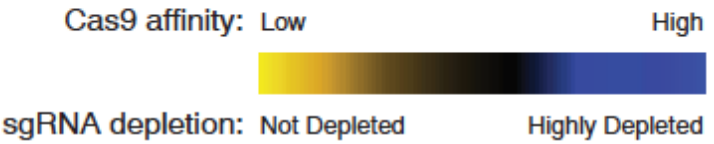
Outline

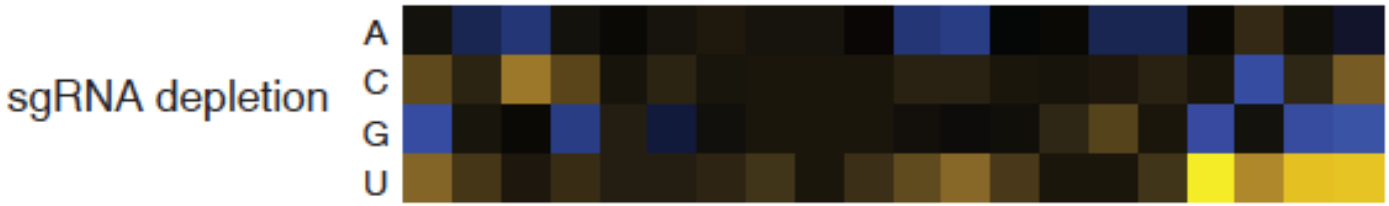
- Define your question
- Know your gene
- Positive selection (on target efficiency)
- Negative selection (off target probability)
- Repair Oligo Design
- Screening

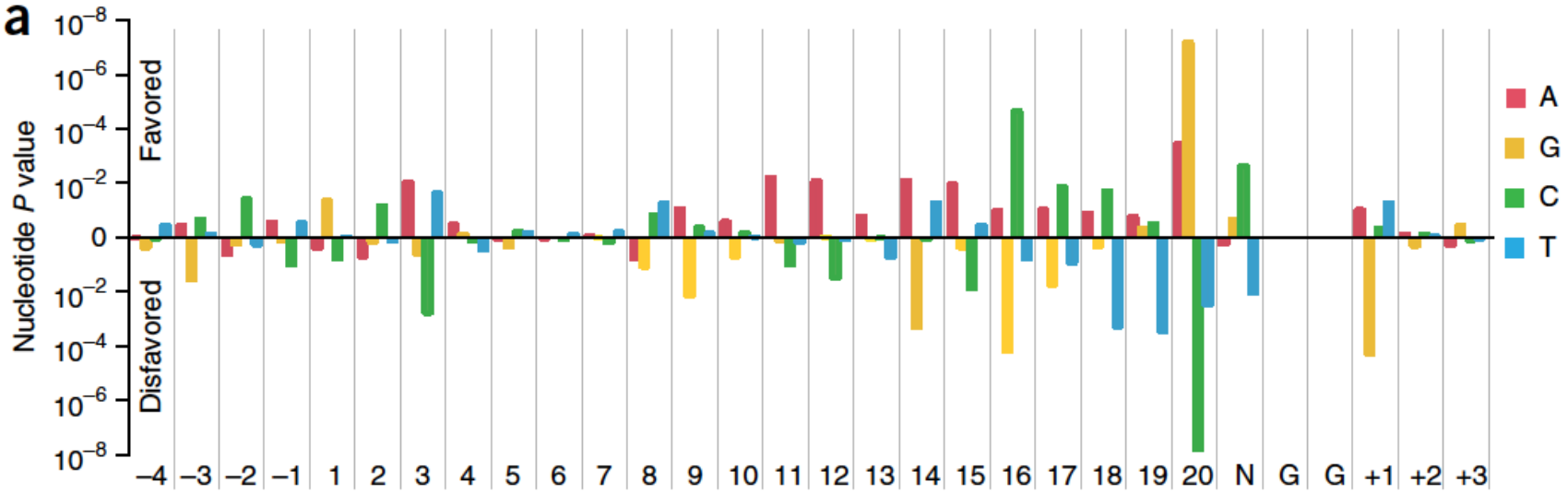
Positive Selection

- G/C content
- Base Preferences

Wang, Science 2014







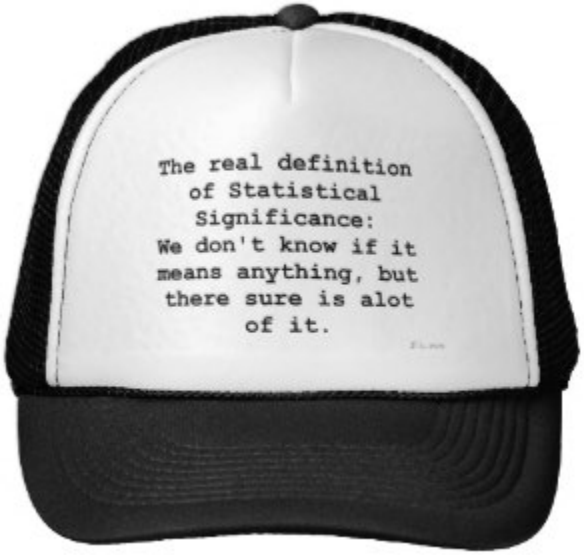
Doench, Hartenian, Nature Biotech. 2014

sgRNA Designer

http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design

Score is from 0 to 1, the higher the better.
Over 0.2 is supposed to work well

“Facts are stubborn things,
but statistics are more pliable.”

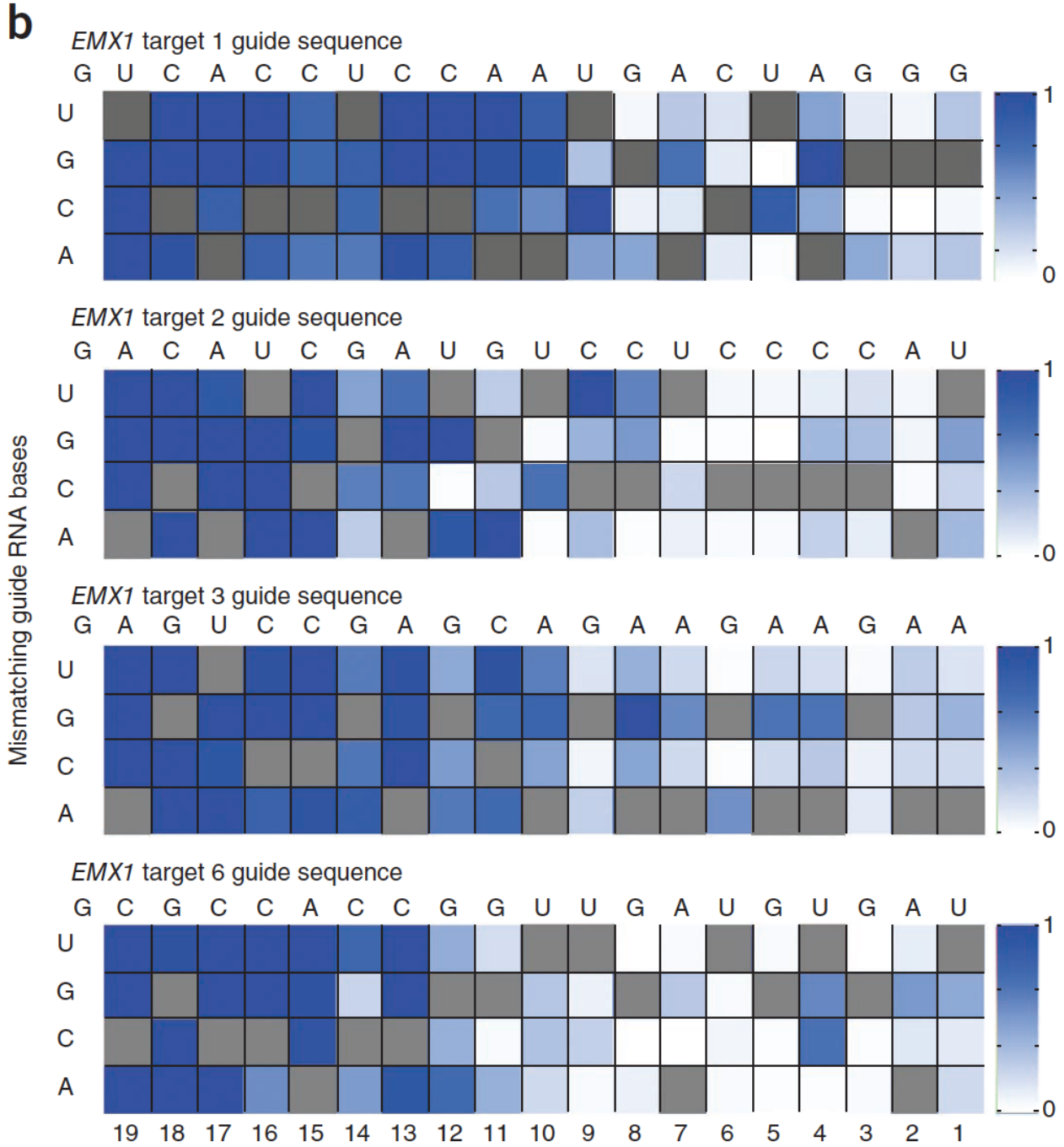


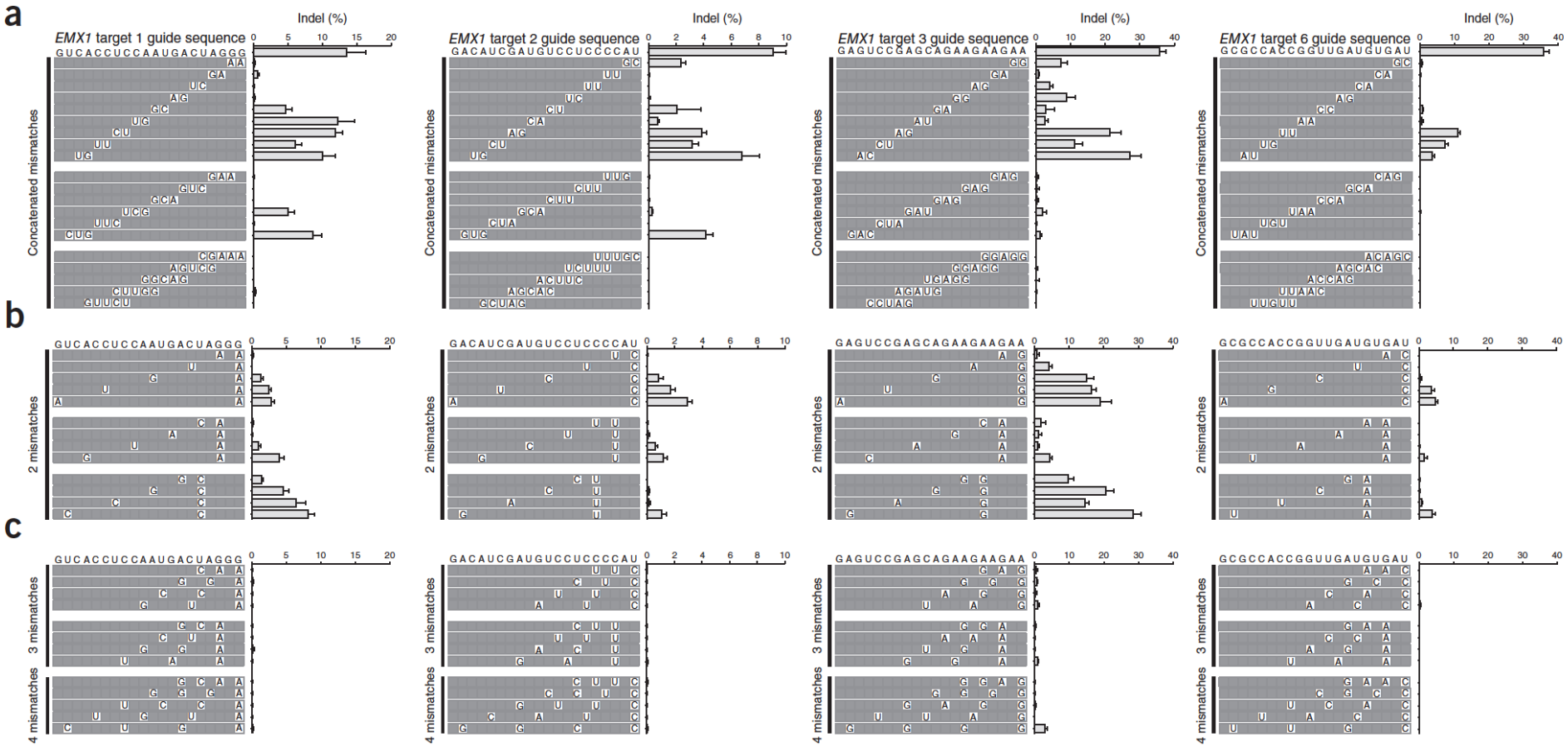
A group at WIS designed CRISPR based on off-target alone, and retroactively checked the positive scores: One scored 0.0629 and the other scored 0.04859. They both worked beautifully.

Outline

- Define your question
- Know your gene
- Positive selection (on target efficiency)
- Negative selection (off target probability)
- Repair Oligo Design
- Screening

OFF TARGET HAPPENS


b
EMX1 target 1 guide sequence
G U C A C C U C C A A U G A C U A G G G
EMX1 target 2 guide sequence
G A C A U C G A U G U C C U C C C C A U
EMX1 target 3 guide sequence
G A G U C C G A G C A G A A G A A G A A
EMX1 target 6 guide sequence
G C G C C A C C G G U U G A U G U G A U
U
G
C
A
Mismatching guide RNA bases
1
0
19 18 17 16 15 14 13 12 11 10 9 8 7 6 5 4 3 2 1


a
b
c
Indel (%)
EMX1 target 1 guide sequence
GUCACCUCCAAUGACUAGGG
EMX1 target 2 guide sequence
GACAUCGAUGUCCUCCCCAU
EMX1 target 3 guide sequence
GAGUCCGAGCAGAAGAAGAA
EMX1 target 6 guide sequence
GCGCCACCGGUUGAUGUGAU
Concatenated mismatches
2 mismatches
3 mismatches
4 mismatches

Off Target programs

Zhang lab	http://crispr.mit.edu/
E-CRISP	http://www.e-crisp.org/E-CRISP/
CHOPCHOP	https://chopchop.rc.fas.harvard.edu/
ZIFIT	http://zifit.partners.org/ZiFiT/
GT-Scan	http://gt-scan.braembl.org.au/
COSMID	https://crispr.bme.gatech.edu/
sgRNAcas9	http://www.biootools.com/
CasOT	http://eendb.zfgenetics.org/casot/
DNA 2.0	https://www.dna20.com/eCommerce/cas9/input

Considerations in off-target

- Mismatches
- PAM
- G/C content
- Is an initial G required?
- What version of the genome is used?
- Are the criteria updated?
- Does it come with primer design?

Mismatches

- How many mismatches are allowed?
  - 2, 3, 4 or more?
- Where are the mismatches allowed?
  - PAM proximal as compared to PAM distal
- What type of mismatches are allowed?
  - Insertions and deletions?

Considerations in off-target

- Mismatches
- PAM
- G/C content
- Is an initial G required?
- What version of the genome is used?
- Are the criteria updated?
- Does it come with primer design?

PAM

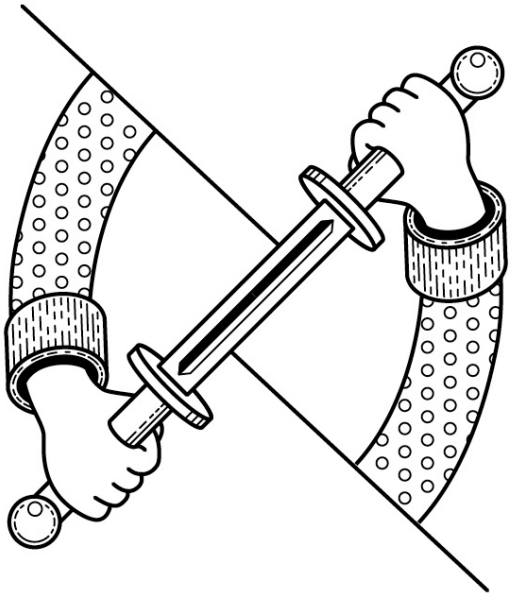
- NGG is the ‘classic’ but others may be acceptable
  - NAG
  - Possibly others

Considerations in off-target

- Mismatches
- PAM
- G/C content
- Is an initial G required?
- What version of the genome is used?
- Are the criteria updated?
- Does it come with primer design?

G/C content

- For activity, a higher G/C content is preferred
- For specificity, a lower G/C content is preferred



Considerations in off-target

- Mismatches
- PAM
- G/C content
- Is an initial G required?
- What version of the genome is used?
- Are the criteria updated?
- Does it come with primer design?

How can you double check?

- Run your own Blast (but change the parameters)
- Make sure you don’t have a genomic repeat sequence
- Check if the genome sequence is masked

Program Selection

Optimize for

- ○ Highly similar sequences (megablast)
- ○ More dissimilar sequences (discontiguous megablast)
- ◉ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Make sure you are doing genome blast, and run against only the reference version of the genome!

BLAST Search database Genome (reference only) - Drosophila melanogaster using Blastn (Optimize for somewhat similar sequences)

☐ Show results in a new window

Algorithm parameters

Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

General Parameters

Max target sequences: 100
Select the maximum number of aligned sequences to display

Short queries: ☐ Automatically adjust parameters for short input sequences

Expect threshold: 1000

Word size: 7

Max matches in a query range: 0

Scoring Parameters

Match/Mismatch Scores: 1,-1

Gap Costs: Existence: 2 Extension: 1

Filters and Masking

Filter

- ☐ Low complexity regions
- ☐ Species-specific repeats for: Drosophila melanogaster (Fruit fly)

Mask

- ☐ Mask for lookup table only
- ☐ Mask lower case letters

BLAST Search database Genome (reference only) - Drosophila melanogaster using Blastn (Optimize for somewhat similar sequences)

☐ Show results in a new window

Outline

- Define your question
- Know your gene
- Positive selection (on target efficiency)
- Negative selection (off target probability)
- Repair Oligo Design
- Screening

Repair Oligo Design

- Keep in mind:
  - We don't want the repair oligo to be cleaved, if possible mutate the PAM, if not, introduce mismatches
  - Add or delete cleavage site for screening
  - Don't add mutations to far from each other, need at least 60bp flanking, and a total length of no longer than 200bp

Other issues

Lox sites for conditional mutations may fold over on themselves (partly palindromic)

CRISPR Screening Considerations

So, you've done CRISPR… now we have see if it worked

- Screening of CRISPR experiments is as variable as the types of use of CRISPR
- Different uses will have different methods of screening
- Some types have more options than others

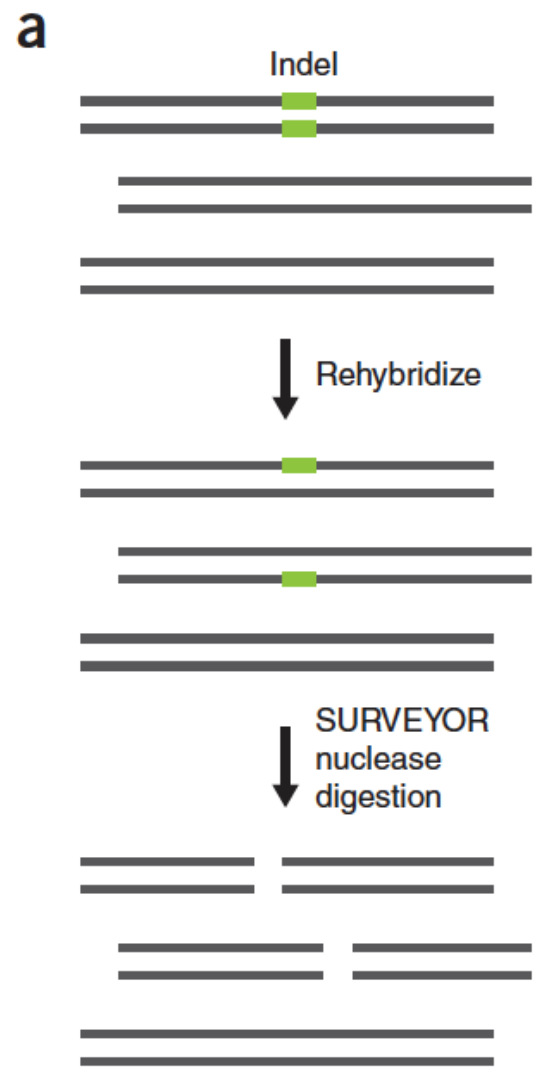
Activity vs Mutation

In cases where we are trying to knockout or mutate a protein, do we check the DNA for mutation, or do we ‘go for the gold’ and check protein presence or activity?

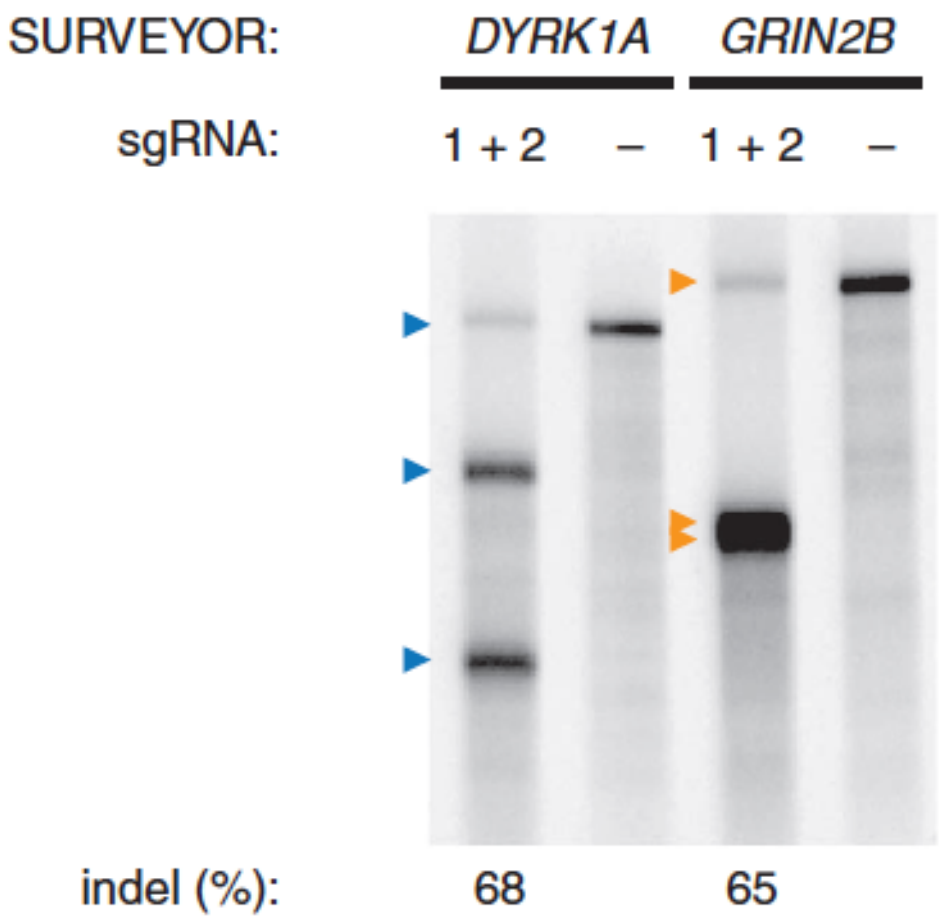
Types of indel screening

- EMC (Enzyme Mismatch Cleavage)
  - Surveyor, T7E1, etc
- HRM (High Resolution Melt)
- IDAA (Indel Detection by Amplicon Analysis)
- PCR
- Sequencing
  - NGS
  - Sanger

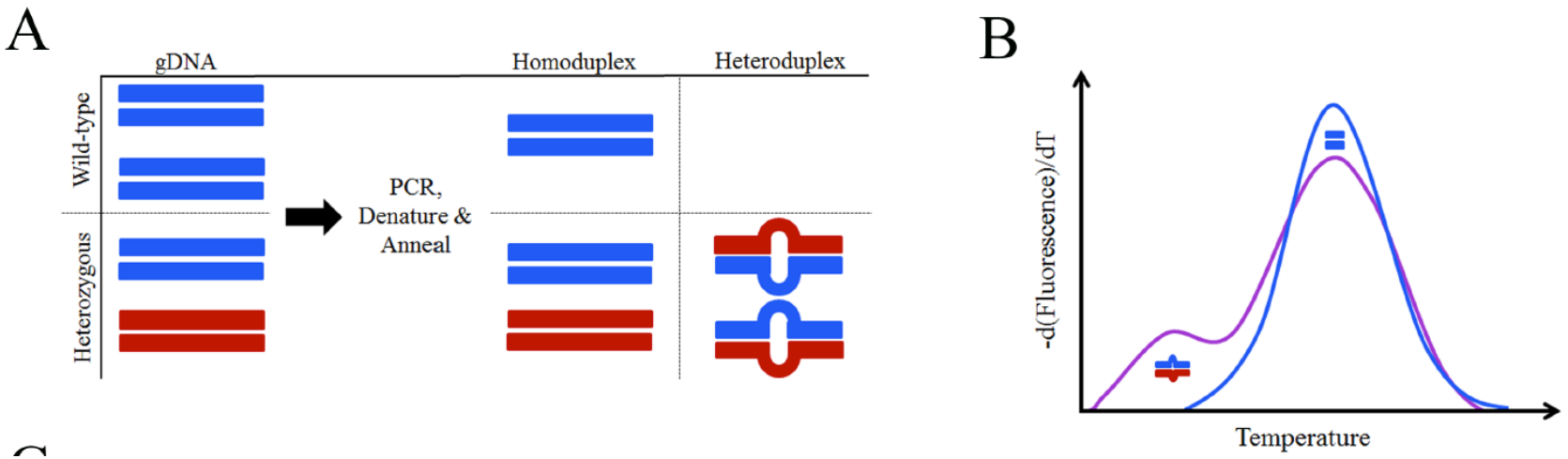
EMC



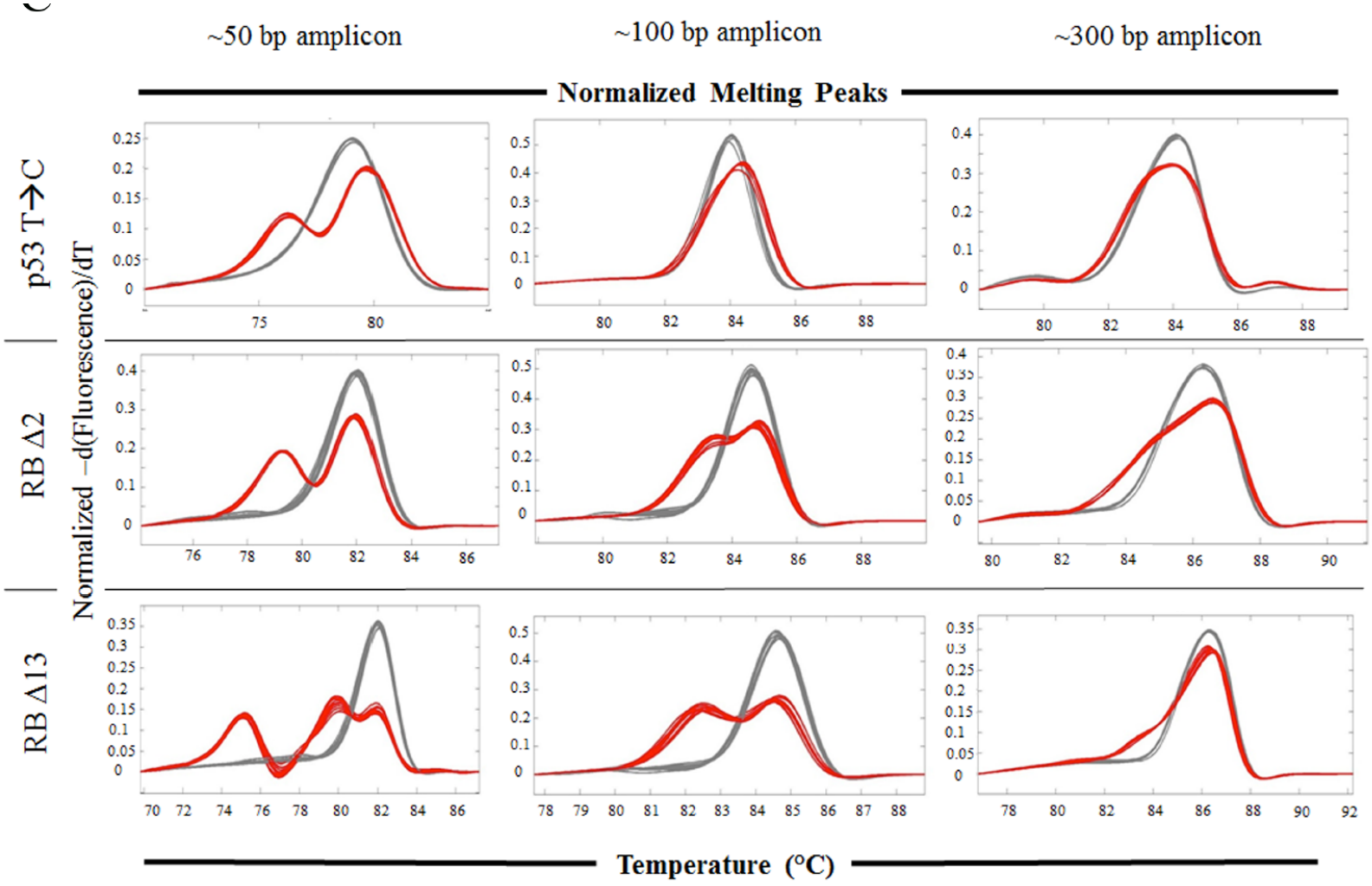
Ran FA et al, Genome engineering using the CRISPR-Cas9 system, Nature Protocols (2013)


SURVEYOR:
DYRK1A
GRIN2B
sgRNA:
1 + 2
–
1 + 2
–
indel (%):
68
65

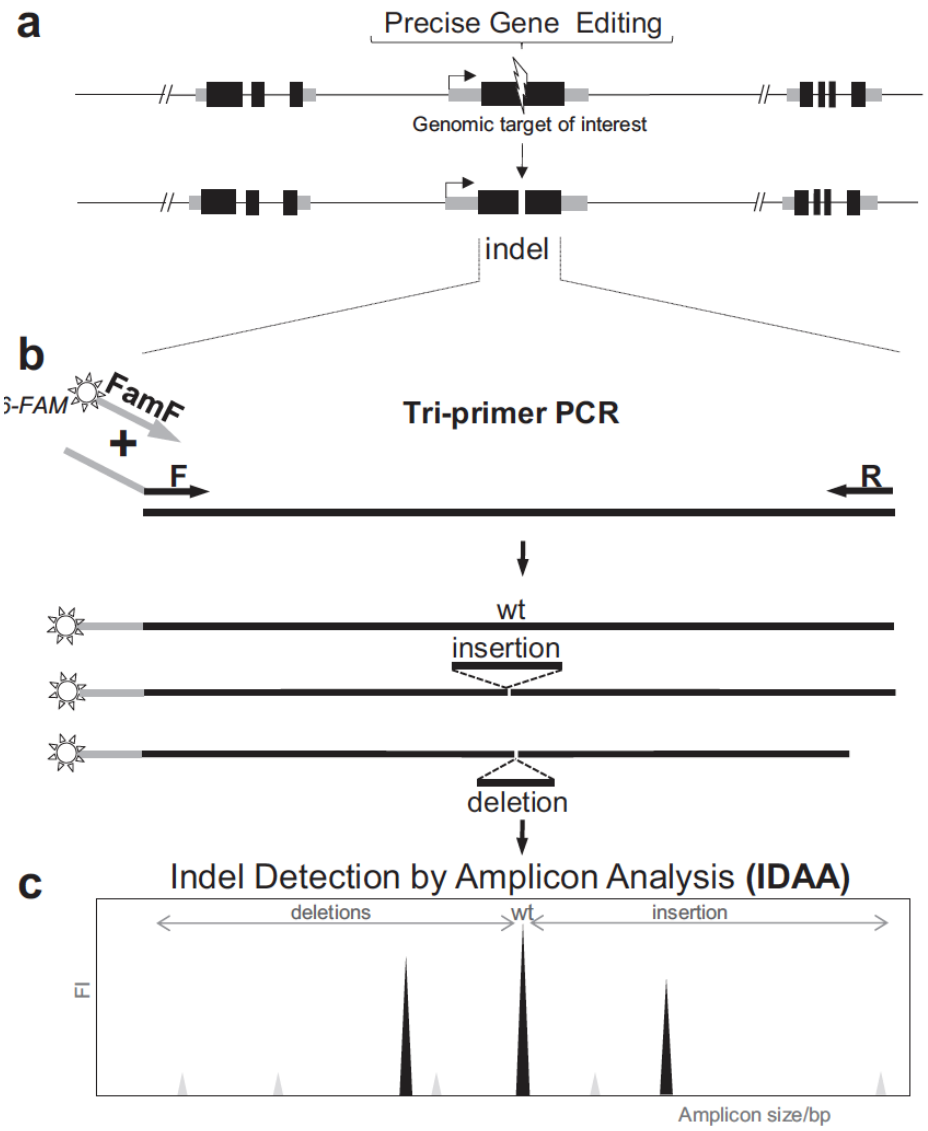
HRM



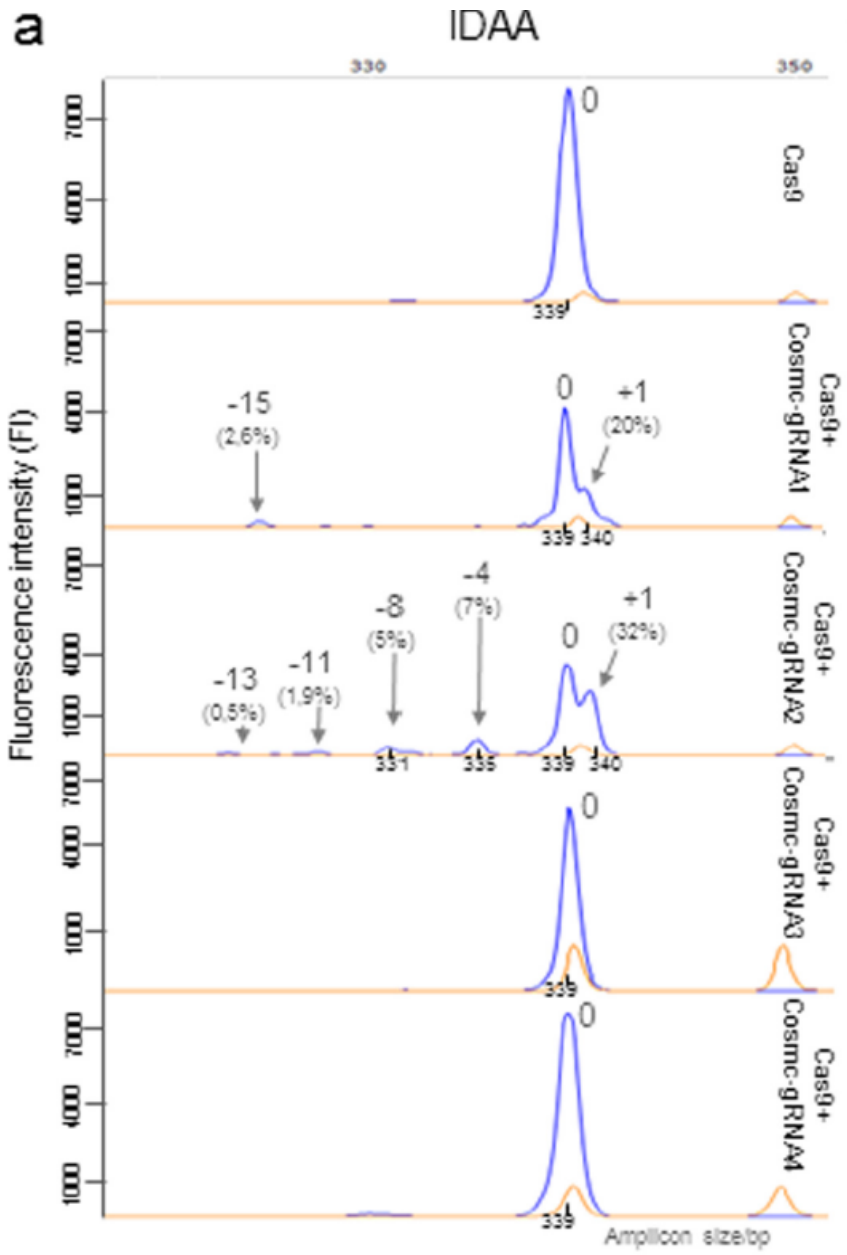
Thomas, HR et al. High-Throughput Genome Editing and Phenotyping Facilitated by High Resolution Melting Curve Analysis. PlosOne 2014

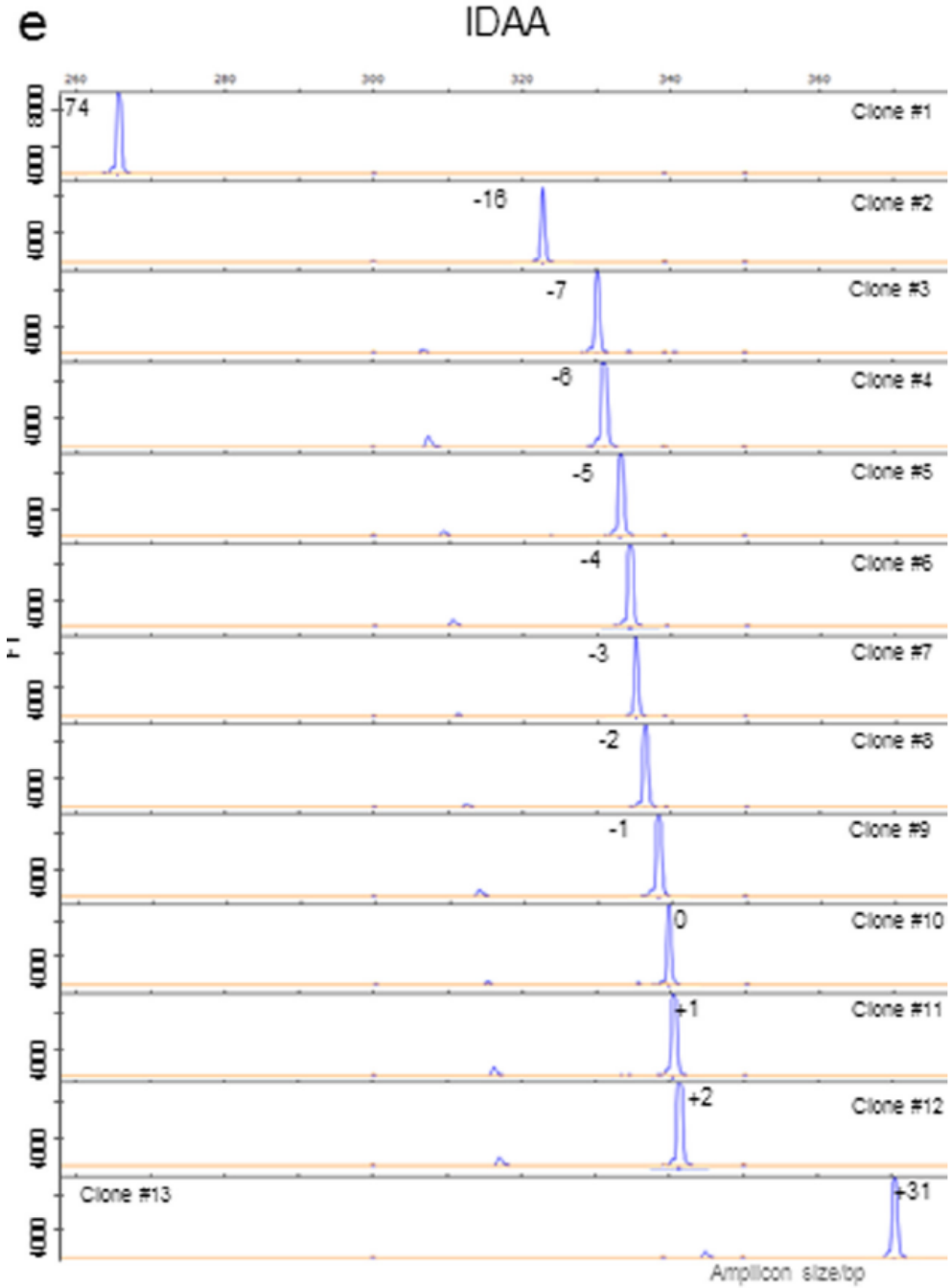

~50 bp amplicon
~100 bp amplicon
~300 bp amplicon
Normalized Melting Peaks
p53 T→C
RB Δ2
RB Δ13
Normalized −d(Fluorescence)/dT
Temperature (°C)

IDAA



Yang Z et al, Fast and sensitive detection of indels induced by precise gene targeting. NAR 2015


a
IDAA
330
350
Fluorescence intensity (FI)
7000
4000
1000
0
339
Cas9
-15
(2.6%)
0
+1
(20%)
339 340
Cosmc-gRNA1
-13
(0.5%)
-11
(1.9%)
-8
(5%)
-4
(7%)
0
+1
(32%)
331
335
339 340
Cosmc-gRNA2
0
339
Cosmc-gRNA3
0
339
Cosmc-gRNA4
Amplicon size/bp


e
IDAA
260
280
300
320
340
360
74
Clone #1
-16
Clone #2
-7
Clone #3
-6
Clone #4
-5
Clone #5
-4
Clone #6
-3
Clone #7
-2
Clone #8
-1
Clone #9
0
Clone #10
+1
Clone #11
+2
Clone #12
Clone #13
+31
Amplicon size/bp

Types of indel screening

- EMC (Enzyme Mismatch Cleavage)
  - Surveyor, T7E1, etc
- HRM (High Resolution Melt)
- IDAA (Indel Detection by Amplicon Analysis)
- PCR
- Sequencing
  - NGS
  - Sanger

Amplicon size

- EMC: Surveyor: 400-800
- HRM: 50-100 bp
- IDAA: 250-450
- NGS: Depends on kit

Which is better?

- EMC, HRM, IDAA only tell you of the presence of a change, and sometimes the length, but not the exact mutation

- Sequencing gives the exact mutation, but is more work and more expensive

Indel vs Deletion

- Current design recommendations are to use two guides where possible for easier screening, as if both work, you can easily see a size difference in a regular PCR
- But if only one works, there may still be an indel based mutation in one site, but it won't be detected by a standard PCR

Other issues

- Large deletions
  - Long Range PCR is very tricky
- Blastocysts give messy PCR/sequence
- G/C rich sequence

Contact Info

Phone: 2470

Email: shifra.ben-dor@weizmann.ac.il

http://bip.weizmann.ac.il/toolbox/target/dna/cripsr.html