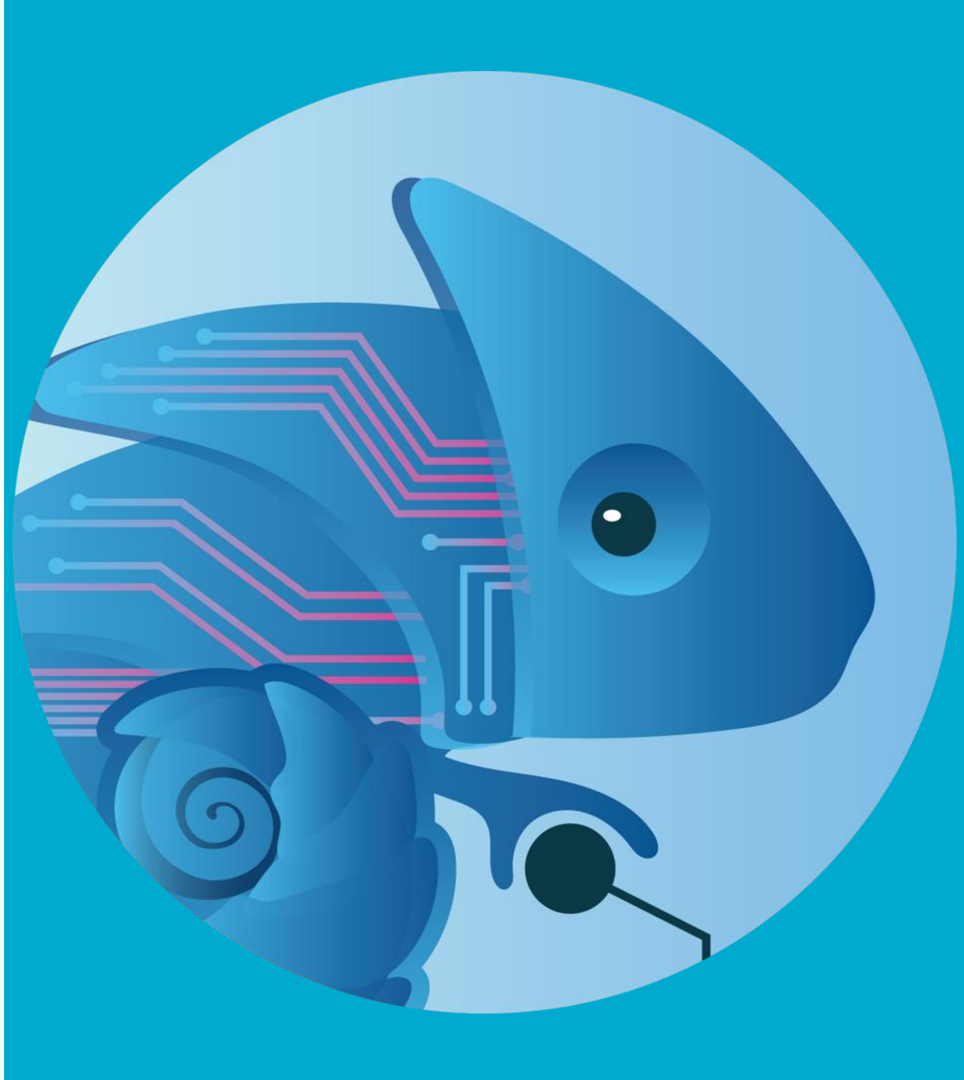
Australian e-Health
Research Centre

Data Visualisation

CRISPR cas9 guide RNA design tools

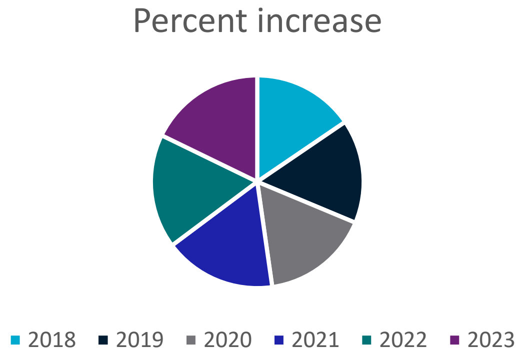
Yatish Jain |@yatish0833 @Tbioinf

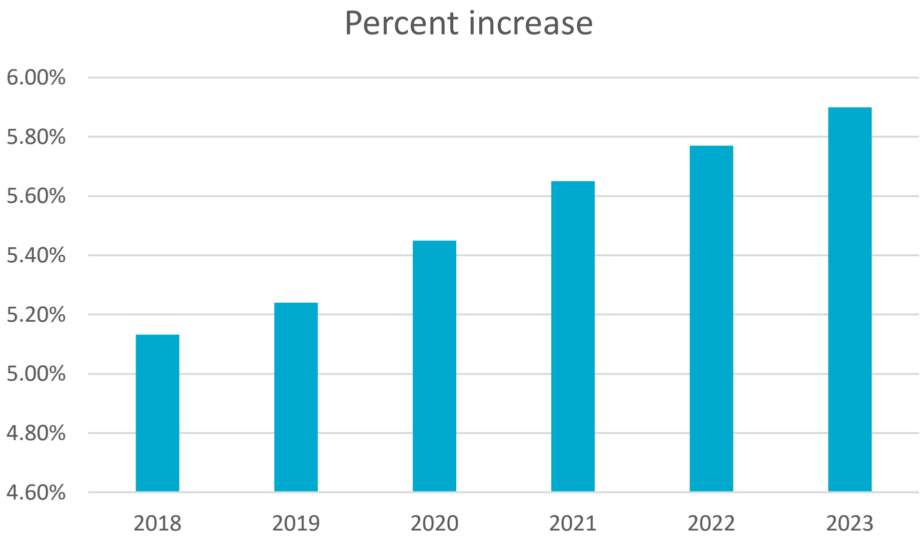
Australia's National Science Agency

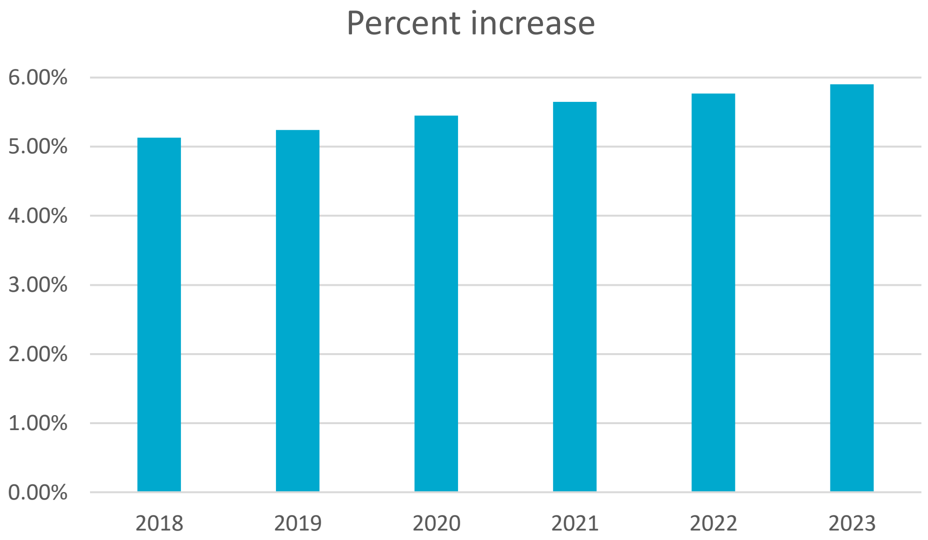


Australian e-Health
Research Centre

Data visualization is important


Percent increase


Percent increase


Percent increase

Guide RNA design pipeline

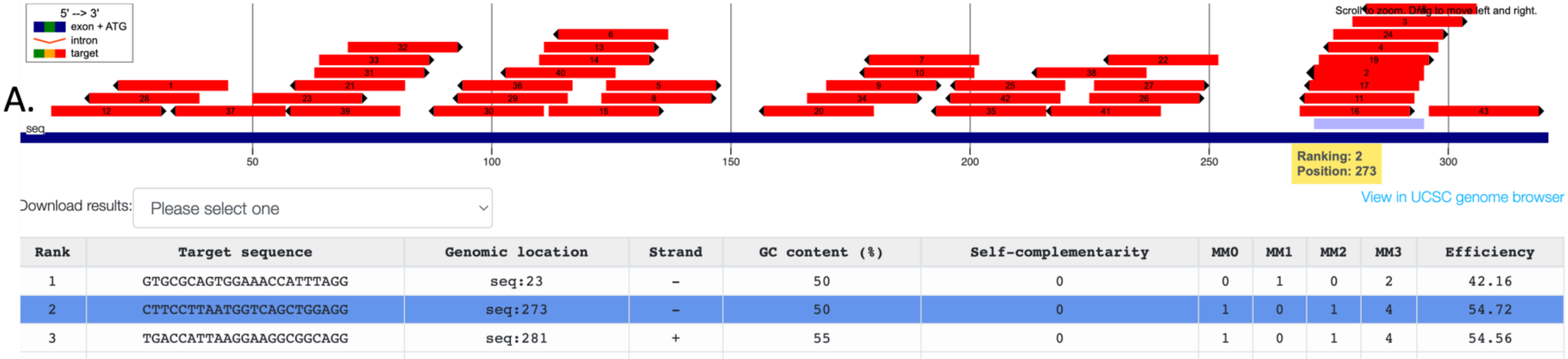
- Identify all CRISPR interference targets
- Filter targets based on scoring metrics/off-targets/orientation
- Predict knockdown efficacies of mismatch gRNA

Important features

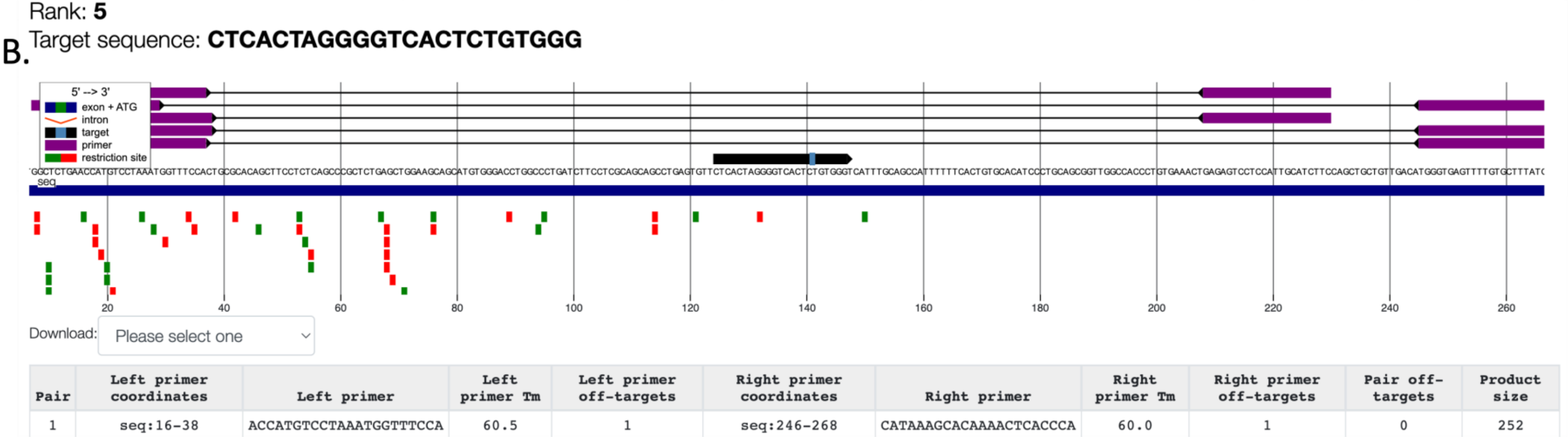
A. Location of targets along with upstream and downstream sequence information.
B. Targets strand/orientation.
C. Represent the scoring metric used and visually represent better targets.
D. Make visualization interactive to show more information about the targets.
E. Demonstrate adherence to specific rules like Protospacer Adjacent Motif (PAM) sequence presence.

Australian e-Health Research Centre

CHOPCHOP



Rank	Target sequence	Genomic location	Strand	GC content (%)	Self-complementarity	MM0	MM1	MM2	MM3	Efficiency
1	GTGCGCAGTGGAAACCATTTAGG	seq:23	-	50	0	0	1	0	2	42.16
2	CTTCCTTAATGGTCAGCTGGAGG	seq:273	-	50	0	1	0	1	4	54.72
3	TGACCATTAAGGAAGGCGGCAGG	seq:281	+	55	0	1	0	1	4	54.56



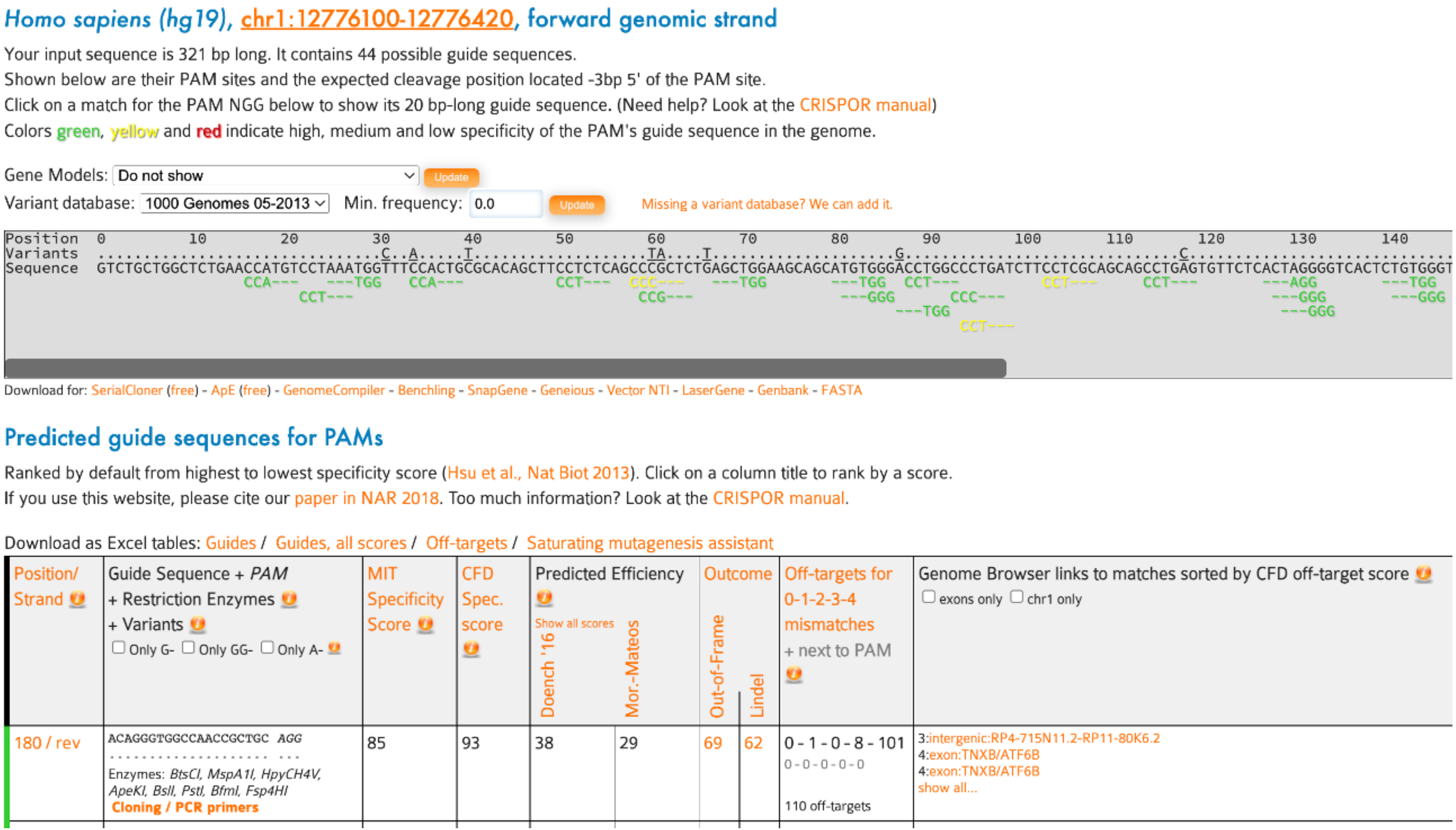
Rank: **5**

Target sequence: **CTCACTAGGGGTCACTCTGTGGG**

Pair	Left primer coordinates	Left primer	Left primer Tm	Left primer off-targets	Right primer coordinates	Right primer	Right primer Tm	Right primer off-targets	Pair off-targets	Product size
1	seq:16-38	ACCATGTCCTAAATGGTTTCCA	60.5	1	seq:246-268	CATAAAGCACAAAACTCACCCA	60.0	1	0	252

Australian e-Health Research Centre

CRISPOR



Homo sapiens (hg19), chr1:12776100-12776420, forward genomic strand

Your input sequence is 321 bp long. It contains 44 possible guide sequences.
Shown below are their PAM sites and the expected cleavage position located -3bp 5' of the PAM site.
Click on a match for the PAM NGG below to show its 20 bp-long guide sequence. (Need help? Look at the CRISPOR manual)
Colors green, yellow and red indicate high, medium and low specificity of the PAM's guide sequence in the genome.

Gene Models: Do not show ⌄ Update
Variant database: 1000 Genomes 05-2013 ⌄ Min. frequency: 0.0 Update Missing a variant database? We can add it.

```
Position  0         10        20        30        40        50        60        70        80        90        100       110       120       130       140
Variants  ...............................C..A.....T...................TA....T....................G..............................C.............................
Sequence  GTCTGCTGGCTCTGAACCATGTCCTAAATGGTTTCCACTGCGCACAGCTTCCTCTCAGCCCGCTCTGAGCTGGAAGCAGCATGTGGGACCTGGCCCTGATCTTCCTCGCAGCAGCCTGAGTGTTCTCACTAGGGGTCACTCTGTGGGT
                          CCA---   ---TGG   CCA---          CCT---  CCC---   ---TGG       ---TGG  CCT---         CCT---     CCT---       ---AGG       ---TGG
                                CCT---                               CCG---                ---GGG      CCC---                             ---GGG       ---GGG
                                                                                                 ---TGG                                    ---GGG
                                                                                                        CCT---
```

Download for: SerialCloner (free) - ApE (free) - GenomeCompiler - Benchling - SnapGene - Geneious - Vector NTI - LaserGene - Genbank - FASTA

Predicted guide sequences for PAMs

Ranked by default from highest to lowest specificity score (Hsu et al., Nat Biot 2013). Click on a column title to rank by a score.
If you use this website, please cite our paper in NAR 2018. Too much information? Look at the CRISPOR manual.

Download as Excel tables: Guides / Guides, all scores / Off-targets / Saturating mutagenesis assistant

Position/ Strand	Guide Sequence + *PAM* + Restriction Enzymes + Variants ☐ Only G- ☐ Only GG- ☐ Only A-	MIT Specificity Score	CFD Spec. score	Predicted Efficiency Show all scores		Outcome		Off-targets for 0-1-2-3-4 mismatches + next to PAM	Genome Browser links to matches sorted by CFD off-target score ☐ exons only ☐ chr1 only
				Doench '16	Mor.-Mateos	Out-of-Frame	Lindel		
180 / rev	ACAGGGTGGCCAACCGCTGC *AGG* .... Enzymes: *BtsCI, MspA1I, HpyCH4V, ApeKI, BslI, PstI, BfmI, Fsp4HI* **Cloning / PCR primers**	85	93	38	29	69	62	0 - 1 - 0 - 8 - 101 0 - 0 - 0 - 0 - 0 110 off-targets	3:intergenic:RP4-715N11.2-RP11-80K6.2 4:exon:TNXB/ATF6B 4:exon:TNXB/ATF6B show all...

Australian e-Health
Research Centre

GT-Scan

Target Details

Location:
1:12776224-12776246

Sequence:
CTCACTAGGGGTCACTCTGTGGG

GC content:
61%

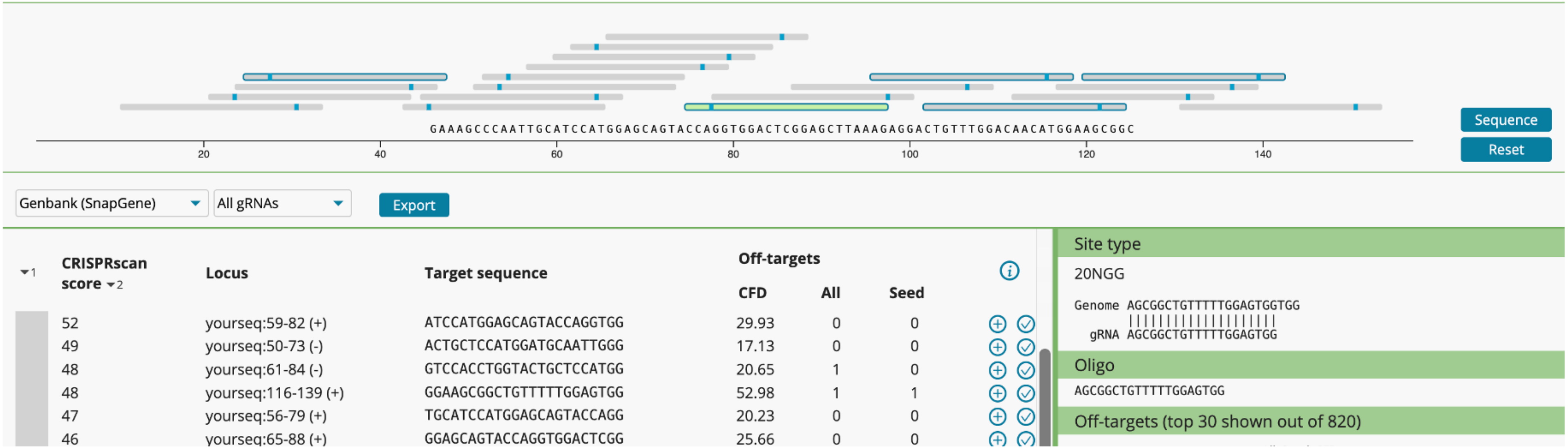
Histones:
H4K91ac
H3K4me1
H4K5ac
H3K4me3
H3K27ac
H3K4me2
H4K8ac
H3K9ac
H4K12ac

Transcription:
0

Location	Strand	Candidate Target Sequence	Predicted Activity	0	1	2	3	Experimental Results
12,776,278	-	CAGGGTGGCCAACCGCTGCAGGG	High	0	0	0	10	No
12,776,224	+	CTCACTAGGGGTCACTCTGTGGG	High	0	0	0	11	No
12,776,150	+	CCTCTCAGCCCGCTCTGAGCTGG	High	0	0	0	16	No

Australian e-Health
Research Centre

CRISPRSCAN



▾1	CRISPRscan score ▾2	Locus	Target sequence	Off-targets CFD	All	Seed
	52	yourseq:59-82 (+)	ATCCATGGAGCAGTACCAGGTGG	29.93	0	0
	49	yourseq:50-73 (-)	ACTGCTCCATGGATGCAATTGGG	17.13	0	0
	48	yourseq:61-84 (-)	GTCCACCTGGTACTGCTCCATGG	20.65	1	0
	48	yourseq:116-139 (+)	GGAAGCGGCTGTTTTTGGAGTGG	52.98	1	1
	47	yourseq:56-79 (+)	TGCATCCATGGAGCAGTACCAGG	20.23	0	0
	46	yourseq:65-88 (+)	GGAGCAGTACCAGGTGGACTCGG	25.66	0	0

Site type

20NGG

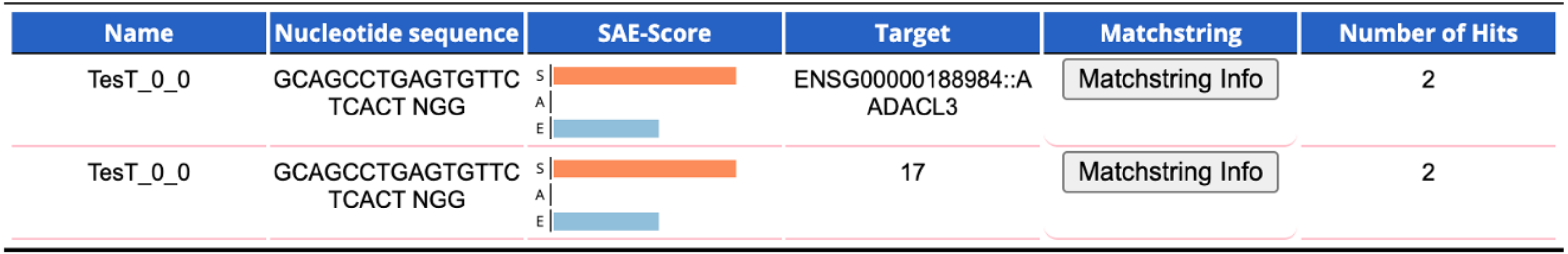
```
Genome AGCGGCTGTTTTTGGAGTGGTGG
       ||||||||||||||||||||
  gRNA AGCGGCTGTTTTTGGAGTGG
```

Oligo

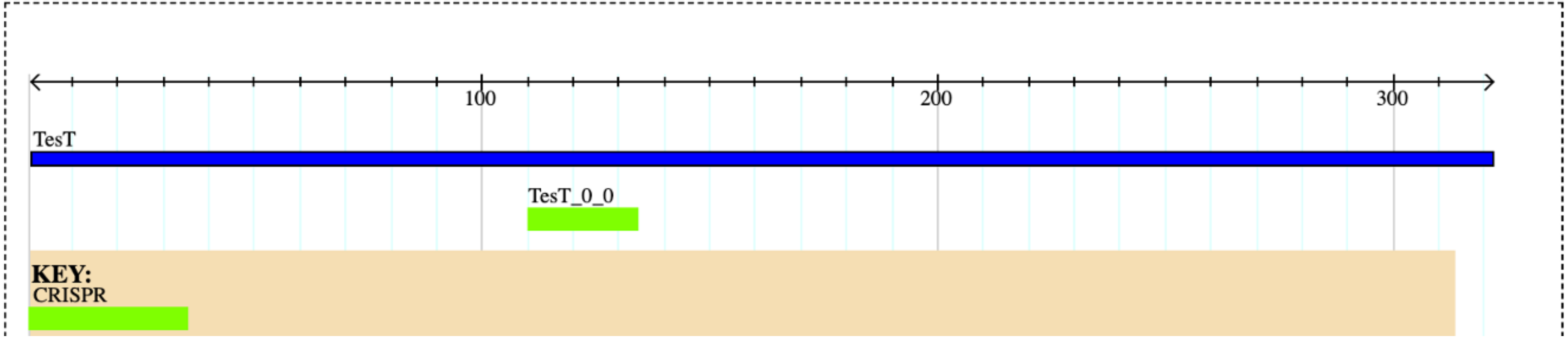
AGCGGCTGTTTTTGGAGTGG

Off-targets (top 30 shown out of 820)

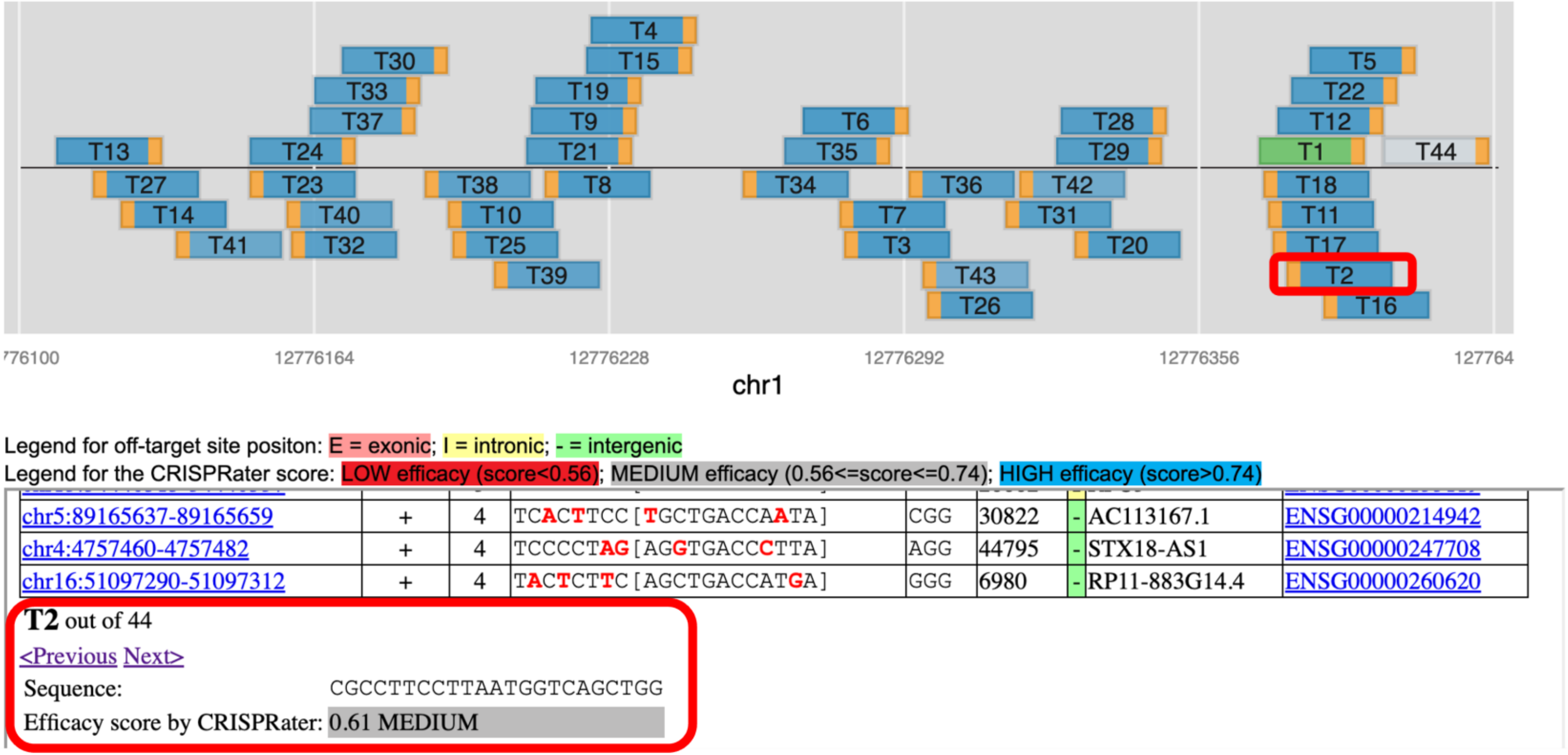
E-CRISP



Name	Nucleotide sequence	SAE-Score	Target	Matchstring	Number of Hits
TesT_0_0	GCAGCCTGAGTGTTC TCACT NGG	S A E	ENSG00000188984::AADACL3	Matchstring Info	2
TesT_0_0	GCAGCCTGAGTGTTC TCACT NGG	S A E	17	Matchstring Info	2



CCTOP



Legend for off-target site positon: E = exonic; I = intronic; - = intergenic

Legend for the CRISPRater score: LOW efficacy (score<0.56); MEDIUM efficacy (0.56<=score<=0.74); HIGH efficacy (score>0.74)

chr5:89165637-89165659	+	4	TCACTTCC[TGCTGACCAATA]	CGG	30822	-	AC113167.1	ENSG00000214942
chr4:4757460-4757482	+	4	TCCCCTAG[AGGTGACCCTTA]	AGG	44795	-	STX18-AS1	ENSG00000247708
chr16:51097290-51097312	+	4	TACTCTTC[AGCTGACCATGA]	GGG	6980	-	RP11-883G14.4	ENSG00000260620

T2 out of 44

<Previous Next>

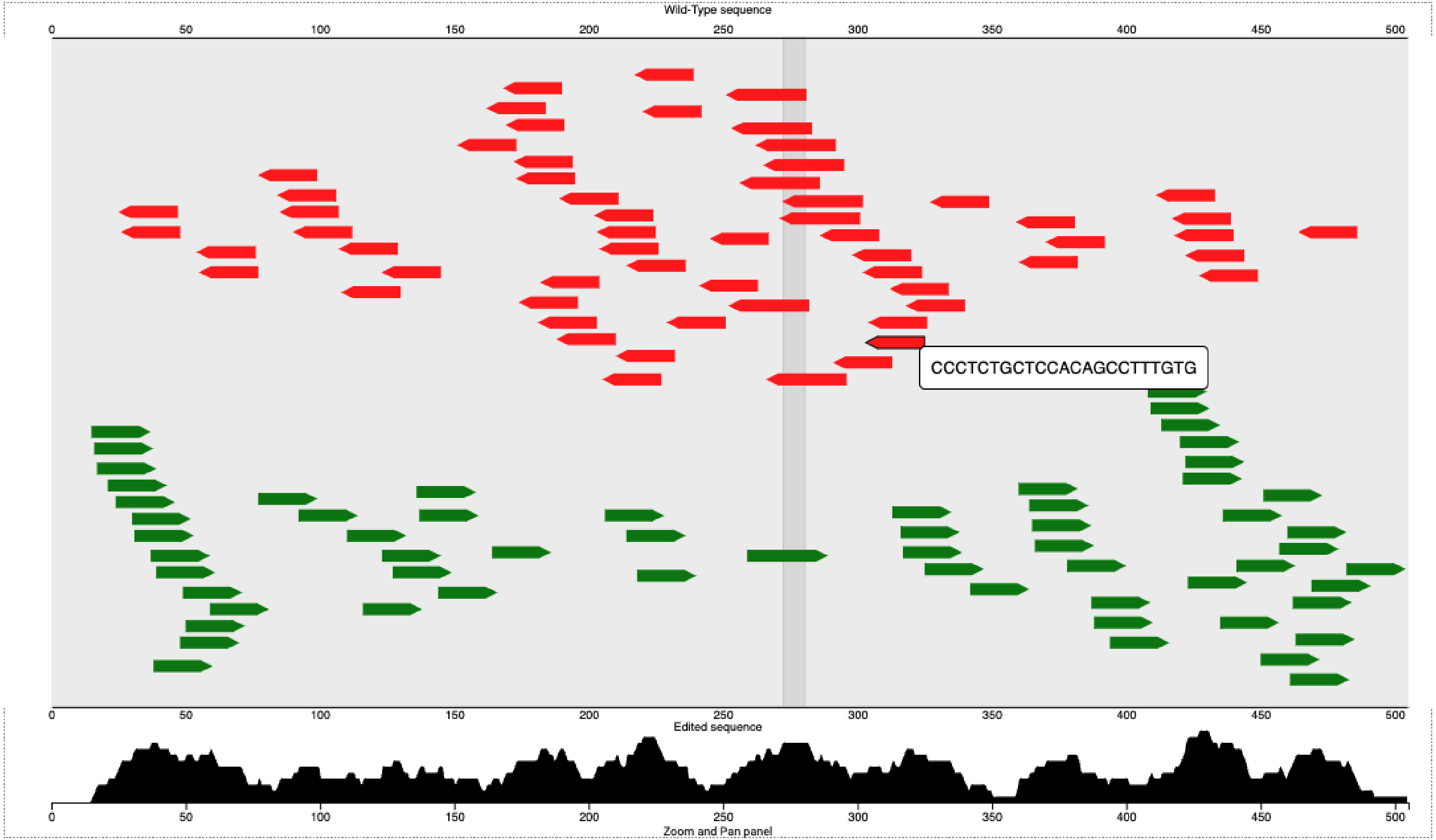
Sequence:	CGCCTTCCTTAATGGTCAGCTGG
Efficacy score by CRISPRater:	0.61 MEDIUM

Challenges

- Depict all available guides in specified search space.
    - Zoomed out pane with coordinates
    - Feature viewer package or D3 libraries – Ex PETAL
- Filter targets using scoring metrics visually.
    - Scoring metric to color code guides
    - Directional arrows representing orientation
- Visualising downstream information.
    - Hover over text with actual sequence, off target information, GC content etc.

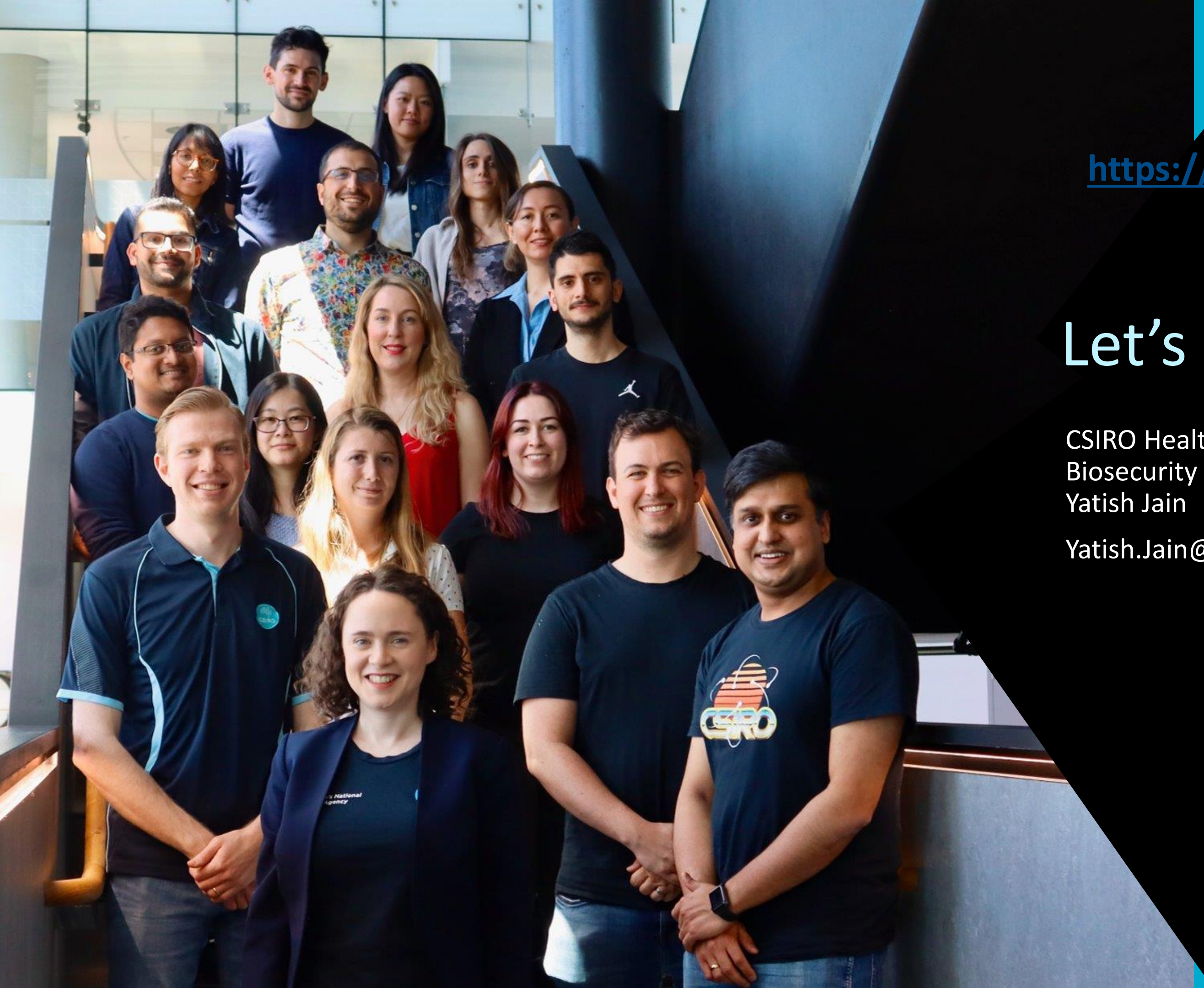
Australian e-Health Research Centre

PETAL



Quick takeaways

- Efficient visualization can save both time and resources.
    - Quick summaries and inferences
- Identifying important features to display can improve the design process.
    - Requirement gathering – what is your visualization summarizing?
- More testing and quality control measures needed.
    - Visualisation regression testing, Multiview visualization
    - User acceptance testing



Visit our site:
https://bioinformatics.csiro.au

Let’s work together

CSIRO Health and
Biosecurity
Yatish Jain

Yatish.Jain@csiro.au

@yatish0833
@Tbioinf

Australian e-Health
Research Centre