Abstract number: 209

CALITAS: a CRISPR/Cas-aware ALigner for In silico off-TArget Search

Tim Fennell[1], Meltem Isik[2], Tongyao Wang[2], Gregory Gotta[2], Deric Zhang[2], Christopher J. Wilson[2], Eugenio Marco*[2]

[1] Fulcrum Genomics; [2] Editas Medicine, 11 Hurley Street, Cambridge, MA 02141; * Presenter

Introduction

Assessing the specificity of CRISPR/Cas medicines is important

In silico off-target prediction methods are used together with experimental methods like GUIDE-seq and Digenome to assemble lists of candidate off-target sites

Current *in silico* off-target prediction tools lack features such as flexible PAM or PAMless searches, addition of multiple bulges, inclusion of variants or returning unique lists of sites

To address these issues we have developed CALITAS

CALITAS: a CRISPR/Cas-aware aligner

We present CALITAS, a **C**RISPR/Cas-aware **AL**igner for **I**n silico off-**TA**rget **S**earch

with the following features:

- User-defined maximum number of gRNA mismatches and gaps
- Mismatches in the PAM are tolerated
- Ability to use multiple PAM sequences or no PAM
- Option to produce either the single best alignment per off-target site or all alignments meeting mismatch/gap limits
- Ability to set base pair overlap cutoff for differentiating unique adjacent alignments
- Similar penalties for mismatches, gRNA and DNA gaps
- Ability to align against alternate alleles in the reference, via user-provided VCF files, for example from the 1000 Genomes Project

CALITAS uses a modified Needleman-Wunsch algorithm

gRNA and DNA gaps result in different number of basepair matches

1 gRNA gap (genome bulge)

```
tttnAGG-AAACTTCTGGCAGGACC
|||||||~||||||||||||||||
TTTCAGGTAAACTTCTGGCAGGACC
```

4 PAM matches + 19 guide matches
1 gRNA gap

1 DNA gap (guide bulge)

```
tttnAGGAAACTTCTGGCAGGACC
||||||||~|||||||||||||||
TTTCAGGA-ACTTCTGGCAGGACC
```

4 PAM matches + 18 guide matches
1 DNA gap

CALITAS uses net cost parameters to give similar penalties for mismatches, gRNA gaps and DNA gaps (which is not found in standard NW aligners)

Key CALITAS parameters

User-tunable net cost parameters

```
guide-mismatch-net-cost = -120

genome-bulge-net-cost = -121

guide-bulge-net-cost = -122

pam-mismatch-net-cost = -260
```

Internal NW penalties

guide match = 60
PAM match = 130
guide mismatch = – 60
gRNA gap = – 61
DNA gap = – 2
PAM mismatch = –130

Similar penalties with biases to standardize alignments scores

Aligner 'preference':

1. Guide mismatch (most preferred)
2. gRNA gap
3. DNA gap
4. PAM mismatch (least preferred)

Examples of CALITAS Alignments and Scores

Alignment example	Internal NW calculation	Net cost	CALITAS score (Higher is better)
Perfect Match tttnAGGAAACTTCTGGCAGGACC ||||||||||||||||||||||| TTTCAGGAAACTTCTGGCAGGACC 4 PAM matches + 20 guide matches	4 * 130 + 20 * 60	0	**1720**
1 Mismatch tttnAGGAAACTTCTGGCAGGACC |||||.||||||||||||| TTTCACGAAACTTCTGGCAGGACC 4 PAM matches + 19 guide matches 1 guide mismatch	4 * 130 + 19 * 60 − 60	− 120	**1600**
1 gRNA gap (genome bulge) tttnAGG-AAACTTCTGGCAGGACC |||||||~||||||||||||| TTTCAGGTAAACTTCTGGCAGGACC 4 PAM matches + 19 guide matches 1 gRNA gap	4 * 130 + 19 * 60 − 61	− 121	**1599**
1 DNA gap (guide bulge) tttnAGGAAACTTCTGGCAGGACC ||||||||~|||||||||||| TTTCAGGA-ACTTCTGGCAGGACC 4 PAM matches + 18 guide matches 1 DNA gap	4 * 130 + 18 * 60 − 2	− 122	**1598**
1 PAM mismatch tttnAGGAAACTTCTGGCAGGACC |.||||||||||||||||||| TATCAGGAAACTTCTGGCAGGACC 3 PAM matches + 20 guide matches 1 PAM mismatch	3 * 130 + 20 * 60 − 130	− 260	**1460**

CALITAS uses a two-step approach to find CRISPR/Cas alignments

Step 1

Align gRNA to DNA sequence without PAM and with user-defined maximum number of mismatches and gaps

Step 2

If PAM(s) present extend alignment, allowing guide-PAM gap and mismatches

Searches can be performed with one or multiple PAMs or PAMless

PAM extension with perfect match

```
ATTGAGATAGTGTGGGGAAG
|.|||||~|||||||||~||
TGTTGTCTGAAATGAGA-AGTGTGGGG-AGTGGAGTAG
```

```
ATTGAGATAGTGTGGGGAAGnrg
|.|||||~|||||||||~|||||
TGTTGTCTGAAATGAGA-AGTGTGGGG-AGTGGAGTAG
```

PAM extension with RNA gap

```
ATTGAGATAGTGTGGGGAAG
|.||||||||||.||||||
AGGAATATCCAATGAGATAGTGTAGGGAAGCCGGGTTT
```

```
ATTGAGATAGTGTGGGGAAG-nrg
|.||||||||||.|||||~|||
AGGAATATCCAATGAGATAGTGTAGGGAAGCCGGGTTT
```

PAM extension with mismatch

```
ATTGAGATAGTGTGGGGAAG
|.||||||||||.||||||
AGGAATATCCAATGAGATAGTGTAGGGAAGCCGGGTTT
```

```
ATTGAGATAGTGTGGGGAAGnrg
|.||||||||||.||||||||.
AGGAATATCCAATGAGATAGTGTAGGGAAGCCTGGTTT
```

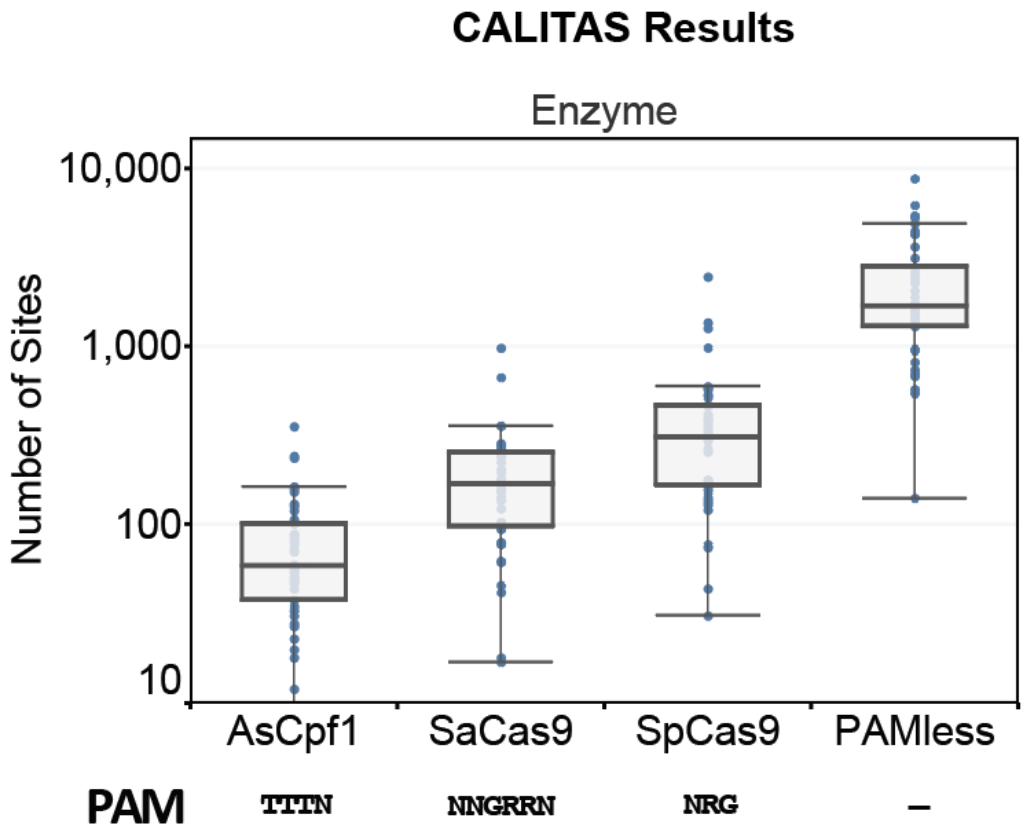
	CALITAS score
PAM extension with perfect match	1226
PAM extension with RNA gap	1229
PAM extension with mismatch	1090

This results in a final unique alignment with the best PAM selected

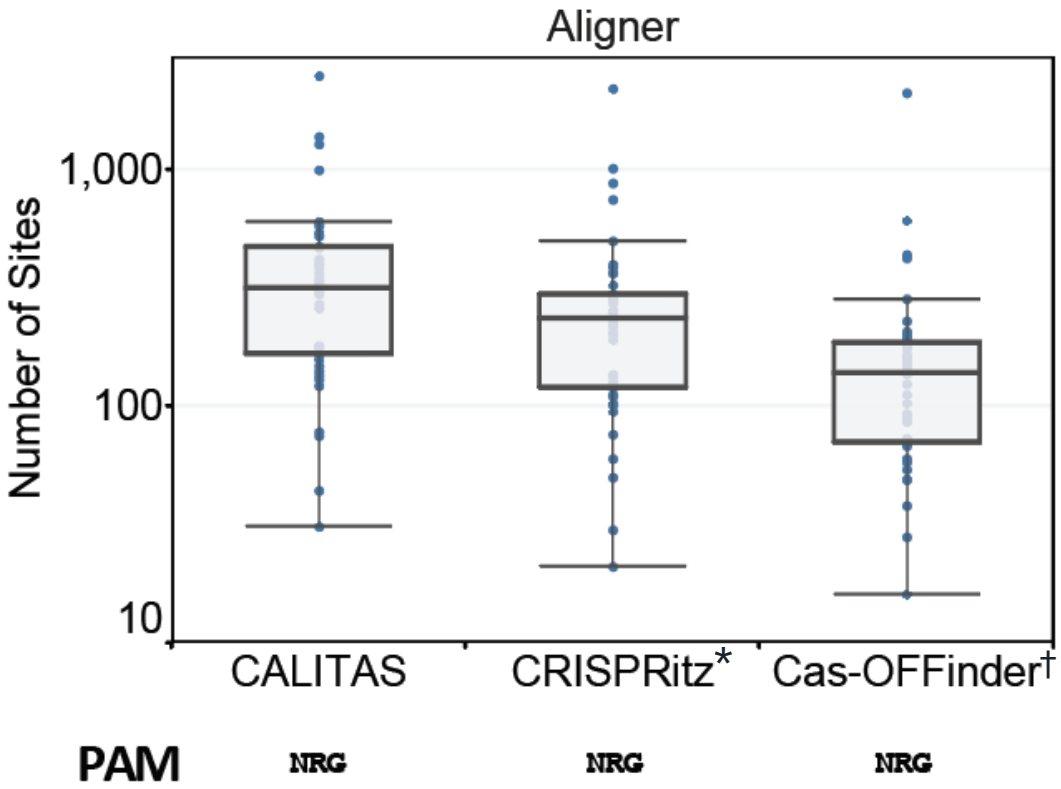
AsCas12a has fewer predicted off-targets by CALITAS, followed by SaCas9 and SpCas9

We used CALITAS to make in silico predictions for 41 gRNAs with PAMs for AsCas12a, SaCas9 and SpCas9

Sites up to 3 Mismatches plus Gaps



CALITAS predicts more off-target sites than CRISPRitz or Cas-OFFinder



*, † for CRISPRitz and Cas-OFFinder redundant sites were removed using `bedtools cluster` and `pandas groupby`

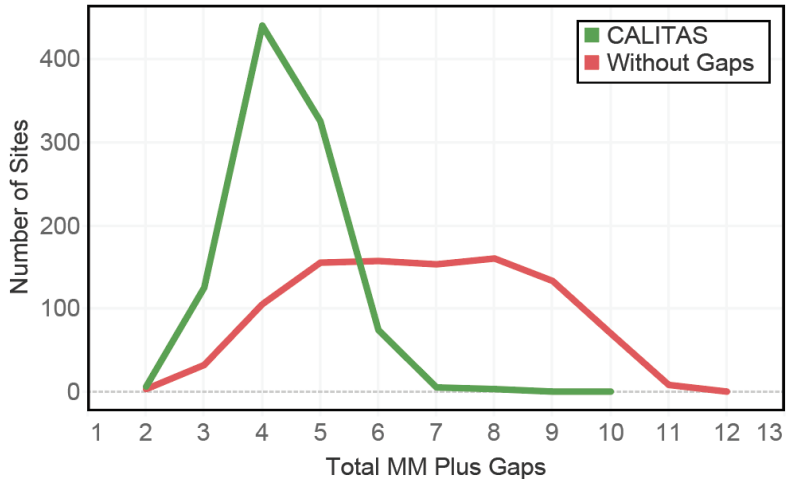
* As an example, CRISPRitz returned a file with 2,379,786 rows, that could be reduced to 396 unique sites with up to 3 mismatches plus gaps

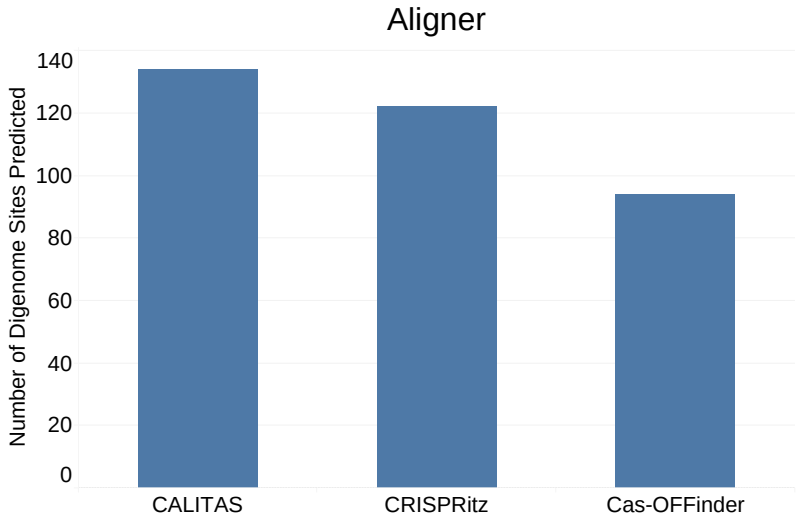
Comparison with other methods shows that CALITAS can predict more off-target sites, allowing for a more comprehensive search

Importantly, CALITAS returns a unique list of sites, suitable for building off-target verification panels

Multiple gaps are present in off-targets

We compared CALITAS alignments to 987 experimental Digenome sites, with or without gaps





Alignments with gaps are better

Alignment with gaps

```
ATTGAGATAGTGTGGGGAAGnrg
|.|||||~|||||||||~|||||
AATGAGA-AGTGTGGGG-AGTGG
```

1 mismatch
2 gaps (RNA bulges)

Alignment without gaps

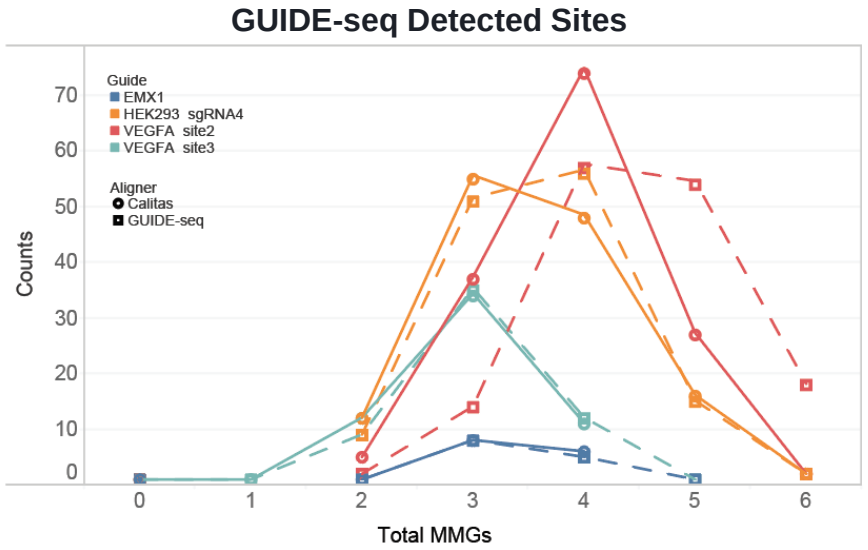
```
ATTGAGATAGTGTGGGGAAGnrg
|.||...||||....|||..|||
AGTGGAGTAGTTCCTGGACAGGG
```

10 mismatches

CALITAS has the highest number of predictions up to 3 mismatches and gaps that are confirmed as detected in Digenome

CALITAS aligns GUIDE-seq sites with fewer mismatches plus gaps

We compared CALITAS alignments with the alignments without gaps in the original GUIDE-seq paper



To evaluate CALITAS off-target predictions, we compared the total number of in silico-predicted CALITAS sites with the ones detected by GUIDE-seq

		Total Mismatches plus Gaps					
Guide		0	1	2	3	4	5
EMX1	CALITAS Predicted	1	0	6	342	8,366	129,574
	Guide-Seq Detected	1	0	3	7	5	0
	Fraction Detected	100.00%		50.00%	2.05%	0.06%	0%
HEK293 sgRNA4	CALITAS Predicted	1	0	41	1,942	16,042	166,747
	Guide-Seq Detected	1	0	14	58	50	9
	Fraction Detected	100.00%		34.15%	2.99%	0.31%	0.01%
VEGFA site2	CALITAS Predicted	1	0	26	483	11,713	72,396
	Guide-Seq Detected	1	0	6	38	76	27
	Fraction Detected	100.00%		23.08%	7.87%	0.65%	0.04%
VEGFA site3	CALITAS Predicted	1	1	89	2,518	45,605	283,577
	Guide-Seq Detected	1	1	13	36	9	0
	Fraction Detected	100.00%	100.00%	14.61%	1.43%	0.02%	0%

CALITAS alignments with gaps have fewer total number of mismatches and gaps, suggesting that CALITAS is better suited for building off-target verification panels

Very few of the CALITAS-predicted off-target sites with 4 mismatches and gaps are detected, suggesting that verification panels should include in silico predictions up to 3 mismatches and gaps

Data from Tsai et al, *Nat Biotechnol.* 2015;33:187

CALITAS can incorporate variants from a standard VCF file

Step 1

Identify individual variants/alleles that are present above some threshold frequency (e.g. 1%)

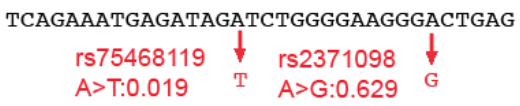
Filtered variants 1000 Genomes

rs75468119:54989257:A>T:0.019
rs2371098:54989270:A>G:0.629

Step 2

Identify putative short-range haplotypes by linking variants that are within 1 guide length of each other

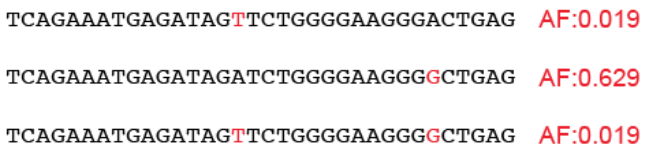
hg38



Step 3

Assemble modified sequences that include the individual variants and/or haplotype variants

hg38 + variants

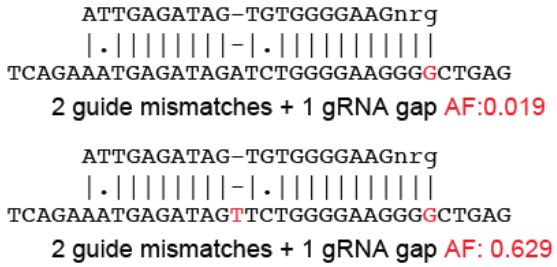


We test all possibilities, not filtered by haplotypes observed in a population

Step 4

Search the modified sequences for putative alignments

1000 Genomes alignment



Summary

CALITAS is a new state-of-the-art aligner useful for *in silico* prediction of CRISPR/Cas off-target sites

Features include:

- User-defined maximum number of gRNA mismatches and gaps
- Mismatches in the PAM are tolerated
- Ability to use multiple PAM sequences or no PAM
- Option to produce single best alignment or all alignments per off-target site
- Similar penalties for mismatches, gRNA and DNA gaps
- Ability to align against alternate alleles in the reference, via user-provided VCF files

Comparison with experimental data shows the importance of including multiple gaps

Comparison with CRISPRitz and Cas-OFFinder shows that CALITAS' off-target site list is more comprehensive and more CALITAS' predicted sites are detected by Digenome

Acknowledgements

We would like to thank the following Editas teams for supporting this project: Sequencing, Screening, Sample Management, Bioinformatics, and Scientific Communications

Disclosures:

Employees and shareholders of Editas Medicine: M.I., T.W., G.G., D.Z., C.J.W., E.M.