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

Tim Fennell¹, Meltem Isik², Tongyao Wang², Gregory Gotta², Deric Zhang², Christopher J. Wilson², Eugenio Marco*²

¹ Fulcrum Genomics; ² Editas Medicine, 11 Hurley Street, Cambridge, MA 02141; * Presenter
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
We present CALITAS, a CRISPR/Cas-aware Aligner for In silico off-Target Search 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

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

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

<table>
<thead>
<tr>
<th>Alignment example</th>
<th>Internal NW calculation</th>
<th>Net cost</th>
<th>CALITAS score (Higher is better)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Perfect Match</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tttAGGAACCTCGGCAGGACC</td>
<td>4 * 130 + 20 * 60</td>
<td>0</td>
<td>1720</td>
</tr>
<tr>
<td>TTTCAAGAAACCTCGGCAGGACC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1 Mismatch</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tttAGGAACCTCTGGCAGGACC</td>
<td>4 * 130 + 19 * 60 - 60</td>
<td>- 120</td>
<td>1600</td>
</tr>
<tr>
<td>TTTCAAGAAACCTCTGGCAGGACC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1 gRNA gap (genome bulge)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tttAGGAAACTCTGGCAGGACC</td>
<td>4 * 130 + 19 * 60 - 61</td>
<td>- 121</td>
<td>1599</td>
</tr>
<tr>
<td>TTTCAAGGAAACTCTGGCAGGACC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1 DNA gap (guide bulge)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tttAGGAACCTCTGGCAGGACC</td>
<td>4 * 130 + 18 * 60 - 2</td>
<td>- 122</td>
<td>1598</td>
</tr>
<tr>
<td>TTTAGGAAACCTCTGGCAGGACC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>1 PAM mismatch</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tttAGGAACCTCTGGCAGGACC</td>
<td>3 * 130 + 20 * 60 - 130</td>
<td>- 260</td>
<td>1460</td>
</tr>
<tr>
<td>TTTAGGAAACCTCTGGCAGGACC</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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**

<table>
<thead>
<tr>
<th>ATTGAGATAGTGTGGGGAAG</th>
<th>ATTGAGATAGTGTGGGGAAGnrg</th>
</tr>
</thead>
<tbody>
<tr>
<td>TGTGTCTGAAATGAGA-AGTGTTGGG-AGTGAGTAG</td>
<td>TGTGTCTGAAATGAGA-AGTGTTGGG-AGTGAGTAG</td>
</tr>
</tbody>
</table>

**PAM extension with RNA gap**

<table>
<thead>
<tr>
<th>ATTGAGATAGTGTGGGGAAG</th>
<th>ATTGAGATAGTGTGGGGAAG-nrg</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGGAATATCCAATGAGATRGGTGGGAGCCGGGT</td>
<td>AGGAATATCCAATGAGATRGGTGGGAGCCGGGT</td>
</tr>
</tbody>
</table>

**PAM extension with mismatch**

<table>
<thead>
<tr>
<th>ATTGAGATAGTGTGGGGAAG</th>
<th>ATTGAGATAGTGTGGGGAAGnrg</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGGAATATCCAATGAGATRGGTGGGAGCCGGGT</td>
<td>AGGAATATCCAATGAGATRGGTGGGAGCCGGGT</td>
</tr>
</tbody>
</table>

CALITAS score

1226

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

1229

1090
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

TTTCAggaaactttctggcaggacc agggaT

CALITAS Results

Sites up to 3 Mismatches plus Gaps

Enzyme

PAM

TTTN

NNGRNN

NRG

PAMless

Number of Sites

10,000

1,000

100

10

AsCpf1

SaCas9

SpCas9

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

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.

* 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.

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

© 2020 Editas Medicine
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**

- **ATTGAGATAGTGGGGAAGnrg**
  - 1 mismatch
  - 2 gaps (RNA bulges)

**Alignment without gaps**

- **ATTGAGATAGTGGGGAAGnrg**
  - 10 mismatches

CALITAS has the highest number of predictions up to 3 mismatches and gaps that are confirmed as detected in Digenome.
We compared CALITAS alignments with the alignments without gaps in the original GUIDE-seq paper.

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

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

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%)

**Step 2**
Identify putative short-range haplotypes by linking variants that are within 1 guide length of each other

**Step 3**
Assemble modified sequences that include the individual variants and/or haplotype variants

**Step 4**
Search the modified sequences for putative alignments

*Filtered variants*  
1000 Genomes  
rs75468119:54989257:A>T:0.019  
rs2371098:54989270:A>G:0.629

*hg38*  
TCAGAAATGAGATAGATCTGGGAAGGGACTGAG  
rs75468119  
A>T:0.019  
rs2371098  
A>G:0.629

*hg38 + variants*  
TCAGAAATGAGATAGATCTGGGAAGGGACTGAG  
AF:0.019  
TCAGAAATGAGATAGATCTGGGAAGGGACTGAG  
AF:0.629  
TCAGAAATGAGATAGATCTGGGAAGGGACTGAG  
AF:0.019

1000 Genomes alignment  
2 guide mismatches + 1 gRNA gap  
AF:0.019

We test all possibilities, not filtered by haplotypes observed in a population
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
We would like to thank the following Editas teams for supporting this project: Sequencing, Screening, Sample Management, Bioinformatics, and Scientific Communications

Disclosures: