# Novel CRISPR/Cas13 systems and uses thereof

## Reference to related applications

This application claims the benefit of Korean Patent Application No. 10-2021-164539 filed on 25 November 2021, the entire contents of which, including any sequence listing and drawings, are incorporated in their entirety herein by reference.

## Field of the invention

The present invention relates to the field of RNA editing, particularly to novel CRISPR effector enzymes and systems and methods for CRISPR based RNA-targeting. In particular the invention relates to a novel CRISPR enzyme (alternatively referred to as a CRISPR protein, a Cas effector protein, a Cas enzyme or a Cas protein) and compositions and systems thereof, and their use in a nucleic acid detection system.

## Background of the invention

CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated protein) have revolutionized biological and biomedical sciences in many ways.

CRISPR–Cas-mediated genome editing has emerged as a versatile molecular biological tool with applications expanded to every corner of life science. In addition, CRISPR-Cas enabled assays have already established critical roles in clinical diagnostics and biosensing. The recent discovery of the *trans*-cleavage activities of Cas12 and Cas13 family proteins, where the target triggered collateral cleavage of non-target ssDNA or ssRNA, made them an ideal toolbox for biosensing. Class 2 CRISPR–Cas systems, including Cas9, Cas12 family, Cas13 family, harness only a single Cas effector to modulate nuclease function and has thus become the most widely studied and utilized CRISPR-Cas systems.

Cas13 (formerly C2c2) is the only family of class 2 Cas enzymes known to exclusively target single-stranded RNA. Cas13, when assembled with a CRISPR RNA (crRNA) forms a crRNA-guided RNA-targeting effector complex. The RNA-guided RNA-targeting CRISPR/Cas13 therefore has a great potential for diagnosis, therapy, and

research owing to its biochemical properties including higher RNA digestion efficiency compared to the traditional RNAi, while simultaneously exhibiting much less off-target cleavage compared to RNAi.

Cas13 proteins are present in at least 21 bacterial genomes (Shmakov S et al. (2015) Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. Mol Cell. 60(3):385–97). CRISPR/Cas13 has been used for the highly efficient and specific degradation of RNAs (referred to as cis-cleavage activity) both *in vitro* and *in vivo*. This relies on the cleavage of the targeted RNAs by the endogenous RNase activity of the dual higher eukaryotes and prokaryotes nucleotide (HEPN) domains of the protein. They all have two HEPN domains, but the size of the full-length protein can vary depending on which bacteria they come from. For this reason, the CRISPR/Cas13 has different subtypes, including Cas13a, Cas13b, Cas13c, and Cas13d. The protospacer flanking sequence (PFS) for Cas13, which is analogous to the PAM sequence for Cas9, is located at the 3' end of the spacer sequence and consists of a single A, G, U, or C base pair.

Cas13 protein can also cleave other RNAs non-specifically when activated. When the Cas13 recognizes its target *in vitro*, it becomes activated and then subsequently promiscuously cleaves RNA species in solution regardless of homology to the crRNA or presence of a PFS. This property, also referred to as collateral cleavage activity, or trans-collateral activity, provides a key characteristic for a nucleic acid detection platform based on the CRISPR/Cas13 system.

As CRISPR-Cas13 systems can be used for a wide range of applications and some CRISPR systems are better suited for certain applications than others, it is critical to pair the appropriate CRISPR/Cas13 systems to the appropriate applications. By finding new CRISPR/Cas13 systems and expanding the CRISPR toolkit, it is possible to enable and improve many of CRISPR/Cas13's applications.

## Summary of the invention

Provided herein are new Cas13 polypeptides and compositions thereof, together with CRISPR/Cas13 systems comprising the new Cas13 polypeptides, methods for CRISPR/Cas13 based RNA-targeting, and a nucleic acid detection system using the Cas 13 polypeptides, compositions and systems of the invention. The compositions include

either the Cas13 polypeptide itself, or a nucleic acid molecule comprising a sequence encoding a Cas13 polypeptide. There is also provided CRISPR-Cas13 systems comprising the Cas13 polypeptide, or a nucleic acid molecule that encodes a Cas13 polypeptide, together with a sequence encoding a CRISPR RNA (crRNA) comprising one or more spacers and one or more Cas13-specific direct repeats.

One aspect of the invention provides a Cas13 polypeptide, or a nucleotide sequence encoding said Cas13 polypeptide. There is also provided a composition comprising a Cas13 polypeptide, or a nucleotide sequence encoding said Cas13 polypeptide. In preferred embodiments of Cas13 polypeptide compositions, CRISPR/Cas13 systems, methods for CRISPR/Cas13 based RNA-targeting, and nucleic acid detection systems, the Cas13 polypeptide is a Cas13a polypeptide, a Cas13b polypeptide, or a Cas13d polypeptide. Similarly, in the Cas13 polypeptide compositions, CRISPR/Cas13 systems, methods for CRISPR/Cas13 based RNA-targeting, and nucleic acid detection systems that include a nucleic acid molecule comprising a sequence encoding a Cas13 polypeptide, the nucleic acid molecule encodes a Cas13a polypeptide, a Cas13b polypeptide, or a Cas13d polypeptide.

The Cas13 polypeptides of the invention have an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30.

The Cas13 polypeptides of the invention are encoded by a nucleic acid molecule selected from the group consisting of SEQ ID NOS: 1-15, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 1-15.

In embodiments wherein the Cas13 polypeptide is a Cas13a polypeptide, the sequence encoding the Cas13a polypeptide is selected from SEQ ID NO: 1, or SEQ ID NO:2, or SEQ ID NO: 3, or SEQ ID NO:4, or SEQ ID NO: 5, or SEQ ID NO:6, or SEQ ID NO:7, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 1, or SEQ ID NO:2, or SEQ ID NO: 3, or SEQ ID NO:4, or SEQ ID NO: 5, or SEQ ID NO:6, or SEQ ID NO:7.

In embodiments wherein the Cas13 polypeptide is a Cas13a polypeptide, the Cas13a polypeptide has an amino acid sequence of SEQ ID NO: 16, or SEQ ID NO:17, or SEQ ID NO: 18, or SEQ ID NO:19, or SEQ ID NO: 20, or SEQ ID NO:21, or SEQ ID NO:22, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 16, or SEQ ID NO:17, or SEQ ID NO: 18, or SEQ ID NO:19, or SEQ ID NO: 20, or SEQ ID NO:21, or SEQ ID NO:22.

In embodiments wherein the Cas13 polypeptide is a Cas13b polypeptide, the sequence encoding the Cas13b polypeptide is selected from SEQ ID NO: 8, or SEQ ID NO:9, or SEQ ID NO: 10, or SEQ ID NO:11, or SEQ ID NO: 12, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 8, or SEQ ID NO:9, or SEQ ID NO: 10, or SEQ ID NO:11, or SEQ ID NO: 12.

In embodiments wherein the Cas13 polypeptide is a Cas13b polypeptide, the Cas13b polypeptide has an amino acid sequence of SEQ ID NO: 23, or SEQ ID NO:24, or SEQ ID NO: 25, or SEQ ID NO:26, or SEQ ID NO: 27, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 23, or SEQ ID NO:24, or SEQ ID NO: 25, or SEQ ID NO:26, or SEQ ID NO: 27.

In embodiments wherein the Cas13 polypeptide is a Cas13d polypeptide, the sequence encoding the Cas13d polypeptide is selected from SEQ ID NO: 13, or SEQ ID NO:14, or SEQ ID NO: 15, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 13, or SEQ ID NO:14, or SEQ ID NO: 15.

In embodiments wherein the Cas13 polypeptide is a Cas13d polypeptide, the Cas13d polypeptide has an amino acid sequence of SEQ ID NO: 28, or SEQ ID NO:29, or SEQ ID NO: 30, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 28, or SEQ ID NO:29, or SEQ ID NO: 30.

In preferred embodiments of this aspect of the invention, the Cas13 polypeptide is a Cas13a, a Cas13b, or a Cas13d polypeptide, with at least trans cleavage activity and preferably both trans cleavage and cis cleavage activity.

The Cas13 polypeptide of the invention is preferably a Cas13a or Cas13d polypeptide, and more preferably, is Cas13a7, Cas13d13, Cas13d14 and Cas13d15.

In another aspect, there is provided a nucleic acid molecule comprising: (a) a sequence encoding a Cas13 polypeptide wherein the Cas13 polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30; and (b) a sequence encoding a CRISPR RNA (crRNA) comprising one or more spacers and one or more Cas13-specific direct repeats, wherein the crRNA is capable of hybridising with one or more target RNA molecules.

In another aspect, there is provided a CRISPR/Cas13 system for targeting RNA molecules, the system comprising

i) at least one Cas13 polypeptide wherein the Cas13 polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30; or a nucleic acid molecule comprising a sequence encoding the Cas13 polypeptide and

ii) at least one CRISPR RNA (crRNA) or a nucleic acid molecule encoding the crRNA, the crRNA comprising one or more spacers and one or more Cas13-specific direct repeats, wherein the crRNA is capable of hybridising with one or more target RNA molecules.

Preferably the nucleic acid molecule comprising a sequence encoding the Cas13 polypeptide is selected from the group consisting of SEQ ID NOS: 1-15, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 1-15.

In another aspect of the invention there is provided a composition comprising a CRISPR/Cas13 system for targeting RNA molecules, the system comprising i) at least

one Cas13 polypeptide wherein the Cas13 polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30; or a nucleic acid molecule comprising a sequence encoding said Cas13 and ii) at least one CRISPR RNA (crRNA) or a nucleic acid molecule encoding said crRNA, the crRNA comprising one or more spacers and one or more Cas13-specific direct repeats, wherein said crRNA is capable of hybridising with one or more target RNA molecules.

Preferably the nucleic acid molecule comprising a sequence encoding the Cas13 polypeptide is selected from the group consisting of SEQ ID NOS: 1-15, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 1-15.

The crRNA in aspects of the invention preferably comprises one or more Cas13 specific direct repeats, and one or more spacers, wherein the spacers are the sequences of the crRNA capable of hybridising with one or more target RNAs. In some embodiments, the crRNA comprises two or more spacers, wherein the two or more spacers are capable of hybridising with different target RNAs.

The nucleic acid molecule components of the compositions and systems described herein may be comprised within one or more vectors. The one or more polynucleotide molecules may further comprise one or more regulatory elements operably configured to express said Cas13 protein, and said guide molecule. Preferably the one or more regulatory elements comprise operably linked promoters.

Accordingly, in another aspect, provided herein are vectors and vector systems comprising any of the nucleic acid molecules of the invention described herein, and in one embodiment of the CRISPR/Cas13 systems of the invention, a system comprising the vector systems.

There is provided a CRISPR/Cas13 system wherein the system comprises a vector system comprising one or more vectors comprising:

i) a first regulatory element operably linked to a nucleic acid molecule comprising a sequence encoding a Cas13 polypeptide wherein the Cas13 polypeptide has an amino

acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30; and

ii) a second regulatory element operably linked to a nucleic acid molecule encoding a CRISPR RNA (crRNA), the crRNA comprising one or more spacers and one or more Cas13-specific direct repeats, wherein the crRNA is capable of hybridising with one or more target RNA molecules;

wherein components (i) and (ii) are located on the same or different vectors of the system.

Preferably the nucleic acid molecule comprising a sequence encoding the Cas13 polypeptide is selected from the group consisting of SEQ ID NOS: 1-15, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 1-15.

There is also provided a kit comprising the one or more vectors of the vector system.

Another aspect of the invention provides an *in vitro* method of modifying a target RNA, the method comprising contacting the target RNA with a ribonucleoprotein (RNP) complex of a CRISPR/Cas13 system, the system comprising:

i) at least one Cas13 polypeptide wherein the Cas13 polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30; and

ii) at least one CRISPR RNA (crRNA), the crRNA comprising one or more spacers and one or more Cas13-specific direct repeats, wherein the crRNA is capable of hybridising with one or more target RNA molecules,

wherein the Cas13 polypeptide and the crRNA form a ribonucleoprotein (RNP) complex, and upon binding of the complex to the target RNA through the one or more spacers, the Cas13 polypeptide modifies the target RNA.

In an alternative embodiment of this aspect of the invention, prior to contacting the target RNA with the RNP complex, the method comprises:

a) expressing from a vector system at least one Cas13 polypeptide and at least one CRISPR RNA (crRNA), the vector system comprising one or more vectors comprising:

i) a first regulatory element operably linked to a nucleic acid molecule comprising a sequence encoding a Cas13 polypeptide wherein the Cas13 polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30; and

ii) a second regulatory element operably linked to a nucleic acid molecule encoding a CRISPR RNA (crRNA), the crRNA comprising one or more spacers and one or more Cas13-specific direct repeats, wherein the crRNA is capable of hybridising with one or more target RNA molecules;

wherein components (i) and (ii) are located on the same or different vectors of the system;

b) isolating the expression products of step (a); and then

c) contacting the target RNA with the isolated expression products of step (b), wherein the Cas13 polypeptide and the crRNA form a complex, and upon binding of the complex to the target RNA through the one or more spacers, the Cas13 polypeptide modifies the target RNA.

Optionally the isolated expression products of step (b) are assembled in to the RNP complex prior to contact with the target RNA in step (c).

Preferably the nucleic acid molecule comprising a sequence encoding the Cas13 polypeptide is selected from the group consisting of SEQ ID NOS: 1-15, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 1-15.

In another aspect of the invention there is provided a nucleic acid detection system for detecting a target RNA in a sample, the system comprising:

i) at least one Cas13 polypeptide wherein the Cas13 polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30, or a nucleic acid molecule comprising a sequence encoding the Cas13 polypeptide and

ii) at least one CRISPR RNA (crRNA) or a nucleic acid molecule encoding the crRNA, the crRNA comprising one or more spacers and one or more Cas13-specific direct repeats, and

iii) a detector RNA

wherein the crRNA is capable of hybridising with one or more target RNA molecules, and the Cas13 polypeptide has at least trans cleavage activity.

The detector RNA can be, for example, a labeled detector RNA such as a fluorescence-emitting dye pair, i.e., a FRET pair and/or a quencher/ fluor pair or RNA molecule generating any other detectable signal after collateral cleavage.

Preferably the nucleic acid detection system is in kit form.

Various embodiments of the features of this disclosure are described herein. However, it should be understood that such embodiments are provided merely by way of example, and numerous variations, changes, and substitutions can occur to those skilled in the art without departing from the scope of this disclosure. It should also be understood that various alternatives to the specific embodiments described herein are also within the scope of this disclosure.

## Brief description of the drawings

Figure 1: Phylogeny analysis of novel CRISPR-Cas13 proteins. Protein sequence alignment of the Cas13s was performed using ClustalW in MEGA11 with default settings. Rumen metagenomics derived Cas13 were compared with previously known (characterised) Cas13a (LbaCas13a and LwaCas13a), Cas13b (PspCas13b), and

Cas13d (EsCas13d and RspCas13d) to generate the analysis shown in Figure 1. The enzymes of the invention are labelled as “C”, followed by a, b or d for the subtype, and a number.

Figure 2: Fluorescence curves of the Cas13 polyprotein trans-cleavage reactions. The fluorescence was measured in 2 min intervals. Values are shown in the graphs as means ± SD (n = 3).

Figure 3: Fluorescence curves of the Cas13 polyprotein trans-cleavage reactions using different ssRNA reporter (ie probe) sequences (Table 6). Figure 3A is Cas13a3; Figure 3B is Cas13a7; Figure 3C is Cas13d13; Figure 3D is Cas13d14 and Figure 3E is Cas13d15. The fluorescence was measured in 2 min intervals. Values are shown in the graphs as means ± SD (*n* = 3).

Figures 4: Fluorescence curves of the Cas13 polyprotein trans-cleavage reactions using different activator (PFS) sequences presented in Table 5. Figure 4A is Cas13a3; Figure 4B is Cas13a7; Figure 4C is Cas13d13; Figure 4D is Cas13d14 and Figure 4E is Cas13d15. The fluorescence was measured in 2 min intervals. Values are shown in the graphs as means ± SD (*n* = 3).

Figure 5: Screening of the buffers. Detailed compositions of the tested buffers are listed in Table 7. Figure 5A is Cas13a3; Figure 5B is Cas13a7; Figure 5C is Cas13d13; Figure 5D is Cas13d14 and Figure 5E is Cas13d15. End-point fluorescence signal measured after 60 min. Values are shown in the graphs as means ± SD (n = 3).

Figure 6: Screening of the buffer pH. Figure 6A is Cas13a3; Figure 6B is Cas13a7; Figure 6C is Cas13d13; Figure 6D is Cas13d14 and Figure 6E is Cas13d15. End-point fluorescence signal measured after 60 min. Values are shown in the graphs as means ± SD (n = 3).

Figure 7: Probe concentration assay. Figure 7A is Cas13a3; Figure 7B is Cas13a7; Figure 7C is Cas13d13; Figure 7D is Cas13d14 and Figure 7E is Cas13d15. The fluorescence was measured in 1 min intervals. Values are shown in the graphs as means ± SD (n = 3).

Figure 8: The limit of detection (LoD) of Cas13-based nucleic acid detection. The reaction systems were comprised of 50 nM crRNA, 100 nM Cas13, 1000 nM reporter,

and different concentrations of target RNA (between 1 nM and 100 fM in 10-fold dilutions), for each enzyme in the corresponding reaction buffer. The fluorescence was measured in 2 min intervals. Values are shown in the graphs as means ± SD (n = 3).

Figure 9: Table summarising the optimisation of cleavage conditions. Reactions were incubated for 1 h at 37 °C fluorescence was measured in 1 or 2 min intervals. Real-time or end-point fluorescence measurements were collected on a microplate reader. All experiments were performed using three independent replicates and results were shown in the graphs as means ± SD (n = 3); Conc.: concentration; NA: not applicable.

Figure 10; Cas13-based detection of SARS-CoV2 N gene using crRNAs presented in Table 11. Cd14-CoV crRNA 1 was used for Cas13d14. End-point fluorescence signal measured after 60 min. Values are shown in the graphs as means ± SD (n = 3).

Figure 11: Cas13d14 based detection of SARS-CoV2 N gene using two different crRNAs (Table 11). Fluorescence curves of the Cas13d14 trans-cleavage reactions. The fluorescence was measured in 2 min intervals. Values are shown in the graphs as means ± SD (n = 3).

Figure 12: Cas13d14 visual-based detection of SARS-CoV2 N gene using Cd14-CoV crRNA 1. The image of strip of reactions were captured using SYNGENE transilluminator (420nm wavelength), 15 minutes after beginning of collateral activity.

Figure 13: Representative denaturing gel showing the targeted *in vitro* RNase cleavage activity of the Cas13 polypeptides when incubated with the ssRNA target of SARS-CoV2 N gene and different crRNAs (labelled as crRNAs in the figure). RNA cleavage activity is most evident with Cd13-crRNA 1 relative to other enzymes and crRNAs and no crRNA control. The presence and absence of crRNAs has been designated as '+' and '-', respectively. Cd14-CoV crRNA 1 and Cd14-CoV crRNA 2 were shown by '+1' and '+2', respectively. The reaction containing Cas13d14 without crRNA was used as control.

## Description of sequences

Table 1. Sequences of the invention

| SEQ ID NO | Description | Sequence |
|---|---|---|
| 1 | Ca1<br>ContigID: k127_1867445 | ATGAAGATTTCGAAGGTAGATCATACTAAGAGTGCGGTGAGTGTACAAAC<br>TGCACAGGGACAGCAAGGTATTCTTTATAAGGATCCGTCCACAGAAGAG<br>ATGAGCGTGGAAGATCGCGTCACCAAAAGAGCAGATGCTACGAAAGCAT<br>TATACGCTGTGTTCAATCAACCTAAGGATAAACGAAGCATTTCAGGGGAA<br>GCAACAACGGTTGCATCTTCTTTTAACTATGTGATCAAAGACCTGAAAAAA<br>AGCAAATCACTTAATGGGAAGTTGAGTGTGGAGAGCCTCTATGAAGCTGT<br>GGGAAATGAACTTAAAGGAAAACATGCGAGTGCTGAAGAGATAGATTTG<br>GCAATTACATTATTGCTGAAGAAAAGTCTGCGTAGAGATTCCTTTATAGA<br>GGCGTTAAAGCTGGTGCTGGGTAAAGCGTATAAAGGAGAAAAGCTAAAT<br>GAGGAAGATAAGAGAATAATCAAGGACGATCTGATAGTTCCTTTGATAAA<br>GGATTATGATAAATCATCTATCAGAGAGCAGGCAGTGGCCTCGATCAAAC<br>ATCAGAACCTTATCGCACAGCCGGATTCCAAGTCTGATGACGCAGTAATG<br>GTTATATCCAATATAGCCGGGGCTTCTGAGAGAAGTACGAATGAGAAAGA<br>AGCTTTAAGGCAGTTTATATCTGAATATGCAGTGCTAGATGACAGTGTTC<br>GTCATGATATGCGTGTTAAACTCCGTCGTCTGGTGATCCTGTATTTTTATG<br>GCATGGATGTAGTTCCGACGGGTGATTTTGACGAATGGGAAGATCATGT<br>CCAGAGGGGAAAGACCGCTGATCTGTTCATAGACTTTGCACCTGTCGGA<br>GGTAAGACAGATGCTGATAGATTAAAGGATGCTATTCGGAAAATGAACAT<br>TGAAAGATACCGATATAGCGTAGATGCGATAGATCAGGATAATACAGAAC<br>TCTTCTTCGAAGATATGATGATCAACAAATTCTTTATTCATCATATTGAAAA<br>CGAAGTAGAGCGGATATATAGGAATACTAAGCCTGGTGATGAGTTCAAG<br>AGAAGCTTAGGCTATATCAGCGAGCGTGTGTGGAAGGGAATTATTAACTA<br>TCTGTGTATTAAGTATATAGCAATAGGAAAGGCTGTATATAATTGCGCTAT<br>GGCTGGGTTGGGCTCTGATCAGCCCGATATTAAATTGGGAGTCATCGAC<br>AGAGTGTATGCTGATGGCATCAGTTCCTTTGATTATGAGATTATCAAGGC<br>ACAGGAAACACTCCAGAGAGAAACCTCTGTATACGTTTCATTTGCGATTA<br>ATCATCTTGGGGCTGCTACTGTTAATCTGACTGAAAAAGAGACCGATTTT<br>CTTACACTTGATAATAAACAGATTAAGGAACTGGCAAAGACAGGTGTTCT<br>CAGGAATATTCTTCAGTTTTTTGGTGGTAAATCCGTATGGAAAAATTTCGA<br>GTTTGCTCCTGAAGGGGGTACCGGCAATGAAGAAATTGTATTGTTATATT<br>ACCTTAAAGACATCCTGTATGCGATGAGAAATGAGAACTTTCATTTTTCAA<br>CAGCCAGCATTAATGATGGCTCATGGGACACTGATCTGATCGGCAGGAT<br>GTTTGCATACGACTGTACCAGAGCAGGGGTGGGACAAAAGAACAAATTC<br>TATTCCAACAACCTTCCAATGTTCTATAAATCAGAGGACCTGGAACGAGC<br>ATTACATATTTTGTATGATCACTATAGCGAGCGAGCTTCGCAGGTTCCGG<br>CGTTTAATACGGTATTTGTGAGAAAGAACTTTTCTGAAATACTTAAGGGG<br>CAGAATCTGCCTATGCCAACTTCAGCTGAAGAATCTCTTAAATATCAGAAT<br>GCGATCTATTATCTCTATAAAGAGATTTATTATAACGTATTTCTGAGTTCAT<br>CAGAAAGCCGAGATTATTTCATTAAAGCAGTCAAGTCACTTAGGTGGGAA<br>AACTCGAATGAAGAGAATGCCGTAAAAGACTTTCAAAATCGAATTAATGA<br>ACTGACCGGAAAATACAGTTTATCCCAGATTTGTCAGTTGATCATGACAG<br>AATACAATCAGCAAAACAGCGGAAGCAGAAAAAAGAAGACTGCTAAGGA<br>TGAACAGAATAAACCGGATATTTTCAAGCACTATAAGATGCTTTTGTATAA<br>GAGCATTCGGGAAGCAATGCTTAAGTATGTGGACGATAAATCAGAGGAC<br>TTTGGTTTTATAAAAAGTCCGGTATTCGGAAAGGATGACAACTGCATTGC<br>GTTAGAAGAATTCCTCCCTGATTATGAATCGACGCAAAATGCAAAACTGA<br>TAGAACGCGTGAAGTCAGATTTCAGACTTCAGAAATGGTACATTTTGGGA<br>AGACTGCTCAATCCTAAACAGGTAAATCAGCTTGCCGGATCTATCAGGAG<br>TTATATTCAGTATTCAGATGATGTAAAGAGACGTGCAAAAGAAAATGGTAA<br>TAAGATTCATGTATCCACGGAATCGTATCCTTATCAGACTGTTTTGAGAGT<br>CATTGATCTGTGTGCGAAATTGAGTGGACTTACCACCAATAACATAGATG<br>ACTATTTTGATGGTAGTGGGGATTATCTGTCATACCTTGCTCGCTTTGTG<br>GAGTATGATCCGAATGATATACCGAAGATATATCATGATGAAGCTAATCC<br>CATTCTTAATCGCAACATCATTATGGCTAAATTATATGGCGCAGGAGATG<br>TTATTACCAATGCAGTAGAACATGTCAATACAAGTATGATCAGAGATCTTG<br>AATCATATGAGAAGAAAACGTTGGGGTATCGCTCTTCTGGAGTATGTAAG<br>GACAAGGATGAACAGGAAACGCTAAAAAAATATCAGGAATTAAAGAACCG<br>GGTAGAGCTTCGAGAGATAGTGGAGTGTTCTGAAATTATAAATGAACTTC<br>AGGGACAGCTGATCAACTGGTGTTATCTGAGAGAGAGGGATCTGATGTA<br>TTTCCAGCTGGGGTTCCATTATACCTGTTTGAAGAATTCATCCGATAAAC<br>CGGAGATGTATGTAAAAGCGAAAACCGTAGATGGAACTATTGATGGATTT<br>ATCCTGCACCAGATCGCCGCATTATATACGAATGGACTAAAGCTGTATTC<br>TTGCGGTAAAGCTGTAAGAGATGATAATAGAAAGATAATTCATTATGACCT<br>GAGCAGTGGAAAAGAATTAAAGGGCAATGATAAGAGTGCAGCCGGGAAA |

| | | |
|---|---|---|
| | | AAGATTACTGACTTTATGGGGTATACGTCATTAGCGTTAAACAGGACAGA<br>GAATGACATACTCCCTATATCCGGTGATTTTTATTATGCCGGGCTGGAAC<br>TGTTTGAAAACGTAAATGAACACGAGAATATCATAAGCCTCAGGAATTATA<br>TCGATCACTTCCACTATTATGCAAAACATGACAGAAGCATGATCGATATCT<br>ACAGCGAGGTATTTGACAGATTCTTCTCCTATGACATGAAATATCGTAAG<br>AATGTGCCTAATATGATGTACAACATTCTGCTGTCGCATTTTGTGAAAGCA<br>CAATTCGTATTCGGATCAGGAATGAAGGAGTCCGGAGAAAAGACCAAGT<br>CCCAGGCCAGATTTGACTTGAAGGATAAAGCCGGGCTTGAGCCCGAACA<br>ATTGACATATAAGGTTGCGAATTCTGAGAAACCGGTGCAACTTTCTGCTA<br>AAGACAAACAATTCCTGAAGACGGTTGCATTATTGTTGTATTATCCGGAA<br>AAAAAGACGTTTCCGGAGGGGATGTATGCCGATACCAGGTTTGTAGAAG<br>GGACATCATCAAATAAAAGGAATAACAATTCTTCAGGCAACCGACATGGC<br>AATGGTAATCACAATGTAGGAGGACATAACAAAGGGTACAATCAGGGCA<br>GGAAAAACGGCAATTGGTCTAAGGACAAATCCGGAGACAGAAATGCCGG<br>AAAGAAACAAACAAATAAAAACCGGAAAGATAGTACCAGCGTCTATAAAG<br>ATGAAGGATTCAGTAACAGAATAAATATTCCTTCTGAGTATTATTCTCAGA<br>AACCGGGGAAAAAGTAG |
| 2 | Ca2<br><br>ContigID:<br>k127_4200118 | ATGAAATTATCAAAAACTGGAAAAAACGGCTGGCATCACAGAAACGGAGT<br>AAAGGTTAACAACAGTAAACAGGAGGGATTTGTTTACAGTATTCCTCATA<br>ATGATGGTGAGAGTACAGATAAGTTTGTTGAAGACAGAAAGAAAGATTTT<br>AAGAGACTTTATAAGGTATTTCCTTCTGTTGAAAAAGCCAGAAACATAAGT<br>GAAGAGATTGCTGCAGTCATAGATAAAACGATCAGAAATAAAAGAACAGA<br>AATATGGACCGGAAAAAATGATTATTCTGAAATGGCCTGCAGATTCAGAA<br>ATCTGCTTCAGCGCGAATCTATGTTCAGACAGCCTGTAGAAGTAAAAACC<br>GCTGAATATATGGTTTATGGTCTTTTGCGATCAAGCCTGCGTTCTGAAAA<br>AACGGAGAAGGATCTTATAGATTTTTTATGCCATGTGAATGACAAGTCCT<br>CATCAGCTGGTGCAATATTTATGCAGGAGCTGACCAGGGATTATATGGGT<br>GAAAAGATCAATTATAAGTCGATAATAAACCAGAATCTCGTAATCCAGCCT<br>GTTAAGACAGAATCCTTAAAAGATAACATTGATCAGGAGGATGTTTTACTA<br>ACTGTTTCCGAAAGGAAAAATCCTGAAGACTCGGCAAAGAATTATAAAAG<br>TATTGAAAATAAAGCATTACGGAGCTTTTTATTGGAATATGCCAGCCTTGA<br>TGATAATAAGCGTAAGGATCTGAGAAAAAAGCTCCGCAGGATAGTTGTAC<br>TCTATTTTTATGGAAAGTCCGAAGCAGACGGCTTAGGAGAAAACTTTGAT<br>GAGTGGAATGACCATGAATCCAGGCGTGCATGTGAGGAGAAGTTCATTG<br>AATTTGATGAGAGTACCAAGGATAATTTCAGATCCAAAACATCTAAGCTG<br>ATAAGAAATGCCAATATCACAGCATACAGGACTTCAAAAGAAATCATTGA<br>AAACAATCATGACGGGCTGTATTTTGCCAATCCTGATTATAATTTCCTTTG<br>GTTAAGGCATTTATCCAGGGAGGTTGAGCGTCTGACATCAAATATCAATG<br>CTGATAAGACGTATAAGCTGAATAAGGGCTATTTAAGTGAAAAGTCCTGG<br>AAGGGAATAATCAATTATCTTAGCATAAAGTATATAGCTATAGGCAAATCT<br>GTATTCAATTTCGCATCTTCCGGAGTAGATTCTGATGGCAGTGATATACA<br>GATTGGCGAAGTAAACAGGGAGTTCCAAAACGGAATAAGTTCATTTGAAT<br>ATGAGAGGATAAAAGCCGAAGAAACACTTCAGAGAGAAAGTGCGGTAAA<br>AGTTGCCTTTGCTGCAAGACATCTTGCATCGGCTACGATGAATCTTACAC<br>CGGAAGATTCAGATATGCTCCTGTTTGATAAGGACAAAATGTCTCAGAAT<br>CTAAAAGATACAGGCAGGGTACTGGCTGATGTATTGCAGTTTTTTGGAGG<br>ACAGTCTGTCTGGAAGGATTATATTATAAAAGAAACTGAAAAGTATTCAAG<br>TGAAGAAGAATTCGGAACTGATTTATTATATAATCTGAAAAAATGCGTCTA<br>TGCCTTAAGAAATGACAGTTTCCATTTCAAAACTTTGAATAATAAGGCTGA<br>TTGGGATACTGACCTTGTAGCAGGATTGTTTGAAAAAGACTGTGAAAACA<br>TGGTAGGCCTTGATAAAGATAAGTTTTATGATAATAACCTTTACAAGTTCT<br>TTAAACAGGAAGACCTGAAAAAAGTATTAGACAAACTGTATGATAAAATCC<br>ATGACAGGGCATCTCAGATACCTGCTTTCAATACAGTGTTTGTAAGAAAT<br>AATTTCAGCAGGTTCCTGTTGTCTAAAGGTATCACACGCTCATTTGCGTC<br>AGAGGAGCAGGGAAGACAGTTTGCAAGTGCAGTATATTATCTTTTCAAAG<br>AAATCTATTACAATGATTTTATCCAGTCAGGAAATGCTAAAACGCTGTTCC<br>TGAATTATGTTAATTCAATCAAAATCGAGAAAGCAATTAACAGATACGGGA<br>AAGAAGAAACTAAAAGAGAATGCAAACCGGCAGAAGATTTCAAGACATAT<br>ATAACGTTGTGCCGGAATATGAGTTTCTCTGAGATATGTCAGGCAGTAAT<br>GACTGAGTATAACCAGCAGAACAACCAGTCCAGAAAGAAAAAATCAGCTT<br>TTGATACAGCAAAAAACAAAGATAAATTCAGACATTATGTAGATATTTTAC<br>ATGAAGGGATCAGGGAAGCATTTGCAGCATATATTGGTCTCAATGATCAG<br>AAAAATTATGATGGTATATATGGCTTTGTAAAGTCATTTAACAGCAGCGAT<br>GTGTTCACAGTCGAAAAGGACAAGTTTATTGAAGGGTACAGATCTGAACG<br>TTTCCAAAATCTGATCAGTAAAGTGAGACAGAATCCTGAACTTCAAAAGT<br>GGTATATTGTGGCTAGACTTCTGAATCCCAAACAGGTAAATGAACTTTCC |

| | | |
|---|---|---|
| | | GGATCGATAAGAAGCTATAAGCAGTATATTGAAGATGTATGCCGCAGGTC<br>GATCGCTGAAAAATGTCCTGTAAGGAAAAATGATGGAAAAGAGGTAAGCA<br>AAGCTTCTTTTGATAATGATGTAAAAGAACTGATGTCCATCGATTATATGG<br>GTGTTGTAGCGATATTGGAGATATGTATACGTCTTAACGGAAGATTTTCA<br>ACAGTATGCGATGATTATTTTAAGGATGGTGCTGACGGCTATGCTGAATA<br>TCTGGAGCAGTATCTTGATTATCAAGATGAAAAAACTAAGGATGCCGGTG<br>TAAGTCCTTCAACCATGCTGTCGATGTTTTCAGAAGAAGTCTCAGCGGAT<br>AATACTGATAAGAATCAGGGAATCATATATCATGACGGGACTAATCCCAT<br>TATGAACAGAAATATTCTCTTATCGAAACTCTATGGTGGAGCAAATTCGGT<br>AATACATTCTGTAAAGAAGGTGGATAACCGCCTGATAGCAGATTTTATTAA<br>AAGCGGTAAGCTTATACAGGAATACAATAAGAGAGGATACTGTATTAATG<br>AGGAGGAACAGAAAAATCTGAAGAGATATCAGGCTCTTAAGAACAGGGT<br>TGAGTTCCGAGATATAGTAGAATATGGCGAAATCCTGGATGAACTGCAAG<br>GGCAACTGATTAACTGGAGTTATTTAAGGGAACGTGATCTGATGTATTTC<br>CAGCTTGGTTTCCATTATACGAGCCTTCATAATTCTGAAAGAAAGTTTGAA<br>GGGTACAGATATATAACTAAAGAAGACGGAAGTGTTATAGAGAATGCAGT<br>ACTGCACCAGATTCTATCATTATATATTAACGGAATACCGTTCTATTACAG<br>CTACGCCGATGTGGAAGGCAGAGACCGGTTTATTTGCTGTGCACTAAAG<br>AAGAAAGAACCTGTTGATGGTACCAGCAACAAGTTTGAAGATACCGGAAC<br>AAAGATGAGATATGTAGGATATTACTGCAAAGAAGGTGATAATTATCTGG<br>GTGAAGGGATATATCTTGCAGGTCTGGAACTGTTTGAGAATATTGCTGAA<br>CATGATAATATCATCAAGCTTAGAAACTACATAGATCATTTCCATTATTATA<br>TTGAAGACGACAGGTCGATGCTGGATGTTTACAGTGAAGTTTTTGACAGA<br>TATTTTTCATATGATATTAAATATCAGAAGAATGTTGTTAATATGCTGTATA<br>ATGTCCTTTTGAGTCATTTCGCTAAAGCCGGATTTGAATTTGGCGAAGGA<br>ATAAAACAGATAGGAAGCAGCAAAAAGAACGAAATGCCGTTAACCAAAAA<br>GATGGCGCGCATTCTGCTGAAATCACTTGAATCGGATGATTTTACTTATA<br>AGATTGGTAATACATCTGAAGCAAAGAATCAGGAAACGGTCGTTCTTCCT<br>GCCAGAGATGATCTTTTCCTTGATGCTCTTGGAAAAGTACTTAATTGGGA<br>TGGCAGTATTACTGATGAAAACGCTTTACAGAAAACAGAGATTATTACAG<br>GAAGAAGTGGCAATTATCTGAAGAAATCAAGGAACAGTGATAAGAAGAAT<br>GATGGTAATAAAAGAAGTGATATAAAAAAATCATTCAATAAAGATAAAAAC<br>ATACCCGAAAAGAAAGAAGCACTGACCAGTACTCCGTTTGCAAACCTGTT<br>TAATAATTTACATATGGATTTTGATTAA |
| 3 | Ca3<br><br>ContigID:<br>k127_751200 | ATGAAGATATCCAAGGTTAATCATACTAAGTCAGCCGTTAGTGTTTCTGAA<br>GGTTCTCCCAAGGGAATACTGTATGAGGATCCGACCAAAAGCGGAACGA<br>AGGATCTTGAAACTCGAATACTAGAGCGTAACGAAGCTGCGAAATTACTG<br>TATAACCCTATAAATACATCGAGAAGCAGAAAGAAAACACATAAAATAATC<br>AATAGAAGTCTTAGGGCTTTCTTCAATCGAGTAAAGAAGAAAACCGGTGG<br>GAGTTTCTCTTGGGATGAATTGAAAAGAGTATCGTATGATTCTTCTTTGGA<br>TGCAGAGCGTGATAAGATTACTGATTCTGATATAGATTCTGTTGTTGAAG<br>CCTGTTTGAAGAAGAGTCTTTCAACCCCGGAATGTATAGAGGCGGTAAA<br>GCAGATTACAAGAGTTTTGTGTGGCAAGAATACCTCCCATGATCTGGATG<br>ACAAACTTATAGGGAAATTGTCTTCCAAATTACATGATGACTACTCGAAG<br>GAACGATTGCTGGGTAACATCAAGAAGTCGATTGAAAACCAGAATATGGT<br>AGTTCAACCCGGACAGGTTGACGGAGAAAGCATTTTTAAATTAACTGGTG<br>ATGATTCTTTGAAAGAGAATCCGGAGAAGGTTTCTTTCGAAAGATTCCTTA<br>TTAGCTATGCCAATCTCGACAAGAAATTCCGTGATTGTGAATTGCGCAAG<br>TTAAGAAGGCTGATAGTTCTATATTTTTATGGTGAGACAGAAGTAGATAC<br>GACAGATGATTTTGACGTTTGGGCAGATCACAAAAAGCAGCGTAATTTTA<br>AATGGTTTATTTCAGATATCGAGTTTATTGCTACATATGAAAAATACCTTAA<br>AGAATTACAGTTTGAGGATAGAAGGAATCATCATACAAAGATATCGGAAC<br>CTGAGTTCAGGGAGAAGATAAGACAAGAAAACATAAACAGATATAGGAAT<br>TCAATTGCAGTAATTAACAAATCAAATGAAGTGTACTTTGATGATCCTGTC<br>CTTAACAAATTCTGGATCCATCATATAGAAAATTCTGTTGAAAAACTGCTG<br>AAAAGGGTAAATCCCGCCGATTCATTTAAGCTAAATGTTGCATATATAGG<br>CGAGAAAGTCTGGAAAGAGGTTATTAACTATCTGAGTATTAAATATATTGC<br>TGTTGGTAAGGCTGTCTATCGTTTTGCAGTTGATGATATGACCTATGGCG<br>TCATACCGGATCTGTATAAGTCAGGAATAAGCTCATTTGATTATGAACTCA<br>TAAAAGCCGACGAATCGCTTCAAAGGGATATTGCCGTTTCGGTAGCATTT<br>GCTGCCAATAATATGGCACGAGCTACTGTTGTGCTCGATGAGAAGAGCA<br>GCGATTTTCTTGTTGAGAGTTTTGATCTGGAAAAATCCATTAGAACAGATG<br>TTGCGCTCGATATGGCGATACTGCAATTTTTCGGGGGCAAATCGTCTTGG<br>AAACAGTGCGACGCATTGAAAGATTGTAAATATATCGATCTTCTTTATGAC<br>ATGAAGAAAATGCTCTATTCCATTCGTAATATGAGTTTTCATTTTATTTCAT<br>CTGAAGAAGGAGATAACGGTTATAAGACTAACGGGATAATCCCGGCGAT<br>GTTCAATCAAGAAATAACAACCTATACGACGATTCTTAAGAGCAAGTTCTA |

| | | |
|---|---|---|
| | | TTCTAACAATCTGCCTGCTTTTTATAATGATTCAGATTTGGAAGGAGAATT<br>TAAACTTCTATATAAAAATTACGTAGAAAGAGCATCTCAAGTTCCTTCTTTT<br>AATAGTGTTGTTGTCAGAAAGAGTCTTCCGGATTTTGTAAAAAGAGATCT<br>GAAGATTAAAACAGCTCTCTCGGGAGATGATCTCACAAAGTGGCATAGC<br>GCTCTATACTATCTTCTCAAGGAAATCTATTACAATCTCTTCCTTGCAAGT<br>GATGATGCAAAGATTTTGTTTTTGAAAGCAGTTGAAAATAATAAGAATTCT<br>AATAATAGCAGTGTTTCTGATAAGAATGACCATCGAAGAGAAGCTGGAAT<br>TGATTTTGCCGAGCGAATAGAGAGTATAAAAGATCATAGTCTTTCAGAGA<br>TATGTCAGATCATTATGACCGAGTATAACCAGCAGAATCAAAGCAGAAAG<br>GTTAAGACGGCTCAAGATGAGAAGAATAAGAAATCTTTATTCATTCACTAC<br>AAGATGTTGCTTAATCTTTGCCTTCGGAATGCATTTAAGATGTTCCTTGAT<br>CGCAACGAGTTCTCATTCTTAAAGAGTATTCATAATCGCGAAGTTAAGAG<br>TTCCTCTGATGAATGGACCGCTGCTTTTTGCGCGGATTGGACATCAAATG<br>CATATAGCATGATTCAAGATGAGATTAATAAAAATCCTTCGTTGCAAAGCT<br>GGTATATACTATCAAGATTCATTACAACAAAACAGCTTAACCATCTTAGTG<br>GAGATATCAGACATTTTATTCAGTATGTAGAGGACGTTAAGAGGCGAGCC<br>AAAGAGACAGGAAATGCATGTAAATACGATCTTGATAATAAGGTTTGTATA<br>TACCGTAAAGTACTTCAGGTTCTGGATTTTTGTAACAAAACAAGCGGAAT<br>AGTATCTTCAGAGATTGGTGACTATTTTAAAGATGATGATGAATATGCAAA<br>GTTTGTTAGTAACTATCTTGATTTTGGTGGAACGACGAAGCTGGAATTGA<br>TTGCGTTTACAAACCAGACTGTTGGAGATGATCAGATCAATATATATTGCA<br>ACGATTCAAAGCCAATCCTCAACAGGAATATTGTAATGGCAAAACTGTTT<br>GCCCCTACAGATACGATAAGTAAGGCGATAGCTGCAAATGGAAACCGAG<br>TAACTGTCGATGATATTGAGGAGTTCTATTCTATTAAACCGATTGCTCAGA<br>AATTTCTTTCTGATGGTGATAGTGTTGCCAAGAAAGAAAAGAAGCAACTT<br>ATAGAGGAACTTAAGAAAACAAAGCGATATCAAGAGATAGTTAATCGCAT<br>TGAGTTCCGCAATATTGTAGATTATGCAGAGATGATAAATGACTTGCTTG<br>GTCAGTTGGTCAGTTGGTCATATCTGAGGGAAAGAGATCTGTTGTATTTC<br>CAGTTGGGTTTCCATTATCTTTGTCTTATCAATGATTCTTACAAGCCGGAT<br>AAGTATAGAGTCCTTCGTGATGGAGAGCGGATAATAAACAATGCCGCACT<br>ATATCAGATCGTTTCGTTGTATTCCTTTAATGTAGATACTTTTAGAGATGA<br>TAATGATAAAGATAAAGGTAAAAAATATAATAATATATGTGAGTACAGCCT<br>GAATATTGGTTTAGATGAAGAATGGCAATTCTATACTGCCGGGCTTGAAC<br>TATTCGAGACTATCACTGAACATGATTCTATTAAGAAGTTTAGGGATTATA<br>TTGATCACTTCCATTACTACACAAATCAGGATAGAAGCATTCTGGATATGT<br>ATAGCGAGGTCTTTGACCGATTCTTCTCTTATGATATGAAATTCCGTAAAA<br>ACACTGTCGTTATCTTGCAGAATATACTGAAATCTTATTTAGTAATAATGC<br>CGGTTAAATTTAATTCGAAGTATAAATCAACAGATAACGGTAGTAGTAAAA<br>TGCGTGCCAATGTTGACATGGGTGAAAAAGGCTTAAAGTCAGAGGTGTTT<br>ACTTATAAGTATTCTGACAGCTGTAAGGTCATCTTGCCTGCAAGGTCGAT<br>TAATTATCTTAAGGATGTTGCATCTATCCTGTATTATCCTCATAAAACGCC<br>TCGTGATGCTGTTGATATGGAAGATTTTAAAAAGAACTATGAGGTTGCAC<br>AAACTTTAAATAAGGCAAAAGACAATCATCATAAGTCTAAAAATGATTATA<br>AGAATAATGATCGTCCTAAAGATAATTATTCTCCGTTTAAGAGCCAGTTTG<br>ATAAATTGAAAAAGAAGGGTATTACATTTGAAGATAACTGA |
| 4 | Ca4<br><br>ContigID:<br>k127_5935133 | ATGAAAATATCAAAAGTAGATCATACAAGAACGGCAGTTGGGGTGAATGA<br>GAATGGACCATTGGGAATTGTTTATTCAGATCCATCTCAGAATGCTGTTC<br>AAAATCCAGAGATTAGAGTTAGGACTAGAATTAAAAAAGCAAATATGCTTT<br>ATACGGTTTTTGGACCTACGAATGATGAGATGGATTCTCAACGAGAGAAT<br>GGAATAGCAAAAGAATTCAATAAAATAATCAAAAGATATAATAATAAAATA<br>GATCCCAAGAGGGGAGAAAAGGAAACTGATAAAAAGATATATAAAATGAA<br>TTCTGATGAACTGATTAAAGATATTAAATCGGTTTTTGGCAATTATTCTTTA<br>AATGAATCAACGAGAAAAGAAATTGATGAAGCGTTAAATGTATTAATTAAA<br>CGTTCTCTTAGAAAAAAGGAAACTATTGAATCATTAAATCTCTTATTTGAA<br>AAAACAATAAAGGGAGAGGAGTTTAAGGCGGAAGAGAAGGATAAGATTC<br>AAAAGTATGTGGTTGATCGTATAGTTGCTGACTATTCAAAGAATACGCTAT<br>CAAAAAATACTATAAAATCAATAAAAAATCAAAATCTTGTGGTCCAACCTC<br>AAAATAAAAATGGCGAATTTGTTTTTACACAGGCAAAAAATAGGATGAATG<br>GGAAAGTAAATCAAGGAAGTATAAGGATTTCAAAAGCTCAGGAAAAAGAT<br>GCTTTAAATGATTTTCTTGATGGTTTTGCAGTGCTTGATAAACAAATGAGA<br>GATAAGCAACTCATGAAAATAAGACGGTTAGTTGATTTATACTTTTATGGG<br>ATTGATGAGGTTGTAAAGGAAGACTTTTCAGTTTGGGAAAGACATGAAAA<br>AACAAAAGGTAATGATAAAAAAATCATTCCTTTTTCTCGTACTGATATATC<br>GACTTTACAAATAAAACGAGGAGATTCTGAGGACGAGAAAAAAAGGAAG<br>AATAGAGAAAAGAAAATAATTAAAAAATCTGATAGCGCTAAGTTGGACGA<br>TATGATAAGAAGGTGGAACATAGATAGATTCCGTGAATCGTTTAGTGCTA<br>TTGATAAAAGTGATAACAGTTTGTTTTTTGATGATAAAAACATTTCTAAATT |

| | | |
|---|---|---|
| | | CTTTATTCATCATATTGAAAATGAAGTAGAAAGGTTATTTAATTCAGAGAG<br>ATTGGATGATTATAAAATGCATATTGGTTATGTTAGCGAGAAGGTATGGA<br>AGGGAATTATTAATTATCTAAGTATAAAGTATATATCAATAGGCAAATCTG<br>TTTATAATTATGCAATGGAGGAGTTAAATAATTCATCAGGAGATGTAAATC<br>TTGGTGTGATAGATAGTAGATATTTGACAGGTATTAGTTCCTTCGATTATG<br>AAAAGATAAGTGCTGAAGAAACTCTTCAACGAGAGACAGCAGTATATGTT<br>TCATTTGCTTCTAGTAATTTGTCGAGAGCTGTTTTTAAAGATGGAGTAGAT<br>TGTGACTTAATGTCGACTAAGATTATAGATAATCACGATAAGTTTGATGAA<br>AGCAAAGTAAAAAAAAGAGTTCTACAGTTTTTTGGAGGAGAAAGCTCTTG<br>GGATGGTTTTGGAAAGACGTTTTTATCTGAAGAATATAATGAATTCGATTT<br>TTTGGAAGATCTAAAGACACTTATTTACCAGATGAGAAATGAGAGTTTTCA<br>TTTTAATACTGAAAAGAAAAATGTAGATATAAAAAATCCTAAACTTTTTTCA<br>GATATGTTTGCTTATGAATGCAGTAAAGCGTGTGTGTCTGAAAAAGATAA<br>ATTTTATTCGAATAATTTACCTCTCTTCTATTCAGAAAAACCTCTTGAGAAA<br>GTTCTGAATAAACTATACACCAAGTATAATGATCGTAAGTCGCAAGTTCC<br>GTCTTTTGAGAAGGTAATGAAAAGAAGTGAATTTGGAAAATATCTCATAAA<br>ATCTGGAGTTGCCACCAACTTTAATAAAGAGGATACAGATAAACTTGAGT<br>CGGGACTATATTACCTGTATAAGCAGATATATTATAATGATTTCCTAGTTA<br>ACGATATGATTGCAAAGGGGATATTTGTAGATAATATAAACAATAAGAAAC<br>TAAGAAGAAATGAGAATAATAAAGTAATCAAAGCTGATAAAGGGCTTGAA<br>GATTTTAAGAAACGCTTGAATGAAATAAAAAATTATTCTCTTTCTGAAATAT<br>GCCAAATTATAATGACAGAATATAATCAGCAAAATAATCAAAAGAAGAAAT<br>CTCAGAAAAATGAGGAGATATTCCAGCATTATAAATTGGGACTTTATTCAT<br>ACCTTCGCGAGGCGTTAATTATATATATTAATAATAATAGTGATATTTATG<br>GTTTTATAAAACAACCTACTATTAAATCTGAAGGTAAAATGCCGAATATTA<br>ATGAGTTTCTACCAGATTATTCTTCAAGTCAGTATGATGATTTGATAGCAA<br>AGGTCTCCGATTCTTTTGAACTGAAAAAGTGGTATGTAATGACTAGATTCT<br>TAAATCCTAAACAGACTAATCATTTGGTTGGAGCGTTAAGGAATTATATAC<br>AGTATGTGGAAAGCATAAAAAGAAGGGCGGAAGAAACAGGAAATAAAAT<br>ATATATAGATTGTCAGATTTTAGAATCTGTAAAAGATATCACAAAAGTAGT<br>AGATATGTGTACCAGAATATGTGGTAATACTTCTAATGAAATCTCTGATTA<br>CTTTGATGATAACGATGACTATGCAGGTTATCTAGAACGTTTTTTAGACTT<br>TGAATATAAAGAATCTTTGGGCTCGAAATCATCAATGCTTGGAGCATTCT<br>GTATGACCAAAATTAATAGTGAAGAGATAAAGATTTACCATGATGGAACA<br>AATCCTATACTTAATCGAAATATCGTATTATCTAAACTATATGGAGCAAAT<br>AGTATAATTTCAGAGGCTGTTCCAAAAGTTGATCAAAATATGATAAAAGAA<br>TACTATATTGTGGCTGATAAAATTAAAGAATATCGAAAGAGTGGCGATTGT<br>AAAAACATTGATGAGATTAAACAGCTTAAAGAATACCAAGAATTAAAGAAT<br>AGGGTTGAATTTAGAGATATTGTTGAATATTCAGAAATACTTAGTGAGCTG<br>CAAGGACAGCTCGTTAATTGGGCATATCTTAGGGAACGAGATTTGATGTA<br>TTTCCAATTAGGGTTTCATTACGTATGTCTCAAGAATGATAGTCAAAAGCC<br>AGAAGCATATAAAATGATTGAGGTGCCTTGCGTTGATGGGTCTTCTCGAA<br>TGATAAATGGTGCTATTCTTTATCAGATAGTCGCAATGTATACTTATGGAA<br>TGAATATATACTATAGGGGGCATAAAAAAGATGAGGAGTACAATGATAGT<br>GAAAATCGATGGGAAGCCTTCAATGGTTCAATAGGTGAGAGAATTCCAAG<br>ATTTGCCTTGTATTCTGGATATATGATTAAGGGAGACAATGCTAAGTACAA<br>ACTATCATATAATATTTATACATCAGGGTTGGAACTATTTGAAGTTCTTGA<br>AGAACATGGTAACATAGTTGATTTTAGGAACGATATAGACCATTTTAACTA<br>CTATCAAAAGAAGGATAGAAGTATGCTTGATTATTATAGTGAGGCTTTTGA<br>TAGATTTTTTACATATGATATGAAGTATCGGAAAAATGTTCCTAATACATTA<br>TCTAATATTTTAGCGTCTCATTTTCTTGTTCCTTCTTTTGTGTTTGGGACTT<br>CTTCAAAAAAAGTGGGGAACAAAAACTATATTGAAAAGAAGTGTGCTCAC<br>ATTAGATTTAATACTAAGAATCCGTTAAAACCAGGCAGTTTTACATATGTA<br>ATTTCGGAAGATAAACGTGTAGTAGGACCAGCGAGACTGAAGGGGTATG<br>TAAAAAATGTATTGAATATTTTATATTATCCAGAAGTGCCTGAAATGGAGC<br>TTTTAGATTCCTCTTATATATTTAAAGAAGAAAAAAAGAGAAAACTTCTCAA<br>ATAA |
| 5 | Ca5<br><br>ContigID:<br>k141_14579520 | ATGAAGATTTCTAAAGTAAGAGGAACTCAGGGCAAAGGAAGTAAACTTAC<br>AATTAATGCAAAGGCAGCAGTTGTAATTAACCCAACCGGTCAGGAGGGT<br>ATTCTCTATGATGATCCGTCAAGAATGGGCGAATCAAGAAAAAATGACAA<br>GCAGAGAGAATCATACATCAAGGATCGTATTCGTGCCTCTCAAAAATTGT<br>ATTCAATTTTCAATAGTAATCAAAAAATCCCCAAAAACAAAAAAACTGAAT<br>CTGAAAAAGCAATTGATATGATTATTGCTGGTTTTTCATCAGAAGATGGC<br>GCCAGTTTTCGTTTAATGTTTAAAGATTTCGCTGAAATTCTGGATAAATAT<br>GCAGAAAAAAGTTATGAAAACAGAAGAAATCATATAGACGAATCTCCCGA<br>ATTATCAAAGCTTGGGGTAAATATAAGTGACAATCAGATAAACGCCCTTT |

| | | |
|---|---|---|
| | | CTAATCTATTAAGTGAAGAAAGTATAGCTATAAAAATAAAAAAAGGAACTG<br>AATCAGTAAAAGATAAGGTAAAGGTTAGCGAGAGGGATATTGATTCGGCA<br>ATATCAAATTGCCTAAAGAAATGTATGTGCAGGGTAAAAACAAAGAAAGC<br>GTTAAAGGCTCTTCTTATGAAAGTTTTTGATATCCCATATACTTTAGATGG<br>AGACGTCAATATTAGAAGAGATTTTATTGATTATGCGATAGAAGATTATTG<br>CCGTATTCGTGTAAAAAACAGTGTCTCTGAATCAATTAAAAAGAATAATAT<br>GCCGGTTCAGCCGACGAGTTCGGAAGGTGTTACTGTTTTTCAGATGCCT<br>TCTTTGCAGGAAACAAAAAGTACCAAGAGTAAAGAAAGGGAAGCATTTAA<br>TCATTTCCTGTCAGAATATGCAGATTTAGATGAGAATAAGAGAAAGTCGTT<br>ACGAATAAAACTCCGAAGACTTAATGATTTATATTTTTATGGAAAAGACGC<br>AACTATGGCATTGGCTGATAACGAGGATGTGGACGTTTGGGAAGACCAT<br>GCAAAACATGGCGATATTAAAGAGTTATTTATAAAAGTTCAAAAACCACAA<br>ATCACAGGTGACGGAAAAGCGGATAAGCTGGCTATGAGTCAGTATGAAG<br>ATAATATTCGAACTAAATATAGAGAAGCAAATATTACTTGTTACAGAAAAG<br>CAGTAGAAGAAATAGATAATGACAAGAGTCTATTCTTTGAAGATAATATGC<br>TCAATATGTTTGTGCTTCACAGAATTGAGAGTGGTGTAGAACGAATATATT<br>CGCATATTAAAGCCAATGAGGAGTACAAGCTTCAAACGGGATATGTAAGT<br>GAAAAAGTATGGAAAGATCTGATTAACTATATTTCCATAAAATATATTGCT<br>ATTGGAAAAGCGGTGTATAACTATGCCATGGATGAACTTGTCAGTGGAGA<br>TAAGAGTATTGAAATGGGCAAGATTAATGACAATTATATTTCCGGTATAAG<br>TTCCTTTGACTATGAGCTTATTAAGGCTGAGGAGATGCTTCAGAGAGAAA<br>CTGCTGTTTATGTTGCATTTGCAGCAAGACATCTTGCTCATCAGACAGTC<br>GACTTAGACGAGAAAAATTCAGATTTTTTACTATTTCCGGATAAGAGTAGA<br>AAAGATAAAGATGGTAAAAATATAAATGATTTTATAAAAGAAGGTATTAAT<br>CTTCGTTCTACTATATTGCAGTATTTTGGTGGGGCATCTTCATGGAGTGA<br>TTTTTCATTTGAAAAGTATATGACAGATGGTCGCGATGATGTAGATCTGCT<br>TACCGATTTACAGAAAGCAATTTATTCTATGCGAAATGACAGTTTTCATTA<br>TACATCTAAAAACCATAATAATGATGGTTGGAATAAAGAATTAATTGGAGC<br>ATTGTTTGAATATGAGGCTAATCGGCTGACTATAATACAGAAAGATAAATT<br>TTATTCAAACAATTTACCAATGTTTTATGATGAAAGTAATTTGAAGGAATTA<br>TTATCATCTCTATACAGTAAATCTGTAGAAAGGGCTTCTCAGGTTCCTTCA<br>TTTAACAGTGTATTTGTCAGAAAATCATTCCCAAAAGTCTGTACACAAGAC<br>TTATCTATTGATGTAAAGACAATGAATGAAGAGGATAAACTTAAGTTTTAT<br>AATGCTCTTTATTTTATGTTTAAAGAGATTTACTACAATCTTTTTCTTAATG<br>ATTCAAACGTTTTAAATCGTTTTATTGATATTTCGACAAAAACAAAAAAGAA<br>CGGAAAGGGAGATGAAGGTACTCATTATTGGGCAGAAAAAGATTTCAGA<br>CAGAGGATTCTCTCTATAATAGAGAGCAGAAAAAATTATACTCTTTCACAG<br>ATATGTCAGTTAATTATGACTGAATATAATCAGCAGAATACAGGCAATATG<br>AGACATAAGTCTGCTGATAAAAACGGTAAAAATCCTGACAGTTATCAGCA<br>TTACAAAATGCTTCTATTGTCATATCTTGGAGAAGCGTTTGTTGAATTTGT<br>AAAAGAAAAATATGATTTTGTATTCACACCGGTTAAAAGGGATTTGATGGA<br>TAAAGAGGCATTTTTGCCTGATTTCGCAAAGACAGTCAATCCTCTTGGCG<br>ATTTAATTGAAAGAGTGAAAGAATCCGGTGTATTGCAAAAATGGTATATAG<br>TAGGCAGATTCCTCAGTCCCAAACAGGCTAATCAGATGCTTGGTTCTTTG<br>CACAGTTATAAGCAGTATGTGTGGGATATATATCGAAGAGCAGAAGAGAC<br>TGGTACGAAAATTAATAAACGTGTTTCAGAAGATACCATATCAGGAGTTG<br>CCATTAGAGATATAGACAGCGTGCTTGATCTTTGTGTAAAGATGTCCGGA<br>ACTATTACGAATAATCTGACAGATTATTTCAAAGACAAGGAAGAATATGCA<br>GCTTATATTAACGATTTTCTTGATTTTGAGTATAAAACCGGAGATTACAAT<br>TGGGCTCTTAAAGACTTTTGTAAAGAAATAACGGATGAAGATGACAAAGA<br>AGGTATTTATTATGACGGGGAAAATCCGATAATAAATCGTAATATTGTTAT<br>ATCAAAACTGTATGGCGAAGCGGAATTTGTTTCAAAAATCTTCAAAAGAG<br>TAAATAAAGAAGATATAAAAGTATATAAAGACCTTAAGAAAAATATCGAAC<br>CATATCAGAATATGGGAACATTTGAAACAAAAGAGCAGCAGGAAAATGTT<br>AAACGTTTCCAGGAATTAAAGAATCATATAGAATTCAGAGATCTTGTTGAT<br>TACAGCGAAATAACAAATGAACTTCAGGGGCAGCTTGTAAACTGGATTTA<br>TTTGCGTGAACGAGATTTAATGTATTTCCAGCTTGGATTCCATTATTTATG<br>TTTGAATAATAACAGTGAGAAACCGGAGCTATATAAGAAAATAGAATTCAA<br>GGACGAAAAAGTCATTGATAATGCCGTACTTTATCAAATTTGTGCTATGTA<br>TACCAACGGTTTGCCATTATACTATAGCAGTACCAAAAATGCAAACATAAA<br>GGAAGTGAGTGCAAAAGCGGGAACGAGCACAAAAGTAGATAAATTCTAT<br>TCATCTGGAATCAGAGCTAATGGTGAAAGTTATTCGAGAGACTATACAAC<br>TTATATGGCTGGTTTGGAGCTATTCGAAAATACAAAGGAACATATAAATAT<br>TACTATGTTTAGAAATGATATTGAGCATTTCAGATATTTAGTTTCAAACACA<br>AGAAGTATGTTGGATGTATACAGTGAAATATTTGATCGGTTCTTTACATAT<br>GATATGAAATACAGGAAAAATATACCAAACATTTTATATAATATTTTATTGG<br>CCCATTTTGTGAATGTTCAATTCGATTTTTCTACAGGGAAAAAGAATATTG |

| | | |
|---|---|---|
| | | GAACAGGGGAGAATATATATGAAAAGAAATGCGCGAAAATAAATATCCAA<br>AATAATGGTGGAATTGTATCTGAGAAGTTTACTTATAAATTAAAGGACGAG<br>AAAACGATTGATTTGCCTGCAAGGGGACGAAGGTACATGGAAACCGTAG<br>CAAGGTTATTATACTATCCTGAAACAGTTGATGAGGAGAAAATGGTAAAA<br>GATTTGGTCATTAAAGATAATAAGCCATTTGGAAAGAAACGAAATAATAAG<br>TATAGTAACAGAAAAGAAGGTGCTTCGGATAGAAAAAAATATGAAGAGAA<br>TAAAGCCAGGAAAAAAGATAATAGTTTTATGTCCGGAATGGACGGTGTAG<br>ATTGGTCTAAATTAAATTTTAAATAA |
| 6 | Ca6<br><br>ContigID:<br>k141_10995992 | ATGAAAATATCAAAAGTTGATCACACCAGGATGGCGGTTGCTAAAGGTAA<br>TGAACTTAGGAGAGATGAGATCAGTGGAATCCTCTATAAGGATCCGACAA<br>AGGCAGGGAGTATAAACTTTGATGAACGGTTCAATAAATTGAATCAATCG<br>GCAAAAATCCTGTATCACGTGTTCAATGGAGTTGTTACAGGAAACAAACA<br>TTTTATTAATACTGTTAAAAGGGTTAATGACAATTTAGACAGGGTATTATT<br>CACAGGTAGGAACGATGAAAGAAAATCTATCACAGATACAGATGTTGTTC<br>TGAGAAATGCGGATAGGATCAATGCATTCGATAGGATTTCAACAGACGAG<br>AGAAAACAGATAATTGATGAGTTATTGGAGATCCAACTGAGAAAGGGCTT<br>GAGAAAGGGAAAGACCGGACTTAGAGAGATATTGCTGATAGGTGCCGGA<br>GTAAAAGGCAGAACTGACAGGAAACAGGATATAGCTAAGTTCCTTGAGAT<br>CTTGGATGAAGATTTCAATAAGACAAAGCAGGCTAAGAATATAAAGTTGT<br>CCATAGAGAATCAGGGATTGGTAGTAGCGCCTGTAGAAAAAGGAGAGGA<br>CAGGATCTTTGATGTCAGCGGGGTTCAGAAAGGAAAAAGCAGCAAAAAA<br>GCTCAGGAGAAAGAAGCTCTGTCTGCATTTCTGTCAGATTATGCTGATCT<br>GGACAAGAGCGTCAGGACTGAGTATCTTCGTAAGATCAGAAGACTGATA<br>AATCTATATTTCTACGTCAAAAACGATGACGATCTGTCTTCAGCAGAAATT<br>CCGGCAGAAGTGAATCTGGAAAAGGACTTTGATATCTGGAGAGATCACG<br>AACAAAAAAAGGGAGAAAAAGGAGACTTTGTTGACTACCCGGACATACTT<br>TTGGCAGATCGTGATGAGAAGAAAAGAAACAGTAAACAGGTAAAAATTGC<br>AGAGAAGCAATTAAGGGAGTCAATACGCGAAAATAATATAAAACGGTATA<br>GATTTAGCATAAAGACAATCGAAAAAGATGATGGAACATACTTCTTTGCA<br>GATAAGCAGATAAGTGCATTCTGGATTCACCATATCGAAAATGCGGTTGA<br>ACGAATATTAGGATCGATTAATGACAAAAAACTGTACAGATTGCATTTAGG<br>ATATCTTGGAGAAAAAGTCTGGAAGGATATACTTAATTTCCTTAGCATAAA<br>ATACATCGCTGTGGGTAAGGCGGTATTTAATTTTGCAATGGATGATCTGC<br>AGGAGAAGGATAGAGATATCGAACCCGGCAAGATATCAGAAAAAGCATT<br>AAATGGATTGACATCATTTGATTATGAACAGATAAAGGCTGATGAAATGCT<br>GCAGAGGGAAGTAGCTGTCAATGTGGCATTCGCAGCAAATAATCTGGCC<br>AGGGTAACTGTAGATATTCCTCAAGATGAAAACAAGGACAAAGAGGATAT<br>CCTTCTTTGGAATAAGCAGGACATACACAAATACAAAAAGAAGTCTCAGA<br>AAGGTATTCTGAAGTCTACTCTGCAGTTTTTTGGAGGCGCTTCAACCTGG<br>GATCTTAAAATGTTTGAGAAGGCATATCCGGACCAGAAAGAGGATTACGA<br>AGAAGAATATCTATATGACATTATCCGGATCATTTATGCACTCAGGAATAA<br>GAGCTTTCATTTCAAGACATATGATCAGGGTGACAGGAATTGGAACAGCA<br>AACTGATCGGAATGATGATCGAGCATGATGCTGAGAAAGTTGTTTCTGTT<br>GAGAGAGAAAAGTTCCATTCCAATAATCTGCCGATGTTTTATAAAGACGC<br>TGATCTAGAGAAGATGTTGGATCTCTTATACAGCGACTATACAGGACGAG<br>CATCGCAGGTTCCGGCATTTAACACTGTTTTGGTTCGAAAGAATTTCCCG<br>GAATTTCTTAGGAAAGACATGGGCTATAAGGTTCATTTCAGCAATCCTGA<br>GGTAGAGAATCAGTGGCACAGTGCGGTGTATTACCTATATAAAGAGATTT<br>ATTACAATCTGTTTTTGAGGGATAAAGATGTAAAGAATCTTTTTTATACTTC<br>GTTAAAGAATATAGGCAATGAAGTTTCGGACAAAAAACAAAAGCTGGCTT<br>CGGATGATTTTGCGTCCAGATGTAAAGAAATAAAGGATAGAGACCTTTCG<br>GAAATCTGTCAGATGATAATGACAGAATATAACGCTCAGAACTCCGGCAA<br>TAGAAAAGTTAAATCTCAGCGTATGATCGAGAAAAATAAGGATATTTTCAG<br>ACATTATAAAATGCTGTTGATAAAGACTCTATCCGGTGCTTTTGCACTTTA<br>CTTAAAGCAGGAAAAATTCGCATTTATCGGAAATGCGGCAACGATACCGT<br>ATGAAACAACTGATGTGAAGGAATTTTTGCCTGAATGGAAATCCGGAATG<br>TATGCATCGCTTGTAGATGAGATAAAGGAGAATCTTGATCTTCAAGAATG<br>GTATATCACTGGGCGATTCCTCAATGGAAGGATGCTTAATCAGTTGGCAG<br>GAAGCCTGCGTTCATACATACAGTATGCAGAAGACATAGAACGTCGTGC<br>AGCAGAAAATAGGAATAAGCTTTTCTATAAGTCTGACGAAAAGATTGAGA<br>CATGTAAAAAGGCAGTTAGAGTACTTGATCTCTGCATAAAAATTTCCACAA<br>GAATATCTGCAGAGTTTACAGACTATTTTGATAGCGAAGATGATTATGCA<br>GATTATCTTGAAAATTATCTTAGCTATCAGGATGATACGATTAAGGAATTA<br>TCCGGATCTTCGTATGCTGCGCTTGATCATTTTTGCAACAAAGATGATCT<br>GAAATTTGATATCTATGTAAATGCTGGACAGAAGCCGATCCTGCAGAGAA<br>ATATCGTGATGGCAAAGCTTTTCGGACCTGATAGTATTTTACCGGAGGTT<br>ATGGAAAAGGTCACAGAAAGTGACATACGGGAATACTATGACTATCTGAA |

| | | |
|---|---|---|
| | | AAAAGTATCAGGTTATCGCGTAAAGGGAAAATGCAGTACCGTGAAAGAAC<br>AGGATGATCTGCTGAAGTTTCAGAGATTGAAAAATGCAGTAGAATTCCGG<br>GATGTTACTGAGTATGCAGAGGTTATCAATGAGCTTTTAGGACAGCTGAT<br>CAGCTGGTCATATCTTAGAGAGAGGGATCTGCTGTATTTCCAGTTGGGAT<br>TCCATTATATGTGTCTGAAAAACAAATCTTTCAAGCCGGCAGAATATATGG<br>ATATTAAGAGAAAGAATGGTACAACTATACATAATGCGATTTTGTACCAGA<br>TCGTATCTATGTATATTAATGGATTAGATTTCTACAGCTGTGAAAAAGATA<br>ATGACAAGCTAGAAGTGGCGGCAGCAGGAAAGGGAGTAGGAAGTAAGA<br>TATCGCTTTTTATAAAGTACTCAGAGTATTTGTATAATGATCCGTCATATAA<br>GTATGAGATCTATAATGCAGGATTAGAAGTTTTTGAAAACAATGATGAGC<br>ATGATAATATTACGGATCTTAGAAAGTATGTAGATCATTTTAAGTATTATG<br>CGTCGGATGATTCTGATAAAAAAATGAGCCTGCTTGATCTTTATAGTGAAT<br>TCTTTGATCGATTCTTTACATATGATATGAAGTATCAGAAGAATGTGGTGA<br>ATGTGTTAGAAAACATCCTTTTGAGGCATTTTGTCATTTTCTATCCAAAGT<br>TTGGATCAGGAACAAAAGAGGTTGGAGTCAAGAACTGTAAAAAAGAAAAA<br>GATAGGGCTCAGATTGAAATAAGTGAACAGAGCCTTACTTCGGAAGACTT<br>TATGTTTAAGCTTGATGACAAATCGGAAGGAGAACCAAAGAAGTTTCCGG<br>CAAGGGATGAACGTTATCTCCAGACGATCGCCAAGTTGCTTTATTATCCT<br>AAAAAAGATGTTGATTTGAACAAATTCATGACAAAAGAAGAATCAATGAAT<br>AAAAAAGTTCAGTTCAATAGAAAAAAGGAAACAAACAGGAGACAACAGAA<br>TAATTCATCAAGCGGAGCATTATCTTCATCTATGGGTGATTTATTAAAGAA<br>CATCAAATTGTAA |
| 7 | Ca7<br><br>ContigID:<br>k141_12677984 | ATGAAAATATCTAAAGTTAATCATGTGAGAACAGGAACGAGAATTAAAGA<br>AAACAATGGTGAAGGAGTATTATATGCTAATCCTTCAAAACAGACAAATG<br>CCGTAAAAGATTTATCTAAGCATATCCAGGATGTAAATCAGAAAGCTCAA<br>GGATTGTATTCTCCATTAAACCCGGTAAAATCTCTTATTAATCCTAAAATG<br>CCAAAAGAAAAGAAGGATGAGATTAATGGTTCATACAAAGCTTTTAAGAG<br>CGTTGTTATCGGTATTGTAAAAGAAAATGAAACAGGAATTCCGGATTCTG<br>CTTCTGTAATTAGGACTTTATATGAAAAAGCTAAAAAAATAGATCTAAAAG<br>TTTCGGATGCATCTTATTTGTCTTCGAAGCTGATAGACAAGTGCTTAAGAA<br>AAAGTCTCGAGTCAAAATCTGAGATTGCAAAAGAAATATTAAAGGCAATTA<br>TTTCTACGGATAAAAGTGCGGTTAATTCGCTTAATGCCGAAGAAGTAAAG<br>GCTTTCTTTGAACTGGTTCATAAGGATTATTATAAGAAAGAACAACTCAAA<br>GCAATTGAAAAGTCTATAGAAAATAAAGATGTCAAAGTTCAGGTAAAAAC<br>AGGACAAAATGGCGAAAATCATCTTGTTCTTTCAAATGCTGATAGTGCGA<br>AAAAGCATTATTATTTTGATTTTGTAAAAGAATTTGCCACAAAAGACAAAG<br>CTGAAAGAGAAGAAATGATTATCAGATTTCGTCAGTTGATTATTCTTTTTT<br>ATTCGGGTTCAGAGTCTTATAAACTTTCGATTGGCTCTGATGTTGGGGCT<br>TGGACTTTTGGTTCGTCTCTTCCTGAAGTTACAGCCAATGTCGACGATGA<br>AATTGCTTCTTTGATTGCAGAATATAATGAAAATATTGCGCGCAAAAACGA<br>TATTCAAAAATCGATTGATTTGAAATCCAATCAAATGAAGAATTATAAGTTT<br>AATTCTCCTGAATATAAAAAATTAGATGATCAGGTTTCAAAACTAAAGGAT<br>GAACAGGGAGATTGTAAGCATGCAATATCCGACGCAAAAAGAAAGATTAA<br>AGCTCTTGTTGAGAATCTGATATGCACAAAATATCGTGATGCTGTTAAGG<br>CAGAGGGCTTAACTGATTCTGATATATTCTGGATAGGATATATTCAGCAG<br>GTTGCTCAAAAACAGTTTAGCAAGAAGGACGCATATAACAATTACAGAAT<br>ATCAACTAAATACCTGTATGAAGTTACATTTAATGAGTGGATTTCGTTCAT<br>GGCATCAAAATACATTGATCTTGGAAAGGCAGTATATCATTTTGCGATGC<br>CTGACTTTAGCGATATTAAATCAGGTAAGGAAGTCCATGCGGGAAAAGTA<br>CAACCCGCTTTTGAAGATGGAATTACAAGCTTCGATTATGAAAGAATTAAA<br>GCCAAGGAAACATTGGCAAGAGATTTTTCAGTGTATGCCACTTATTCATC<br>AGGCATTTTCTCCAATGCTGTTACAGATAGCGAATATAGGCTAAAAGATG<br>AAAAAGAAGATGCTTTGTTTTATAAGCAGGAAGATTGGGAGCAAGCGCTT<br>TTGCCCAATGCTAAGAAGAAGCTTCTTATGTATTTTGGCGGTCAAACAAA<br>ATGGGAGGACTCTGAAATTGAAAAGTTGTCGGATCTAGAGATGACTAAAG<br>CATTTCAGGATATGATAAACGTCATCCGTAACTCGAACTATCATTATGCAG<br>GAAGTGTGTTGGAACCCGGTGAGCAAAGCGTTAACATTGCAAAAATGCTT<br>TTTGAAAAAGAGTTTTCCCAGCTTGGAAGAATAATAAGGGAAAAGTATTTA<br>TCCAATAATGTTCCCGTATATTACAACGTTGAAGATATTAATAAGATGATG<br>ACTTATCTATATCAGGGTGAATCCAAGAGAGAAGCACAGATTCCATCATT<br>TGGCAATGTTCTTAAAAAGAAAGAAATGCCCGGATTTGTATCTAAGTATAT<br>TCCCGGAAACTTACTTGCTAAGTTTGATTCCGAAGGTATGGACAAGTTCA<br>GAGCATCTCTTTACTTTGTATTGAAGGAAGCGTATTATTATGGATTTTTGA<br>ATGAGACTAATCTTAAAGACAGATTTATTATGGCATTTAAAAATTCCGAAA<br>AAGATGCCAAGAATCCTGAAGCTATTGAAAACTTCAAAGCACGAATCGCT<br>GATATGGATGATTCATGTTCGTTTGGTGAGATATGTCAGATTCTTATGACA<br>GATTACAATCAGCAAAATCAGGGTGAATATAAGGTAAAGTCTCAGATAAA |

| | | |
|---|---|---|
| | | ACAAAATCAAGATGAGAAAGACAACAAAGGTCATAAGTATTCTCATTTCAA<br>AATGCTTTTGTATGTAACTCTGCAAAAAGCGTTTATTGATTATATTTTTGAA<br>AAACAGGATATATATGGCTATATTAAAGCTCCGATTTTCAAAAGTAATTTC<br>TTTGACGGAGATGAACCTCAAAAGTTTGTAGAATCATGGGAAGCCAATCT<br>GTTTGGCGATGTAAAGAAAACAACAGAAACAGACTCGTACTATTTGGCAT<br>GGTATGTGCTTTCTCATATGCTTCCGGCAAAGCAGGTTAATCAACTTCAA<br>GGCGGAATAAAGAGTTATATTCAATTTGTCACAGATATAAACAGACGTGA<br>AAAGAGTGTACTCGGAACAGAAAAAGATAATAGCTTGGTCAATAATATAG<br>ATTATTATCAAAATATACTTAAGGTCCTTGAGTTTGTAATGTGTTTTGTTGG<br>AAAAACATCCAATGTTTTGACAGATTACTTTGCTGATGAAGATGACTATGC<br>ATTGCATTTGTATTCCTATGTCGGATTCGCTGGTAAGAAAGAGGAGAAAA<br>CCAATTCTACTCTTTCCGGTTTTTGCAGTAAATCTATAACAAAAGCAGGAA<br>AAGTATTAACGGACAGAATAGGAATATATCATGATGGAACTAACCCTATT<br>GTAAATAATAATGTCGTTAAAGCGCTCATGTATGGCAACGAGAACGTTCT<br>TTCTGAAGCGGTTACCAGAGTTTCAGCTGATTTGATTAATGGAGAAATAA<br>CAAAATACTATGAAGTAAAGAATAAACTTGAAAAGGTATTTGAAAAAGGC<br>GAATGCTCAAATATAGAAGAGCAGAAAGAATTAAGAGAGTTCCAAAACCT<br>TAAGAACAGAATAGAACTTCAAGACATATCCATATTTACGGAAATAATTAA<br>TGACTATATGTCTGAACTTGTAAATATGGCTTATCTTCGTGAAAGAGATCT<br>TATGTATTATCAGCTTGGCTATAATTATATTAGATTGGAATACGGTAATGT<br>TGAGGATAAGTATAAAGAACTTCAAGGCGACAATATAAATATCAAATCAG<br>GAGCTCTTTTGTACCAGATAGTTGCACTTTACACACATGAGTTGCCGATT<br>GTTTATAAGGACAAGGATAGCTATAAATATACTAATAACGGTAAAATTGGC<br>AGATTTGTTAAATCATACTGCGAAGAAGAATTCAATGATTTAGATAATACT<br>TATTTGAAGGGCTTGGAATTGTTTGAAGATATAAAACTTCATGACGACTTG<br>CATATGTTTAGAAATGAAATAGATCATCTTAAGTACTTTATTCGTGCAGAC<br>AAATCTATTCTTCAAATGTATAGTCGTATTTACAATGGATTCTTTAGCTAC<br>GATTTGAAGCTTAAGAAGAGTGTATCATACATATTTGCCAATATCCTTGCG<br>AAGTATTTTATTATTGCCGATACAGAGATGAAGAGTTCGGTTGAAAACGG<br>AAAAAGAGTAGCCATGCTTTCAGTCAAGGGATTGGAATCGGATGTATTTA<br>CTTATAAGGGCAAAAAACGTGATAAAGAAGGAAAAGAAAGAGACAGCAA<br>GTATACATTGCCGGTTAGAAGTGATGAATTCTTAAAAGAAGTCAAGAAAC<br>TTCTTGGCTATAAGAGTATGTGA |
| 8 | Cb8<br><br>ContigID:<br>k127_4804511 | ATGCCAAACGTAAAATTCACCTTAGTCCCTGTGGACTATTCAAAGCCTTA<br>CGACGAACAACCCGATTGTAAACGTCACGTCATTGGAGCTTATGCTAATC<br>TGGCTCGTCACAACATGACACTAACCATCAACACCATCATGCAGGCCATC<br>GGCATGCCTCTTTTTAATGAGAACGAAATAGAAAATGCTTTCAACAAGTC<br>CCATCGCAAAAAATTGGAAGCTCTTGACAACATCCAGAAAGTGAAACTGC<br>AGAAACGCCTCTATCGTCATTTCCCGTTCTTCAAGCGAATGAAGTTGGAG<br>GATGAAGAGAAAAAGACGGTTCAACTGAAATCGCTCATGACAGTCATGTC<br>ACTTTTTACAAGTCTGATGGCTGATATACGCAACAATTACACACATTACCG<br>ACCTTATAATAACAAAGAAGAACAAAACAGACAATTAGAACTTAAAAAAGA<br>AGTAGGCAAAAAACTACAATATTTGTACGAGAATAGCAGCCAGACATTTA<br>AGAGCATGGAGGAACTTGATCATTCCAGCAATGAGGTGCTTTCAGCTCTT<br>CGCATTCCTGAAGACGTTGTGGAACGTTTCTCGCCAGACGATCCTGATTA<br>CAAGAAACTACTCAACACGCTACATGATTCCAATATACCGAAATGGAAAA<br>AATCAGGTTTGAAACTCGACATGAAAACCCAGATAATCACCAAGAAATCT<br>GTACGCTATGTGCGAAACCCTAACTATCAAGCCTATATGATGGATGAAGA<br>GAAAGGCTTGTCTGATATAGGCATTATTTACTTCTTGTGCCTTTTTCTGGA<br>CAAACAAGTATCTTTCTCACTAATGGATGAAGTTGGCTTCAATCAGCAGA<br>TTAAGTTCACAGGTGAGCATGCAGAACAGCAGTTAATGTATGTCAAAGAA<br>ATCATGTGCATGAACCGTATCCGGATGGTGAAGGCCAGGATAGACAGTG<br>AGATGTCAGACACGGCACTGGCATTGGATATGTTAAGTGAACTGCGCAA<br>ATGTCCTCGTCCGTTGTATGATGTATTCTGTAAAGAGGCACGTAACGAAT<br>TCAAAGATGATGCCACAGTAGTTTGGGAAAACACTCACGGCGAGGAGGC<br>TGTTATTACTGAAGAGCAAGGTGATATAGGGGAAGAGACAGATGCTATTG<br>CCGCAAATACTACGGGAAAAAATACTCCTCGCAGTACCTTTGTACGCTGG<br>GAAGACCGCTTTCCCCAGTTGGCACTCAAGTATATAGACTTGACAGGTAT<br>GTTCGACCATCTTCGTTTTCAACTTAATTTAGGTAAATATCGCTTTGCCTT<br>TTACCAACACGACAAAGCATACAGCGTTGATAATGCTGAGCGCCTGCGT<br>ATTCTTCAAAAGGAATTGCATGGTTTTGGACGCATCCAGGAGGTGAACGA<br>AATGATGAAGGAAAAATGGCAGGATGTCATGGAGATCAAGAACGTTGAG<br>GATGGACAAATATATAAGGAACCCGATGTAGCAGGGCAGAAGCCTTACG<br>TGACCCAGCAGAATGCCCAATATGACTTTGACACCAAGAGCCACTCCATC<br>GGCATCCGATGGGAGGGATGGCACAACAACCACTCTGACAATCATTATG<br>GCGATTTGGATAGGAGGGATATGTTTATCCCAAGATTGCCCGCTAACCCT<br>GCATCGCCTGAGGGCGACAAGCGTCAGACTAATCAGGCAGAAGAACTGT |

| | | |
|---|---|---|
| | | TACCGCCTCAGTGTATGCTGAGTCTCTATGAGTTGCCAGCTATTTTGTTC<br>TATCACTATCTGCTTAAAAAATATCAGAAAAATACTGGATTGGTGGAGAAG<br>AAAATCTCCGACTTTTACACCAACATGAAAAACTTCCTGACAGAAGTCAG<br>CGAGGGTAACATATTACCTGCCGATGAGACCACGTTAATCCGTGAATTAC<br>AGACAAGAGGGCTTAAGTTTTCTGACATTCCTGTCAAACTGAAGAAACTG<br>CTCAAGGGCGAAGTGACAGACAATGCAAAGCGCATGGAGGAATCAGCC<br>CTCCTGCGTCTGCATGAGCGTAAAGATAAGAAAAGACGGGCACTTGAAA<br>GTTTCATTGCCAAATGCAAGATGATTGGCACCAAAGAAAACAAATTTAATA<br>AAATACGTGCCGTCGTCAAGACTGGATCCTTAGGACAGCTACTGGCTCG<br>CGACATCATGGAATGGCTAACGACAGACACAAAGAAACGTATGAACCTG<br>ACAGGTCAAAACTATGTTGCCATGCAAACGGCACTTTCGATGATGGGACA<br>AAGTTTCGAGTTGGCACCTGAAGCAAAGGTGACTTGTGAAAAGATGAGA<br>AACATCTTTGTTAAGGCTAATATCCTGCCCATGAACGATGACGACTTTGA<br>TGCAGATTTTCATCATCCGTTCCTACTTGATGTATTTGACGAAGAGCCTGT<br>TTCTATCGAAGACTTCTACAAGATTTACTTGGAGAAAGAAATCTTCTACAT<br>CGACTATCTAACCGAACATTTCAAAAAGTACAAAGCCAAAGGGGCTGCTC<br>TGTACATACCATTCCTGCATTGCGAAAGGCTTCGATGGAAAAATACTGAA<br>CAGAACGGTCTGAAGGAACTGGCAGCTCGCTACCTGCAGCGTCCTTTGC<br>AATTGCCTAACGGACTCTTTACTGATGACATTTTCCATTTATTGGAGGATA<br>TAGCCACTAAGAATGCCGATTTTGCAAAAGTGCTTGAAAAACAGAAAAAA<br>GACAACCATCAGTTGCAGCAGAATGTAGCCTACATGATTCGAATCTACAT<br>GAAGACAGTAGAGTATGATCAGCCACAAAACTTCTATAACACGATGCCCG<br>TCGGAGATACCAATAGTCCATATCGTCACATCTATCGCATCTTCAAGAAA<br>TTCTTCGGTGAATCCATCCCCAAGACAAACAAGACGACATCGCCAGCCTA<br>CACCATCGAGGAGATTCGTGCCATTTTGAATAATAAGCAGCTATTGAAGG<br>ATAAGATAGACTTCTTTGTCAAGGAAGAAAAAGAGAAACTGAAAAAACAA<br>CAGATACGTGATTTCAGAAACTATGAAAAAAAGCAGTGGAAACTGCTAAA<br>AGCAAAAAACGAAGCAGCCCCCAAAGGCCAGCACTTTAATGTGAAAGCT<br>GAAGTTCAGAAACGATTGAACGAGAAACGCGAGGAGCAGCGCAAAGCAT<br>TAGATCTTTTAGTTATGGATGTTAAACAGAAGTTGGAGGGTAAACTGAGA<br>AAAGTGAATGATAATGAGCGAGCAATACGCCGTTATAAAACACAGGATAT<br>CCTATTGCTCTTCATGGCTCGTGAAATTCTAAAGGCAAAGAGTCAGGACG<br>AGGACTTTACCAAGGGATTCTGTCTGAAATATGTTATGAGCGACTCACTG<br>CTCGACAAACCAATCGACTTCCAGTGGACGGTGAATTTCCAGAATAAGGA<br>AAAGAAAACTATCGCCAAGACTATTGAGCAAAAGGACATGAAGATGAAAA<br>ACTATGGTCAGTTCTACAAGTTTGCCAGCGACCACCAACGTCTGTCGTCG<br>CTACTATCACGGCTACCTGCCGATATATTCGAGCGTGCAGAAATCGAAAA<br>TGAGTTTGCATACTATGACACCAACCGAAGTGAGGTATTCCGCCAGGCAT<br>ATATCATCGAGAGCAAAGCCTACCAGTTAAAGCCAGAGTTGACTGATGAT<br>GCGAATGCCAATGAAGAGTGGTTTACCTATCTTGATAAAAAAACAAAGAA<br>GCTTCGCGCTAAGCGCAACAACTTTGGCGAGCTGCTGAAGATTCTGGCC<br>GCTGGTGGCGATGGCGTACTGGACGATGTCGAGAAGAGTCTGCTGCAA<br>AGCACACGCAATGCCTTCGGACACAACACCTACGATGTAGATATGCCTG<br>TTATCTTCAGCGGGAAACTGGATAAAATGAATATCCCTGAGGTGGCTAAC<br>GGTATTAAGGACAAAATCATAGAACAGACTGAACAATTGAAGAAAAACGT<br>TTAA |
| 9 | Cb9<br><br>ContigID:<br>k127_1483864 | ATGGCTATATTTGTAATTAAGGTAGAAATCCAACGACAAATGACACTCGC<br>ATTCAAAAACGAAACCGAACAGAAACCCGTTCTCGGAGCATACGCCGCC<br>ATGGCAAGAAACAATGCCTTCCTTACCGTGATGGATATCATGGACCAATT<br>GCATATCCCCCGGACTGTGCTCACCGATAATTCCGGAAAAGAAGTGGAT<br>CCGGAATCCCATATATGGAGACTCAACTTATTCCCGAAGAACTATAGGCT<br>TCTTCCCGAACAGGAGGCCAGGGCTTCCCGGCTCCTTCGCAATCATTTC<br>CCTTTCCTCGATCTCGCGGATGACCTCAAGGATGTTGAGAATGACCATG<br>GACAGGCAGCAAGGAAAAATGTCTCCTATGAGGACTTGTGCAATTCCTTT<br>CTTACCATGATGGAAGTGCTGACGCACCTTAGAGATGTCAACCTTCACTA<br>CAAGATCAAGGACGAAAGAATCGCGGATTTCTATTTCCGCCGTGCGGAG<br>AAAGAGACCGGGCACATCCTCCGGGAAGTGCTGAAAGCAGCACCCCGA<br>AAGATCAAGGACCGTTACAAGGGAACAGCGCTCATGGATGAAACCTCGT<br>TGTCCTTCTTCACGGACGGGAATTACATCCAAAAGGGAAGGAAATATTCT<br>TTCAATTCCCGATGGGCCTTCAATCCGCAGCGGCAACCGCAGCCGTCTG<br>AAGTGAAAGTCTTGAAGAACAACGCTCCCGTCATTGACAGAAACACGGG<br>CATACCGCGCATGTTCGAAAGGCTATCCACTTTCGGCGAGATCCTCTTCA<br>TCGCCCTTTTCACGGAGAAACGGTATATCCCCGACCTTCTTAGAGACAGT<br>GGGCTTGACAACAACTTCATGGCATCAGGCGATAACGGAAAAATGTCCC<br>AGCAGAGGATTATTCGGGAGATCATCTCCGCCTACAGCATCCGGCTCCC<br>GGAACGCAAGTTGGATATCGAGACCGGGGCGACACAAATCATGCTGGAC<br>ATGCTGAACGAACTGGCGCGCTGCCCCTCTGAACTGTTCGATGTTCTCC |

| | | |
|---|---|---|
| | | CGGAAAGCGAACGAAGATCCTTTGAGATAACGGGCAGTGACGGCTCCCA<br>GGTCCTTATGAAGCGTTATTCGGACAGGTATGTCCCGCTGGTGTTGCGC<br>TATTTAGACGTGACTGAAGAATTCAAACGGCTCAGGTTCCAGGTCAATAC<br>CGGTCTGCTCCGGTACGAGCACCACGGACCGAAGGAGTATATGGACGG<br>TGTCGCCCGGTTCCGTATCGTCCAGTCAAGCATCAATGGATTCGGGAGA<br>ATCCAGGAGATGGAAGCGGCACGGACAGCCGGCGCGACTTACCTCGGC<br>TTCCCCCTCCTGAAAACTGATGACGATGGAAACATGACCGAGATGCCCT<br>ACATTACCGACTCGGCGGCCCGGTATGTCCTGAACGGCGATCTGATCGG<br>CTTATCATTTGGGGACGCCGCTCCGAAGATTGATACCCTCCCTAATGGG<br>GCGGGATTTAAGTACAAGGTGTCCTGTCCTCAACCGGACTGCTGGCTAA<br>GCCGGTATGAGCTCCCGGCGCTGGCCTTCTATACGTTTCTTTCCAGAAA<br>ATACCATATCTCACGGAGTACCGAGGATATCGTCGAGGATGCTTTAAATG<br>AATACCGGGCCTTCTTTGCAGGTATCGCTGATGGAAGCATCACATCCATG<br>GACGGGGTCGGCATCCCCAGGAAGAATATTCCGGAGAAGTTGCTTGACT<br>ATCTGGAGGGGAGGGGCAAGAGAACGGATTTCAAGAAATACAAGGAGG<br>GTCTGGTCGCAAAGATGCTCCTCAATACGGAAGCACTGCTGTCCCGCCT<br>CAAAGAAGACTTGAAGGTAATTGGAACGAAGGATAACCGTATCGGAAAA<br>AAATCCTACGTGCGCATCCGCCCCGGTAAGCTTGCCGAGTTCCTCGCAG<br>AGGATATCGTCCGGTTTCAAGGTCATCCTGCGGGGATGCCTGAAAAGAA<br>ACTGACAGGACAGCAGTACAGCATCCTACAGGGAATGATTGCCACTTTC<br>CATGAAGGACTGGCCGACGCGTGTCGGAATGCCGGCCTGCTAGACGGA<br>GATTCCGCCCATCCATTCCTTTCCTTGGTATTCACACGCCATGCACAGGG<br>AATGACATCCACAGTCGATTTCTATCGGGCATACCTTGAGGAGCGGCGT<br>ACATACCTAAAGGGCGTGGTTCCGGACGAGGCTCCCTTCCTTCATCAGG<br>AAAGAAGGAAATGGTCGGCCAACAAGGATTCTGCGTATTACAAGTCCCT<br>CGCATTGCGATACATTAAGGATGAGAAGACGGGAGACAAGGTCGGTGTC<br>TTCTTACCGCGGGGACTGTTTGACAATGCGGTTCACGCCATCATCAAAGA<br>GCATTGCCCGAACACGTCGAAGGTCATCAATGCTTCAGAACGGGCAAAT<br>ATGGCCTTCATCATACTTACCTACCTGGAGAAAGAACTGGATGACCAAAA<br>CCAGGGATTCTATTTCAACGAAGAAAGGCTGAAGGAGTACGGGTTCTCG<br>AAAGCCATACGGAAAGAGTTGGAAGAGAGCGGCATGAAACGACTGTCTC<br>ATGTCCTGCGTCTTGAGAGGAACACCAATCCTTCTGGACTGTATTACGAA<br>GCTCTCAGGGAAGAGTCCGGCTGGAAAGATGACCGTAGAAAGGGTGGA<br>CAACTGGATAGGAAGACCGAGGAATTCGCTGAAAAACTGCGCCATAGTT<br>ATAAACGTATGTGCGATAATGAAAAGACTATCCGTCGATACATGGTGCAG<br>GACATTACACTTTTCCTTCTGGCAAGAAGTCTTGTCCGCATCGCAGGGAA<br>CTCCGTCAACCTATGGTCAGTAGGGCCAGAAGGGAATGGAATCCTCGAC<br>CAGTGGGTCGACGTAATGACACCGTATAAGAAATACATAATCCGTCAGAA<br>GGGTATCAAGATCAAGGACTACGGTGAGATCTACAAGATTCTTAAGGACA<br>GGCGGATTGACTCCCTCCTCCTGAACCAGAAGAAACGGGTGCCGGAGG<br>CTATCGACCTGGAGGAAATCAAGGAGGAACTCGTCACATACAACCGCAA<br>ACGCGCCTCCATGGTCAGCGCTATACAGGTGTATGAGAAAGGTGTCTTC<br>GAGGAGAATCGGGAGCATTTCGACAGTATGACCAGTCGCTTCGGCTTCA<br>AGGAAATCCTTGAGGCAGACGTAAGGTCAAGTTCCTTGACAAAGGAGGC<br>GGTTAAGAAAGTAAGGAATGCGGTCTCCCACAATCAGTATCCCGACCGT<br>ATGGTGATTAAGGATGGCCGTAGCATGGTACTCTACTCCCCCGATCTTCC<br>CGATATGGCGAAGGGTATCGCCGAAACGACCGATAGACTCACGAAATAT<br>GGAACTGATATTAGTAAGCAAACTGACGAGTGA |
| 10 | Cb10<br><br>Contig ID:<br>Cas13/21_contig<br>-81_616 | ATGTTACAAACAGAAAAAAACGATCGTGGAGCTTTTTGGGCTGCTTATTT<br>CAACACGGCAACGAACAACGTTCAGGCGATTCTTCAATTCGCCGGAAAA<br>AACGTTCAGCTTGAGGAACTTTCAAATCAGGAGTTCAAGCTTTCAGCAAA<br>CGGAATCGGCGATGAAGAGACAAAAGAACTCGAATGGTCGGCCGAAAG<br>CTATCCGGCAATTCAAACGCTGAAAATGACAGGCGATGAAGTCAACATTC<br>CAGAACAAATTCGAATCATGAAGACGCTGTCCAAGCATCTTCCTTTTATG<br>AAGAGAATCTCAACTCGCGTGCAGAACGGCGTTCGCAAAAATGGCAAAA<br>CCGAAAAAGCGGGAGAGGAAATGACGCCGCAGATGTTTGCTGAAATTCT<br>TATTGGGTACGTGAATTATCTTTACGACCTGCGTAATTATTTCACGCACTA<br>CAAACACGTTCCCGTCTCCCGGAAGAATATGCAGTCCGAGTATTTGGGG<br>ATTCTCTTTGACGCCAATGTTGGAACGGCCAAAGAGCGCTTCTATTCGGA<br>AGACAAGATTGCCAAAGACGACAAGCGGTTTAACAATTTCCGTATGCACA<br>AGGGCGCGGAATCCGTACCGGACAAAAAAGGCGGCATGAAAAAACAGC<br>CGAAACTCAACAAGGATTTCCTGTTTTATCTTTGGGAGACGGATCCAATC<br>ACGCCCAAAGGCAAAGACAATCCGTATCTCGAACTGACAGCTCAGGGGC<br>TGGCGTTTTTCCTGTGTCTGTTTTTGGAAAAGAAGTCCGCTAACATGCTTT<br>TGGACTCCGTCGGCATTGAAGAAGACGCGGAAAGTTTGATTGACGGATT<br>CGGTTTCGAGAATAATTCCGGCGATAATCGAACGCTGCTCAAACGGATTT<br>TCACCATTACATGCGCCCGACTTCCGCGCACTCGTCTGGAAAGCGAGAA |

| | | |
|---|---|---|
| | | TCTGATAAGCAATCAGGTTCTTGGCCTGGACATTATGAACTATCTTCATAA<br>ATGCCCGAAAGAGTTTTACAATCTGCTTTCACCTCAAGATCAGCATAAATT<br>CCGGACGCTTTCAGACGATCACGGGACGGAAACGCTGCTGAAGCGTTTT<br>GACGACCGGTTCCCGTATCTGGCTCTGAACGTTATGGACAGATTGGAGT<br>GTTTCGATTCTCTGCGATTCTGCATTGACATGGGAACGTTCTATTTCCGTT<br>GTCATTCTCGCGTTCAGATTGACGGCTCCCGTCTGGAAAACCGCCGATT<br>AAAGAAGAAGCTGACGTGTTACACCCGCCGTCAGGACGCGATCGAGTAT<br>TTTCAAACCGAACGCGCGGCGGAAAACACGTTCTATCAAACGGACAATC<br>TTGCTCCGGCTCCGAAGGCGTATCGAACGGACATGCTTCCGCAGTACGA<br>CCTCGGCCGCGGTCGGAAGCAGGAAAACCGGATTGGAATTGCTCTGAAA<br>TCGTTAGACGACAGCCGGCCGATGTTCAATCAACCGACGCTGGATCCCG<br>CCGGGCAAATCAAGCCGAAAACGTACAAGCCGGACGCGTGGCTGTCCA<br>CGTACGAACTGGCTCCCGCGCTGTTTTTGTCTTTGCATGGGAAAGGCAG<br>CGACGTTGAAAAGCGAATTGAGGAATTCATTCGTTCCTGGAAGGGGTTT<br>GCAGACTGGATGTCCAAGGCGTCAAAAGAACAACTCAAAGAATTGCGTT<br>ACAATTCCGCCAGGGAGGAATTCGAAGTTTTTAAGTCCCGCTTTGAAAAG<br>CAGTTTGGTTTGTCGGTAAACGATATTCCGGACGAATTCCGCTATTATCT<br>TGTCAACGGCAAGATCAAGCCGATTTATATGCGATTTCAAAGAGGCAGC<br>GCTGTTCCAATAACGATGGACGAAGCCGCTCAGATTTGGCTTTACAACGA<br>ATGTGACAAGACCCGATCGATCATTCGGAAATTTAACAATGAGCAAAGCT<br>ATGATTTTAAGCTCGGCAAAGGCAGGCAGCGCCGGTATACATCCGGAAA<br>CATCGCCTTATGGCTTGTTCGCGATTTTATGCGCTTCCAGAAAGCCAACG<br>ACAATCCTCGAGAAGGTTTTGGAAAACTGAAATCGTCTCCAGACTTCAAC<br>GTCCTCCAGTCGTCTCTGGCGCTGTTCAACAGCCGAAAAGACAGTCTGA<br>AAGAGATTTTGAAAAATGCTCGACTGATTTATAACCCGTCCGGCAACCAC<br>CCGTTTTTTGGAAAACGTAATCAATCGCAGCGGGGCGTTGTTGTCAATTGA<br>AGGATTCTTCAACGCCTATCTGGACGAACGATTGGAGTGGCTAAGCAAT<br>GCCTCTGGACGGGAGGTTTATCAGCTTCGCAAGCTGTACGAACGCGGAG<br>AAAAGAAGCGCAGCGCCGCCAAAGGCCTTCCCCCTGCAAACGTTTATTT<br>GGGCGAAATGGCAAAGCAGTTTCTGGACGAGTCCGTTTGCTTGCCGCGA<br>GGTCTGTTTGACGACCTTGTTCGTTCAGCGCTCAAGGAAATCTATCCGGA<br>ACAGTACGCCAAAGACGTCCCGGACGGTCAGCGCGCCAATTTCACCTGG<br>CTGATGCAAAAGCATCTGAAATGGTCTGGAGACGATCATCAGTGGTTCTA<br>TAAAGAGCTTAAGCAGTCGATGACGGCTGAAGAGTTCAAAAAGCTGTTTG<br>AATTGCTTGGAACAACGGACGTTCGTTACGACAACGCAGGCAAGAAAAC<br>GCTCAATGACAAACTCGAGGAGACGTACTCTCAGCGCTGGTATGATTTG<br>GAAAAAAGCTTGAAAAATGATAAAGAATTTAAGAGCCTTTCACCTAATGAT<br>AAACAGAGCGCTATTGACGACCGCATGGGGCAGGAGAATCAAATGTACC<br>GAAGCCTGCTCAATCGTCTTCGCGATGTGTGGAAGCGCATTCGAATGTC<br>CAGCGTTCAGGACGTTGTACTTCGCGACGCGGTCTGGCAGCTGTTGGGA<br>CTGCCGGAAAAGTCTGTGCGTCTGTGCAACGTCAAGCCGGAATACGACC<br>TGAAAACGCAGCGCATGATACGCGACGACGCTGGAGACGGCAGCATCT<br>CCAACGATGCAATTCTTAATGAGACGCATTCGTTGGAAAATATCATCGAT<br>CTTCCCAAGGGCTTGGGAACAATCGTGCTCAAAGGCGAAATGAAGCTCA<br>AGAATTTTGGAAACTTCTACCGCATGCTTTCCGATCCCAAACTGCCTTCG<br>TTCCTTTGCCTGTACAAACAGTTTGGGTTTAACGCCGTTGATTACAGCTAT<br>CTCGAACTTCAGGAGTTTGAGTTCTACGATCGCAAGTTTCGTCCGGCTGT<br>TTTCAATTGGGTTCATCAGCTTGAGGAGACGGTGTTGGAAAAGTATCCGG<br>ATTTGCCGATGAAATATGGCAAATTTGTTGATTTTTGGACAATTGCCGGTA<br>AAGCAAACGGCGGGTATATTTTCACTCAATTATTGTTGACGTCAATTCGC<br>AACGCCGTTTCGCACCAGTACTATCCGGAATTCTGCGTTCCGTCAGATTG<br>GACAAATCAGAAGGATATCGACACGTACTCGCCGCAGTTCAAAGATCAG<br>CTTTTAAGCGTCAAGGCAGAATTCCTCAAACGACGTTCTGAAGACAAGGA<br>CGCATTGTTGTCAGAAACGATTTACGACTACGCGCAGTCTCTGTTTGAAA<br>ACGCAATCGCCTTTGTTGAACAGCAACCGTAA |
| 11 | Cb11<br><br>ContigID:<br>k141_16137484 | ATGGCAAATTTTCAAACACCCCAAAGGCATATTTTTGGTACATATCTGAAT<br>ATAGCTCGTGCCAATTTTTACAAAACGATTCTTCATGTCTTTTCTGCATCA<br>GGCATTGACTGTTACACAAAAAGGGGTGACCTCTTTGTAAGAGAAGACAC<br>CGTAGACAAAGTCATTAGTGCATTATACTTGATTGTTAACGGAGAGAATG<br>CAGGTTATCACGCCATCAAAGAAATCGTTAGTAAGAGCTATGACAAGCGA<br>TGGAAAGAAGACAAAGCACTACAAGGTAATCTTTCTGGCAGTGAACTGAA<br>AGCTAGAAAAGAAGAGTTTAAATCGCCTCTGAAAGATGAGGGGCCTGAT<br>GGCGAAGATGCAAGAATTTGCAAATCATTCACATTAGGAAGTGAGCAGG<br>AAGAGCGAATGAGGAAATTGCTTTTCCGGCATATACCTCTATTATCACCT<br>ATTATGGCAGATGTTGTAGCCATGCAGTTCAAGGAAACTACCAATGAGCA<br>TCAAGAGGCAAATAAAACGTTACATGATGCAACGCTTGCCGATTGCTTCA<br>AAGAACTGAGCAACATTGCAAGGTGTCTCAGTGAAAGCCGCAATTTCTAC |

| | | |
|---|---|---|
| | | ACACATAAAAATCCATACAACTCAATAGAAGCCCAAAGAACTCAGCTCCA<br>ATTACAGAAAGTCATCGCAAACAATTTGGATAAAGCTTTCATTGGCAGTC<br>GTAGAATTGCGAAGAAACGTAACAGCTATTCTGAGAAAGACCTGGCATTC<br>TTGACTGGTCATGATAATGATTGCAGAATGGAAGAGATATTTGTACTTGAT<br>GAAAATGGTAACAAAATATGGAAAGTAGAGAAGGATAAAAACGGAAAAGA<br>CAAACTCGATAAAGACAAGAATCCCATTTATGTCTATAAGAAGGTTAAAG<br>CAAAAGACGGGAAAGGTAGAGAGAAACTTGATGAAAAAGGGAAGCCAGT<br>TTATGAGACTCTACGTGAAAATGGGGAACCTGTACATGAATATGAGAAAA<br>AATTTGTAGAACGGAAAGATTTCTATTTCCGTATTAGGGGTAAGCGTGAA<br>GTATTAGCTCCTGATCTTACCCCAACAGGAGAAGAGTGCGATGGCCTTTC<br>TGCATTTGGAATGCTTTATTTTTGTTCTTTATTCTTGTCAAAAGAGCAAAC<br>GGCACAATTGTGTACAGAATCTCGTGTATTCGTCACGAGCCCGTATCAGC<br>CTGCTGGTAATTTAAAGAATAATATCATTCTTAATATGATGTTCGTATATG<br>CCATACATATTCCGCGTGGCAAGCGTCTTGACAGCGAGACTGACAGCCA<br>GGCATTGGGAATGGATATGCTAAACGAACTACGTCGCTGCCCTATAGAA<br>CTTTACGATGTACTACCATCTATCGGAAAACGGGATTTTGAAGACAATGT<br>TAAACATGAGAACAATAGAACGCCGGAATTGTCCAAACGTATTCGATTAA<br>AAGACCGTTTCCCATATCTTGCCATGCGCTATATTGACCAACAACAATTAT<br>TTAAAAGAATACGCTTCCATGTCCGGCTCGGTAGTTTTCGTTTTTGCTTCT<br>ATGATAAGACTTGTATTGACGGAAAGTCACATCCTCGGCAACTCCACAAA<br>GATATTAATGGCTTTGGCCGGTTGCAGGACATGGAAAAAGAGCGTAAAG<br>AACAATATGGGCACTTTTTCCAACAATCAAGAGAACAGTCCATATGGCAG<br>AAAGACGAAAATGCCTATGTAAATCTAAAGCAATTGGAACCAATTAAAGC<br>TGGTGACCAGCCGCATATCACCGATATGTTTGCCCAGTACAATATCCATC<br>AAAACCGAATCGGGCTCTTTTGGAACACCGATGAAGAATGTAAACTTGTT<br>AATAAAACAAATGCTCAGGGAGAGATAATTTACAACGGATATTATCTTCCA<br>CCGCTTAACTACGTGGATTCTCCAACAGAAAACAACAAACACAAAAGAAA<br>AGCACCTGTTGATATGCCGGCACCACTATGCTCATTGAGCGTTTTTGAGT<br>TGCCTGCTATGTTGTTCTACAATTTTTTAAGAAACACCGATAGTCTTGGGG<br>GAGAGGAATTTCCCGATGTTGAAGAAATTGTTATTAAACAATACGACAATA<br>TCCGAAAATTTTTCTTAGAGGTCAAGGATATTCAGCCTACTGACAATATAG<br>AGAATCTGGCTACCATCCTAAACGCTTATGGTTTGTCTAAACAAAGTGTT<br>CCAAAGAAGATATATGACTATTTGTCTAATAAGAACACATTAATTAGTAAA<br>GACATTCGGAAATCTACAGAGAAAGAGGTTAAAAACCGCCTTAGAAGAG<br>CAATCATTAGAAAGCAACGCTTTGAACAGGATCAGGAGCACATAGAAAAC<br>ACCAAGGATAACAAGATCGGAAAAGATAGCTTTGTCAGCATCCGCTATGC<br>CAAAATTGCTGAGGAACTGGCAAAGTCCATGATGGAATGGCAAAGCGGC<br>AATACAAAGATGACAGGATTGAATTTCAGGGTTCTGACTGCTGCTCTGGC<br>AAAATTTGGCGATGGTGTAATAAAGCGAGATACGATCATTTCAATGCTTC<br>AAAAGGCGGAGATCATGGGAGGAGACAACCCACATTGCTTTATTGAACA<br>AGCAATCGAGCAGGAACAATATGACATTGAAGAATTCTATCTTGATTATAT<br>TAGTGCAGAGATTGAATACCTGAAGCGTTTTCTGATGATTGATGGAAAAA<br>CGATTATATTGAAAGACGAGCAATTGCTTGATGCTCTTAGGAAAGATAAA<br>GAAGTCCATGATGATGTCAGGATTCGTCTTAAGAACGATGTTGATTTCGG<br>CCAACTCCCTTTTATCCACAAGTCACGCTTGCGTTGGCAACAAAGCAAAA<br>TCGAAGAATTGGCAAACAGATATCTTTATGTAAAAGAAGAAGGTGAAGAA<br>ACACTTGGACGTGCCACTCTTTTATTGCCTGATGGTATGTTTTTCCCATAT<br>ATAATGAAGGGATTTCAGAAATGTCATCAGGAGCTAACAAACAGTATAGA<br>AGCCCTTTCTGATGAGCAGAAAAAAGGAATTGAAAACAACGTGGCTTATT<br>TGATAAATCTCTATTTTGAGTCTAAAGGAGAGAAGTCCCAATCGTTTTATG<br>ATTCCACAGAACCTTCACACTATAATGACAATATAAGGCAATTGGCACCA<br>TACAAATATGCTAGATCTTATGAATTTTTCAAGATAATCAAAGGCTGGCAG<br>ATTCACCTGTCTTGCGATGAAATGAAAAAGAGGTTGACTGGAAAAAAGAC<br>CATCATCGACAACAAAATTAATGCTTTAAAAGAGAAAGGAAATTACATTTC<br>TTTAGAGGAAGCCAAAAACGCTCTTCGAAGAAAACTTCACAACACGTTCC<br>GCGACATGCAGGACAATGAGCGTGTCATTCGTCGTTATAAGACACAAGA<br>TCGCATTCTTTTCCTTATGGCGAGAGATATGATGGGCGAAATTGTTAACA<br>AGAAGGCGGACTTGTTCAAGTTGGAAAATGTATGCAAGGATGATTTCCTT<br>AGTCAGAAAGTTAAGGCATCCATCGCGGTACATTTATCTATGGGAGAGGT<br>GTTTAAAATCCAGAAAGATGAGATGGCAATAAAAGATTATGGTAAATTGTA<br>TCGGCTATTGAGAGATGACCGAATAACCAAATTGTTATCTTATGCCTTGTA<br>CGAAACAGGTGAAACAATCGACTATGATGACATCACGGATGAACTGAAG<br>GAATATGAAAGTTGTCGTTCTGCTGCTTTCGAGGCAGTTCAGATGATAGA<br>AGATACAAGATACCAACAAGACAAAGAGGTGCTTTCAAATCCCAATGAAA<br>ACAACTTCTATTGTGGAAACATACGCTATAAAAATGGAAAAGATAACGGA<br>AGAGAGAATGAGGCGAAACGAAACAATTTTAGAACATTATTAGAAGATCT<br>CCAAAAATTTACACCAGAACAAATGGAAATGTTTAGTCAGGAAGATAGAG |

| | | |
|---|---|---|
| | | AATTGATTATATCTATCCGTAACGCGTTTGGGCATAATAGTTACCCAAAGC<br>AAGTCGATTTTGAAAGATTAATCAATCAAGAAAGGAAGAATAATCCCAATT<br>TTAAAATCGAGCTAAAACAAGTTGCATCTTTCATTCTGGATAAATTGGAAG<br>AATATGTAAATCAAGTAAATCCCCAAACATAG |
| 12 | Cb12<br>ContigID:<br>k127_333529 | ATGAGCTACACCCCCTCCTCCCGCCCTCCGCGCCGCCCGCAGATCGTC<br>GAAGGTTCGCGCAACAACGCCCTTCGCATCCTTAAAATCACGCCCGACG<br>AACAGACCGCGTTCGTGACCTACCTCAACTACGCCGTCAACAATCTCTCC<br>GAAGTGGCTGGCGTCGCGTTTTCGGACGAGAGGAAGGTCCGTGCCGAC<br>GTCTTCAGGGGAGAACCGGCCGACATCCAGCGCCGGATTTCCCGTCTG<br>GCGGATTTCCTCTGGTCGTTCCGCGAGAAGGACCCGTCCGCCGACGAAT<br>CGGGCTACAAGGCCAAACTCGGCGGCGGCCACGACGACATGGCGGTTT<br>GGCTAACCGAAAAGATTGTCTCCCTCCGCAACTTCTTCTCCCATGTCAAT<br>CGACAGGACTGCACGCCGCTCGTCATTTCGCATGACGAATACGTCTTCG<br>TCGAGGGCATCCTCGGCGGAGCCGCGCGCGATGCGGCGATGGGGCCG<br>GGGCTCAATCCCGCCAAGGCCCAGAAACTGAAGCTCGCAACGCACCAC<br>GTCAAGGAGGCCCACACCTACGAGTTCACACGCAAGGGGCTGGTGTTC<br>CTGACGTGCCTCGGCCTCTACAAGGACGAAGCCGAGGAGTTCTGCCATC<br>TCTTCCACGAAATGAAGGTCCCGGACCGCATCGAGGACGCGGACCTCG<br>ATGAAGAACTGCCGGACGGCCGCCACCGCGCCGTCCTCGGACTGGAGG<br>ACTTGGACAAGTTCGCCGGTCTCAAGGGCAAGGGCCGCGCTTTGATCGA<br>GCTGTTCTCGTTCTACAGCTACCGCCGCGGACGCCAGTCCCTGGACGC<br>GGCGAACCTCGACTTCATGCAGTTCGCCGACGTCGTCGGCTGCCTCAAC<br>AAGGTTCCGGCCCCCGCGTTCGAATACCTTCCGCTCGAAAAGGAGCACG<br>CGGAACTCGAAGCCCTCAAGGCCGCCTCGACCGAATCCGAGGAAAACA<br>AGCGCACCAAGTACGTCCTGCGCCGCCGCGACCGCAACCGCTTCCTTTC<br>CTTCGCGGCCGCCTACTGCGAGGACTTCGGCGTCCTGCCGTCCGTCCG<br>CTTCAAGCGGCTCGACGTACGCCGGACGACCGGCCGCCACCAGTACGT<br>CTTCGGCCCCGGCGCCGACGCGGAAACGGACGAGGAGGAAAGCAACC<br>GCGTCCGCCTCAACCGCCACTACGCCATCCGCCGTGACGCGATCCCGT<br>TCGAGTTCGAGCCGGACCGCCACTACGGTCCGGTTCGCATCGGCGCGC<br>TCCGCAGCGCCGTCAGCGCCACGGAGATGCGCAGACTTCTCTACCTCCA<br>CGCCCAGGGCGCGGACATCGACGCCGTGCTCCGCGGCTACTTCGAGGC<br>CTACCACCGCGTCCTCGAACGCATGGTCAACGCCGGCTCCCTCGCCGA<br>GATCTCCCTCGACGACGAAACCTTCCTCGCCGACTTCGCGACGATCTGC<br>GGCACGTCGCCGGAAGCGTTCAAGGCGGATCCGGAGGCCTTCCGCAAG<br>TTCGTTCCGGAGAGTCTCCGGCGCTACTTCTCCAAGGACGCCGGACCCC<br>GTTCGGAACAGCGTCTCCACGCGCTTCTCTGTTCGAAGCTCCTGGCCGC<br>GGTCAACCGCACGTGCGACATCCTCGTCCGTCACGACGCCCTCGAGGC<br>GTGGCGCGAGGCGTCCCGTCCCTGGCTCGACGCACGGGACGACTGGC<br>GCGGCCGGATCGACGTCTGGCGCAAGGCCAATCCGAAGAAAGATCTTC<br>CCGACGAACTCAAGACCTTCGAGGCCTATCTTGAAACGCTCCCGGCCGA<br>CGAGCGTCCCGCCCGTCCGGAGGAACCCCGTTGCCGCGTCGGCGAAG<br>GCGAAGGCGAAATCACGAATCCGCCGACCTGGTGCCGGTTCTCCGACG<br>CCGACTACGTCTCGTTCGTCTTCGACTGGTTCAACCTCTTCCTTCCGAAC<br>AACCGCAAGTTCCGCCAGCTCCCCATCTCCGAGCAGCACAGGGAAGGC<br>GTCGAAGACCATCTTTTCCAGATGGTCCACGCCGCCATCGGCAAATTCT<br>CTCTCGACCAGAAGGGGCTCTGGTCCCTGCTCGAAAAGAACCGGGCGG<br>AGCTCAAGCCGTATGCCGACCTCCTGCAATACCGGACGGACATACGTGT<br>CCTTGTGGCTCCGTTCGTTCGGGAACTCAATCATCAACTTCCTGAGAATC<br>GGCGACTTCCGGAATTTGACAATGGTGTTGACAAAGACGAAAAAGAAATT<br>GCAAGCCCGATTCTTAACATCTTCCGTCATTTCAGGCAGAACGAGCTCTG<br>GCAATTTTTGCAGAATTTCCGGCCGGAGTTGAAGTCTTGGTGTCGGAGC<br>ATGAAGAGCAAGCTCCAAGACGCCATGGAAAAGAAACCGTCTCGAAAAG<br>ATGGCAAACCGCACTTCCCGAAAGTTGAAGACTTGTTTGAGGTGTTGAAT<br>CTTCGTCGCCCCTCCCCGACCCTGGAAGACCTCGCCGTGGAGGCTGTC<br>AAGCTCTCGAACGACATGTTCCTCGAAGAGACGAACCGCTGGGGCGCTG<br>CCAATCCAATGTCCGTTTCGCGCGACGATTTGCTCGCCGCCTGCCGCCG<br>CTTCGGCGTCCGTCCCGGAATGCCCGTGGAGTACAAGTCCCTCCTCAAG<br>ACCGTCCTCGGCATCGACTACGATGCCTGGCGCCACGCCTTCGACTACG<br>GGACCGGCCGTCCGTTCGCGGACCGCCGGCTGGAGGACGAGGAGCAC<br>GTCGCCGCCCAGATTCCGCTTCCGAACGGCTTCGCCGACCGTGTCGCG<br>GCGTCTCTCCCGCGCAAGGTCCGCGAGCGGTTCCAGGCTCCCGGCTCC<br>GCCGCGTTCGACTTCAACGCCGCCTTCCGCGCCTGGTCCCCGGATCCG<br>GCCGTTTCGCTCCGCGACTTCTACGACGTCAAGCCGCTCGTCGCGAGCC<br>AGATCGCCCGAAAGCACCCGCCCGAGGGCGGGGCGGCGACCGCCGAC<br>CCCTTCGCCGCCCTTTCCGCAAAGACGCTCGATTCCGCCGTGCGCGAAA |

| | | |
|---|---|---|
| | | TCAAGGACGCCGAAAACCAGGACAAGGTCCTCCTGTTCGTGGCGATGAA<br>ATACTGGGAGCGCTTTCGCGGCGGCGACACCTACTCGACCGGCAAGCT<br>CAAGATGCCGTTCTCGGAGAAGACCACGCTGCGCGAGTTCTTCGACACG<br>CCCGTTCAGATCGAAAAAGATGGCCTGACCGTTTCCTTCCGTCCCAACG<br>ACGTCAACCGTCCGGCGTTCGCGACGCTGTTCGGCAATTCCCGCGCGG<br>CGAAGGACTACCGCGCCAAGATCGCGAAGATCCTTTCTCCGGACGGGTC<br>CCGCACGGCCTTCGACTTCTACGAGATGGTCGTCGCCTTCCGCGAGCAG<br>AAGGCTCGCGACCGCCACGAGCGCCTGGCCTTCACGCCCTACGCCGTC<br>CATTTCGACGCCCTCTGCGAAATCCCCGCCTCCGCCTACGAGAAGGCGG<br>CGGAAGGTCGTTTCGGGGAGGCGAAGACCGAGGCGATCCGCGCCATGG<br>AGTTCGAGCGCTACAAGGCGGTCCTCCCCGGCTTGACCCGCGAGGACT<br>ACGATGCCGTCGCCGACGCCCGCAACGCCGTTTTTCACACAGGTTTCAA<br>GTTCGACTGCTCCAAGGCGATCGCCGTCTGCAAGCGCCTCGGCGTCCT<br>CGGAATGATGCCGGGCGAAGCCGCTCCGGCTGGCCGCAGAACGTCGTC<br>CCATTCTGGGCCGAAATGGTCCGGCCCGCGCCGGAACGGCGGTTCGAG<br>GTGGTGA |
| 13 | Cd13<br><br>ContigID:<br>k127_2411982 | GTGAGCAAGAATGATAACATCAAATCCAAAGCAAAAGCACTTGGCTTGAA<br>GTCAACGTTTCAGGTTGGCGACGAGGTCGTTATGACTTCTTTCGGTAAAG<br>GTAATAAAGCCATCGTTGAAAAGATTGTACGTGGAACCGACGTGCAGTCT<br>GTTCCGACTGAGCCGAACTTCAGCGCTGAAATCGAAGGTAAAAAATTTGA<br>CCTTGTCGGCAGAGCTCATATTCAGACAAAGTCGGACAATCCTCAGTATT<br>CAAAGAAAAGAACCGGCGACGATATGATAGGCGCAAAGGCGGCACTTGA<br>AAAGAGATTCTTCGGCGGTACGTTTGACGACAATATACATATTCAGCTCG<br>CTTACAATGTTCTTGATATTGAAAAGATTCTTTCTGTCCATATCAACAATAT<br>TGTTTACACGCTTAATAACATACGAAGAAAAGACAACGCTGAAGATGACG<br>ATTTTATCGGCTATATGAGTACCAGAAACGACTACGACACCTTTATTGAAC<br>CGAGAAAGCACAACATTAGCGAAGACGCTGCTAAAAGCATTGACAAATCA<br>AGAGCCTCGTTTGAAGAGTATCTTCAGCCGCCGGTAAAGGATTGTCTGC<br>ATTATTTCGGAAACACCTTTTTTGCTCCCCGTGAGATTGAAAAGACATACA<br>CGGATAATCGCGGCTTTGAAAGGAAGAAAAAAGTTACCGTAAACTCGCTT<br>ATAGACGAGAAGGAAATATACTATATTTTTGCCCTCCTCGGCGGACTTCG<br>TCAATTCTGTACCCATGACAACAAAAGCGGCAGAAACTGGCTTTATTCTC<br>TCGAGGAAAACGGAATAAACTCTGATGCAAAGGCGGTACTTGACAAGTA<br>CTATAATTCTGCAGTTGCAAGGATTGACGAAAGCTTTGTTGACAACGCAT<br>CTAAAACTAATTTCAAGCTGATTTTTAATGCAATGGATGTCAGCGACGAA<br>GCGCTTCAGGACAATATAGCAAAGGCGTTTTACAAGTTTACGGTTTGTAA<br>GAGCTTTAAAAACATGGGCTTTTCAATTAAAAAGCTCCGTGAACAGCTGC<br>TCGAACTTCCTGAATATGAAGGGCTTAAAGATAAGCATTACGATTCCGTG<br>CGCTCAAAGCTTTATCAGATTATGGACTTCGTTATTTATCTGAGTTTCAAA<br>GATGAAAAGTTTAAGAAGGATAATGAAGAAAAGATTAATAATATCGTTAAC<br>GAACTCCGTGCGACCTTGAACGAAGAGGATAAAGGCAGAGTATATGCCT<br>CTTATGCCGAAAGGATGAAATCTGAGCTTAAACCGGCAATTGCACGCCTT<br>AAGTCTGATATTGACAAAATTAAGGACAGCAGGGTGAAAGAGTTTGAACT<br>TGATGCGTCCGTAAAATATAGACTCTCAAAGGTCGTAGAAAGTGTGCGCC<br>TTAAGGATAGGGCAACATATTTTACAAAGCTCATATATCTGACTACTCTGT<br>TCCTTGACGGCAAGGAGATTAACGACCTCCTGACTACTCTTATTCATCAG<br>TTTGAGAATATCGCAAGTTTTATTGATGTTATGAACGACAGAGGTATTGAT<br>TGCCGTTTTTCGGATGGATATAAGCTTTTTGAATCCTCAAAGCAAATTGC<br>GTTTGAGCTTCGCAATGTCAATAGCTTTGCTCGTATGACAAGGACCTCAA<br>AGGATGACGAAAACGCAACGCATATGATGTATATAGATGCTGCAGAGATT<br>CTCGGCACCGATTATACCGAAGAGCAGATTGAGGAGCATCTTAACCTGG<br>AAAAGAAAAGAATGATACCAGGCACGAAAAAAGCGGATATGAACTTCCGT<br>AACTTTATAATCAATAACGTTATAAAGTCATCACGCTTCAATTATCTTGTCA<br>GATATTCCAATCCTAAAAAAATCAGAGCGCTCGCCGATAACGAGGGTGTA<br>ATCCGTTTTGTCCTGGGCGAGTTGCCGGATGCGCAGATTGACCGTTATA<br>CTTTACTTTGCGGCTTCAATCCCGATGCCGACCGGCAGGAAAAAACGGA<br>CAAACTTGCAAAGGCAATTACTGGTTTGAGGTTTAACGATTTTGAAAATGT<br>TAAACAGGGCGCAAATACTGAGGGCGAATCTCAGGAGTCAATTGACAAG<br>GCACAGAAGCAGGGACTGATTTCTTTGTACTTAACCGTTCTATATCTTTTG<br>ACAAAGAATCTTGTTTATGTAAACTCACGCTATTTTCTTGCATTCCATTGT<br>CTTGAGCGTGATGCGCAGCTGCTTGGAAGCGGTGCTGGGCATCATGAG<br>CCTTATGTTGCTCTTACTCAGCGCTTTATTAATGAAGATAAGCTCAATGAG<br>CATGCCTGCGAATATCTTAAGACCAATATTGCAAATTCGGATGAATATACA<br>ATTAGAATATTCCGCAATAATGCTGCTCACTTGAGTGCTGTTAGAAATGC<br>AAATCTGTATATTGACAAGTTGAAGGAATTTAAATCGTACTATGAGATTTA<br>TCATTTCCTTTCTCAAGAGAATATTTACGGTAAATATTGCGTTGATAAGAA |

| | | |
|---|---|---|
| | | ATATGTTACAGCCGATGAAAATGGAAGTAAAACAATAAGTGTTAAAATTAG<br>CAAGGATTATTGTCCTCAGGTATATATCGACAAGTCGCTTGAATACTTTGA<br>CAAGCTCAATAAATATGGTACGTATTGCAAGGACTTCACGAAGGCACTCA<br>ATTCGCCATTCGGCTACAACCTCGCAAGATATAAGAATCTGTCTATCGAG<br>GGCTTGTTTGACCGAAACCGTCCCGGAGACAAAGGTGAGAACACGTTTG<br>AGGACTAA |
| 14 | Cd14<br>ContigID:<br>k141_15335538 | ATGGCGAAGAAACTTAGTCCGAAAGAAATAAGGGAAGCTGCAAAAGCAG<br>AAAAGATGAAAAGCATAAAAGCTGCAGAGGCTGAAAGAGAAAAAGCTGC<br>TGAAGAAGCGAAACTTAAAGCTGAAGCAGAAAAGAAAGAAAAAGCAGAA<br>AAAAATGAACGCGAAAAAGCACTGAAGCGCTTCAGACTTGATGAAAAATC<br>ACGAATGGCTCTTCCTAAAAGTGAAAGAAAGTCACTTGCCAAAGCCGCTG<br>GGGTCAAATCAGCTTTTGCTGTTGGAAATGACATTTATCTTACTTCATTCG<br>ATCGCGGTAACGATGCAATCGTAGAAAAGAAAATCACTGATACAGTGGTA<br>ACAAATCTGAGAAGTGATGAATCATTTGAGGTAAATGAAAATACTATTACA<br>GAAATGTCTGTTCCGATAAAAAGCAAACGTATATCTGATCTGTACGCTATA<br>GCTGACAACCCGCTTTACAGAAAAGATTCAGCGACTAAAGTTCAACCGGA<br>CAAGCTTCTTCTCAAGGATACTTTAGAAAAACTATACTTTGGAAAAACTTT<br>TGATGATACACTTCACATCCAGATCATATACAACATTCTGGACATTGAAAA<br>AATACTTACCGTTTACAGCATCAATACGATCTACTGCTTAAACAATCTGTT<br>TGGAAAAGAAAGCGGTGAAAAAGAAGACCTTATTAGTAAACTGACCTATC<br>AGATAACATATGATGAATTTAAAGAAAGTAAAGCTCATAATGAATTCATCG<br>ATTTTTATAATCTGAACACCTTAGGGTATTACGGCAATATTTTTTTCAAAGA<br>AAAAAAGAAACGTTCACAAAAAGAAATCTATGATATCATAGCTCTTATAGC<br>TACCATAAGACAATGGTGTGTTCACTGTGAAGAAGACAAGCGCACATGG<br>CTGTTCAACACAGAAACCGTTTTAAGTAAGGAATTTCTTGATATTCTTGAT<br>GACGTTTACGAAAGTCTTGTTGAAAAGGTAAACAGAAATTTTCTGAAAGA<br>CAACAAGGTTAATCTGCAAATACTCGAAGACGTCCTGGAAATAAAAGACA<br>GTGAATCACGTGAAAAACTTATCCGACAGTACTATCGTTTCATTGTTACCA<br>AAGAACAGAAACTTCTCGGGTTCTCGATTAAAAAGCTTCGTGAAGCAATG<br>CTTGAAGAAACCGAATTCAAAACAGACAAAAAATATGATACTGTTCGTTCA<br>AAGCTTTACAAGCTTATCGATTTTCTTCTGTTTACAGGATACACTACTGAT<br>GAAGCTGAAAAAGAAAAAGCACTTTTTCTTATTAAATCTCTCCGGGAATCA<br>TTGACCGAGGAAACTAAAGACAGAATATACAAATCAGAAGCTGCTCGTTT<br>ATGGCTGAAATACGAAAATACTATAACCAACAAAATCAGAGAAGCTCTCA<br>ATGAAAAATCTATAAGCGAATTAAAAAAAGATAAAAGCTTCGATGATAAAA<br>GTATTACCAGCATCATAACAGATGAAGTATCAGGGAAAAAAGCAACTTAT<br>TTTTCAAAAACAATCTATCTTCTTGCTCAGTTCATCGACGGAAAAGAAGTA<br>AACGATCTGACCACTTCACTTATAAACAAATTTGATAATATACGAAGTCTC<br>ATTGATACTGCCGGACAGATAGGACTTGACTGTAAATTTACAGAAGAATA<br>CAAATTCTTTGAAAACTCCGATCAGATAAGAACAGAACTTCATGTTATAAA<br>AAATCTTACTAACATGGAGTATTACGATACTACTGTAAAAAAACAGATGTA<br>CAAAGACGCCGTTCACATTCTCGGAATACAGGATGATGTTTCTGACGCAG<br>AACTTGAAAAGATCATCAATTCCATTCTTCTTCTTAATGAAAACGGCAAAC<br>CACTCCCTGGAACCAAAGGTAAAAAAGGATTCCGTAATTTTATCATTTCCA<br>ATGTTTTAAAATCACGTAGATTCATTTATCTCATAAAATACTGCAATCCTAA<br>GAAGATAAGAAAAATTGCAGGAAACAGAAAAATCATAAAGTTCGTTTTAA<br>GCCGTATAACAGACTCACAACTGGAAAGATACTACTATTCATGCAATCCG<br>GAACTTAAATCCGGTATCTATCCGGGACGTGATGATGCTGTAAATGATCT<br>TTCTATTCTTATCGCTGACATGAAATTCGAAGATTTTAAAAATGTTGATCA<br>GAGCGCAAATGTTCATGACAACAACAATGCTGCCAGAGAAAAAATGAAGT<br>ATCAGACTATAATCAGTCTTTATCTTACAGTTTGCTATCATCTGGTTAAAA<br>ATCTTGTAAACATAAACGCCAGGTATGCTATGGCCTTCCACGCACTTGAA<br>CGCGACGCTCGCCTTTACCAAATTTTCAGCAGTGAAGAGAATTATGTGGA<br>TAATTTAAATGCTGACTATGCTATACTTACTAAAACACTTCTGAAAGACAA<br>CTATGAAAATGCCGGAAACCTTTATCTCAGAAACAAAAAATGGAATAAACT<br>CACAAGAGAGAATCTTGATAATTACATTCCTCAGGCCGCTGCAAACTTCA<br>GAAACGCAGTAGCTCATCTTAACCCGATAAGAAATGCTGATATGCTGCTT<br>GAAGATATTGAAGATGTATCATCATATTATGCCATCTATCATTACATCATG<br>CAGAAAAGTGTTACAAACAGAACTATAAGGGTTTCTAATACTACTGAAGA<br>CGAAAAGCGAATACTTACAGATTATCAGAACAAAATCAAAAGACATCATG<br>GTTATAACAAAGATTTTGTAAAAGCTCTCTGCGTGCCTTTTGCCTATAATA<br>TAGTCAGGTTCAAATCTCTTTCTATATATGAAATGTTTGATCGTAATTATCA<br>TGAAAAAACGTCACCGGAAACTGATGACAGCCAAAGCCCCTGA |
| 15 | Cd15 | ATGGATAAGGAAAAAACAACTGTAGAAGGAAAAAATACTAATCAAAAATC<br>TGATGTGTTAAAATCGCTGGCAAAAGCAAATGGACTTAAATCTTCTTTTGT<br>TATAGGGAATGAAGTTGTTATGACTTCATTTGGAAGAGGAAACAGTGCGA |

| | | |
|---|---|---|
| | ContigID: Cas13/23_contig-81_4932 | TTCTTGAGAAAAAGATAACGGGCTCGAGAATAGAAAATCTTAATCCGAAC<br>GTTGCGTTTTATGTTAAAAAACATATATCAGATAGTTGTAATCCTGATAGT<br>GGTAAATATGACGTAAAGAGTAAACGAATGAAGGAAAAAGCTGTTGTTGA<br>TGATCCAGTATATGTTTCTCCGGAAAAAGCTTCAAATGTTCATGCTGGTC<br>AGGATCTCATTGGCTGTAAGAATGTTCTGGAAGAACGGTATTTCGGAAAG<br>ACTTTTGATGATAATATCCACATACAGCTGATCTATAATATTCTTGACATA<br>GAAAAGATACTTGCGGTTCATGTTTCAAATGCAACTTTTGCCATTAACAAT<br>ATTTTAGGAATTGAAGGCAAGGAAAACGAAGATTTTATCGGAAATCTATC<br>AGTGCTCAATACTTTTGATGAGTTCGAGAATTATGAAACGCATCCAAAGTT<br>TGCAAACAAAAGTGCGATAAAAGAAAATCTCAGGAAATCCAAAGTTTTTTT<br>TGATAAGATTAAAAAAGGAAATAAACTCGGATATTTCGGTCAGGCTTTTTA<br>TTATGCAACAGGTACTGGTAAAAATCTGATATTCACCAAGAAATCTTCAGA<br>GACGATATATGAATTACTTGCTCTTGTAGGAAGTCTCAGACAATTTTGCGT<br>TCATGACGAGGTAATGGTTGATAATAAAGTAAAATCAAGAAGCTGGCTGT<br>ATAATGCTCAAAAAGAGTTAAAGCCAGATTTTCTAAAAGCACTTGATGAAC<br>TGTATAGTAAAGAAGTTGAAAAAATAGACAGTGATTTTATTGTAAATAATA<br>CAGTTGATTTGCACATTATTCATGATGCAATAGATGTTATTGACGGATCAG<br>CCGACTGGCAGAAAATCACAAATGAATATTATGATTTCATAATAAGAAAAT<br>GTTTTAAGAATATCGGTTTTTCTATAAAAAGACTTCGCGAAACAATGATTG<br>AAGAACAGATGAAAGTATTATGCGGAAAATGCGATAGAGAGAACTGTAAA<br>GGTTATGGCAAATGCTTCAAAAACAAGAAGTATGATAGCGTTCGTTCCAG<br>ATTAAACAGAATCGTTGATTTTATAATATTCAGGCATTATAATGATGAAGC<br>AATCTTAAAGAATGTAAGTCTTCTCAGAACCTGTATGAGTGAAGAAGAAA<br>AACAGAAACGATTTTATTTGCCGGAAGCTAAAGCTTTATGGAAAAAGTAC<br>AATTATGTGTTCAGAAATTATGTTCTTAAAAAACTTAATGGCAAGTCGATA<br>AGCGGGTTAAAAGAGAAAGCAATCGAAAATTCCATAGATATTAACTCAGT<br>AAAGATAAGCCTCGGTGATCCGGATTATTTCTGTAAGTTTATTTATCTTCT<br>TACGTTCTTCCTTGATGGTAAGGAGATAAACGATTTACTTACAACGCTTAT<br>CAATAAGTTTGATAACATTGCAAGTTTTATAAGTGTTATGAAAAATGATAA<br>GTTATCTATTGACTGTGAATTTGTACCGGAATACAGCTTTTTCGCAAACAG<br>TGCTCAAATCACATCTGATCTCAGAGTTATAAACAGTTTTGCAAGAATGCA<br>GGCACCGGCTGAGCCGTCGAAAGATGATATGTACCGTGATGCTCTTGAT<br>ATCCTTGGCATGGATGACCTGTCAGAAGATGGAAAGAAGCAGCTTGAAG<br>ATACAGTTTTATGCAGAGATGAAAACGGTAAGTACATGAAAAAAGAAGAT<br>AATAATCCTAAACGTGATACCAACTTTAGAAATTTTCTTGGAAATAACGTT<br>CTTGCAAGTACACGTTTCAAGTATCTTATTCGCTATAATAATGCCAAGAAG<br>ACACGTGCACTTGCAAATAATAAAGCAGTTATCATCTTTATGTTGAACAAG<br>ATCAATAAACAGAATCCTGAACAGATAGTTAGTTATTATAAGGCATGCAGA<br>GATGACAGTGACCCAGTTGCTTCGGATGCAGAAGCTAAAATTGAGTTTCT<br>TGCCGAAAAGATAATGAATGTCAGCTGTACTCAGTTCAGATACGTTAAAA<br>ACGGAACTAAAGTAAGACCGGATGAGGCTAAGGAAAAAGAAAGATTCAA<br>AGCGATAATCGGTCTTTATCTTACAGTCATGTACCTGATAACCAAGAATAT<br>GGTTTATATCAATTCAAGATATGTGACCGCATTTCATTGCCTTGAAAGGG<br>ACAGCGAACTTCATGGTGTTAAGTTTGATCAGAAGAAACTGCAACCGAAT<br>CTTACAAAAAAGTTTATTGATCCGAAAACTTGTGGTGACTATGGCCTCAG<br>AAATAATAAACGCGCAAGAACTTATATCGAGCAGAATATGGATAAGATGT<br>CCAACTGTACTAGTTACTGGAATGAATACAGAAATGCCGTAGCACATCTT<br>TCGGTTATCAGGAATATGAATCAGTATATTAAAAATGTTAAGAATATAGGC<br>AGTTGTTTTGAGCTGTATCATTATATTATGCAGCGTTTCCTGTTAGACAAA<br>GAGAATATTGCTGAATCACTTAGAGAATATGATGATTTCATAAAGAAGAAG<br>GGATGCTATAGAAAAGACTTCGTTAAGGCTTTGAATACTCCGTTCGGATA<br>CAACCTTGCAAGATATAAGAATCTTTCAATTGCAGAACTGTTTGACAGAAA<br>TGATACTGAGCTGGAGAGAACGAATAAGCTTAGAAATGAAGCAATCAAAG<br>CAGATATAGAAGAAATCTGA |
| 16 | Ca1<br><br>ContigID: k127_1867445 | MKISKVDHTKSAVSVQTAQGQQGILYKDPSTEEMSVEDRVTKRADATKALY<br>AVFNQPKDKRSISGEATTVASSFNYVIKDLKKSKSLNGKLSVESLYEAVGNEL<br>KGKHASAEEIDLAITLLLKKSLRRDSFIEALKLVLGKAYKGEKLNEEDKRIIKDD<br>LIVPLIKDYDKSSIREQAVASIKHQNLIAQPDSKSDDAVMVISNIAGASERSTN<br>EKEALRQFISEYAVLDDSVRHDMRVKLRRLVILYFYGMDVVPTGDFDEWED<br>HVQRGKTADLFIDFAPVGGKTDADRLKDAIRKMNIERYRYSVDAIDQDNTEL<br>FFEDMMINKFFIHHIENEVERIYRNTKPGDEFKRSLGYISERVWKGIINYLCIKY<br>IAIGKAVYNCAMAGLGSDQPDIKLGVIDRVYADGISSFDYEIIKAQETLQRETS<br>VYVSFAINHLGAATVNLTEKETDFLTLDNKQIKELAKTGVLRNILQFFGGKSV<br>WKNFEFAPEGGTGNEEIVLLYYLKDILYAMRNENFHFSTASINDGSWDTDLI<br>GRMFAYDCTRAGVGQKNKFYSNNLPMFYKSEDLERALHILYDHYSERASQV<br>PAFNTVFVRKNFSEILKGQNLPMPTSAEESLKYQNAIYYLYKEIYYNVFLSSS<br>ESRDYFIKAVKSLRWENSNEENAVKDFQNRINELTGKYSLSQICQLIMTEYN |

| | | |
|---|---|---|
| | | QQNSGSRKKKTAKDEQNKPDIFKHYKMLLYKSIREAMLKYVDDKSEDFGFIK<br>SPVFGKDDNCIALEEFLPDYESTQNAKLIERVKSDFRLQKWYILGRLLNPKQV<br>NQLAGSIRSYIQYSDDVKRRAKENGNKIHVSTESYPYQTVLRVIDLCAKLSGL<br>TTNNIDDYFDGSGDYLSYLARFVEYDPNDIPKIYHDEANPILNRNIIMAKLYGA<br>GDVITNAVEHVNTSMIRDLESYEKKTLGYRSSGVCKDKDEQETLKKYQELKN<br>RVELREIVECSEIINELQGQLINWCYLRERDLMYFQLGFHYTCLKNSSDKPE<br>MYVKAKTVDGTIDGFILHQIAALYTNGLKLYSCGKAVRDDNRKIIHYDLSSGK<br>ELKGNDKSAAGKKITDFMGYTSLALNRTENDILPISGDFYYAGLELFENVNEH<br>ENIISLRNYIDHFHYYAKHDRSMIDIYSEVFDRFFSYDMKYRKNVPNMMYNIL<br>LSHFVKAQFVFGSGMKESGEKTKSQARFDLKDKAGLEPEQLTYKVANSEKP<br>VQLSAKDKQFLKTVALLLYYPEKKTFPEGMYADTRFVEGTSSNKRNNNSSG<br>NRHGNGNHNVGGHNKGYNQGRKNGNWSKDKSGDRNAGKKQTNKNRKDS<br>TSVYKDEGFSNRINIPSEYYSQKPGKK |
| 17 | Ca2<br><br>ContigID:<br>k127_4200118 | MKLSKTGKNGWHHRNGVKVNNSKQEGFVYSIPHNDGESTDKFVEDRKKDF<br>KRLYKVFPSVEKARNISEEIAAVIDKTIRNKRTEIWTGKNDYSEMACRFRNLL<br>QRESMFRQPVEVKTAEYMVYGLLRSSLRSEKTEKDLIDFLCHVNDKSSSAG<br>AIFMQELTRDYMGEKINYKSIINQNLVIQPVKTESLKDNIDQEDVLLTVSERKN<br>PEDSAKNYKSIENKALRSFLLEYASLDDNKRKDLRKKLRRIVVLYFYGKSEAD<br>GLGENFDEWNDHESRRACEEKFIEFDESTKDNFRSKTSKLIRNANITAYRTS<br>KEIIENNHDGLYFANPDYNFLWLRHLSREVERLTSNINADKTYKLNKGYLSEK<br>SWKGIINYLSIKYIAIGKSVFNFASSGVDSDGSDIQIGEVNREFQNGISSFEYE<br>RIKAEETLQRESAVKVAFAARHLASATMNLTPEDSDMLLFDKDKMSQNLKDT<br>GRVLADVLQFFGGQSVWKDYIIKETEKYSSEEEFGTDLLYNLKKCVYALRND<br>SFHFKTLNNKADWDTDLVAGLFEKDCENMVGLDKDKFYDNNLYKFFKQEDL<br>KKVLDKLYDKIHDRASQIPAFNTVFVRNNFSRFLLSKGITRSFASEEQGRQFA<br>SAVYYLFKEIYYNDFIQSGNAKTLFLNYVNSIKIEKAINRYGKEETKRECKPAE<br>DFKTYITLCRNMSFSEICQAVMTEYNQQNNQSRKKKSAFDTAKNKDKFRHY<br>VDILHEGIREAFAAYIGLNDQKNYDGIYGFVKSFNSSDVFTVEKDKFIEGYRS<br>ERFQNLISKVRQNPELQKWYIVARLLNPKQVNELSGSIRSYKQYIEDVCRRSI<br>AEKCPVRKNDGKEVSKASFDNDVKELMSIDYMGVVAILEICIRLNGRFSTVC<br>DDYFKDGADGYAEYLEQYLDYQDEKTKDAGVSPSTMLSMFSEEVSADNTD<br>KNQGIIYHDGTNPIMNRNILLSKLYGGANSVIHSVKKVDNRLIADFIKSGKLIQE<br>YNKRGYCINEEEQKNLKRYQALKNRVEFRDIVEYGEILDELQGQLINWSYLR<br>ERDLMYFQLGFHYTSLHNSERKFEGYRYITKEDGSVIENAVLHQILSLYINGIP<br>FYYSYADVEGRDRFICCALKKKEPVDGTSNKFEDTGTKMRYVGYYCKEGDN<br>YLGEGIYLAGLELFENIAEHDNIIKLRNYIDHFHYYIEDDRSMLDVYSEVFDRY<br>FSYDIKYQKNVVNMLYNVLLSHFAKAGFEFGEGIKQIGSSKKNEMPLTKKMA<br>RILLKSLESDDFTYKIGNTSEAKNQETVVLPARDDLFLDALGKVLNWDGSITD<br>ENALQKTEIITGRSGNYLKKSRNSDKKNDGNKRSDIKKSFNKDKNIPEKKEAL<br>TSTPFANLFNNLHMDFD |
| 18 | Ca3<br><br>ContigID:<br>k127_751200 | MKISKVNHTKSAVSVSEGSPKGILYEDPTKSGTKDLETRILERNEAAKLLYNPI<br>NTSRSRKKTHKIINRSLRAFFNRVKKKTGGSFSWDELKRVSYDSSLDAERDK<br>ITDSDIDSVVEACLKKSLSTPECIEAVKQITRVLCGKNTSHDLDDKLIGKLSSK<br>LHDDYSKERLLGNIKKSIENQNMVVQPGQVDGESIFKLTGDDSLKENPEKVS<br>FERFLISYANLDKKFRDCELRKLRRLIVLYFYGETEVDTTDDFDVWADHKKQ<br>RNFKWFISDIEFIATYEKYLKELQFEDRRNHHTKISEPEFREKIRQENINRYRN<br>SIAVINKSNEVYFDDPVLNKFWIHHIENSVEKLLKRVNPADSFKLNVAYIGEKV<br>WKEVINYLSIKYIAVGKAVYRFAVDDMTYGVIPDLYKSGISSFDYELIKADESL<br>QRDIAVSVAFAANNMARATVVLDEKSSDFLVESFDLEKSIRTDVALDMAILQF<br>FGGKSSWKQCDALKDCKYIDLLYDMKKMLYSIRNMSFHFISSEEGDNGYKT<br>NGIIPAMFNQEITTYTTILKSKFYSNNLPAFYNDSDLEGEFKLLYKNYVERASQ<br>VPSFNSVVVRKSLPDFVKRDLKIKTALSGDDLTKWHSALYYLLKEIYYNLFLA<br>SDDAKILFLKAVENNKNSNNSSVSDKNDHRREAGIDFAERIESIKDHSLSEIC<br>QIIMTEYNQQNQSRKVKTAQDEKNKKSLFIHYKMLLNLCLRNAFKMFLDRNE<br>FSFLKSIHNREVKSSSDEWTAAFCADWTSNAYSMIQDEINKNPSLQSWYILS<br>RFITTKQLNHLSGDIRHFIQYVEDVKRRAKETGNACKYDLDNKVCIYRKVLQV<br>LDFCNKTSGIVSSEIGDYFKDDDEYAKFVSNYLDFGGTTKLELIAFTNQTVGD<br>DQINIYCNDSKPILNRNIVMAKLFAPTDTISKAIAANGNRVTVDDIEEFYSIKPIA<br>QKFLSDGDSVAKKEKKQLIEELKKTKRYQEIVNRIEFRNIVDYAEMINDLLGQL<br>VSWSYLRERDLLYFQLGFHYLCLINDSYKPDKYRVLRDGERIINNAALYQIVS<br>LYSFNVDTFRDDNDKDKGKKYNNICEYSLNIGLDEEWQFYTAGLELFETITE<br>HDSIKKFRDYIDHFHYYTNQDRSILDMYSEVFDRFFSYDMKFRKNTVVILQNI<br>LKSYLVIMPVKFNSKYKSTDNGSSKMRANVDMGEKGLKSEVFTYKYSDSCK<br>VILPARSINYLKDVASILYYPHKTPRDAVDMEDFKKNYEVAQTLNKAKDNHHK<br>SKNDYKNNDRPKDNYSPFKSQFDKLKKKGITFEDN |

| | | |
|---|---|---|
| 19 | Ca4<br>ContigID:<br>k127_5935133 | MKISKVDHTRTAVGVNENGPLGIVYSDPSQNAVQNPEIRVRTRIKKANMLYT<br>VFGPTNDEMDSQRENGIAKEFNKIIKRYNNKIDPKRGEKETDKKIYKMNSDEL<br>IKDIKSVFGNYSLNESTRKEIDEALNVLIKRSLRKKETIESLNLLFEKTIKGEEFK<br>AEEKDKIQKYVVDRIVADYSKNTLSKNTIKSIKNQNLVVQPQNKNGEFVFTQA<br>KNRMNGKVNQGSIRISKAQEKDALNDFLDGFAVLDKQMRDKQLMKIRRLVD<br>LYFYGIDEVVKEDFSVWERHEKTKGNDKKIIPFSRTDISTLQIKRGDSEDEKK<br>RKNREKKIIKKSDSAKLDDMIRRWNIDRFRESFSAIDKSDNSLFFDDKNISKFF<br>IHHIENEVERLFNSERLDDYKMHIGYVSEKVWKGIINYLSIKYISIGKSVYNYA<br>MEELNNSSGDVNLGVIDSRYLTGISSFDYEKISAEETLQRETAVYVSFASSNL<br>SRAVFKDGVDCDLMSTKIIDNHDKFDESKVKKRVLQFFGGESSWDGFGKTF<br>LSEEYNEFDFLEDLKTLIYQMRNESFHFNTEKKNVDIKNPKLFSDMFAYECS<br>KACVSEKDKFYSNNLPLFYSEKPLEKVLNKLYTKYNDRKSQVPSFEKVMKR<br>SEFGKYLIKSGVATNFNKEDTDKLESGLYYLYKQIYYNDFLVNDMIAKGIFVD<br>NINNKKLRRNENNKVIKADKGLEDFKKRLNEIKNYSLSEICQIIMTEYNQQNN<br>QKKKSQKNEEIFQHYKLGLYSYLREALIIYINNNSDIYGFIKQPTIKSEGKMPNI<br>NEFLPDYSSSQYDDLIAKVSDSFELKKWYVMTRFLNPKQTNHLVGALRNYIQ<br>YVESIKRRAEETGNKIYIDCQILESVKDITKVVDMCTRICGNTSNEISDYFDDN<br>DDYAGYLERFLDFEYKESLGSKSSMLGAFCMTKINSEEIKIYHDGTNPILNRN<br>IVLSKLYGANSIISEAVPKVDQNMIKEYYIVADKIKEYRKSGDCKNIDEIKQLKE<br>YQELKNRVEFRDIVEYSEILSELQGQLVNWAYLRERDLMYFQLGFHYVCLKN<br>DSQKPEAYKMIEVPCVDGSSRMINGAILYQIVAMYTYGMNIYYRGHKKDEEY<br>NDSENRWEAFNGSIGERIPRFALYSGYMIKGDNAKYKLSYNIYTSGLELFEVL<br>EEHGNIVDFRNDIDHFNYYQKKDRSMLDYYSEAFDRFFTYDMKYRKNVPNT<br>LSNILASHFLVPSFVFGTSSKKVGNKNYIEKKCAHIRFNTKNPLKPGSFTYVIS<br>EDKRVVGPARLKGYVKNVLNILYYPEVPEMELLDSSYIFKEEKKRKLLK |
| 20 | Ca5<br>ContigID:<br>k141_14579520 | MKISKVRGTQGKGSKLTINAKAAVVINPTGQEGILYDDPSRMGESRKNDKQR<br>ESYIKDRIRASQKLYSIFNSNQKIPKNKKTESEKAIDMIIAGFSSEDGASFRLM<br>FKDFAEILDKYAEKSYENRRNHIDESPELSKLGVNISDNQINALSNLLSEESIAI<br>KIKKGTESVKDKVKVSERDIDSAISNCLKKCMCRVKTKKALKALLMKVFDIPY<br>TLDGDVNIRRDFIDYAIEDYCRIRVKNSVSESIKKNNMPVQPTSSEGVTVFQM<br>PSLQETKSTKSKEREAFNHFLSEYADLDENKRKSLRIKLRRLNDLYFYGKDA<br>TMALADNEDVDVWEDHAKHGDIKELFIKVQKPQITGDGKADKLAMSQYEDNI<br>RTKYREANITCYRKAVEEIDNDKSLFFEDNMLNMFVLHRIESGVERIYSHIKA<br>NEEYKLQTGYVSEKVWKDLINYISIKYIAIGKAVYNYAMDELVSGDKSIEMGKI<br>NDNYISGISSFDYELIKAEEMLQRETAVYVAFAARHLAHQTVDLDEKNSDFLL<br>FPDKSRKDKDGKNINDFIKEGINLRSTILQYFGGASSWSDFSFEKYMTDGRD<br>DVDLLTDLQKAIYSMRNDSFHYTSKNHNNDGWNKELIGALFEYEANRLTIIQK<br>DKFYSNNLPMFYDESNLKELLSSLYSKSVERASQVPSFNSVFVRKSFPKVCT<br>QDLSIDVKTMNEEDKLKFYNALYFMFKEIYYNLFLNDSNVLNRFIDISTKTKKN<br>GKGDEGTHYWAEKDFRQRILSIIESRKNYTLSQICQLIMTEYNQQNTGNMRH<br>KSADKNGKNPDSYQHYKMLLLSYLGEAFVEFVKEKYDFVFTPVKRDLMDKE<br>AFLPDFAKTVNPLGDLIERVKESGVLQKWYIVGRFLSPKQANQMLGSLHSYK<br>QYVWDIYRRAEETGTKINKRVSEDTISGVAIRDIDSVLDLCVKMSGTITNNLTD<br>YFKDKEEYAAYINDFLDFEYKTGDYNWALKDFCKEITDEDDKEGIYYDGENPI<br>INRNIVISKLYGEAEFVSKIFKRVNKEDIKVYKDLKKNIEPYQNMGTFETKEQQ<br>ENVKRFQELKNHIEFRDLVDYSEITNELQGQLVNWIYLRERDLMYFQLGFHY<br>LCLNNNSEKPELYKKIEFKDEKVIDNAVLYQICAMYTNGLPLYYSSTKNANIKE<br>VSAKAGTSTKVDKFYSSGIRANGESYSRDYTTYMAGLELFENTKEHINITMF<br>RNDIEHFRYLVSNTRSMLDVYSEIFDRFFTYDMKYRKNIPNILYNILLAHFVNV<br>QFDFSTGKKNIGTGENIYEKKCAKINIQNNGGIVSEKFTYKLKDEKTIDLPARG<br>RRYMETVARLLYYPETVDEEKMVKDLVIKDNKPFGKKRNNKYSNRKEGASD<br>RKKYEENKARKKDNSFMSGMDGVDWSKLNFK |
| 21 | Ca6<br>ContigID:<br>k141_10995992 | MKISKVDHTRMAVAKGNELRRDEISGILYKDPTKAGSINFDERFNKLNQSAKI<br>LYHVFNGVVTGNKHFINTVKRVNDNLDRVLFTGRNDERKSITDTDVVLRNAD<br>RINAFDRISTDERKQIIDELLEIQLRKGLRKGKTGLREILLIGAGVKGRTDRKQ<br>DIAKFLEILDEDFNKTKQAKNIKLSIENQGLVVAPVEKGEDRIFDVSGVQKGK<br>SSKKAQEKEALSAFLSDYADLDKSVRTEYLRKIRRLINLYFYVKNDDDLSSAE<br>IPAEVNLEKDFDIWRDHEQKKGEKGDFVDYPDILLADRDEKKRNSKQVKIAE<br>KQLRESIRENNIKRYRFSIKTIEKDDGTYFFADKQISAFWIHHIENAVERILGSI<br>NDKKLYRLHLGYLGEKVWKDILNFLSIKYIAVGKAVFNFAMDDLQEKDRDIEP<br>GKISEKALNGLTSFDYEQIKADEMLQREVAVNVAFAANNLARVTVDIPQDEN<br>KDKEDILLWNKQDIHKYKKKSQKGILKSTLQFFGGASTWDLKMFEKAYPDQK<br>EDYEEEYLYDIIRIIYALRNKSFHFKTYDQGDRNWNSKLIGMMIEHDAEKVVS<br>VEREKFHSNNLPMFYKDADLEKMLDLLYSDYTGRASQVPAFNTVLVRKNFP<br>EFLRKDMGYKVHFSNPEVENQWHSAVYYLYKEIYYNLFLRDKDVKNLFYTSL |

| | | |
|---|---|---|
| | | KNIGNEVSDKKQKLASDDFASRCKEIKDRDLSEICQMIMTEYNAQNSGNRKV<br>KSQRMIEKNKDIFRHYKMLLIKTLSGAFALYLKQEKFAFIGNAATIPYETTDVK<br>EFLPEWKSGMYASLVDEIKENLDLQEWYITGRFLNGRMLNQLAGSLRSYIQY<br>AEDIERRAAENRNKLFYKSDEKIETCKKAVRVLDLCIKISTRISAEFTDYFDSE<br>DDYADYLENYLSYQDDTIKELSGSSYAALDHFCNKDDLKFDIYVNAGQKPIL<br>QRNIVMAKLFGPDSILPEVMEKVTESDIREYYDYLKKVSGYRVKGKCSTVKE<br>QDDLLKFQRLKNAVEFRDVTEYAEVINELLGQLISWSYLRERDLLYFQLGFH<br>YMCLKNKSFKPAEYMDIKRKNGTTIHNAILYQIVSMYINGLDFYSCEKDNDKL<br>EVAAAGKGVGSKISLFIKYSEYLYNDPSYKYEIYNAGLEVFENNDEHDNITDL<br>RKYVDHFKYYASDDSDKKMSLLDLYSEFFDRFFTYDMKYQKNVVNVLENILL<br>RHFVIFYPKFGSGTKEVGVKNCKKEKDRAQIEISEQSLTSEDFMFKLDDKSE<br>GEPKKFPARDERYLQTIAKLLYYPKKDVDLNKFMTKEESMNKKVQFNRKKE<br>TNRRQQNNSSSGALSSSMGDLLKNIKL |
| 22 | Ca7<br><br>ContigID:<br>k141_12677984 | MKISKVNHVRTGTRIKENNGEGVLYANPSKQTNAVKDLSKHIQDVNQKAQG<br>LYSPLNPVKSLINPKMPKEKKDEINGSYKAFKSVVIGIVKENETGIPDSASVIR<br>TLYEKAKKIDLKVSDASYLSSKLIDKCLRKSLESKSEIAKEILKAIISTDKSAVNS<br>LNAEEVKAFFELVHKDYYKKEQLKAIEKSIENKDVKVQVKTGQNGENHLVLS<br>NADSAKKHYYFDFVKEFATKDKAEREEMIIRFRQLIILFYSGSESYKLSIGSDV<br>GAWTFGSSLPEVTANVDDEIASLIAEYNENIARKNDIQKSIDLKSNQMKNYKF<br>NSPEYKKLDDQVSKLKDEQGDCKHAISDAKRKIKALVENLICTKYRDAVKAE<br>GLTDSDIFWIGYIQQVAQKQFSKKDAYNNYRISTKYLYEVTFNEWISFMASKY<br>IDLGKAVYHFAMPDFSDIKSGKEVHAGKVQPAFEDGITSFDYERIKAKETLAR<br>DFSVYATYSSGIFSNAVTDSEYRLKDEKEDALFYKQEDWEQALLPNAKKKLL<br>MYFGGQTKWEDSEIEKLSDLEMTKAFQDMINVIRNSNYHYAGSVLEPGEQS<br>VNIAKMLFEKEFSQLGRIIREKYLSNNVPVYYNVEDINKMMTYLYQGESKREA<br>QIPSFGNVLKKKEMPGFVSKYIPGNLLAKFDSEGMDKFRASLYFVLKEAYYY<br>GFLNETNLKDRFIMAFKNSEKDAKNPEAIENFKARIADMDDSCSFGEICQILM<br>TDYNQQNQGEYKVKSQIKQNQDEKDNKGHKYSHFKMLLYVTLQKAFIDYIFE<br>KQDIYGYIKAPIFKSNFFDGDEPQKFVESWEANLFGDVKKTTETDSYYLAWY<br>VLSHMLPAKQVNQLQGGIKSYIQFVTDINRREKSVLGTEKDNSLVNNIDYYQ<br>NILKVLEFVMCFVGKTSNVLTDYFADEDDYALHLYSYVGFAGKKEEKTNSTL<br>SGFCSKSITKAGKVLTDRIGIYHDGTNPIVNNNVVKALMYGNENVLSEAVTRV<br>SADLINGEITKYYEVKNKLEKVFEKGECSNIEEQKELREFQNLKNRIELQDISI<br>FTEIINDYMSELVNMAYLRERDLMYYQLGYNYIRLEYGNVEDKYKELQGDNI<br>NIKSGALLYQIVALYTHELPIVYKDKDSYKYTNNGKIGRFVKSYCEEEFNDLD<br>NTYLKGLELFEDIKLHDDLHMFRNEIDHLKYFIRADKSILQMYSRIYNGFFSYD<br>LKLKKSVSYIFANILAKYFIIADTEMKSSVENGKRVAMLSVKGLESDVFTYKGK<br>KRDKEGKERDSKYTLPVRSDEFLKEVKKLLGYKSM |
| 23 | Cb8<br><br>ContigID:<br>k127_4804511 | MPNVKFTLVPVDYSKPYDEQPDCKRHVIGAYANLARHNMTLTINTIMQAIGM<br>PLFNENEIENAFNKSHRKKLEALDNIQKVKLQKRLYRHFPFFKRMKLEDEEK<br>KTVQLKSLMTVMSLFTSLMADIRNNYTHYRPYNNKEEQNRQLELKKEVGKK<br>LQYLYENSSQTFKSMEELDHSSNEVLSALRIPEDVVERFSPDDPDYKKLLNT<br>LHDSNIPKWKKSGLKLDMKTQIITKKSVRYVRNPNYQAYMMDEEKGLSDIGII<br>YFLCLFLDKQVSFSLMDEVGFNQQIKFTGEHAEQQLMYVKEIMCMNRIRMV<br>KARIDSEMSDTALALDMLSELRKCPRPLYDVFCKEARNEFKDDATVVWENT<br>HGEEAVITEEQGDIGEETDAIAANTTGKNTPRSTFVRWEDRFPQLALKYIDLT<br>GMFDHLRFQLNLGKYRFAFYQHDKAYSVDNAERLRILQKELHGFGRIQEVN<br>EMMKEKWQDVMEIKNVEDGQIYKEPDVAGQKPYVTQQNAQYDFDTKSHSI<br>GIRWEGWHNNHSDNHYGDLDRRDMFIPRLPANPASPEGDKRQTNQAEELL<br>PPQCMLSLYELPAILFYHYLLKKYQKNTGLVEKKISDFYTNMKNFLTEVSEGN<br>ILPADETTLIRELQTRGLKFSDIPVKLKKLLKGEVTDNAKRMEESALLRLHERK<br>DKKRRALESFIAKCKMIGTKENKFNKIRAVVKTGSLGQLLARDIMEWLTTDTK<br>KRMNLTGQNYVAMQTALSMMGQSFELAPEAKVTCEKMRNIFVKANILPMND<br>DDFDADFHHPFLLDVFDEEPVSIEDFYKIYLEKEIFYIDYLTEHFKKYKAKGAA<br>LYIPFLHCERLRWKNTEQNGLKELAARYLQRPLQLPNGLFTDDIFHLLEDIAT<br>KNADFAKVLEKQKKDNHQLQQNVAYMIRIYMKTVEYDQPQNFYNTMPVGDT<br>NSPYRHIYRIFKKFFGESIPKTNKTTSPAYTIEEIRAILNNKQLLKDKIDFFVKEE<br>KEKLKKQQIRDFRNYEKKQWKLLKAKNEAAPKGQHFNVKAEVQKRLNEKRE<br>EQRKALDLLVMDVKQKLEGKLRKVNDNERAIRRYKTQDILLLFMAREILKAKS<br>QDEDFTKGFCLKYVMSDSLLDKPIDFQWTVNFQNKEKKTIAKTIEQKDMKMK<br>NYGQFYKFASDHQRLSSLLSRLPADIFERAEIENEFAYYDTNRSEVFRQAYII<br>ESKAYQLKPELTDDANANEEWFTYLDKKTKKLRAKRNNFGELLKILAAGGDG<br>VLDDVEKSLLQSTRNAFGHNTYDVDMPVIFSGKLDKMNIPEVANGIKDKIIEQ<br>TEQLKKNV |
| 24 | Cb9 | MAIFVIKVEIQRQMTLAFKNETEQKPVLGAYAAMARNNAFLTVMDIMDQLHIP<br>RTVLTDNSGKEVDPESHIWRLNLFPKNYRLLPEQEARASRLLRNHFPFLDLA |

| | | |
|---|---|---|
| | ContigID: k127_1483864 | DDLKDVENDHGQAARKNVSYEDLCNSFLTMMEVLTHLRDVNLHYKIKDERIA DFYFRRAEKETGHILREVLKAAPRKIKDRYKGTALMDETSLSFFTDGNYIQKG RKYSFNSRWAFNPQRQPQPSEVKVLKNNAPVIDRNTGIPRMFERLSTFGEIL FIALFTEKRYIPDLLRDSGLDNNFMASGDNGKMSQQRIIREIISAYSIRLPERK LDIETGATQIMLDMLNELARCPSELFDVLPESERRSFEITGSDGSQVLMKRY SDRYVPLVLRYLDVTEEFKRLRFQVNTGLLRYEHHGPKEYMDGVARFRIVQ SSINGFGRIQEMEAARTAGATYLGFPLLKTDDDGNMTEMPYITDSAARYVLN GDLIGLSFGDAAPKIDTLPNGAGFKYKVSCPQPDCWLSRYELPALAFYTFLS RKYHISRSTEDIVEDALNEYRAFFAGIADGSITSMDGVGIPRKNIPEKLLDYLE GRGKRTDFKKYKEGLVAKMLLNTEALLSRLKEDLKVIGTKDNRIGKKSYVRIR PGKLAEFLAEDIVRFQGHPAGMPEKKLTGQQYSILQGMIATFHEGLADACRN AGLLDGDSAHPFLSLVFTRHAQGMTSTVDFYRAYLEERRTYLKGVVPDEAP FLHQERRKWSANKDSAYYKSLALRYIKDEKTGDKVGVFLPRGLFDNAVHAII KEHCPNTSKVINASERANMAFIILTYLEKELDDQNQGFYFNEERLKEYGFSKA IRKELEESGMKRLSHVLRLERNTNPSGLYYEALREESGWKDDRRKGGQLD RKTEEFAEKLRHSYKRMCDNEKTIRRYMVQDITLFLLARSLVRIAGNSVNLW SVGPEGNGILDQWVDVMTPYKKYIIRQKGIKIKDYGEIYKILKDRRIDSLLLNQ KKRVPEAIDLEEIKEELVTYNRKRASMVSAIQVYEKGVFEENREHFDSMTSR FGFKEILEADVRSSSLTKEAVKKVRNAVSHNQYPDRMVIKDGRSMVLYSPDL PDMAKGIAETTDRLTKYGTDISKQTDE |
| 25 | Cb10<br>ContigID: Cas13/21_contig -81_616 | MLQTEKNDRGAFWAAYFNTATNNVQAILQFAGKNVQLEELSNQEFKLSANG IGDEETKELEWSAESYPAIQTLKMTGDEVNIPEQIRIMKTLSKHLPFMKRISTR VQNGVRKNGKTEKAGEEMTPQMFAEILIGYVNYLYDLRNYFTHYKHVPVSR KNMQSEYLGILFDANVGTAKERFYSEDKIAKDDKRFNNFRMHKGAESVPDK KGGMKKQPKLNKDFLFYLWETDPITPKGKDNPYLELTAQGLAFFLCLFLEKK SANMLLDSVGIEEDAESLIDGFGFENNSGDNRTLLKRIFTITCARLPRTRLES ENLISNQVLGLDIMNYLHKCPKEFYNLLSPQDQHKFRTLSDDHGTETLLKRF DDRFPYLALNVMDRLECFDSLRFCIDMGTFYFRCHSRVQIDGSRLENRRLKK KLTCYTRRQDAIEYFQTERAAENTFYQTDNLAPAPKAYRTDMLPQYDLGRG RKQENRIGIALKSLDDSRPMFNQPTLDPAGQIKPKTYKPDAWLSTYELAPAL FLSLHGKGSDVEKRIEEFIRSWKGFADWMSKASKEQLKELRYNSAREEFEV FKSRFEKQFGLSVNDIPDEFRYYLVNGKIKPIYMRFQRGSAVPITMDEAAQIW LYNECDKTRSIIRKFNNEQSYDFKLGKGRQRRYTSGNIALWLVRDFMRFQKA NDNPREGFGKLKSSPDFNVLQSSLALFNSRKDSLKEILKNARLIYNPSGNHP FLENVINRSGALLSIEGFFNAYLDERLEWLSNASGREVYQLRKLYERGEKKR SAAKGLPPANVYLGEMAKQFLDESVCLPRGLFDDLVRSALKEIYPEQYAKDV PDGQRANFTWLMQKHLKWSGDDHQWFYKELKQSMTAEEFKKLFELLGTTD VRYDNAGKKTLNDKLEETYSQRWYDLEKSLKNDKEFKSLSPNDKQSAIDDR MGQENQMYRSLLNRLRDVWKRIRMSSVQDVVLRDAVWQLLGLPEKSVRLC NVKPEYDLKTQRMIRDDAGDGSISNDAILNETHSLENIIDLPKGLGTIVLKGEM KLKNFGNFYRMLSDPKLPSFLCLYKQFGFNAVDYSYLELQEFEFYDRKFRPA VFNWVHQLEETVLEKYPDLPMKYGKFVDFWTIAGKANGGYIFTQLLLTSIRN AVSHQYYPEFCVPSDWTNQKDIDTYSPQFKDQLLSVKAEFLKRRSEDKDAL LSETIYDYAQSLFENAIAFVEQQP |
| 26 | Cb11<br>ContigID: k141_16137484 | MANFQTPQRHIFGTYLNIARANFYKTILHVFSASGIDCYTKRGDLFVREDTVD KVISALYLIVNGENAGYHAIKEIVSKSYDKRWKEDKALQGNLSGSELKARKEE FKSPLKDEGPDGEDARICKSFTLGSEQEERMRKLLFRHIPLLSPIMADVVAM QFKETTNEHQEANKTLHDATLADCFKELSNIARCLSESRNFYTHKNPYNSIE AQRTQLQLQKVIANNLDKAFIGSRRIAKKRNSYSEKDLAFLTGHDNDCRMEE IFVLDENGNKIWKVEKDKNGKDKLDKDKNPIYVYKKVKAKDGKGREKLDEKG KPVYETLRENGEPVHEYEKKFVERKDFYFRIRGKREVLAPDLTPTGEECDGL SAFGMLYFCSLFLSKEQTAQLCTESRVFVTSPYQPAGNLKNNIILNMMFVYAI HIPRGKRLDSETDSQALGMDMLNELRRCPIELYDVLPSIGKRDFEDNVKHEN NRTPELSKRIRLKDRFPYLAMRYIDQQQLFKRIRFHVRLGSFRFCFYDKTCID GKSHPRQLHKDINGFGRLQDMEKERKEQYGHFFQQSREQSIWQKDENAYV NLKQLEPIKAGDQPHITDMFAQYNIHQNRIGLFWNTDEECKLVNKTNAQGEII YNGYYLPPLNYVDSPTENNKHKRKAPVDMPAPLCSLSVFELPAMLFYNFLR NTDSLGGEEFPDVEEIVIKQYDNIRKFFLEVKDIQPTDNIENLATILNAYGLSK QSVPKKIYDYLSNKNTLISKDIRKSTEKEVKNRLRRAIIRKQRFEQDQEHIENT KDNKIGKDSFVSIRYAKIAEELAKSMMEWQSGNTKMTGLNFRVLTAALAKFG DGVIKRDTIISMLQKAEIMGGDNPHCFIEQAIEQEQYDIEEFYLDYISAEIEYLK RFLMIDGKTIILKDEQLLDALRKDKEVHDDVRIRLKNDVDFGQLPFIHKSRLR WQQSKIEELANRYLYVKEEGEETLGRATLLLPDGMFFPYIMKGFQKCHQELT NSIEALSDEQKKGIENNVAYLINLYFESKGEKSQSFYDSTEPSHYNDNIRQLA PYKYARSYEFFKIIKGWQIHLSCDEMKKRLTGKKTIIDNKINALKEKGNYISLE EAKNALRRKLHNTFRDMQDNERVIRRYKTQDRILFLMARDMMGEIVNKKAD |

| | | |
|---|---|---|
| | | LFKLENVCKDDFLSQKVKASIAVHLSMGEVFKIQKDEMAIKDYGKLYRLLRDD<br>RITKLLSYALYETGETIDYDDITDELKEYESCRSAAFEAVQMIEDTRYQQDKE<br>VLSNPNENNFYCGNIRYKNGKDNGRENEAKRNNFRTLLEDLQKFTPEQME<br>MFSQEDRELIISIRNAFGHNSYPKQVDFERLINQERKNNPNFKIELKQVASFIL<br>DKLEEYVNQVNPQT |
| 27 | Cb12<br>ContigID:<br>k127_333529 | MSYTPSSRPPRRRPQIVEGSRNNALRILKITPDEQTAFVTYLNYAVNNLSEVA<br>GVAFSDERKVRADVFRGEPADIQRRISRLADFLWSFREKDPSADESGYKAK<br>LGGGHDDMAVWLTEKIVSLRNFFSHVNRQDCTPLVISHDEYVFVEGILGGAA<br>RDAAMGPGLNPAKAQKLKLATHHVKEAHTYEFTRKGLVFLTCLGLYKDEAE<br>EFCHLFHEMKVPDRIEDADLDEELPDGRHRAVLGLEDLDKFAGLKGKGRALI<br>ELFSFYSYRRGRQSLDAANLDFMQFADVVGCLNKVPAPAFEYLPLEKEHAE<br>LEALKAASTESEENKRTKYVLRRRDRNRFLSFAAAYCEDFGVLPSVRFKRLD<br>VRRTTGRHQYVFGPGADAETDEEESNRVRLNRHYAIRRDAIPFEFEPDRHY<br>GPVRIGALRSAVSATEMRRLLYLHAQGADIDAVLRGYFEAYHRVLERMVNA<br>GSLAEISLDDETFLADFATICGTSPEAFKADPEAFRKFVPESLRRYFSKDAGP<br>RSEQRLHALLCSKLLAAVNRTCDILVRHDALEAWREASRPWLDARDDWRG<br>RIDVWRKANPKKDLPDELKTFEAYLETLPADERPARPEEPRCRVGEGEGEIT<br>NPPTWCRFSDADYVSFVFDWFNLFLPNNRKFRQLPISEQHREGVEDHLFQ<br>MVHAAIGKFSLDQKGLWSLLEKNRAELKPYADLLQYRTDIRVLVAPFVRELN<br>HQLPENRRLPEFDNGVDKDEKEIASPILNIFRHFRQNELWQFLQNFRPELKS<br>WCRSMKSKLQDAMEKKPSRKDGKPHFPKVEDLFEVLNLRRPSPTLEDLAVE<br>AVKLSNDMFLEETNRWGAANPMSVSRDDLLAACRRFGVRPGMPVEYKSLL<br>KTVLGIDYDAWRHAFDYGTGRPFADRRLEDEEHVAAQIPLPNGFADRVAAS<br>LPRKVRERFQAPGSAAFDFNAAFRAWSPDPAVSLRDFYDVKPLVASQIARK<br>HPPEGGAATADPFAALSAKTLDSAVREIKDAENQDKVLLFVAMKYWERFRG<br>GDTYSTGKLKMPFSEKTTLREFFDTPVQIEKDGLTVSFRPNDVNRPAFATLF<br>GNSRAAKDYRAKIAKILSPDGSRTAFDFYEMVVAFREQKARDRHERLAFTPY<br>AVHFDALCEIPASAYEKAAEGRFGEAKTEAIRAMEFERYKAVLPGLTREDYD<br>AVADARNAVFHTGFKFDCSKAIAVCKRLGVLGMMPGEAAPAGRRTSSHSG<br>PKWSGPRRNGGSRW |
| 28 | Cd13<br>ContigID:<br>k127_2411982 | VSKNDNIKSKAKALGLKSTFQVGDEVVMTSFGKGNKAIVEKIVRGTDVQSVP<br>TEPNFSAEIEGKKFDLVGRAHIQTKSDNPQYSKKRTGDDMIGAKAALEKRFF<br>GGTFDDNIHIQLAYNVLDIEKILSVHINNIVYTLNNIRRKDNAEDDDFIGYMSTR<br>NDYDTFIEPRKHNISEDAAKSIDKSRASFEEYLQPPVKDCLHYFGNTFFAPRE<br>IEKTYTDNRGFERKKKVTVNSLIDEKEIYYIFALLGGLRQFCTHDNKSGRNWL<br>YSLEENGINSDAKAVLDKYYNSAVARIDESFVDNASKTNFKLIFNAMDVSDEA<br>LQDNIAKAFYKFTVCKSFKNMGFSIKKLREQLLELPEYEGLKDKHYDSVRSKL<br>YQIMDFVIYLSFKDEKFKKDNEEKINNIVNELRATLNEEDKGRVYASYAERMK<br>SELKPAIARLKSDIDKIKDSRVKEFELDASVKYRLSKVVESVRLKDRATYFTKL<br>IYLTTLFLDGKEINDLLTTLIHQFENIASFIDVMNDRGIDCRFSDGYKLFESSKQ<br>IAFELRNVNSFARMTRTSKDDENATHMMYIDAAEILGTDYTEEQIEEHLNLEK<br>KRMIPGTKKADMNFRNFIINNVIKSSRFNYLVRYSNPKKIRALADNEGVIRFVL<br>GELPDAQIDRYTLLCGFNPDADRQEKTDKLAKAITGLRFNDFENVKQGANTE<br>GESQESIDKAQKQGLISLYLTVLYLLTKNLVYVNSRYFLAFHCLERDAQLLGS<br>GAGHHEPYVALTQRFINEDKLNEHACEYLKTNIANSDEYTIRIFRNNAAHLSA<br>VRNANLYIDKLKEFKSYYEIYHFLSQENIYGKYCVDKKYVTADENGSKTISVKI<br>SKDYCPQVYIDKSLEYFDKLNKYGTYCKDFTKALNSPFGYNLARYKNLSIEG<br>LFDRNRPGDKGENTFED |
| 29 | Cd14<br>ContigID:<br>k141_15335538 | MAKKLSPKEIREAAKAEKMKSIKAAEAEREKAAEEAKLKAEAEKKEKAEKNE<br>REKALKRFRLDEKSRMALPKSERKSLAKAAGVKSAFAVGNDIYLTSFDRGN<br>DAIVEKKITDTVVTNLRSDESFEVNENTITEMSVPIKSKRISDLYAIADNPLYRK<br>DSATKVQPDKLLLKDTLEKLYFGKTFDDTLHIQIIYNILDIEKILTVYSINTIYCLN<br>NLFGKESGEKEDLISKLTYQITYDEFKESKAHNEFIDFYNLNTLGYYGNIFFKE<br>KKKRSQKEIYDIIALIATIRQWCVHCEEDKRTWLFNTETVLSKEFLDILDDVYE<br>SLVEKVNRNFLKDNKVNLQILEDVLEIKDSESREKLIRQYYRFIVTKEQKLLGF<br>SIKKLREAMLEETEFKTDKKYDTVRSKLYKLIDFLLFTGYTTDEAEKEKALFLI<br>KSLRESLTEETKDRIYKSEAARLWLKYENTITNKIREALNEKSISELKKDKSFD<br>DKSITSIITDEVSGKKATYFSKTIYLLAQFIDGKEVNDLTTSLINKFDNIRSLIDT<br>AGQIGLDCKFTEEYKFFENSDQIRTELHVIKNLTNMEYYDTTVKKQMYKDAV<br>HILGIQDDVSDAELEKIINSILLLNENGKPLPGTKGKKGFRNFIISNVLKSRRFIY<br>LIKYCNPKKIRKIAGNRKIIKFVLSRITDSQLERYYYSCNPELKSGIYPGRDDAV<br>NDLSILIADMKFEDFKNVDQSANVHDNNNAAREKMKYQTIISLYLTVCYHLVK<br>NLVNINARYAMAFHALERDARLYQIFSSEENYVDNLNADYAILTKTLLKDNYE<br>NAGNLYLRNKKWNKLTRENLDNYIPQAAANFRNAVAHLNPIRNADMLLEDIE<br>DVSSYYAIYHYIMQKSVTNRTIRVSNTTEDEKRILTDYQNKIKRHHGYNKDFV<br>KALCVPFAYNIVRFKSLSIYEMFDRNYHEKTSPETDDSQSP |

| | | |
|---|---|---|
| 30 | Cd15<br><br>ContigID: Cas13/23_contig -81_4932 | MDKEKTTVEGKNTNQKSDVLKSLAKANGLKSSFVIGNEVVMTSFGRGNSAIL<br>EKKITGSRIENLNPNVAFYVKKHISDSCNPDSGKYDVKSKRMKEKAVVDDPV<br>YVSPEKASNVHAGQDLIGCKNVLEERYFGKTFDDNIHIQLIYNILDIEKILAVHV<br>SNATFAINNILGIEGKENEDFIGNLSVLNTFDEFENYETHPKFANKSAIKENLR<br>KSKVFFDKIKKGNKLGYFGQAFYYATGTGKNLIFTKKSSETIYELLALVGSLR<br>QFCVHDEVMVDNKVKSRSWLYNAQKELKPDFLKALDELYSKEVEKIDSDFIV<br>NNTVDLHIIHDAIDVIDGSADWQKITNEYYDFIIRKCFKNIGFSIKRLRETMIEEQ<br>MKVLCGKCDRENCKGYGKCFKNKKYDSVRSRLNRIVDFIIFRHYNDEAILKN<br>VSLLRTCMSEEEKQKRFYLPEAKALWKKYNYVFRNYVLKKLNGKSISGLKEK<br>AIENSIDINSVKISLGDPDYFCKFIYLLTFFLDGKEINDLLTTLINKFDNIASFISV<br>MKNDKLSIDCEFVPEYSFFANSAQITSDLRVINSFARMQAPAEPSKDDMYRD<br>ALDILGMDDLSEDGKKQLEDTVLCRDENGKYMKKEDNNPKRDTNFRNFLGN<br>NVLASTRFKYLIRYNNAKKTRALANNKAVIIFMLNKINKQNPEQIVSYYKACRD<br>DSDPVASDAEAKIEFLAEKIMNVSCTQFRYVKNGTKVRPDEAKEKERFKAIIG<br>LYLTVMYLITKNMVYINSRYVTAFHCLERDSELHGVKFDQKKLQPNLTKKFID<br>PKTCGDYGLRNNKRARTYIEQNMDKMSNCTSYWNEYRNAVAHLSVIRNMN<br>QYIKNVKNIGSCFELYHYIMQRFLLDKENIAESLREYDDFIKKKGCYRKDFVK<br>ALNTPFGYNLARYKNLSIAELFDRNDTELERTNKLRNEAIKADIEEI |

## Detailed description of the embodiments.

It will be understood that the invention disclosed and defined in this specification extends to all alternative combinations of two or more of the individual features mentioned or evident from the text or drawings. All of these different combinations constitute various alternative aspects of the invention.

General and definitions

For purposes of interpreting this specification, the following definitions will apply and whenever appropriate, terms used in the singular will also include the plural and vice versa. In the event that any definition set forth conflicts with any document incorporated herein by reference, the definition set forth below shall prevail.

Unless specifically defined otherwise, all technical and scientific terms used herein shall be taken to have the same meaning as commonly understood by one of ordinary skill in the art (for example, in cell culture, molecular genetics, immunology, immunohistochemistry, protein chemistry, and biochemistry).

In the following, the invention will be described in greater detail. The examples and preferred embodiments described throughout the specification should not be construed to limit the present invention to only the explicitly described embodiments. This description should be understood to support and encompass embodiments which combine the explicitly described embodiments with any number of the disclosed and/or preferred elements.

Throughout this specification, unless specifically stated otherwise or the context requires otherwise, reference to a single step, composition of matter, group of steps or group of compositions of matter shall be taken to encompass one and a plurality (i.e. one or more) of those steps, compositions of matter, groups of steps or groups of compositions of matter. Thus, as used herein, the singular forms “a”, “an” and “the” include plural aspects, and vice versa, unless the context clearly dictates otherwise. For example, reference to “a” includes a single as well as two or more; reference to “an” includes a single as well as two or more; reference to “the” includes a single as well as two or more and so forth.

As used herein, except where the context requires otherwise, the term "comprise" and variations of the term, such as "comprising", "comprises" and "comprised", are not intended to exclude further additives, components, integers or steps.

The term “and/or”, e.g., “X and/or Y” shall be understood to mean either “X and Y” or “X or Y” and shall be taken to provide explicit support for both meanings or for either meaning.

Unless otherwise indicated, the term "at least" preceding a series of elements is to be understood to refer to every element in the series. The term “at least one” refers to a minimum of one, but also encompasses two, three or more such as four, five, six, seven, eight, nine, ten and more.

The terms “about” or “approximately” as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/−10% or less, +/−5% or less, +/−1% or less, and +/−0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention.

Where values are described in terms of ranges, it should be understood that the description includes the disclosure of all possible sub-ranges within such ranges, as well as specific numerical values that fall within such ranges irrespective of whether a specific numerical value or specific sub-range is expressly stated.

As used herein, the term “polypeptide” is intended to encompass a singular “polypeptide” as well as plural “polypeptides,” and refers to a molecule composed of monomers (amino acids) linearly linked by amide bonds (also known as peptide bonds). The term “polypeptide” refers to any chain or chains of two or more amino acids, and does not refer to a specific length of the product. Thus, peptides, “protein,” “amino acid chain,” or any other term used to refer to a chain or chains of two or more amino acids, are included within the definition of “polypeptide,” and the term “polypeptide” may be used instead of, or interchangeably with any of these terms.

“Cas13 polypeptide”, and the specific subtypes Cas13a, 13b and 13d, may be used interchangeably with Cas13 protein, Cas13 peptide, Cas13 effector enzyme and Cas13 enzyme. At times throughout the specification, the phrase “Cas13 polypeptide” may also be abbreviated to “Cas13”.

As used herein, “Cas13” is a CRISPR-Cas effector enzyme that displays at least collateral cleavage activity upon binding to a cognate target RNA complementary to the spacer sequences in the guide sequence. The collateral activity of the Cas13 enzyme enables it to cleave RNase or endonuclease activity against a non-target RNA. This property of the Cas13 polypeptides of the invention is also referred to as trans-collateral activity or trans-cleavage activity and is used interchangeable in the specification.

The term “cis cleavage”, or “cis cleavage activity” or “cis activity” will be understood to mean specific, targeted degradation of target RNAs (referred to as cis-cleavage activity) both *in vitro* and *in vivo*.

The Cas13 polypeptides of the invention, while naturally occurring in bacteria, are all isolated from bacteria or are expressed from a nucleic acid molecule that is itself engineered or recombinant. Unless the Cas13 polypeptide and/or nucleic acid molecule encoding it are being discussed in the context of the bacteria in which they are naturally found, all references in this specification are to the non-natural versions of both and may be described as non-natural or engineered.

More specifically, the term "isolated” in relation to a nucleic acid sequence, or a polypeptide, is used herein to define that by virtue of its origin or source of derivation, it is not associated with naturally-associated components that accompany it in its native state; it is substantially free of other proteins from the same source. A nucleic acid

sequence or polypeptide may be rendered substantially free of naturally associated components or substantially purified by isolation, using techniques known in the art. The term “isolated” is also used herein to recombinant nucleic acids and polypeptides.

The term “engineered”, which can be used synonymously with terms such as “non-naturally occurring” is used herein with respect to CRISPR/Cas systems that are not found to occur naturally in eukaryotic cells. That is, the components of the system do not exist together naturally in eukaryotic cells.

A “functional variant”, as used herein in respect of the Cas13 polypeptides of the invention, refers to a sequence variant of the Cas13 polypeptide which retains at least 95% of the cleavage of the reference Cas13 polypeptide, when both are subjected to the same methodology of assessing cleavage activity. The amino acid sequence of a functional variant has at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity amino acid sequence. The cleavage function may be cis activity, trans activity or preferably both.

A “target RNA” or “RNA of interest” as used herein refers to an RNA polynucleotide being, or comprising, the target sequence. The target RNA may be an RNA that is endogenous in a target host cell (derived from mammalian, yeast, insect or bacterial cells) or it may be a foreign source of RNA, such as viral RNA. The target RNA may be RNA that has been isolated and purified; or may be an RNA that has been synthesised.

A “CRISPR RNA”, or “crRNA”, also referred to in the literature as “guide RNA” or “gRNA”, includes one or more spacers and one or more Cas13-specific direct repeats. The spacers are capable of specifically hybridising with one or more target RNAs. Hybridisation promotes the formation of a CRISPR complex the Cas13 polypeptide at the site of potential cleavage of the RNA.

The Cas13 proteins and crRNAs are assembled in to a ribonucleoprotein (RNP) complex.

As used herein, “direct repeat” or “direct repeat sequence” refers to the RNA sequence in the crRNA that folds to form a secondary hairpin RNA structure, to which the Cas13 polypeptide binds.

The “spacer” within a crRNA can include a nucleotide sequence that is fully or partially complementary to a specific sequence within a target RNA.

The terms “hybridising” or “hybridise” can refer to the pairing of substantially complementary or complementary nucleic acid sequences within two different molecules. Pairing can be achieved by any process in which a nucleic acid sequence joins with a substantially or fully complementary sequence through base pairing to form a hybridization complex.

“Homology”, “identity” or “similarity” refers to sequence similarity between two polypeptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences.

As used herein, “contacting” a cell with a nucleic acid molecule, can be allowing the nucleic acid molecule to be in sufficient proximity with the cell such that the nucleic acid molecule can be introduced into the cell. Similarly, “contacting” a nucleic acid with an RNP complex, or the Cas13 proteins and crRNAs components of an RNP complex, means bringing them in to sufficient proximity for the RNP complex to bind to a target sequence in the nucleic acid if the target is present. Conducting a binding and cleavage reaction in a single vessel is sufficient proximity.

As used herein, “detector RNA” is used synonymously with the term probes, detection molecules and reporter sequence. It is any RNA sequence that is cleaved by the trans cleavage activity of the Cas13 polypeptides of the invention, and which is in turn indicative of the presence of the target sequence in a sample with a nucleic acid.

As used herein, the Cas13 polyprotein of the invention is “activated” upon the presence of a target sequence to which the spacers of the crRNA of the RNP complex binds.

As used herein, the term “operably linked to" means positioning a promoter or other regulatory element relative to a nucleic acid such that expression of the nucleic acid is controlled by the promoter or other regulatory element.

As used herein, the term “promoter” is to be taken in its broadest context as a naturally occurring sequence, or a recombinant, synthetic or fusion nucleic acid, or derivative which confers, activates or enhances the expression of a nucleic acid to which

it is operably linked that initiates transcription of a gene. Suitable promoters include but are not limited to tissue-specific promoters, inducible promoters, and constitutive promoters.

A used herein, a “vector” is DNA molecule (often plasmid or virus) that is used as a vehicle to carry a particular DNA segment into a host cell. The vector typically assists in replicating and/or expressing the inserted DNA sequence inside the host cell.

Where values are described in terms of ranges, it should be understood that the description includes the disclosure of all possible sub-ranges within such ranges, as well as specific numerical values that fall within such ranges irrespective of whether a specific numerical value or specific sub-range is expressly stated.

i) Cas13 Polypeptide

One aspect of the invention provides a Cas13 polypeptide, or a nucleotide sequence encoding the Cas13 polypeptide. There is also provided a composition comprising a Cas13 polypeptide, or a nucleotide sequence encoding the Cas13 polypeptide.

The Cas13 polypeptides of the invention may be isolated from bacterial species in which they naturally occur, or may be expressed in vivo or in vitro from non-naturally occurring nucleic acid molecules.

In a preferred embodiment of the invention, the Cas13 polypeptide is a Cas13a polypeptide, a Cas13b polypeptide, or a Cas13d polypeptide. Alternatively, in a preferred embodiment of the invention, the nucleic acid molecule comprising a sequence encoding a Cas13 polypeptide is a nucleic acid molecule encoding a Cas13a polypeptide, a Cas13b polypeptide, or a Cas13d polypeptide. The Cas13a, Cas13b, and Cas13d polypeptide have at least trans cleavage activity and preferably both trans cleavage and cis cleavage activity. The Cas13 polypeptide of the invention is preferably a Cas13a or Cas13d polypeptide, and more preferably, is Cas13a7, Cas13d13, Cas13d14 and Cas13d15.

In embodiments wherein the Cas13 polypeptide is a Cas13a polypeptide, the Cas13a polypeptide has an amino acid sequence of SEQ ID NO: 16, or SEQ ID NO:17, or SEQ ID NO: 18, or SEQ ID NO:19, or SEQ ID NO: 20, or SEQ ID NO:21, or SEQ ID NO:22, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 16, or SEQ ID NO:17, or

SEQ ID NO: 18, or SEQ ID NO:19, or SEQ ID NO: 20, or SEQ ID NO:21, or SEQ ID NO:22.

In embodiments wherein the Cas13 polypeptide is a Cas13a polypeptide, preferably the nucleic acid molecule encoding the Cas13a polypeptide comprises a sequence selected from SEQ ID NO: 1, or SEQ ID NO:2, or SEQ ID NO: 3, or SEQ ID NO:4, or SEQ ID NO: 5, or SEQ ID NO:6, or SEQ ID NO:7, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 1, or SEQ ID NO:2, or SEQ ID NO: 3, or SEQ ID NO:4, or SEQ ID NO: 5, or SEQ ID NO:6, or SEQ ID NO:7.

In embodiments wherein the Cas13 polypeptide is a Cas13b polypeptide, the Cas13b polypeptide has an amino acid sequence of SEQ ID NO: 23, or SEQ ID NO:24, or SEQ ID NO: 25, or SEQ ID NO:26, or SEQ ID NO: 27, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 13, or SEQ ID NO:14, or SEQ ID NO: 15.

In embodiments wherein the Cas13 polypeptide is a Cas13b polypeptide, preferably the nucleic acid molecule encoding the Cas13b polypeptide comprises a sequence selected from SEQ ID NO: 8, or SEQ ID NO:9, or SEQ ID NO: 10, or SEQ ID NO:11, or SEQ ID NO: 12, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 8, or SEQ ID NO:9, or SEQ ID NO: 10, or SEQ ID NO:11, or SEQ ID NO: 12.

In embodiments wherein the Cas13 polypeptide is a Cas13d polypeptide, the Cas13d polypeptide has an amino acid sequence of SEQ ID NO: 28, or SEQ ID NO:29, or SEQ ID NO: 30, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 28, or SEQ ID NO:29, or SEQ ID NO: 30.

In embodiments wherein the Cas13 polypeptide is a Cas13d polypeptide, preferably the nucleic acid molecule encoding the Cas13d polypeptide comprises a sequence selected SEQ ID NO: 13, or SEQ ID NO:14, or SEQ ID NO: 15, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 13, or SEQ ID NO:14, or SEQ ID NO: 15.

In the abovementioned embodiments, wherein the Cas13 polypeptide is described as being a sequence with at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%,

96%, 97%, 98%, or 99% identity (ie a functional variant), it is intended that the function of the Cas13 polypeptide be substantially the same as a polypeptide of the defined, reference SEQ ID NO despite the variation in the sequence. By “substantially the same”, in this context, it is meant that the Cas13 functional variant has cleavage activity within +/- 5% compared to the reference sequence, when both are subjected to the same methodology of assessing cleavage activity.

In that regard, the polypeptide sequence may vary in preferred embodiments only by way of a “conservative amino acid substitution”. As would be understood by the skilled person, that means the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art, including basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine).

Similarly, as a result of the redundancy in the genetic code, whereby most amino acids are specified by more than one mRNA codon, a nucleotide sequence can be altered while maintaining the same amino acid sequence of the encoded protein. The sequence encoding a Cas13 polypeptide of the invention may be at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the reference sequence encoding the polypeptide of the defined SEQ ID NO. Accordingly, in embodiments of the invention, a sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to:

SEQ ID NO: 1 encodes a Cas13 polypeptide of SEQ ID NO:16

SEQ ID NO: 2 encodes a Cas13 polypeptide of SEQ ID NO:17

SEQ ID NO: 3 encodes a Cas13 polypeptide of SEQ ID NO:18

SEQ ID NO: 4 encodes a Cas13 polypeptide of SEQ ID NO:19

SEQ ID NO: 5 encodes a Cas13 polypeptide of SEQ ID NO:20

SEQ ID NO: 6 encodes a Cas13 polypeptide of SEQ ID NO:21

SEQ ID NO: 7 encodes a Cas13 polypeptide of SEQ ID NO:22

SEQ ID NO: 8 encodes a Cas13 polypeptide of SEQ ID NO:23

SEQ ID NO: 9 encodes a Cas13 polypeptide of SEQ ID NO:24

SEQ ID NO: 10 encodes a Cas13 polypeptide of SEQ ID NO:25

SEQ ID NO: 11 encodes a Cas13 polypeptide of SEQ ID NO:26

SEQ ID NO: 12 encodes a Cas13 polypeptide of SEQ ID NO:27

SEQ ID NO: 13 encodes a Cas13 polypeptide of SEQ ID NO:28

SEQ ID NO: 14 encodes a Cas13 polypeptide of SEQ ID NO:29

SEQ ID NO: 15 encodes a Cas13 polypeptide of SEQ ID NO:30

Two Higher Eukaryotes and Prokaryotes Nucleotide-binding domains (HEPN) provide the RNAse activity of Cas13.

The Cas13 polypeptides of the invention have at least trans-collateral cleavage activity, and preferably also retain their cis cleavage activity. The Cas13 polypeptide of the invention is preferably a Cas13a or Cas13d polypeptide, and more preferably, is Cas13a3 of SEQ ID NO:18, Cas13a7 of SEQ ID NO:22, Cas13d13 of SEQ ID NO:28, Cas13d14 of SEQ ID NO: 29 and Cas13d15 of SEQ ID NO:30.

The Cas13 polypeptides of the invention may be isolated from, or expressed from nucleic acid molecules engineered on the basis of the sequence of nucleic acid molecules within a bacterial organism. To assist with expression in cells, the coding sequence may be may be codon-optimised for expression in a eukaryotic cell.

The protospacer flanking sequence (PFS) for Cas13, which is analogous to the PAM sequence for Cas9, consists of a single base pair. In some embodiments, the Cas13 polypeptides of the invention are sensitive to, and require the presence of a PFS in the target sequence. In turn, the Cas13 polypeptides may have a preference for, and preferentially cleave, targets with a particular PFS. This characteristic of the Cas13 polypeptide can be exploited in, for example, nucleic acid detection assays, particularly where the sample contains multiple targets.

In alternative embodiments, the Cas13 polypeptides of the invention are capable of binding to target RNAs in a protospacer flanking sequence (PFS)-independent manner. This is particularly advantageous for use of the Cas13 in a CRISPR system as the crRNA can be designed without having to ensure the presence of a PFS in the target sequence.

Preferably, the Cas13 of the invention that is PFS-independent is a Cas13d, most preferably Cas13d13, Cas13d14 and Cas13d15.

ii) CRISPR RNA (crRNA)

Cas13 only binds and cuts ssRNA. Cas13 finds its target with the help of a CRISPR-RNA (crRNA) also known as a CRISPR-RNA (crRNA). The crRNA consists of:

- a sequence which, via the presence of at least 2 direct repeat sequences of at least 16 nucleotides and up to 50 nucleotides in length, forms a double stranded, hairpin-like structure that is bound by Cas13; and

- a variable “spacer” sequence of approximately 15 to 40 nucleotides that is complementary to, and capable of hybridising, the target RNA. The spacer need not be 100% complementary, but must be sufficiently complementary to permit hybridisation. In each embodiment of the invention, the spacer sequence is designed for the specific target/s of interest.

In some embodiments, the Cas13-specific direct repeats are at least 16 nucleotides in length, but preferably no longer than 50 nucleotides in length. In some embodiments, the Cas13-specific direct repeats are 30 to 40 nucleotides in length. In some embodiments, the Cas13-specific direct repeats are about 35-37 nucleotides in length.

In some embodiments, the spacers are 15 to 40 nucleotides in length. In some embodiments, the spacers are 25 to 35 nucleotides in length. In some embodiments, the spacers are about 30 nucleotides in length. As noted above, if the Cas13 is PFS-sensitive, the spacer sequence is designed to include a PFS.

The direct repeat sequences may be located upstream of the spacer sequence, or the direct repeat sequence may be located downstream from spacer sequence.

The ability of a crRNA sequence to direct sequence-specific binding of a CRISPR/Cas13 complex to a target nucleic acid sequence may be assessed by any suitable assay.

The crRNA can include multiple spacer sequences to target multiple RNAs or to target multiple sequences within the same target RNA. The crRNA preferably has at least

1 spacer sequence, but may include 2 spacer sequences or 3 spacer sequences or more. The crRNA may be a precursor crRNA, that is processed in to the individual crRNAs. Alternatively, when being used for targeting multiple RNAs, multiple crRNAs can be used.

iii) CRISPR systems

The at least one Cas13 polypeptide of the invention forms a complex with the at least one crRNA via binding to the direct repeats that have formed a double stranded, hairpin-like structure, and wherein the at least one crRNA directs the complex to the one or more target RNA molecules by way of the engineered spacer sequences, thereby targeting the one or more target RNA molecules.

In one aspect, there is provided a CRISPR/Cas13 system for targeting RNA molecules, the system comprising

i) at least one Cas13 polypeptide wherein the Cas13 polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30; or a nucleic acid molecule comprising a sequence encoding the Cas13 polypeptide and

ii) at least one CRISPR RNA (crRNA) or a nucleic acid molecule encoding the crRNA, the crRNA comprising one or more spacers and one or more Cas13-specific direct repeats, wherein the crRNA is capable of hybridising with one or more target RNA molecules.

Preferably the nucleic acid molecule comprising a sequence encoding the Cas13 polypeptide is selected from the group consisting of SEQ ID NOS: 1-15, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 1-15.

As defined above, the Cas13 polypeptide in the CRISPR system of the invention may be a protein, but alternatively may in the form of a nucleic acid molecule that encodes the protein. It will be appreciated that the nucleic acid molecule encodes the Cas13 polypeptide in expressible form such that expression results in a functional Cas13 polypeptide. Similarly, the crRNA component of the CRISPR system may be in the form of RNA itself, or a nucleic acid molecule that encodes the crRNA.

In one embodiment, the CRISPR system for targeting RNA molecules comprises (i) at least one Cas13 polypeptide and ii) at least one CRISPR RNA (crRNA), the crRNA comprising one or more spacers and one or more Cas13-specific direct repeats, wherein said crRNA is capable of hybridising with one or more target RNA molecules,

wherein the at least one Cas13 polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30 (ie functional variants). Specifically:

- when the Cas13 polypeptide is a Cas13a polypeptide, the Cas13a has an amino acid sequence selected from SEQ ID NO: 16, or SEQ ID NO:17, or SEQ ID NO: 18, or SEQ ID NO:19, or SEQ ID NO: 20, or SEQ ID NO:21, or SEQ ID NO:22, or is a functional variant thereof having 80-99% sequence identity to SEQ ID NOS: 16-22;
- when the Cas13 polypeptide is a Cas13b polypeptide, the Cas13b has an amino acid sequence selected from SEQ ID NO: 23, or SEQ ID NO:24, or SEQ ID NO: 25, or SEQ ID NO:26, or SEQ ID NO: 27, or is a functional variant thereof having 80-99% sequence identity to SEQ ID NOS: 16-22; or
- when the Cas13 polypeptide is a Cas13d polypeptide, the Cas13d has an amino acid sequence selected from SEQ ID NO: 28, or SEQ ID NO:29, or SEQ ID NO: 30 or is a functional variant thereof having 80-99% sequence identity to SEQ ID NOS: 28-30.

The Cas13 proteins and crRNAs are assembled in to ribonucleoprotein (RNP) complexes by standard methods known to the skilled person, by typically mixing the purified Cas13 with crRNA, together with an RNase inhibitor in a cleavage buffer. The RNP complexes are then mixed with the target RNA, wherein if the target it present, the RNP complex recognises and binds the target via the spacer sequences.

In an alternative embodiment, the CRISPR system comprises a nucleic acid molecule comprising a sequence encoding the Cas13 and ii) a nucleic acid molecule encoding said crRNA, the crRNA comprising one or more spacers and one or more Cas13-specific direct repeats, wherein said crRNA is capable of hybridising with one or more target RNA molecules. In this embodiment, the system further comprises one or more vectors for delivering and/or expressing the nucleic acid molecules. The vectors preferably comprise:

i) a first regulatory element operably linked to a nucleic acid molecule comprising a sequence encoding a Cas13 polypeptide wherein the Cas13 polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30; and

ii) a second regulatory element operably linked to a nucleic acid molecule encoding a CRISPR RNA (crRNA), the crRNA comprising one or more spacers and one or more Cas13-specific direct repeats, wherein the crRNA is capable of hybridising with one or more target RNA molecules.

In these embodiments of the invention, the at least one nucleic acid molecule comprising a sequence encoding said Cas13 (element (i)) is preferably selected from the group consisting of SEQ ID NOS: 1-15, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 1-15. (i.e. a sequence encoding functional variants). Specifically:

- when the Cas13 polypeptide is a Cas13a polypeptide, the nucleic acid molecule encoding the Cas13a polypeptide comprises a sequence selected from SEQ ID NO: 1, or SEQ ID NO:2, or SEQ ID NO: 3, or SEQ ID NO:4, or SEQ ID NO: 5, or SEQ ID NO:6, or SEQ ID NO:7, or is a sequence that encodes a functional variant thereof;
- when the Cas13 polypeptide is a Cas13b polypeptide, the nucleic acid molecule encoding the Cas13b polypeptide comprises a sequence selected from SEQ ID NO: 8, or SEQ ID NO:9, or SEQ ID NO: 10, or SEQ ID NO:11, or SEQ ID NO: 12, or is a sequence that encodes a functional variant thereof; or
- when the Cas13 polypeptide is a Cas13d polypeptide, the nucleic acid molecule encoding the Cas13d polypeptide comprises a sequence selected from SEQ ID NO: 13, or SEQ ID NO:14, or SEQ ID NO: 15 or is a sequence that encodes a functional variant thereof.

Any of the above-mentioned methods can utilise two or more Cas13 polypeptides of the invention and two or more Cas13 CRISPR RNA in order to target different target sites of the same target RNA and/or can target different target RNA molecules e.g., based on variation such as single-nucleotide polymorphisms etc., and such could be used for example to target multiple strains of a virus such as Coronavirus variants, influenza virus variants, HIV variants, and the like.

The crRNA can include multiple spacer sequences to target multiple RNAs or to target multiple sequences within the same target RNA and/or the crRNA may be a precursor crRNA, that is processed in to the individual crRNAs. Or in a further alternative, when being used for targeting multiple RNAs, multiple crRNAs can be used.

iv) Methods

There is also provided use of a CRISPR/Cas13 system in an *in vitro* method of modifying a target RNA, the method comprising contacting the target RNA with a ribonucleoprotein (RNP) complex of a CRISPR/Cas13 system, the system comprising

i) at least one Cas13 polypeptide wherein the Cas13 polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30; and

ii) at least one CRISPR RNA (crRNA), the crRNA comprising one or more spacers and one or more Cas13-specific direct repeats, wherein the crRNA is capable of hybridising with one or more target RNA molecules,

wherein the Cas13 polypeptide and the crRNA form a ribonucleoprotein (RNP) complex, and upon binding of the RNP complex to the target RNA through the one or more spacers, the Cas13 polypeptide modifies the target RNA.

In an alternative embodiment of this aspect of the invention, the *in vitro* method of modifying a target RNA the CRISPR/Cas13 system includes a vector system, and the method includes the preliminary step of:

a) expressing from the vector system at least one Cas13 polypeptide and at least one CRISPR RNA (crRNA), the vector system comprising one or more vectors comprising:

i) a first regulatory element operably linked to a nucleic acid molecule comprising a sequence encoding a Cas13 polypeptide wherein the Cas13 polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30; and

ii) a second regulatory element operably linked to a nucleic acid molecule encoding a CRISPR RNA (crRNA), the crRNA comprising one or more spacers and one or more Cas13-specific direct repeats, wherein the crRNA is capable of hybridising with one or more target RNA molecules;

wherein components (i) and (ii) are located on the same or different vectors of the system;

b) isolating the expression products of step (a).

The isolated expression products of step (b) can first be assembled in to ribonucleoprotein (RNP) complex as described above, and the RNP complex contacted with the target RNA, or the target RNA can be contacted with the isolated expression products of step (b), wherein the Cas13 polypeptide and the crRNA form the RNP complex, and the complex binds to the target RNA. In either embodiment, binding occurs through the one or more spacers, and once bound, the Cas13 polypeptide modifies the target RNA.

In these embodiments of the invention, the at least one nucleic acid molecule comprising a sequence encoding said Cas13 (element (i) is preferably selected from the group consisting of SEQ ID NOS: 1-15, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 1-15. (i.e. a sequence encoding functional variants). Specifically:

- when the Cas13 polypeptide is a Cas13a polypeptide, the nucleic acid molecule encoding the Cas13a polypeptide comprises a sequence selected from SEQ ID NO: 1, or SEQ ID NO:2, or SEQ ID NO: 3, or SEQ ID NO:4, or SEQ ID NO: 5, or SEQ ID NO:6, or SEQ ID NO:7, or is a sequence that encodes a functional variant thereof;
- when the Cas13 polypeptide is a Cas13b polypeptide, the nucleic acid molecule encoding the Cas13b polypeptide comprises a sequence selected from SEQ ID NO: 8, or SEQ ID NO:9, or SEQ ID NO: 10, or SEQ ID NO:11, or SEQ ID NO: 12, or is a sequence that encodes a functional variant thereof; or
- when the Cas13 polypeptide is a Cas13d polypeptide, the nucleic acid molecule encoding the Cas13d polypeptide comprises a sequence selected from SEQ ID NO: 13, or SEQ ID NO:14, or SEQ ID NO: 15 or is a sequence that encodes a functional variant thereof.

In the above-mentioned methods of the invention, the trans-cleavage activity of the Cas13 polypeptides of the invention can be exploited to detect the cleavage of the target RNA. Because the Cas13 polyprotein of the invention cleaves non-targeted RNA once activated, which occurs when a Cas13 CRISPR RNA hybridizes with a target RNA in the presence of a Cas13 protein, a detectable signal can be any signal that is produced when the non-target RNA is cleaved.

Detection methods include a step of measuring a detectable signal produced by Cas13 trans cleavage of non-target RNA. The step of measuring can include one or more of: nanoparticle based detection, fluorescence or chemiluminescent detection, lateral-flow immunochromatography colloid phase transition/dispersion, electrochemical detection, semiconductor-based sensing, and detection of a RNA reporter (RNA molecules carrying a fluorophore and a quencher. Trans-cleavage activity on target detection causes the spatial separation of the fluorophore and the quencher, and the resulting fluorescence signal).

The readout of such detection methods can be any convenient readout. Examples of possible readouts include but are not limited to: a measured amount of detectable fluorescent signal; a visual analysis of bands on a gel (e.g., bands that represent cleaved product versus uncleaved substrate), a visual or sensor based detection of the presence or absence of a color (i.e., color detection method), and the presence or absence of (or a particular amount of) an electrical signal.

In embodiments of the methods of the invention that include a further detection step, the method can utilise two or more Cas13 polypeptides of the invention and two or more Cas13 CRISPR RNA in order to target and then detect, different target sites of the same target RNA (e.g., which can increase sensitivity of detection) and/or can target different target RNA molecules e.g., based on variation such as single-nucleotide polymorphisms etc., and such could be used for example to detect multiple strains of a virus such as Coronavirus variants, influenza virus variants, HIV variants, and the like.

In some embodiments for detecting multiple RNAs, two or more Cas13 crRNAs can be provided in the method. A precursor Cas13 crRNA can also be provided which can be processed into individual Cas13 crRNAs.

In alternative embodiments for detecting multiple RNAs, two or more Cas13 polypeptides cleaving different RNA probe sequences may be used. For example, one Cas13 polyprotein cleaves polyA RNA probe sequences but not polyU. The other polyprotein can cleave polyU RNA probe but not polyA probes. You can therefore tell which target is present by assessing which probe is cleaved.

v) Vectors

In certain aspects the invention involves vectors for delivering or as already described, expressing the Cas13 polypeptide and the CRISPR RNA. The vectors may therefore be a part of a CRISPR/Cas13 system.

Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded. The vectors can be nucleic acid molecules, plasmids, or viral vectors (e.g., AAV, adenovirus, lentivirus). Such vectors are also referred to herein as “expression vectors”, and may be gene expression vectors, or protein expression vectors.

The gene and protein expression vectors can comprise a nucleic acid molecule of the invention for expression in a host cell. The vectors therefore contain one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed.

In embodiments wherein the Cas13 polypeptide is encoded by a nucleic acid molecule, and the nucleic acid molecule is part of a vector, the nucleic acid molecule will be operably linked to a promoter. Suitable promoters include but are not limited to ubiquitous promoters (e.g., ubiquitin promoter), tissue-specific promoters, inducible promoters, and constitutive promoters.

vi) Utility in nucleic acid detection assays/kits

The trans-cleavage activity of the Cas13 polypeptides of the invention can be exploited and used in nucleic acid detection assays. Upon complex formation between the Cas13 polypeptide and the crRNA, and binding of the complex to the target RNA, the Cas13 polypeptide is “activated”. It is this activated form of Cas13 polypeptide that then binds and cleaves non-target RNA. Ie without any reliance on the spacer sequences in

the crRNA. Detection of cleavage of this non-target RNA is therefore indicative of the presence of the target and can be both qualitative and quantitative.

The CRISPR/Cas13 polypeptide system of the invention can be utilised in any assay platform that requires detection of RNA. These include, but are not limited to:

- Assays for detecting the presence of microbial agents in a biological sample from an animal, or in environmental samples. Eg to screen for microbial contamination in water, or contamination in food samples, or agricultural pathogens
- Screening for mutations or single nucleotide polymorphisms, which more specifically, may be a diagnostic assay
- Cancer screening and diagnosis, via early detection and monitoring of cancer markers
- Screening for drug resistance including chemotherapy treatment resistance
- Research tools

In a preferred embodiment, the CRISPR/Cas13 polypeptide system of the invention can be utilised in an assay to detect genes and mutations associated with cancer, or mutations associated with cancer drug resistance. Thus, the CRISPR/Cas13 polypeptide system of the invention provides low-cost, rapid, multiplexed cancer detection panels for circulating DNA, such as tumour DNA, particularly for monitoring disease recurrence or the development of common resistance mutations.

In an alternative embodiment, the CRISPR/Cas13 polypeptide system of the invention can be utilised in an assay to detect the presence of Coronavirus variants in an assay for diagnosing Covid infection.

The target RNA can therefore be from any biological, environmental or agricultural source including but not limited to water, soil, blood, human and plant tissue, or can be artificially created. In some embodiments, the target RNAs in the sample may be isolated and/or purified and/or amplified prior to contact with either ribonucleoprotein (RNP) complex or the Cas13 polypeptide and the crRNA. Any suitable RNA amplification technique may be used. Similarly, the target RNA in the sample may first be enriched prior to detection or amplification. This enrichment may be achieved by binding of the target nucleic acids by a CRISPR effector system.

The sample may also contain a target DNA. In these circumstances, the DNA can first be reverse transcribed using any suitable technique to produce RNA, and the transcribed RNA detected. If necessary, the DNA can also be isolated and/or purified and/or amplified and/or enriched just using any suitable technique known to the skilled person, prior to reverse transcription.

Multiple target sequences may be contained within the one sample.

In the nucleic acid detection method provided by the present invention, the method may further include determining binding and cleavage of the target RNA by means of a probe or detection molecule ("detector RNA") that is itself cleaved via the trans cleavage activity of the Cas13 polypeptide of the invention.

SHERLOCK and DETECTR are two assays that already exploit the trans cleavage activity of different Cas13 polypeptides, and Cas12a polypeptides respectively. The concept behind them both is similar: reporters sequences/probes are mixed with the sample, such that when the Cas polypeptide within the Cas/crRNA complex is activated after binding to the specific target sequence, the reporter sequences are then indiscriminately cleaved. By using reporter sequences bound to a fluorophore at one end, and a quencher on the other, degradation of the reporter sequence releases fluorophores and results in stable and strong fluorescent signal detected by a fluorimeter. The presence and intensity of the fluorescent signal thus indicates the amount of the target in the biological sample.

In another aspect of the invention there is provided a nucleic acid detection system for detecting a target RNA in a sample, which may come in a kit form, the system comprising:

i) at least one Cas13 polypeptide wherein the Cas13 polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30; or a nucleic acid molecule comprising a sequence encoding the Cas13 polypeptide and

ii) at least one CRISPR RNA (crRNA) or a nucleic acid molecule encoding the crRNA, the crRNA comprising one or more spacers and one or more Cas13-specific direct

repeats, wherein the crRNA is capable of hybridising with one or more target RNA molecules; and preferably

iii) a detector RNA.

Preferably the nucleic acid molecule comprising a sequence encoding the Cas13 polypeptide is selected from the group consisting of SEQ ID NOS: 1-15, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 1-15.

When the detection assay comprises a nucleic acid molecule comprising a sequence encoding the Cas13 polypeptide and a nucleic acid molecule encoding the crRNA, both nucleic acid molecules are preferably expressed from vectors, the vectors comprising:

i) a first regulatory element operably linked to a nucleic acid molecule comprising a sequence encoding a Cas13 polypeptide wherein the Cas13 polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NOS: 16-30, or a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 16-30; and

ii) a second regulatory element operably linked to a nucleic acid molecule encoding a CRISPR RNA (crRNA), the crRNA comprising one or more spacers and one or more Cas13-specific direct repeats, wherein the crRNA is capable of hybridising with one or more target RNA molecules.

Preferably the nucleic acid molecule comprising a sequence encoding the Cas13 polypeptide is selected from the group consisting of SEQ ID NOS: 1-15, or is selected from a sequence being at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NOS: 1-15.

In some embodiments, when there are two or more Cas13 crRNAs, the crRNAs can be provided via a nucleic acid molecule encoding a precursor Cas13 crRNA which can be processed into individual Cas13 crRNAs each of which has a different spacer sequence within it designed against different targets.

The detector RNA can be, for example, a labeled detector RNA such as a fluorescence-emitting dye pair, i.e., a FRET pair and/or a quencher/ fluor pair or RNA molecule generating any other detectable signal after collateral cleavage.

The step of measuring can include one or more of: nanoparticle based detection, fluorescence or chemiluminescent detection, lateral-flow immunochromatography colloid phase transition/dispersion, electrochemical detection, semiconductor-based sensing, and detection of a RNA reporter (RNA molecules carrying a fluorophore and a quencher. Trans-cleavage activity on target detection causes the spatial separation of the fluorophore and the quencher, and the resulting fluorescence signal).

The readout of such detection methods can be any convenient readout. Examples of possible readouts include but are not limited to: a measured amount of detectable fluorescent signal; a visual analysis of bands on a gel (e.g., bands that represent cleaved product versus uncleaved substrate), a visual or sensor based detection of the presence or absence of a color (i.e., color detection method), and the presence or absence of (or a particular amount of) an electrical signal.

The systems, assays/kits and methods disclosed herein may also be adapted for detection of polypeptides (or other molecules) in addition to detection of nucleic acids, via incorporation of a specifically configured polypeptide detection aptamer. Accordingly, in certain example embodiments, the systems, assays/kits and methods may further comprise one or more detection aptamers.

Embodiments disclosed herein can detect both RNA and DNA with comparable levels of sensitivity and can differentiate targets from non-targets based on single base pair differences. Moreover, the embodiments disclosed herein can be prepared in freeze-dried format for convenient distribution and point-of-care (POC) applications.

The nucleic acid detection system of the invention for detecting a target RNA in a sample, and as already mentioned above in the context of the methods of the invention, can utilise two or more Cas13 polypeptides of the invention and two or more Cas13 crRNA in order to target and detect different target sites of the same target RNA (e.g., which can increase sensitivity of detection) by designing the spacer of the crRNA appropriately, and/or can target different target RNA molecules e.g., based on variation such as single-nucleotide polymorphisms etc., and such could be used for

example to detect multiple strains of a virus such as Coronavirus variants, influenza virus variants, HIV variants, and the like.

In some embodiments, two or more Cas13 CRISPR RNAs can be present on an array. For example, a precursor Cas13 crRNA array which can be cleaved into individual Cas13 crRNAs with different spacer sequences in order to bind different targets. In alternative embodiments, two or more Cas13 polypeptides cleaving different RNA probe sequences may be used. For example, one Cas13 polypeptides cleaves polyA RNA probe sequences but not polyU. The other polyprotein can cleave polyU RNA probe but not polyA probes and two or more probes (e.g., with different RNA sequences).

The nucleic acid detection systems and kits of the invention can be applied to any sample being assessed for the presence of a target RNA. Or as already noted, a sample containing DNA that has been reverse transcribed to RNA. The sample can therefore be from any biological, environmental or agricultural source. In some embodiments, the target RNAs in the sample may be isolated and/or purified and/or amplified prior to contact with either ribonucleoprotein (RNP) complex or the Cas13 polypeptide and the crRNA. Any suitable RNA amplification technique may be used. Similarly, the target RNA in the sample may first be enriched prior to detection or amplification. This enrichment may be achieved by binding of the target nucleic acids by a CRISPR effector system.

Multiple target sequences may be contained within the one sample.

The invention also seeks to provide a method of using the nucleic acid detection systems and kits of the invention as described herein for detecting a target RNA in a sample, the method including one or more of the following steps:

- Obtaining a sample if one has not already been supplied
- Processing the sample if need be to isolate any nucleic acids present in the sample
- Purifying and/or enriching and/or amplifying the nucleic acids if need be
- Reverse transcribing DNA from the sample to RNA if need be
- Contacting the sample/nucleic acid with either (i) a pre-formed RNP complex as described herein; or (ii) a Cas13 polypeptide and a crRNA of the invention; or (iii) vectors from which a Cas13 polypeptide and a crRNA of the invention can be expressed

- Adding one or more detector RNAs to the reaction
- Detecting cleavage of the one or more probe/reporter RNAs when the target RNA is present in the sample.

As will be demonstrated in the examples, the Cas13 polypeptides of the invention represent important new sensitive and specific enzymes that will be particularly useful a CRISPR/Cas system in nucleic acid detection assays.

## Examples

### 1. Metagenomics data analysis and computational identification of novel Cas13a, Cas13b, and Cas13d

Method

Metagenome sequencing of camel, cattle, and sheep rumen was performed using Illumina HiSeq 2500 system with 150bp paired-end sequencing. All metagenome sequences of each species were co-assembled using MEGAHIT v1.2.9 with the following options: --k-min 31, --k-max 141, --k-step 10, --min-count 2, and --min- contig-len 200. The FASTAsplitter software (Version 0.2.6) was used to split initial camel, cattle, and sheep rumen metagenomic contigs into files with <50M size. CRISPRone and CRISPRCasFinder 1.1.2 was then used for CRISPR/Cas prediction and region analysis. CRISPRDetect was then used to predict the orientation of the direct repeat in the Cas13 CRISPR array. In order to detect the transcription direction of the CRISPR region, the sequence alignment and secondary structure prediction were performed for CRISPR repeats from all 15 selected Cas13 proteins.

Protein sequence alignment of Cas13s was performed using ClustalW in MEGA11 with default settings. Rumen metagenomics derived Cas13 were compared with previously known (characterized) Cas13a, Cas13b, and Cas13d. The proteins similarity search was performed by BLASTp search against the nonredundant protein database curated by the National Centre for Biotechnology Information (NCBI)(http//blast.ncbi.nlm.nih.gov/blast.cgi).

References for contigs used to discover Cas13 polyproteins:

Gharechahi J, et al. NPJ Biofilms Microbiomes. 2022 Jun 8;8(1):46.

Gharechahi J, et al. ISME J. 2021 Apr;15(4):1108-1120.

Results

a) Identification

Fifteen (15) candidate Cas13 proteins encoding intact CRISPR-Cas13 systems and containing active site domains and residues were selected (Table 2 and phylogenic analysis Figure 1):

Table 2: Novel Cas13 enzymes subtyping and nomenclature

| **Subtype** | **Enzyme of the invention** |
|---|---|
| Cas 13a | Ca1, Ca2, Ca3, Ca4, Ca5, Ca6 and Ca7 |
| Cas13b | Cb8, Cb9, Cb10, Cb11 and Cb12 |
| Cas13d | Cd13, Cd14 and Cd15. |

A Blast search against the non-redundant protein database curated by NCBI showed that all of the Cas13a enzymes of the invention showed less than 60 percent identity to other Cas13a proteins deposited in the data, most of the Cas13b enzymes of the invention showed less than 40-50 percent identity to proteins deposited in this database, and all 3 of the Cas13d enzymes of the invention showed less than 50 percent identity to proteins deposited in this database (Table 3).

This demonstrates how different the Cas13 polypeptides of the invention are from those already known in the art.

Table 3: The % identity of novel Cas13 proteins of the invention to proteins deposited in GenBank. The Expect value (E) is a parameter that describes the number of hits one can "expect" to see by chance when searching a database of a particular size.

| Subtype | SEQ ID NO: | ID | Contig ID | Protein length | % Identity | E value |
|---|---|---|---|---|---|---|
| *A* | 16 | Ca1 | k127_1867445 | 1338 | <56% | $2e^{-11}$ |
| | 17 | Ca2 | k127_4200118 | 1382 | <40% | 0.0 |
| | 18 | Ca3 | k127_751200 | 1307 | <39% | 0.0 |
| | 19 | Ca4 | k127_5935133 | 1319 | <59% | 0.0 |
| | 20 | Ca5 | k141_14579520 | 1406 | <53% | 0.0 |
| | 21 | Ca6 | k141_10995992 | 1341 | <49% | 0.0 |
| | 22 | Ca7 | k141_12677984 | 1302 | <44% | $7e^{-05}$ |

| | | | | | | |
|---|---|---|---|---|---|---|
| *B* | 23 | Cb8 | k127_4804511 | 1313 | <68% | $7e^{-148}$ |
| | 24 | Cb9 | k127_1483864 | 1125 | <32% | $1e^{-144}$ |
| | 25 | Cb10 | Cas13/21_contig-18_616 | 1215 | <79% | 0.0 |
| | 26 | Cb11 | k141_16137484 | 1380 | <35% | $1e^{-176}$ |
| | 27 | Cb12 | k127_333529 | 1246 | <45% | $3e^{-07}$ |
| *D* | 28 | Cd13 | K127_2411982 | 921 | <44% | 0.0 |
| | 29 | Cd14 | K141_15335538 | 956 | <44% | $2e^{-63}$ |
| | 30 | Cd15 | Cas13/23_contig-81_4932 | 948 | <43% | $1e^{-60}$ |

b) crRNA modelling

Sequence alignment and secondary structure modelling was performed for the CRISPR repeats (ie the direct repeats) of the crRNA for all 15 selected Cas13 polypeptides to identify the transcription direction of CRISPR region. All the repeats in the crRNA for the Cas13a and Cas13d proteins formed the hairpin at the 5' end just before the spacer and all of the crRNA for the Cas13b proteins formed the hairpin at the 3' end.

In the following tables, a number of crRNAs (SEQ ID NOS: 32-54) with direct repeats of 35-37 nucleotides (underlined), are provided for each Cas13 polyprotein of the invention. The consensus direct repeat of the crRNA for each is provided in Table 4A; exemplary variants of the direct repeat is provided in Tables 4B to 4L. All are provided as the coding DNA sequence. As would be understood, the spacer will need to be designed for the target RNA; in Table 4A, an exemplary spacer sequence is used which is designed against an artificial target sequence.

Table 4A: List of crRNAs with consensus direct repeat; LWA = *Leptotrichia wadei* (LwaCas13a) as reference sequence. All sequences given as the coding DNA sequence.

| SEQ ID NO: | **ID** | **Direct repeat + exemplary spacer (listed as the coding DNA sequence)** | **Length** |
|---|---|---|---|
| | Cas13a | | |
| 31 | LWA | GATTTAGACTACCCCAAAAACGAAGGGGACTAAAACTAGATTGCTGTTCTACCAAGTAATCCAT | 64 |
| 32 | Ca1 | GTGGCATGAAAAAAGCCCGACATAGCGGGCAATCACTAGATTGCTGTTCTACCAAGTAATCCAT | 64 |
| 33 | Ca2 | GTAGAAAAGAAGATAGTCCAACATAGTGGATAATCATAGATTGCTGTTCTACCAAGTAATCCAT | 64 |
| 34 | Ca3 | GGAGATGAAAAAAGCCCGACATAGCGGGCAATCGAATAGATTGCTGTTCTACCAAGTAATCCAT | 64 |
| 35 | Ca4 | GTTAAAAGAAAACAGCCCGACATAGCGGGCGATAACTAGATTGCTGTTCTACCAAGTAATCCAT | 64 |
| 36 | Ca5 | GAATTGGAGAAGATCCCGAGAAAGTGGGAAATAACTAGATTGCTGTTCTACCAAGTAATCCAT | 63 |

| | | | |
|---|---|---|---|
| 37 | Ca6 | GTTTGGAAAACAGCCCGACATAGAGGGCAATAGACTAGATTGCTGTTCTACCAAGTAATCCAT | 63 |
| 38 | Ca7 | GTTAGATGAGAACACTCCGAGATAACGGAGAATAACTAGATTGCTGTTCTACCAAGTAATCCAT | 64 |
| | Cas 13b | | |
| 39 | Cb8 dir | TAGATTGCTGTTCTACCAAGTAATCCATATGTTGTCATACCCATCCAAACGATAGGCTTCTACAAC | 66 |
| 40 | Cb8 rev | TAGATTGCTGTTCTACCAAGTAATCCATATGTTGTAGAAGCCTATCGTTTGGATGGGTATGACAAC | 66 |
| 41 | Cb9 dir | TAGATTGCTGTTCTACCAAGTAATCCATATGTTGTAAAGACCTTCATTTCGGAAGGAAGAGACAAC | 66 |
| 42 | Cb9 rev | TAGATTGCTGTTCTACCAAGTAATCCATATGTTGTCTCTTCCTTCCGAAATGAAGGTCTTTACAAC | 66 |
| 43 | Cb10 dir | TAGATTGCTGTTCTACCAAGTAATCCATATGTTGTAGAAGCCCCTTCTTCGTGGGGTAAGTGCAAC | 66 |
| 44 | Cb10 rev | TAGATTGCTGTTCTACCAAGTAATCCATATGTTGCACTTACCCCACGAAGAAGGGGCTTCTACAAC | 66 |
| 45 | Cb11 dir | TAGATTGCTGTTCTACCAAGTAATCCATATGTTGTAGAAGCCTATCGTTTGGATAGGTATGACAAC | 66 |
| 46 | Cb11 rev | TAGATTGCTGTTCTACCAAGTAATCCATATGTTGTCATACCTATCCAAACGATAGGCTTCTACAAC | 66 |
| 47 | Cb12 dir | TAGATTGCTGTTCTACCAAGTAATCCATATGCTGTGAACTCCTGCCGAAATGGCAGGCCGAACAGC | 66 |
| 48 | Cb12 rev | TAGATTGCTGTTCTACCAAGTAATCCATATGCTGTTCGGCCTGCCATTTCGGCAGGAGTTCACAGC | 66 |
| | Cas13d | | |
| 49 | Cd13 dir | GGTTTCAGACCCTTACAAAAAGGGTGTAGTACGTTTCTAGATTGCTGTTCTACCAAGTAATCCAT | 65 |
| 50 | Cd13 rev | GAAACGTACTACACCCTTTTTGTAAGGGTCTGAAACCTAGATTGCTGTTCTACCAAGTAATCCAT | 65 |
| 51 | Cd14 dir | GTTTCAGAACCCTGTAATTTGACAGGGTTGTAGTTGTAGATTGCTGTTCTACCAAGTAATCCAT | 64 |
| 52 | Cd14 rev | CAACTACAACCCTGTCAAATTACAGGGTTCTGAAACTAGATTGCTGTTCTACCAAGTAATCCAT | 64 |
| 53 | Cd15 dir | GATCTATAACCCTGCATTTATGTAGGGCTCTAAAACTAGATTGCTGTTCTACCAAGTAATCCAT | 64 |
| 54 | Cd15 rev | GTTTTAGAGCCCTACATAAATGCAGGGTTATAGATCTAGATTGCTGTTCTACCAAGTAATCCAT | 64 |

Table 4B: Cas13 Ca2

- DR consensus : GTAGAAAAGAAGATAGTCCAACATAGTGGATAATCA (SEQ ID NO:86)

| SEQ ID NO: | ts |
|---|---|
| 55 | AGTCCAACATAGTGAATAATCA |
| 56 | AGTCCAACATAGCGGATAATCA |
| 57 | AGTCCAACACAGTGGATAATCA |
| 58 | AGTCCAACATAGTGGACACTCA |
| 59 | AGTCCAACATGGTGGATAATCA |
| 60 | GTAGAAATGAGGATAGTCCAACATAGCAACTAATTA |

Table 4C: Cas13 Ca3

- DR consensus : GGAGATGAAAAAAGCCCGACATAGCGGGCAATCGAA (SEQ ID NO: 94)

| SEQ ID NO | ts |
|---|---|
| 61 | CCCGACATAGCGGGCAATCGAAAT |
| 62 | GGGGATGAAAAAAGCCCGACATAGCGGGCAATCGAACT |

| | |
|---|---|
| 63 | CCCGACATAGCGGGCAATCGAACT |
| 64 | GGAGATGAAAAAAGTCCAACATTATGAAGACTATAGAG |

Table 4D: Cas13 Ca4

- DR consensus : GTTAAAAGAAAACAGCCCGACATAGCGGGCGATAAC (SEQ ID NO:87)

| **SEQ ID NO** | |
|---|---|
| 65 | CCGACATAGTGGGCGATAAC |
| 66 | CCGACATAGCGGGCGATAAC |
| 67 | GTTAAAAGAAAACAGCCCGACATAACGAGCGATAAC |

Table 4E: Cas13 Ca5

- DR consensus : GAATTGGAGAAGATCCCGAGAAAGTGGGAAATAAC(SEQ ID NO: 95)

| **SEQ ID NO** | |
|---|---|
| 68 | CCGAGAAAGTGGGAAATAAC |
| 69 | GAATTGGAGAAGATCCCGAGATTAAACACAGAAAA |

Table 4F: Cas13 Ca6

- DR consensus : GTTTGGAAAACAGCCCGACATAGAGGGCAATAGAC (SEQ ID NO:96)

| **SEQ ID NO** | **ts** |
|---|---|
| 70 | GTTTGGAAAACAGCCCGACAAAGAGGGCAATAGAC |

Table 4G: Cas13 Cb9

- DR consensus : GTTGTAAAGACCTTCATTTCGGAAGGAAGAGACAAC (SEQ ID NO: 97)

| **SEQ ID NO** | |
|---|---|
| 71 | TTTTGGAAGGAAGAGACAAC |
| 72 | GTTGTAAAGACCTTCATTTTGGAAGGAGGAGACATC |

Table 4H: Cas13 Cb10

- DR consensus : GTTGTAGAAGCCCCTTCTTCGTGGGGTAAGTGCAAC (SEQ ID NO: 98)

| **SEQ ID NO** | **Direct Repeat variants** |
|---|---|
| 73 | GTTGTAGAAGCCCCTTCTTTGTGGGGTAGGTGCAAC |

| | |
|---|---|
| 74 | GTTGTAGAAGCCCCTTCTTTGTGGGGTAAGTGCAAC |
| 75 | GTTGTAGAAGCCCCTTCTTTGTGGGGTATGTGCAAC |

Table 4I: Cas13 Cb11

- DR consensus : GTTGTAGAAGCCTATCGTTTGGATAGGTATGACAAC (SEQ ID NO: 99)

| SEQ ID NO | Direct Repeat variants |
|---|---|
| 76 | CGTGTGGATAGGTATGACAAC |
| 77 | GTTGTAGAAGCCTATCGTTTGGGTAGGTACGACAAA |

Table 4J: Cas13 Cd13

- DR consensus : GGTTTCAGACCCTTACAAAAAGGGTGTAGTACGTTTC (SEQ ID NO: 100)

| SEQ ID NO | |
|---|---|
| 78 | AAAGGGTGTAGTACGTTTCT |
| 79 | AAAGGGTGTAGTACGTTTCG |
| 80 | AAAGGGTGTAGTACAAAACA |
| 81 | AATGGGTGTAGTACAAAACC |
| 82 | TTGTCTTTTCCCAACAAAAAAGGGTGTAGTACGTTTC |

Table 4K: Cas13 Cd14

- DR consensus : GTTTCAGAACCCTGTAATTTGACAGGGTTGTAGTTG (SEQ ID NO:101)

| SEQ ID NO | |
|---|---|
| 83 | TCTAACAGGGTTGTAGTTG |
| 84 | GTTTCAGAATCCTGTAATTTGACAGGGTTGTAGTTG |

Table 4L: Cas13 Cd15

- DR consensus : GATCTATAACCCTGCATTTATGTAGGGCTCTAAAAC (SEQ ID NO:102)

| SEQ ID NO | Direct Repeat variants |
|---|---|
| 85 | GATCTATAACCCTGCATTTATGCAGGGCTCTAAAAT |

**2. Expression and purification of Cas13 enzymes**

Cas13 sequences were N-terminally tagged with a His6-MBP-TEV (His6, six-histidine affinity tag; MBP, maltose binding protein to enhance solubility; TEV, TEV

protease recognition site) and C-terminally tagged with enterokinase (EK) cleavage site (EK-His6). His6-MBP-TEV-Cas13-EK-His6 were synthesised and cloned between *Nde*I and *Not*I restriction endonuclease site of pET21a(+) by Genscript (USA). After dilution of lyophilized synthesis vector and agarose gel analysis, 50ng of His6-MBP-TEV-Cas13-EK-His6 pET21a(+) plasmid was transformed into *E. coli* host strain BL21 (DE3).

First, the expression was performed at a small (20 ml Terrific broth (TB)) scale in 100cc flask to screen optimum conditions for soluble protein expression.

Cells were grown aerobically in a shaker at 37°C at 180rpm up to an optical density of 0.6-0.8, before inducing Cas13 production with inductor concentrations (400 µM of IPTG) in an overnight culture at (16 and 18°C). Then, the optimum conditions determined previous step were used to grow a large amount of biomass in 1-4 L of Terrific broth (TB) media. After induction, cells were pelleted and re-suspended in lysis buffer (50 mM Tris-HCl pH 7, 500 mM NaCl, 5% glycerol and 1 mM TCEP, 0.5 mM PMSF, EDTA-free protease inhibitor (Roche)). 0.1-1 mg/ml lysozyme and 500mg-1G/Lit protamine sulphate were added, then incubated on ice for 30 min and mixed 2–3 times by gently swirling the cell suspension. Cells were then sonicated (60% Amplitude, 0.5 Cycle, Sonication Pulse Rate: 360 seconds ON) on ice. Cell debris was removed by centrifugation at 14,000g for 30 min at 4°C.

Cell debris was removed by 0.45 micron filter. Then the filtered cell lysate was loaded into the column including 1cc Ni-NTA resin (Sigma) per 10cc filtered cell (equivalent to 1/10 of the volume of cell supernatant), and subsequently washed with 4ml wash buffer (50 mM sodium phosphate pH 8, 300 mM NaCl, and 10 mM Imidazole). The protein was then eluted with 2 mL elution buffer each (50 mM sodium phosphate pH 8, 300 mM NaCl, 300 mM Imidazole and 1 mM TCEP). Lysate, flow-through, wash and elution fractions were collected individually and analysed on a 10% SDS gel.

Proteins in 2-14 ml elute fraction were concentrated with 3 or 30k Amicon filter (Merck Millipore). Concentrated proteins were incubated with TEV and EK proteases at 4 °C overnight while dialyzing into ion exchange buffer (50 mM Tris-HCl pH 7.0, 250 mM KCl, 5% glycerol, 1 mM TCEP) in order to cleave off the N-terminal His6-MBP and C-terminal His6 tags, respectively. Cleaved Cas13d enzymes were loaded onto Amicon Ultra-0.5ml centrifugal Filter Ultracel-50k (50000NMWL)(Merck Millipore) and cleaved

Cas13a and Cas13b enzymes were loaded onto micon Ultra-15 centrifugal Filter Ultracel-100k (100000NMWL)(Merck Millipore) centrifugal filters. Concentrated proteins were diluted with storage buffer (20 mM Tris-HCl pH 7.0, 200 mM KCl, 5% glycerol, 1 mM TCEP) for subsequent enzymatic assays.

The concentrated factions were analysed on a 10% SDS gel.

**3. Nucleic Acid Preparation**

Artificial target sequence with either a C, G, A or U protospacer flanking sequence (PFS) were designed. DNA oligo templates for T7 transcription were synthesised and cloned in pUC57 by GenScript (Table 5).

**Table 5. DNA fragments for production of RNA substrates (ie target sequences)**

| Name | Sequence |
|---|---|
| ssRNA 1 (C PFS) | GGCCAGTGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAAATATGGATTACTTGGTAGAAC AGCAATCTA**C**TCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGT GTTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAG (SEQ ID NO: 88) |
| ssRNA 1 (G PFS) | GGCCAGTGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAAATATGGATTACTTGGTAGAAC AGCAATCTA**G**TCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGT GTTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAG (SEQ ID NO: 89) |
| ssRNA 1 (A PFS) | GGCCAGTGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAAATATGGATTACTTGGTAGAAC AGCAATCTA**A**TCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGT GTTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAG (SEQ ID NO: 90) |
| ssRNA 1 (U PFS) | GGCCAGTGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAAATATGGATTACTTGGTAGAAC AGCAATCTA**T**TCGACCTGCAGGCATGCAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGT GTTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAG (SEQ ID NO: 91) |

Phusion™ High-Fidelity PCR Master Mix (2X) (Thermo Fisher Scientific) with forward (TAATACGACTCACTATAG (SEQ ID NO:92)) and reverse (CTTTATGCTTCCGGCTCG (SEQ ID NO:93)) primers were used for amplification of template DNA. PCR assembly reaction in a 25-µL volume was set up according to manufacturer's instructions and using the cycling parameters according to manufacturer's recommendations, with an annealing temperature of 55°C and 35 cycles. The integrity of amplified sequence was checked by Sanger sequencing. PCR product was run on 2% agarose gel and template band was recovered from gel by GeneJET™ Gel Extraction Kit (Thermo Fisher Scientific). 3µl of gel recovered DNA was run on 2% agarose gel. The concentration of purified DNA was determined by Picodrop and Qubit Fluorometer.

TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher Scientific) was used for ssRNA template synthesis using the T7 RNA polymerase forward primer.

The ssRNA reaction components were combined at room temperature according to manufacturer's instructions using 1µg of template, mixed thoroughly, and then centrifuged briefly to collect all drops and incubated at 37°C for 4 h in water bath. DNase I digestion directly after the IVT reaction was performed to prevent the template DNA from interfering with downstream applications of the RNA transcript. 1µL of DNase I (at 1 U/µL; included in the kit) was added into the reaction mix immediately after the IVT reaction and incubate at 37°C for 15 minutes. 0.5µL of the IVT product was diluted in 10µL of DEPC-treated water and 10µL of the diluted sample was mixed with 10µL of 2X RNA Loading Dye Solution of the TranscriptAid T7 High Yield Transcription Kit and incubated at 70°C for 10 minutes and then was chilled on ice prior to loading. Samples were run on a 2% Agarose Gel against an RNA Ladder.

RNA purification from IVT reaction was performed using GeneJET RNA Cleanup and Concentration kit (Thermo Fisher Scientific) as instructed by manufacturer. The purified RNA was stored at -80°C until use. 0.5 µL of the purified IVT product was diluted in 10 µL of DEPC-treated water and 10 µL of the diluted sample was mixed with 10 µL of 2X RNA Loading Dye Solution and was heated the sample at 70°C for 10 minutes and was chilled on ice prior to loading on a 2% Agarose Gel against an RNA Ladder.

**4. In Vitro Cis Cleavage Assays and activity of CRISPR-Cas13**

The ability of the Cas13 proteins of the invention to exhibit cis ssRNAse activity was assessed.

Method

All crRNAs (Table 4A) used in this study were synthesised by GenScript. The cleavage reactions of Cas13 enzymes were performed at 37°C with the in vitro-transcribed RNA targets shown in Table 5.

Briefly, cleavage reactions were carried out in 20µL reaction volume with 100 nM Cas13 protein of the invention, 50 nM crRNAs (unless otherwise indicated), and 1000 nM in vitro-transcribed target RNA in a 1× cleavage buffer (Buffer 1, Table 7); the reactions were then incubated at 37 °C for 1 h (unless otherwise indicated). The samples were then boiled at 70°C for 3 min in a 2× RNA Loading Dye (NEB) and cooled down on the ice for 3 min before loading onto a 10% polyacrylamide- urea denaturing gel. The gel was

stained with the SYBR Gold Nucleic Acid Gel Stain (Thermo Fisher Scientific) and visualized using a Bio-Rad Molecular Imager Gel Doc system.

Results

The corresponding gene sequence of the Cas13 proteins was synthesised and the proteins expressed in Escherichia coli. The proteins were purified and their *in vitro* cis catalytic activities tested. The *in vitro* cleavage activity of the Cas13 proteins of the invention with the respective crRNAs listed in Table 4 targeting the single-stranded RNA substrates of Table 5 harbouring target sequences complementary to the crRNA spacers was evaluated. The activity of Cas13 proteins was evaluated by denaturing gel showing the targeted *in vitro* RNase cleavage activity of the Cas13 proteins of the invention when incubated with the ssRNA target and different crRNAs. The ssRNA cleavage activity of 15 Cas13 enzymes, all of them showed at least some ssRNA cleavage activity.

**5. *In vitro* trans cleavage assays and collateral ssRNAse activity of Cas13 polypeptides**

The ability of the Cas13 proteins of the invention to induce collateral (ie trans) ssRNAse activity was assessed.

Methods

Collateral cleavage assays of Cas13 enzymes were performed in a 20 µL final reaction volume. Cas13 proteins and crRNAs according to Table 4 were first assembled in to ribonucleoprotein (RNP) complexes by mixing 50 nM purified Cas13 with 25 nM crRNA in the 1× cleavage buffer (Buffer 1, Table 7);) followed by incubation at 37 °C for 60 min (unless otherwise indicated).

Next, the assembled RNP was combined on ice with 2µL of 250 nM in vitro-transcribed target RNA (Table 5) and 250 nM RNA reporter (UUUUU) (Integrated DNA Technologies) (unless otherwise indicated), and reactions were incubated for 1 h at 37 °C (unless otherwise indicated). Real-time or end-point fluorescence measurements were collected on a microplate reader (Synergy HTX Multi-Mode Reader) or ABI Real-time PCR (Applied Biosystems, CA, USA) at 2 min intervals for real-time measurements. To

allow comparisons between different conditions, fluorescence for background conditions (no target ssRNA) were subtracted from samples to generate background subtracted fluorescence.

Results

Reactions were performed in the presence of fluorescently labelled ssRNA probe (UUUUU) and equally mixed ssRNA targets (ie a mixture of the 4 ssRNAs of Table 5). It is possible to predict the direct repeat direction only for Cas13a proteins; for Cas13b and Cas13d it should be determined experimentally. Accordingly, two crRNA variants (direct (dir) or reverse (rev)) were designed and tested. Trans cleavage (collateral activity) was detected by all proteins; the measurement of real-time background subtracted fluorescence output for a subset of them (Cas13d13, Cas13d14 and Cas13d15) are presented in Figure 2.

The results showed that Cas13d13 and Cas13d14 were functional using reverse crRNA whereas Cas13d15 showed collateral cleavage activity using direct crRNA.

**Optimisation of collateral ssRNAse activity of Cas13 polypeptides**

In order to develop a nucleic acid detection platform based on the CRISPR/Cas13 system of the invention, particular elements of the platform and/or assay can be optimised. The following examples demonstrate the elements of a platform that can optionally be optimised for each of the Cas13 polypeptides of the invention, and how to do so.

**6. Probe Types**

The length and sequence of the ssRNA probes/reporter sequences may also influence Cas13 trans-cleavage activities. Various ssRNA probes containing FAM at 5' end and 3IABkFQ quencher at the 3' end (Table 6) were tested. When Cas13 cleaves ssRNA, an increase of fluorescence signal is observed. The experimental conditions are is described in Figure 9. The fluorescence was measured in 2 min intervals. The measurement of real-time background subtracted fluorescence output are shown in the graphs as means ± SD (n = 3).

Table 6: Probes/reporters tested for their effect on Cas13 collateral activity

| Reporter RNAs | Probe type |
|---|---|
| FAM------3IABkFQ | /56-FAM/rUrUrUrUrU/3IABkFQ/ |
| FAM------3IABkFQ | /56-FAM/TArUrUGC /3IABkFQ/ |
| FAM------3IABkFQ | /56-FAM/rUrG rArCrG rU/3IABkFQ/ |
| FAM------3IABkFQ | /56-FAM/rArA rArArA /3IABkFQ/ |
| FAM------3IABkFQ | /56-FAM/rNrN rNrNrN /3IABkFQ/ |

Results

Cas13 polypeptides of the invention showed different activities with various probes (Figure 4). Cas13a3, Cas13a7, Cas13d13, Cas13d14 and Cas13d15 preferred probes rUrUrUrUrU, rNrNrNrNrN, rArArArArA, rUrUrUrUrU, and rArA rArArA, respectively. The results for a subset of them (Cas13a3, Cas13a7, Cas13d13, Cas13d14 and Cas13d15) are presented in Figure 3.

**7. PFS analysis of Cas13 polypeptides**

Since the activity of Cas13 enzymes may depend on the presence or absence of a PFS in the target sequence, and they may be more or less sensitive to one PFS over another, experiment 6 was repeated but using the single RNA substrates of Table 5 rather than the initial experiment that tested all 4 RNA substrates at once.

The ability of Cas13a3, Cas13a7, Cas13d13, Cas13d14 and Cas13d15 to cleave a target ssRNA with each possible protospacer flanking site (PFS) nucleotide (A, U, C or G) was assayed.

Method

Test conditions as per Figure 9, comprising 50 nM crRNA, 100 nM Cas13, 500 250 nM target ssRNA, 250 nM reporter and 1× buffer (20 mM HEPES-Na pH 6-8, 50 mM NaCl, 10 mM MgCl2, and 1 mM TCEP Buffer 1, Table 7). The fluorescence was measured in 2 min intervals. End-point activity of enzymes Values are shown in the graphs as means ± SD (n = 3).

Results

Cas13a3 can also robustly cleave a target with A, C, or G PFS, with less activity on the ssRNA with a U PFS. Cas13a7 can also robustly cleave a target with A, U, or C PFS, with less activity on the ssRNA with a G PFS. Cas13d13, Cas13d14 and Cas13d15 can robustly cleave all four targets with slightly less activity on the ssRNA with a A, C, and U PFS, respectively. The results for a subset of them (Cas13a3, Cas13a7, Cas13d13, Cas13d14 and Cas13d15) are presented in Figure 4.

**8. Buffers and pH**

For the optimisation of buffer compositions, a variety of buffers (Table 7) and pH ranges were tested. The procedure was as described in Figure 9.

a) Buffer

First, the buffers in Table 9 were tested:

Table 7: Components of buffers used for optimisation

| **Buffer 1** | **Buffer 2** | **Buffer 3** | **Buffer 4** |
|---|---|---|---|
| 20 mM HEPES | 20 mM HEPES | 20 mM HEPES | 20 mM HEPES |
| pH7.0 | pH7.0 | pH6.8 | pH6.8 |
| 50 mM NaCl | 50 mM NaCl | 60 mM NaCl | 50 mM KCl |
| 10 mM MgCl2 | 10 mM MgCl2 | 6 mM MgCl2 | 5 mM MgCl2 |
| 5% glycerol | 5% glycerol | 5% glycerol | 5% glycerol |
| - | 1µg/ml BSA | - | - |

Results

The results from the comparison of the Cas13 reaction buffers are presented in Figure 6 for a representative sample of cleavage reactions using Cas13a3, Cas13a7, Cas13d13, Cas13d14 and Cas13d15. End-point activity of enzymes are shown in the graphs as means ± SD (n = 3).

Cas13d13, Cas13d14 and Cas13d15 generally preferred buffer 2 and exhibited the highest collateral activities in the presence of Buffer 2, whereas Cas13a3 and Cas13a7 showed the highest activity in the presence of Buffer 4.

b) pH

As the *pH* values may influence the Cas13 *trans*-cleavage activities, the cleavage assay was repeated in the buffer determined to be the best for each of Cas13a3, Cas13a7, Cas13d13, Cas13d14 and Cas13d15 (see Figure 5) but with *pH* values ranging from 4.8 to 8.8 in increments of 1.0.

Results

The results from the comparison of the pH ranges are presented in Figure 6 for a representative sample of cleavage reactions using Cas13a3, Cas13a7, Cas13d13, Cas13d14 and Cas13d15. End-point activity of enzymes are shown in the graphs as means ± SD (n = 3).

Cas13a3, Cas13a7, Cas13d14 and Cas13d15 had the highest collateral activity at the pH 6.8, whereas Cas13d13 demonstrated the highest activity at alkaline pH 8.8.

**9. Probe concentration**

Since the fluorescent probe is another key component that influences the reaction, we optimized the assay by incubating increasing amounts of probes with constant concentrations of Cas13-crRNA and target ssRNA with the optimum buffer and pH for each Cas13 enzyme.

Method

The procedure was as described in Figure 9. The cleavage assay was performed in the buffer and pH using probe sequences determined to be the best for each of Cas13a3, Cas13a7, Cas13d13, Cas13d14 and Cas13d15 (see Figure 5 and 6) but with probe concentration of 125, 250, 500, 100, 2000 and 4000 nM.

Results

The results from the comparison of the probe concentrations are presented in Figure 7 for a representative sample of cleavage reactions using Cas13a3, Cas13a7, Cas13d13, Cas13d14 and Cas13d15. The measurement of real-time background subtracted fluorescence output are shown in the graphs as means ± SD (n = 3). The

results demonstrated that CRISPR-mediated fluorescence signal intensities increased with increasing amounts of probes (from 125 nM to 4000 nM).

**Summary of optimisation experiments**

Figure 9 presents the results from examples 6 to 9 for a representative sample of cleavage reactions using Cas13a3, Cas13a7, Cas13d13, Cas13d14 and Cas13d15.

Initial test: this is example 5 which confirms trans cleavage (collateral activity), and the direct repeat direction for the Cas13d polypeptides tested (see column labelled crRNA type). This information is taken in to the next example: probe types.

Probe types: this is example 6 and confirms the preference for probe type of each of Cas13a3, Cas13a7, Cas13d13, Cas13d14 and Cas13d15 (see column labelled probe type). The preferred probe is used in the next example: PFS test.

PFS test: This is example 7 and assesses whether each of the activity of Cas13a3, Cas13a7, Cas13d13, Cas13d14 and Cas13d15 depend on the presence or absence of a PFS in the target sequence, and if so, if they may be more or less sensitive to one PFS over another (see column labelled PFS). If there is a dependence on a particular PFS that is used in the next experiment; if there is no dependence on, or preference for a PFS, any PFS or mixtures thereof can be used in subsequent examples: Buffer types.

Buffer types: This is example 8a and assesses which buffer gives the best cleavage activity for each of Cas13a3, Cas13a7, Cas13d13, Cas13d14 and Cas13d15 (see column labelled buffer). The next example is conducted using the preferred buffer: pH test.

pH: this is example 8b and assesses the best buffer from the previous example at a range of pHs (see column labelled pH). The next experiment is conducted in the preferred buffer at the preferred pH: probe concentration.

Probe concentration: this is example 9, to optimise how much probe to use in each reaction (see column labelled Probe).

By the end of this process, which can be applied to any Cas13 polyprotein, the optimal conditions for a representative subset of Cas13 polypeptides of the invention have been determined as set out in Table 8:

Table 8: The optimal conditions for a representative subset of Cas13 polypeptides

| Cas13 protein | Direction | PFS | Probe type | Buffer | pH |
|---|---|---|---|---|---|
| Cas13a3 | NA | G=A>C>U | rNrN rNrNrN | 4 | 6.8 |
| Cas13a7 | NA | A=U>C>G | rNrN rNrNrN | 4 | 6.8 |
| Cas13a13 | Rev | U=C=G>A | rArA rArArA | 2 | 8.8 |
| Cas13a14 | Rev | G=A=U>C | rUrU rUrUrU | 2 | 6.8 |
| Cas13a15 | Dir | A=G=C>U | rArA rArArA | 2 | 6.8 |

**10. Limit of Detection (LoD) experiments**

An important feature of a nucleic acid detection platform is the detection sensitivity. The detection sensitivities of the Cas13 polypeptides were therefore investigated. Having determined the optimal reaction conditions for the Cas13 polypeptides of the invention as per Table 8, different concentrations of target RNA were tested to evaluate the sensitivity of the detection system.

Methods

To determine the LoD for each Cas13 enzyme, a 20µL detection system was prepared that consisted of 100 nM Cas13, 50 nM crRNA, 1000nM ssRNA reporter, and different concentrations of target RNA (between 1nM and 0.0001nM in 10-fold dilutions), for each enzyme in the corresponding reaction buffer. Fluorescence readouts were taken in 2 min intervals for a total of 60 min on the ABI Real-time PCR (Applied Biosystems, CA, USA).

Results

As shown in Figure 7, the fluorescence signal intensities of samples composed of 1 nM, 100 pM, 10 pM, 1 pM, and 100 fM target ssRNA were analysed. Cd13 and Cd14 could specifically and stably detect 1 pM target RNA; Ca3, Ca7 and Cd15 could detect up to 10pM.

Altogether, these data indicated that the identified Cas13 polypeptides are catalytically active with a robust trans cleavage activity and good detection sensitivity, and thus are suitable for developing nucleic acid detection platforms.

**11. Covid detection**

The ability of Cas13 polypeptides to rapidly detect nucleic acids with high sensitivity aids in disease diagnosis and monitoring, epidemiology, and general laboratory tasks. To address this, the cis and trans activities *in vitro* of the Cas13 polypeptides of the invention were evaluated for use in a Cas13 SARS-CoV-2-based detection assays. Furthermore, two different crRNAs for Cas13d14 were designed, targeting two different regions in the SARS-CoV-2 nucleocapsid (N) gene.

Methods

Fragment of DNA template of SARS-CoV-2 N gene (Target) (Broughton JP, et al. CRISPR-Cas12-based detection of SARS-CoV-2. Nat Biotechnol. 2020 Jul;38(7):870-874) for T7 in-vitro transcription (IVT) was synthesised and cloned in pUC57 by GenScript (Table 9). The amplification of N gene DNA oligo template were performed using forward and reverse primers as described in Example 3. TranscriptAid T7 High Yield Transcription Kit (Thermo Fisher Scientific) was used for ssRNA template (Target) synthesis. After DNaseI treatment of IVT reaction, RNA purification from IVT reaction were performed as described in Example 3.

**Table 9.** Sense DNA template of SARS-CoV-2 N gene for production of RNA substrates (ie target sequences)

| Name | Sequence |
|---|---|
| Fragment of SARS-CoV-2 N gene | 5´CCAAATTGGCTACTACCGAAGAGCTACCAGACGAATTCGTGGTGGTGACGGTAAA ATGAAAGATCTCAGTCCAAGATGGTATTTCTACTACCTAGGAACTGGGCCAGAAGCT GGACTTCCCTATGGTGCTAACAAAGACGGCATCATATGGGTTGCAACTGAGGGAGC CTTGAATACACCAAAAGATCACATTGGCACCCGCAATCCTGCTAACAATGCTGCAAT CGTGCTACAACTTCCTCAAGGAACAACATTGCCAAAAGGCTTCTACGCAGAAGGGAG CAGAGGCGGCAGTCAAGCCTCTTCTCGTTCCTCATCACGTAGTCGCAACAGTTCAAG AAATTCAACTCCAGGCAGCAGTAGGGGAACTTCTCCTGCTAGAATGGCTGGCAATG GCGGTGATGCTGCTCTTGCTTTGCTGCTGCTTGACAGATTGAACCAGCTTGAGAGCA AAATGTCTGGTAAAGGCCAACAACAACAAGGCCAAACTGTCACTAAGAAATCTGCTG CTGAGGCTTCTAAGAAGCCTCGGCAAAAACGTACTGCCACTAAAGCATACAATGTAA CACAAGCTTTCGGCAGACGTGGTCCAGAACAAACCCAAGGAAATTTTGGGGACCAG GAACTAATCAGACAAGGAACTGATTACAAACATTGGCCGCAAATTGCACAATTTGCC CCCAGCGCTTCAGCGTTCTTCGGAATGTCGCGCATTGGCATGGAAGTCACACCTTC GGGAACGTGGTTGACCTACACAGGTGCCATCAAATTGGATGACAAAGATCCAAATTT CAAAGATCAAGTCATTTTGCTGAATAAGCATATTGACGCATACAAAACATTCCCACCA ACAGAGCCTAAAAAGGACAAAAAGAAGAAGGCTGATGAAACTCAAGCCTTACCGCA GAGACAGAAGAAACAGCAAACTGTG-3´ (SEQ ID NO:103) |

Short 81 or 82-bp dsDNA made by annealing of complementary oligos were used to generate crRNAs dsDNA coding sequence (Fozouni P, et al. Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy. Cell. 2021 Jan 21;184(2):323-333). First, concentrated complementary oligonucleotides (Table 10) were mixed together at a 1:1 molar ratio in a microcentrifuge tube and were diluted to a final concentration of 1 pmol/µl with a Tris or phosphate buffer ( 10 mM Tris,

1 mM EDTA, 50 mM NaCl (pH 8.0) or 100 mM sodium phosphate, 150 mM NaCl, 1 mM EDTA (pH 7.5)). To generate crRNA dsDNA, complementary oligonucleotides were annealed using thermocyler with a denaturation step of 95 ◦C for 5 min; then, the temperature gradually decreased by 1 degree per minute to 25 ◦C. The resultant 81 or 82-bp crRNA dsDNA was used to synthesize crRNA (Table 11) using T7 RNA polymerase primer as described in Example 3.

**Table 10.** Complementary oligonucleotides for synthesis of crRNA dsDNA for SARS-CoV-2. Complementary oligonucleotides are shown as a and b for each crRNA.

| Name | Sequence |
|---|---|
| Ca3-CoV crRNAa | TAATACGACTCACTATAGGAGATGAAAAAAGCCCGACATAGCGGGCAATCGAATTGGTGTATTCAAGGCTCCCTCAGTTGC (SEQ ID NO:104) |
| Ca7-CoV crRNAb | GCAACTGAGGGAGCCTTGAATACACCAATTCGATTGCCCGCTATGTCGGGCTTTTTTCATCTCCTATAGTGAGTCGTATTA (SEQ ID NO:105) |
| Ca7-CoV crRNAa | TAATACGACTCACTATAGTTAGATGAGAACACTCCGAGATAACGGAGAATAACTTGGTGTATTCAAGGCTCCCTCAGTTGC (SEQ ID NO:106) |
| Ca7-CoV crRNAb | GCAACTGAGGGAGCCTTGAATACACCAAGTTATTCTCCGTTATCTCGGAGTGTTCTCATCTAACTATAGTGAGTCGTATTA (SEQ ID NO:107) |
| Cd13-CoV crRNAa | TAATACGACTCACTATAGAAACGTACTACACCCTTTTTGTAAGGGTCTGAAACCTTGTGCAATTTGCGGCCAATGTTTGTAA (SEQ ID NO: 108) |
| Cd13-CoV crRNAb | TTACAAACATTGGCCGCAAATTGCACAAGGTTTCAGACCCTTACAAAAAGGGTGTAGTACGTTTCTATAGTGAGTCGTATTA (SEQ ID NO:109) |
| Cd14-CoV crRNA 1a | TAATACGACTCACTATAGCAACTACAACCCTGTCAAATTACAGGGTTCTGAAACTTGTGCAATTTGCGGCCAATGTTTGTAA (SEQ ID NO:110) |
| Cd14-CoV crRNA 1b | TTACAAACATTGGCCGCAAATTGCACAAGTTTCAGAACCCTGTAATTTGACAGGGTTGTAGTTGCTATAGTGAGTCGTATTA (SEQ ID NO:111) |
| Cd14-CoV crRNA 2a | TAATACGACTCACTATAGCAACTACAACCCTGTCAAATTACAGGGTTCTGAAACTTGGTGTATTCAAGGCTCCCTCAGTTGC (SEQ ID NO:112) |
| Cd14-CoV crRNA 2b | GCAACTGAGGGAGCCTTGAATACACCAAGTTTCAGAACCCTGTAATTTGACAGGGTTGTAGTTGCTATAGTGAGTCGTATTA (SEQ ID NO:113) |
| Cd15-CoV crRNAa | TAATACGACTCACTATAGATCTATAACCCTGCATTTATGTAGGGCTCTAAAACTTGGTGTATTCAAGGCTCCCTCAGTTGC (SEQ ID NO:114) |
| Cd15-CoV crRNAb | GCAACTGAGGGAGCCTTGAATACACCAAGTTTTAGAGCCCTACATAAATGCAGGGTTATAGATCTATAGTGAGTCGTATTA (SEQ ID NO:115) |

**Table 11.** crRNA sequences (5' to 3') for for SARS-CoV-2 detection.

| Name | Sequence |
|---|---|
| Ca3- CoV crRNA | GGAGAUGAAAAAAGCCCGACAUAGCGGGCAAUCGAAUUGGUGUAUUCAAGGCUCCCUCAGUUGC (SEQ ID NO:116) |
| Ca7-CoV crRNA | GUUAGAUGAGAACACUCCGAGAUAACGGAGAAUAACUUGGUGUAUUCAAGGCUCCCUCAGUUGC (SEQ ID NO:117) |
| Cd13-CoV crRNA | GAAACGUACUACACCCUUUUUGUAAGGGUCUGAAACCUUGUGCAAUUUGCGGCCAAUGUUUGUAA (SEQ ID NO:118) |
| Cd14-CoV crRNA 1 | GCAACUACAACCCUGUCAAAUUACAGGGUUCUGAAACUUGUGCAAUUUGCGGCCAAUGUUUGUAA (SEQ ID NO:119) |

| | |
|---|---|
| Cd14-CoV crRNA 2 | GCAACUACAACCCUGUCAAAUUACAGGGUUCUGAAACUUGGUGUAUUCAAGGCUCC CUCAGUUGC (SEQ ID NO:120) |
| Cd15-CoV crRNA | GAUCUAUAACCCUGCAUUUAUGUAGGGCUCUAAAACUUGGUGUAUUCAAGGCUCCC UCAGUUGC (SEQ ID NO:121) |

A 20µL detection system was prepared that consisted of 100 nM Cas13, 50 nM crRNA, 250 nM ssRNA reporter, and 250 nM target ssRNA, for each enzyme in the corresponding reaction buffer. Fluorescence readouts were taken in 2 min intervals for a total of 60 min on the ABI Real-time PCR (Applied Biosystems, CA, USA).

Results

*In vitro* cis- and trans-cleavage activities of Cas13 polypeptides tested revealed the that all 5 could detect SARS-CoV-2 N gene, with Cas13d14 exhibiting the highest efficiency (Figure 10). These results confirm the ability for the Cas13 polypeptides to detect different targets and their use in detection assays, such as a Cas13 SARS-CoV-2-based detection assay.

In addition, Cas13d14 showed different collateral (trans) cleavage activity with two crRNAs, with crRNA 1 (labelled as crRNA 1) mediating the highest efficiency relative to crRNA 2 (labelled as crRNA 2) and controls. This result demonstrated the efficiency of Cas13 detection can be impacted by the crRNAs (Figure 11A and B).

Furthermore, the results from the visual-based detection assay were compared to the results in Figure 11A. The image of strip of reactions were captured using SYNGENE transilluminator (420nm wavelength), 15 minutes after beginning of collateral activity (Figure 12). The results showed concordance between the two assays, indicating that the developed Cas13 visual detection assay is reliable. Visual-based detection assays can therefore be used for home detection kits.

The results of *in vitro* cis-cleavage activity of Cas13 polypeptides showed different efficiencies, with Cas13d14 exhibiting the highest relative efficiency enzymes (Figure 13). The presence and absence of crRNAs has been designated as '+' and '-', respectively. Cd14-CoV crRNA 1 and Cd14-CoV crRNA 2 were shown by '+1' and '+2', respectively. The reaction containing Cas13d14 without crRNA was used as control.