(12) UK Patent (19) GB (11) 2582482 (13) B

(45) Date of B Publication 17.05.2023

(54) Title of the Invention: CASZ compositions and methods of use

(51) INT CL: ***C12N 9/22*** (2006.01) ***C12Q 1/6809*** (2018.01) ***C12Q 1/6816*** (2018.01)

(21) Application No: **2008124.6**

(22) Date of Filing: **31.10.2018**

Date Lodged: **29.05.2020**

(30) Priority Data:
(31) **62580395** (32) **01.11.2017** (33) **US**

(86) International Application Data:
PCT/US2018/058545 En 31.10.2018

(87) International Publication Data:
WO2019/089820 En 09.05.2019

(43) Date of Reproduction by UK Office **23.09.2020**

(56) Documents Cited:
EP 3665279 A1
WO 2017/120410 A1
WO 2016/028843 A2
CN 106701830 A
US 20170175104 A1
US 20170051276 A1
US 20170037432 A1
US 20160208243 A1
HOOTEN et al. "The Bacteriophage Carrier State of Campylobacter jejuni Features Changes in Host Non-coding RNAs and the Acquisition of New Host-derived CRISPR Spacer Sequences", Frontiers in Microbiology, 23 March 2016 (23.03.2016), Vol 7, Article 355, Pgs. 1-8. entire document
HARRINGTON et al. "Programmed DNA destruction by miniature CRISPR-Cas14 enzymes," Science, 16 November 2018 (16.11.2018), Vol. 362, No. 6416, Pgs. 839-842. entire document
Nucleic Acids Res, 2020, 48(9); 5016 - 5023, Karvelis et. al., 'PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage'

(58) Field of Search:
As for published application 2582482 A viz:
INT CL **C12N, C12Q**
Other: **Questel Orbit, Google Patents, PubMed, Google, Google Scholar**
updated as appropriate

Additional Fields
Other: **WPI, EPODOC, BIOSIS, MEDLINE**

(72) Inventor(s):
Jennifer A. Doudna
David Burstein
Janice S Chen
Lucas Benjamin Harrington
David Paez-Espino
Jillian F. Banfield

(73) Proprietor(s):
The Regents of the University of California (Incorporated in USA - California)
1111 Franklin Street, 12th Floor, Oakland, California 94607-5200, United States of America

(74) Agent and/or Address for Service:
Mewburn Ellis LLP
Aurora Building, Counterslip, Bristol, BS1 6BX, United Kingdom

GB 2582482 B

FIG. 1

>CasZa.1|rifcsphigho2_02_scaffold_2167_39 \N id=61050970
partial=0 scaf=rifcsphigho2_02_scaffold_2167
bin=RIFCSPHIGHO2_02_FULL_Archaea_Woesearchaeota_57_22
sample=Archaea-RifleO2 proj=Rifle Groundwater Metagenome
mEVQKTVMKTLSLRILRPLYSQEIEKEIKEEKERRKQAGGTGELDGGFYKKLEKKHSEMFSFDR
LNLLLNQLQREIAKVYNHAISELYIATIAQGNKSNKHYISSIVYNRAYGYFYNAYIALGICSKV
EANFRSNELLTQQSALPTAKSDNFPIVLHKQKGAEGEDGGFRISTEGSDLIFEIPIPFYEYNGE
NRKEPYKWVKKGGQKPVLKLILSTFRRQRNKGWAKDEGTDAEIRKVTEGKYQVSQIEINRGKKL
GEHQKWFANFSIEQPIYERKPNRSIVGGLDVGIRSPLVCAINNSFSRYSVDSNDVFKFSKQVFA
FRRRLLSKNSLKRKGHGAAHKLEPITEMTEKNDKFRKKIIERWAKEVTNFFVKNQVGIVQIEDL
STMKDREDHFFNQYLRGFWPYYQMQTLIENKLKEYGIEVKRVQAKYTSQLCSNPNCRYWNNYFN
FEYRKVNKFPKFKCEKCNLEISADYNAARNLSTPDIEKFVAKATKGINLPEK
(SEQ ID NO: 1)

>CasZa.2|gwa2_scaffold_18027_12 Putative transposase DNA-binding
domain family id=2625699 partial=0 scaf=gwa2_scaffold_18027
bin=GWA2 Unbinned sample=GWA2 proj=Rifle Groundwater Metagenome
mEEAKTVSKTLSLRILRPLYSAEIEKEIKEEKERRKQGGKSGELDSGFYKKLEKKHTQMFGWDK
LNLMLSQLQRQIARVFNQSISELYIETVIQGKKSNKHYTSKIVYNRAYSVFYNAYLALGITSKV
EANFRSTELLMQKSSLPTAKSDNFPILLHKQKGVEGEEGGFKISADGNDLIFEIPIPFYEYDSA
NKKEPFKWIKKGGQKPTIKLILSTFRRQRNKGWAKDEGTDAEIRKVIEGKYQVSHIEINRGKKL
GDHQKWFVNFTIEQPIYERKLDKNIIGGIDVGIKSPLVCAVNNSFARYSVDSNDVLKFSKQAFA
FRRRLLSKNSLKRSGHGSKNKLDPITRMTEKNDRFRKKIIERWAKEVTNFFIKNQVGTVQIEDL
STMKDRQDNFFNQYLRGFWPYYQMQNLIENKLKEYGIETKRIKARYTSQLCSNPSCRHWNSYFS
FDHRKTNNFPKFKCEKCALEISADYNAARNISTPDIEKFVAKATKGINLPDKNENVILE
(SEQ ID NO: 2)

>CasZa.3|gwa1_scaffold_1795_31 Transposase, IS605 OrfB family
id=2784443 partial=0 scaf=gwa1_scaffold_1795
bin=GWA1_Archaeon_57_19 sample=GWA1 proj=Rifle Groundwater
Metagenome
mAKNTITKTLKLRIVRPYNSAEVEKIVADEKNNREKIALEKNKDKVKEACSKHLKVAAYCTTQV
ERNACLFCKARKLDDKFYQKLRGQFPDAVFWQEISEIFRQLQKQAAEIYNQSLIELYYEIFIKG
KGIANASSVEHYLSDVCYTRAAELFKNAAIASGLRSKIKSNFRLKELKNMKSGLPTTKSDNFPI
PLVKQKGGQYTGFEISNHNSDFIIKIPFGRWQVKKEIDKYRPWEKFDFEQVQKSPKPISLLLST
QRRKRNKGWSKDEGTEAEIKKVMNGDYQTSYIEVKRGSKIGEKSAWMLNLSIDVPKIDKGVDPS
IIGGIDVGVKSPLVCAINNAFSRYSISDNDLFHFNKKMFARRRILLKKNRHKRAGHGAKNKLKP
ITILTEKSERFRKKLIERWACEIADFFIKNKVGTVQmENLESMKRKEDSYFNIRLRGFWPYAEM
QNKIEFKLKQYGIEIRKVAPNNTSKTCSKCGHLNNYFNFEYRKKNKFPHFKCEKCNFKENADYN
AALNISNPKLKSTKEEP (SEQ ID NO: 3)

FIG. 1 (Cont.)

>CasZa.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_2 \N
id=132542232 partial=0
scaf=CG10_big_fil_rev_8_21_14_0.10_scaffold_20906
bin=CG10_big_fil_rev_8_21_14_0_10_UNK
sample=CG10_big_fil_rev_8_21_14_0_10 proj=Crystal Geyser
mERQKVPQIRKIVRVVPLRILRPKYSDVIENALKKFKEKGDDTNTNDFWRAIRDRDTEFFRKEL
NFSEDEINQLERDTLFRVGLDNRVLFSYFDFLQEKLMKDYNKIISKLFINRQSKSSFENDLTDE
EVEELIEKDVTPFYGAYIGKGIKSVIKSNLGGKFIKSVKIDRETKKVTKLTAINIGLMGLPVAK
SDTFPIKIIKTNPDYITFQKSTKENLQKIEDYETGIEYGDLLVQITIPWFKNENKDFSLIKTKE
AIEYYKLNGVGKKDLLNINLVLTTYHIRKKKSWQIDGSSQSLVREMANGELEEKWKSFFDTFIK
KYGDEGKSALVKRRVNKKSRAKGEKGRELNLDERIKRLYDSIKAKSFPSEINLIPENYKWKLHF
SIEIPPMVNDIDSNLYGGIDFGEQNIATLCVKNIEKDDYDFLTIYGNDLLKHAQASYARRRIMR
VQDEYKARGHGKSRKTKAQEDYSERMQKLRQKITERLVKQISDFFLWRNKFHMAVCSLRYEDLN
TLYKGESVKAKRMRQFINKQQLFNGIERKLKDYNSEIYVNSRYPHYTSRLCSKCGKLNLYFDFL
KFRTKNIIIRKNPDGSEIKYMPFFICEFCGWKQAGDKNASANIADKDYQDKLNKEKEFCNIRKP
KSKKEDIGEENEEERDYSRRFNRNSFIYNSLKKDNKLNQEKLFDEWKNQLKRKIDGRNKFEPKE

YKDRFSYLFAYYQEIIKNESES (SEQ ID NO: 4)

>CasZa.5|rifcsplowo2_01_scaffold_34461_9 \N id=58382281
partial=0 scaf=rifcsplowo2_01_scaffold_34461
bin=RifCSPlowO2_01_full_UNK sample=RIFCSPLOWO2_01_FULL
proj=RifleCSP2_LowO2
mVPTELITKTLQLRVIRPLYFEEIEKELAELKEQKEKEFEETNSLLLESKKIDAKSLKKLKRKA
RSSAAVEFWKIAKEKYPDILTKPEMEFIFSEMQKMMARFYNKSMTNIFIEMNNDEKVNPLSLIS
KASTEANQVIKCSSISSGLNRKIAGSINKTKFKQVRDGLISLPTARTETFPISFYKSTANKDEI
PISKINLPSEEEADLTITLPFPFFEIKKEKKGQKAYSYFNIIEKSGRSNNKIDLLLSTHRRQRR
KGWKEEGGTSAEIRRLMEGEFDKEWEIYLGEAEKSEKAKNDLIKNMTRGKLSKDIKEQLEDIQV
KYFSDNNVESWNDLSKEQKQELSKLRKKKVEELKDWKHVKEILKTRAKIGWVELKRGKRQRDRN
KWFVNITITRPPFINKELDDTKFGGIDLGVKVPFVCAVHGSPARLIIKENEILQFNKMVSARNR
QITKDSEQRKGRGKKNKFIKKEIFNERNELFRKKIIERWANQIVKFFEDQKCATVQIENLESFD

RTSYK (SEQ ID NO: 5)

>CasZa.6|037|3300012359.a:Ga0137385_10000156_01776
Zr6.cluA_1a.mostc10like protein
MKSDTKDKKIIIHQTKTLSLRIVKPQSIPMEEFTDLVRYHQMIIFPVYNNGAIDLYKKLFKAKI
QKGNEARAIKYFMNKIVYAPIANTVKNSYIALGYSTKMQSSFSGKRLWDLRFGEATPPTIKADF
PLPFYNQSGFKVSSENGEFIIGIPFGQYTKKTVSDIEKKTSFAWDKFTLEDTTKKTLIELLLST
KTRKMNEGWKNNEGTEAEIKRVMDGTYQVTSLEILQRDDSWFVNFNIAYDSLKKQPDRDKIAGI
HMGITRPLTAVIYNNKYRALSIYPNTVMHLTQKQLARIKEQRTNSKYATGGHGRNAKVTGTDTL
SEAYRQRRKKIIEDWIASIVKFAINNEIGTIYLEDISNTNSFFAAREQKLIYLEDISNTNSFLS
TYKYPISAISDTLQHKLEEKAIQVIRKKAYYVNQICSLCGHYNKGFTYQFRRKNKFPKMKCQGC

LEATSTEFNAAANVANPDYEKLLIKHGLLQLKK (SEQ ID NO: 6)

FIG. 1 (Cont.)

>CasZa.7|127|3300009029.a:Ga0066793_10010091_01298
Zr6.cluA_1.c10.inclusive
MSTITRQVRLSPTPEQSRLLMAHCQQYISTVNVLVAAFDSEVLTGKVSTKDFRAALPSAVKNQA
LRDAQSVFKRSVELGCLPVLKKPHCQWNNQNWRVEGDQLILPICKDGKTQQERFRCAAVALEGK
AGILRIKKKRGKWIADLTVTQEDAPESSGSAIMGVDLGIKVPAVAHIGGKGTRFFGNGRSQRSM
RRRFYARRKTLQKAKKLRAVRKSKGKEARWMKTINHQLSRQIVNHAHALGVGTIKIEALQGIRK
GTTRKSRGAAARKNNRMTNTWSFSQLTLFITYKAQRQGITVEQVDPAYTSQDCPACRARNGAQD
RTYVCSECGWRGHRDTVGAINISRRAGLSGHRRGATGA (SEQ ID NO: 7)

>CasZb.1|bjp_ig2599_sub10_scaffold_488_2 transposase id=60827220
partial=0 scaf=bjp_ig2599_sub10_scaffold_488
bin=BJP_IG2599_SUB10_Micrarchaeota_NOVEL_41_24
sample=BJP_IG2599_SUB10 proj=Borehole JP
mIAQKTIKIKLNPTKEQIIKLNSIIEEYIKVSNFTAKKIAEIQESFTDSGLTQGTCSECGKEKT
YRKYHLLKKDNKLFCITCYKRKYSQFTLQKVEFQNKTGLRNVAKLPKTYYTNAIRFASDTFSGF
DEIIKKKQNRLNSIQNRLNFWKELLYNPSNRNEIKIKVVKYAPKTDTREHPHYYSEAEIKGRIK
RLEKQLKKFKMPKYPEFTSETISLQRELYSWKNPDELKISSITDKNESmNYYGKEYLKRYIDLI
NSQTPQILLEKENNSFYLCFPITKNIEMPKIDDTFEPVGIDWGITRNIAVVSILDSKTKKPKFV
KFYSAGYILGKRKHYKSLRKHFGQKKRQDKINKLGTKEDRFIDSNIHKLAFLIVKEIRNHSNKP
IILMENITDNREEAEKSMRQNILLHSVKSRLQNYIAYKALWNNIPTNLVKPEHTSQICNRCGHQ
DRENRPKGSKLFKCVKCNYMSNADFNASINIARKFYIGEYEPFYKDNEKMKSGVNSISM (SEQ
ID NO: 8)

>CasZb.2|rifcsplowo2_01_scaffold_239_52 IS605 OrfB family
transposase id=57672634 partial=0
scaf=rifcsplowo2_01_scaffold_239 bin=RifCSPlowO2_01_full_UNK
sample=RIFCSPLOWO2_01_FULL proj=RifleCSP2_LowO2
LKLSEQENITTGVKFKLKLDKETSEGLNDYFDEYGKAINFAIKVIQKELAEDRFAGKVRLDENK
KPLLNEDGKKIWDFPNEFCSCGKQVNRYVNGKSLCQECYKNKFTEYGIRKRMYSAKGRKAEQDI
NIKNSTNKISKTHFNYAIREAFILDKSIKKQRKERFRRLREMKKKLQEFIEIRDGNKILCPKIE
KQRVERYIHPSWINKEKKLEDFRGYSMSNVLGKIKILDRNIKREEKSLKEKGQINFKARRLMLD
KSVKFLNDNKISFTISKNLPKEYELDLPEKEKRLNWLKEKIKIIKNQKPKYAYLLRKDDNFYLQ
YTLETEFNLKEDYSGIVGIDRGVSHIAVYTFVHNNGKNERPLFLNSSEILRLKNLQKERDRFLR
RKHNKKRKKSNMRNIEKKIQLILHNYSKQIVDFAKNKNAFIVFEKLEKPKKNRSKMSKKSQYKL
SQFTFKKLSDLVDYKAKREGIKVLYISPEYTSKECSHCGEKVNTQRPFNGNSSLFKCNKCGVEL
NADYNASINIAKKGLNILNSTN
(SEQ ID NO: 9)

FIG. 1 (Cont.)

>CasZb.3|rifcsplowo2_01_scaffold_282_93 IS605 OrfB family transposase id=57678680 partial=0
scaf=rifcsplowo2_01_scaffold_282 bin=RifCSPlowO2_01_full_UNK sample=RIFCSPLOWO2_01_FULL proj=RifleCSP2_LowO2
mEESIITGVKFKLRIDKETTKKLNEYFDEYGKAINFAVKIIQKELADDRFAGKAKLDQNKNPIL
DENGKKIYEFPDEFCSCGKQVNKYVNNKPFCQECYKIRFTENGIRKRMYSAKGRKAEHKINILN
STNKISKTHFNYAIREAFILDKSIKKQRKKRNERLRESKKRLQQFIDMRDGKREICPTIKGQKV
DRFIHPSWITKDKKLEDFRGYTLSIINSKIKILDRNIKREEKSLKEKGQIIFKAKRLMLDKSIR
FVGDRKVLFTISKTLPKEYELDLPSKEKRLNWLKEKIEIIKNQKPKYAYLLRKNIESEKKPNYE
YYLQYTLEIKPELKDFYDGAIGIDRGINHIAVCTFISNDGKVTPPKFFSSGEILRLKNLQKERD
RFLLRKHNKNRKKGNMRVIENKINLILHRYSKQIVDmAKKLNASIVFEELGRIGKSRTKMKKSQ
RYKLSLFIFKKLSDLVDYKSRREGIRVTYVPPEYTSKECSHCGEKVNTQRPFNGNYSLFKCNKC
GIQLNSDYNASINIAKKGLKIPNST
(SEQ ID NO: 10)

>CasZb.4|rifcsphigho2_01_scaffold_36781_5 transposase, IS605 OrfB family id=55842010 partial=0
scaf=rifcsphigho2_01_scaffold_36781 bin=RifCSPhighO2_01_full_UNK sample=RIFCSPHIGHO2_01_FULL proj=RifleCSP2_HighO2
LWTIVIGDFIEMPKQDLVTTGIKFKLDVDKETRKKLDDYFDEYGKAINFAVKIIQKNLKEDRFA
GKIALGEDKKPLLDKDGKKIYNYPNESCSCGNQVRRYVNAKPFCVDCYKLKFTENGIRKRMYSA
RGRKADSDINIKNSTNKISKTHFNYAIREGFILDKSLKKQRSKRIKKLLELKRKLQEFIDIRQG
QMVLCPKIKNQRVDKFIHPSWLKRDKKLEEFRGYSLSVVEGKIKIFNRNILREEDSLRQRGHVN
FKANRIMLDKSVRFLDGGKVNFNLNKGLPKEYLLDLPKKENKLSWLNEKISLIKLQKPKYAYLL
RREGSFFIQYTIENVPKTFSDYLGAIGIDRGISHIAVCTFVSKNGVNKAPVFFSSGEILKLKSL
QKQRDLFLRGKHNKIRKKSNMRNIDNKINLILHKYSRNIVNLAKSEKAFIVFEKLEKIKKSRFK
MSKSLQYKLSQFTFKKLSDLVEYKAKIEGIKVDYVPPEYTSKECSHCGEKVDTQRPFNGNSSLF
KCNKCRVQLNADYNASINIAKKSLNISNN (SEQ ID NO: 11)

>CasZb.5|cg1_0.2_scaffold_785_c_37 transposase, IS605 OrfB family id=91826491 partial=0 scaf=cg1_0.2_scaffold_785_c bin=CG1_02_FULL_Micrarchaeota_47_40_curated sample=CG1_02_FULL proj=Crystal Geyser
mSKTTISVKLKIIDLSSEKKEFLDNYFNEYAKATTFCQLRIRRLLRNTHWLGKKEKSSKKWIFE
SGICDLCGENKELVNEDRNSGEPAKICKRCYNGRYGNQMIRKLFVSTKKREVQENMDIRRVAKL
NNTHYHRIPEEAFDMIKAADTAEKRRKKNVEYDKKRQMEFIEMFNDEKKRAARPKKPNERETRY
VHISKLESPSKGYTLNGIKRKIDGMGKKIERAEKGLSRKKIFGYQGNRIKLDSNWVRFDLAESE
ITIPSLFKEMKLRITGPTNVHSKSGQIYFAEWFERINKQPNNYCYLIRKTSSNGKYEYYLQYTY
EAEVEANKEYAGCLGVDIGCSKLAAAVYYDSKNKKAQKPIEIFTNPIKKIKMRREKLIKLLSRV
KVRHRRRKLMQLSKTEPIIDYTCHKTARKIVEMANTAKAFISMENLETGIKQKQQARETKKQKF
YRNMFLFRKLSKLIEYKALLKGIKIVYVKPDYTSQTCSSCGADKEKTERPSQAIFRCLNPTCRY
YQRDINADFNAAVNIAKKALNNTEVVTTLL (SEQ ID NO: 12)

FIG. 1 (Cont.)

>CasZb.6|rifcsphigho2_02_scaffold_55589_5 transposase, IS605 OrfB family id=61682131 partial=0
scaf=rifcsphigho2_02_scaffold_55589 bin=RifCSPhighO2_02_full_UNK
sample=RIFCSPHIGHO2_02_FULL proj=RifleCSP2_HighO2
mARAKNQPYQKLTTTTGIKFKLDLSEEEGKRFDEYFSEYAKAVNFCAKVIYQLRKNLKFAGKKE
LAAKEWKFEISNCDFCNKQKEIYYKNIANGQKVCKGCHRTNFSDNAIRKKMIPVKGRKVESKFN
IHNTTKKISGTHRHWAFEDAADIIESMDKQRKEKQKRLRREKRKLSYFFELFGDPAKRYELPKV
GKQRVPRYLHKIIDKDSLTKKRGYSLSYIKNKIKISERNIERDEKSLRKASPIAFGARKIKMSK
LDPKRAFDLENNVFKIPGKVIKGQYKFFGTNVANEHGKKFYKDRISKILAGKPKYFYLLRKKVA
ESDGNPIFEYYVQWSIDTETPAITSYDNILGIDAGITNLATTVLIPKNLSAEHCSHCGNNHVKP
IFTKFFSGKELKAIKIKSRKQKYFLRGKHNKLVKIKRIRPIEQKVDGYCHVVSKQIVEMAKERN
SCIALEKLEKPKKSKFRQRRREKYAVSMFVFKKLATFIKYKAAREGIEIIPVEPEGTSYTCSHC
KNAQNNQRPYFKPNSKKSWTSMFKCGKCGIELNSDYNAAFNIAQKALNMTSA
(SEQ ID NO: 13)

>CasZb.7|CG03_land_8_20_14_0.80_scaffold_2214_9 transposase, IS605 OrfB family id=135165146 partial=0
scaf=CG03_land_8_20_14_0.80_scaffold_2214
bin=CG03_land_8_20_14_0_80_cor_UNK
sample=CG03_land_8_20_14_0_80_cor proj=Crystal Geyser
mDEKHFFCSYCNKELKISKNLINKISKGSIREDEAVSKAISIHNKKEHSLILGIKFKLFIENKL
DKKKLNEYFDNYSKAVTFAARIFDKIRSPYKFIGLKDKNTKKWTFPKAKCVFCLEEKEVAYANE
KDNSKICTECYLKEFGENGIRKKIYSTRGRKVEPKYNIFNSTKELSSTHYNYAIRDAFQLLDAL
KKQRQKKLKSIFNQKLRLKEFEDIFSDPQKRIELSLKPHQREKRYIHLSKSGQESINRGYTLRF
VRGKIKSLTRNIEREEKSLRKKTPIHFKGNRLMIFPAGIKFDFASNKVKISISKNLPNEFNFSG
TNVKNEHGKSFFKSRIELIKTQKPKYAYVLRKIKREYSKLRNYEIEKIRLENPNADLCDFYLQY
TIETESRNNEEINGIIGIDRGITNLACLVLLKKGDKKPSGVKFYKGNKILGMKIAYRKHLYLLK
GKRNKLRKQRQIRAIEPKINLILHQISKDIVKIAKEKNFAIALEQLEKPKKARFAQRKKEKYKL
ALFTFKNLSTLIEYKSKREGIPVIYVPPEKTSQMCSHCAINGDEHVDTQRPYKKPNAQKPSYSL
FKCNKCGIELNADYNAAFNIAQKGLKTLMLNHSH (SEQ ID NO: 14)

>CasZb.8|058|3300002172.a:JGI24730J26740_1002785_01697 Zr6.newTrpA
MLQTLLVKLDPSKEQYKMLYETMERFNEACNQIAETVFAIHSANKIEVQKTVYYPIREKFGLSA
QLTILAIRKVCEAYKRDKSIKPEFRLDGALVYDQRVLSWKGLDKVSLVTLQGRQIIPIKFGDYQ
KARMDRIRGQADLILVKGVFYLCVVVEVSEESPYDPKGVLGVDLGIKNLAVDSDGEVHSGEQTT
NTRERLDSLKARLQSKGTKSAKRHLKKLSGRMAKFSKDVNHCISKKLVAKAKGTLMSIALEDLQ
GIRDRVTVRKAQRRNLHTWNFGLLRMFVDYKAKIAGVPLVFVDPRNTSRTCPSCGHVAKANRPT
RDEFRCVSCGFAGAADHIAAMNIAFRAEVSQPIVTRFFVQSQAPSFRVG
(SEQ ID NO: 15)

FIG. 1 (Cont.)

>CasZb.9|114|3300013125.a:Ga0172369_10000737_00842
Zr6.cluA_2b.c10_like protein
MDEEPDSAEPNLAPISVKLKLVKLDGEKLAALNDYFNEYAKAVNFCELKMQKIRKNLVNIRGTY
LKEKKAWINQTGECCICKKIDELRCEDKNPDINGKICKKCYNGRYGNQMIRKLFVSTNKRAVPK
SLDIRKVARLHNTHYHRIPPEAADIIKAIETAERKRRNRILFDERRYNELKDALENEEKRVARP
KKPKEREVRYVPISKKDTPSKGYTMNALVRKVSGMAKKIERAKRNLNKRKKIEYLGRRILLDKN
WVRFDFDKSEISIPTMKEFFGEMRFEITGPSNVMSPNGREYFTKWFDRIKAQPDNYCYLLRKES
EDETDFYLQYTWRPDAHPKKDYTGCLGIDIGGSKLASAVYFDADKNRAKQPIQIFSNPIGKWKT
KRQKVIKVLSKAAVRHKTKKLESLRNIEPRIDVHCHRIARKIVGMALAANAFISMENLEGGIRE
KQKAKETKKQKFSRNMFVFRKLSKLIEYKALMEGVKVVYIVPDYTSQLCSSCGTNNTKRPKQAI

FMCQNTECRYFGKNINADFNAAINIAKKALNRKDIVRELS (SEQ ID NO: 16)

>CasZb.10|115|3300013125.a:Ga0172369_10010464_01540
Zr6.cluA_2b.c10_like protein
MEKNNSEQTSITTGIKFKLKLDKETKEKLNNYFDEYGKAINFAVRIIQMQLNDDRLAGKYKRDE
KGKPILGEDGKKILEIPNDFCSCGNQVNHYVNGVSFCQECYKKRFSENGIRKRMYSAKGRKAEQ
DINIKNSTNKISKTHFNYAIREAFNLDKSIKKQREKRFKKLKDMKRKLQEFLEIRDGKRVICPK
IEKQKVERYIHPSWINKEKKLEEFRGYSLSIVNSKIKSFDRNIQREEKSLKEKGQINFKAQRLM
LDKSVKFLKDNKVSFTISKELPKTFELDLPKKEKKLNWLNEKLEIIKNQKPKYAYLLRKENNIF
LQYTLDSIPEIHSEYSGAVGIDRGVSHIAVYTFLDKDGKNERPFFLSSSGILRLKNLQKERDKF
LRKKHNKIRKKGNMRNIEQKINLILHEYSKQIVNFAKDKNAFIVFELLEKPKKSRERMSKKIQY
KLSQFTFKKLSDLVDYKAKREGIKVIYVEPAYTSKDCSHCGERVNTQRPFNGNFSLFKCNKCGI

VLNSDYNASLNIARKGLNISAN (SEQ ID NO: 17)

>CasZb.11|134|3300013127.a:Ga0172365_10004421_00828
Zr6.cluA_2b.c10_like protein
MAEEKFFFCEKCNKDIKIPKNYINKQGAEEKARAKHEHRVHALILGIKFKIYPKKEDISKLNDY
FDEYAKAVTFTAKIVDKLKAPFLFAGKRDKDTSKKKWVFPVDKCSFCKEKTEINYRTKQGKNIC
NSCYLTEFGEQGLLEKIYATKGRKVSSSFNLFNSTKKLTGTHNNYVVKESLQLLDALKKQRSKR
LKKLSNTRRKLKQFEEMFEKEDKRFQLPLKEKQRELRFIHVSQKDRATEFKGYTMNKIKSKIKV
LRRNIEREQRSLNRKSPVFFRGTRIRLSPSVQFDDKDNKIKLTLSKELPKEYSFSGLNVANEHG
RKFFAEKLKLIKENKSKYAYLLRRQVNKNNKKPIYDYYLQYTVEFLPNIITNYNGILGIDRGIN
TLACIVLLENKKEKPSFVKFFSGKGILNLKNKRRKQLYFLKGVHNKYRKQQKIRPIEPRIDQIL
HDISKQIIDLAKEKRVAISLEQLEKPQKPKFRQSRKAKYKLSQFNFKTLSNYIDYKAKKEGIRV
IYIAPEMTSQNCSRCAMKNDLHVNTQRPYKNTSSLFKCNKCGVELNADYNAAFNIAQKGLKILN

S (SEQ ID NO: 18)

FIG. 1 (Cont.)

>CasZc.1|CG08_land_8_20_14_0.20_scaffold_1609_10 Tax=CG_Micra_03
id=133381002 partial=0 scaf=CG08_land_8_20_14_0.20_scaffold_1609
bin=CG08_land_8_20_14_0_20_UNK sample=CG08_land_8_20_14_0_20
proj=Crystal Geyser
mISLKLKLLPDEEQKKLLDEMFWKWASICTRVGFGRADKEDLKPPKDAEGVWFSLTQLNQANTD
INDLREAMKHQKHRLEYEKNRLEAQRDDTQDALKNPDRREISTKRKDLFRPKASVEKGFLKLKY
HQERYWVRRLKEINKLIERKTKTLIKIEKGRIKFKATRITLHQGSFKIRFGDKPAFLIKALSGK
NQIDAPFVVVPEQPICGSVVNSKKYLDEITTNFLAYSVNAMLFGLSRSEEMLLKAKRPEKIKKK
EEKLAKKQSAFENKKKELQKLLGRELTQQEEAIIEETRNQFFQDFEVKITKQYSELLSKIANEL
KQKNDFLKVNKYPILLRKPLKKAKSKKINNLSPSEWKYYLQFGVKPLLKQKSRRKSRNVLGIDR
GLKHLLAVTVLEPDKKTFVWNKLYPNPITGWKWRRRKLLRSLKRLKRRIKSQKHETIHENQTRK
KLKSLQGRIDDLLHNISRKIVETAKEYDAVIVVEDLQSMRQHGRSKGNRLKTLNYALSLFDYAN
VMQLIKYKAGIEGIQIYDVKPAGTSQNCAYCLLAQRDSHEYKRSQENSKIGVCLNPNCQNHKKQ

IDADLNAARVIASCYALKINDSQPFGTRKRFKKRTTN (SEQ ID NO: 19)

>CasZc.2|CG_4_10_14_0.8_um_filter_scaffold_20762_4
Tax=CG_Micra_03 id=144037526 partial=0
scaf=CG_4_10_14_0.8_um_filter_scaffold_20762
bin=CG_4_10_14_0_8_um_filter_cor_UNK
sample=CG_4_10_14_0_8_um_filter_cor proj=Crystal Geyser
mETLSLKLKLNPSKEQLLVLDKMFWKWASICTRLGLKKAEMSDLEPPKDAEGVWFSKTQLNQAN
TDVNDLRKAMQHQGKRIEYELDKVENRRNEIQEMLEKPDRRDISPNRKDLFRPKAAVEKGYLKL
KYHKLGYWSKELKTANKLIERKRKTLAKIDAGKMKFKPTRISLHTNSFRIKFGEEPKIALSTTS
KHEKIELPLITSLQRPLKTSCAKKSKTYLDAAILNFLAYSTNAALFGLSRSEEMLLKAKKPEKI
EKRDRKLATKRESFDKKLKTLEKLLERKLSEKEKSVFKRKQTEFFDKFCITLDETYVEALHRIA
EELVSKNKYLEIKKYPVLLRKPESRLRSKKLKNLKPEDWTYYIQFGFQPLLDTPKPIKTKTVLG
IDRGVRHLLAVSIFDPRTKTFTFNRLYSNPIVDWKWRRRKLLRSIKRLKRRLKSEKHVHLHENQ
FKAKLRSLEGRIEDHFHNLSKEIVDLAKENNSVIVVENLGGMRQHGRGRGKWLKALNYALSHFD
YAKVMQLIKYKAELAGVFVYDVAPAGTSINCAYCLLNDKDASNYTRGKVINGKKNTKIGECKTC

KKEFDADLNAARVIALCYEKRLNDPQPFGTRKQFKPKKP (SEQ ID NO: 20)

FIG. 1 (Cont.)

>CasZc.3|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_2 IS605 OrfB family transposase id=130740989 partial=0 scaf=CG22_combo_CG10-13_8_21_14_all_scaffold_2003 bin=CG22_combo_CG10-13_8_21_14_all_UNK sample=CG22_combo_CG10-13_8_21_14_all proj=Crystal Geyser

mKALKLQLIPTRKQYKILDEMFWKWASLANRVSQKGESKETLAPKKDIQKIQFNATQLNQIEKD
IKDLRGAMKEQQKQKERLLLQIQERRSTISEMLNDDNNKERDPHRPLNFRPKGWRKFHTSKHWV
GELSKILRQEDRVKKTIERIVAGKISFKPKRIGIWSSNYKINFFKRKISINPLNSKGFELTLMT
EPTQDLIGKNGGKSVLNNKRYLDDSIKSLLMFALHSRFFGLNNTDTYLLGGKINPSLVKYYKKN
QDMGEFGREIVEKFERKLKQEINEQQKKIIMSQIKEQYSNRDSAFNKDYLGLINEFSEVFNQRK
SERAEYLLDSFEDKIKQIKQEIGESLNISDWDFLIDEAKKAYGYEEGFTEYVYSKRYLEILNKI
VKAVLITDIYFDLRKYPILLRKPLDKIKKISNLKPDEWSYYIQFGYDSINPVQLMSTDKFLGID
RGLTHLLAYSVFDKEKKEFIINQLEPNPIMGWKWKLRKVKRSLQHLERRIRAQKMVKLPENQMK
KKLKSIEPKIEVHYHNISRKIVNLAKDYNASIVVESLEGGGLKQHGRKKNARNRSLNYALSLFD
YGKIASLIKYKADLEGVPMYEVLPAYTSQQCAKCVLEKGSFVDPEIIGYVEDIGIKGSLLDSLF
EGTELSSIQVLKKIKNKIELSARDNHNKEINLILKYNFKGLVIVRGQDKEEIAEHPIKEINGKF
AILDFVYKRGKEKVGKKGNQKVRYTGNKKVGYCSKHGQVDADLNASRVIALCKYLDINDPILFG

EQRKSFK (SEQ ID NO: 21)

>CasZc.4|rifcsphigho2_01_scaffold_82367_2 Tax=CG_Micra_03 id=56177664 partial=0 scaf=rifcsphigho2_01_scaffold_82367 bin=RifCSPhighO2_01_full_UNK sample=RIFCSPHIGHO2_01_FULL proj=RifleCSP2_HighO2

mVTRAIKLKLDPTKNQYKLLNEMFWKWASLANRFSQKGASKETLAPKDGTQKIQFNATQLNQIK
KDVDDLRGAMEKQGKQKERLLIQIQERLLTISEILRDDSKKEKDPHRPQNFRPFGWRRFHTSAY
WSSEASKLTRQVDRVRRTIERIKAGKINFKPKRIGLWSSTYKINFLKKKINISPLKSKSFELDL
ITEPQQKIIGKEGGKSVANSKKYLDDSIKSLLIFAIKSRLFGLNNKDKPLFENIITPNLVRYHK
KGQEQENFKKEVIKKFENKLKKEISQKQKEIIFSQIERQYENRDATFSEDYLRAISEFSEIFNQ
RKKERAKELLNSFNEKIRQLKKEVNGNISEEDLKILEVEAEKAYNYENGFIEWEYSEQFLGVLE
KIARAVLISDNYFDLKKYPILIRKPTNKSKKITNLKPEEWDYYIQFGYGLINSPMKIETKNFMG
IDRGLTHLLAYSIFDRDSEKFTINQLELNPIKGWKWKLRKVKRSLQHLERRMRAQKGVKLPENQ
MKKRLKSIEPKIESYYHNLSRKIVNLAKANNASIVVESLEGGGLKQHGRKKNSRHRALNYALSL
FDYGKIASLIKYKSDLEGVPMYEVLPAYTSQQCAKCVLKKGSFVEPEIIGYIEEIGFKENLLTL
LFEDTGLSSVQVLKKSKNKMTLSARDKEGKMVDLVLKYNFKGLVISQEKKKEEIVEFPIKEIDG
KFAVLDSAYKRGKERISKKGNQKLVYTGNKKVGYCSVHGQVDADLNASRVIALCKYLGINEPIV

FGEQRKSFK (SEQ ID NO: 22)

FIG. 1 (Cont.)

>CasZc.5|gwc1_scaffold_8732_6 Transposase, IS605 OrfB family id=3520800 partial=0 scaf=gwc1_scaffold_8732 bin=GWC1 Unbinned sample=GWC1 proj=Rifle Groundwater Metagenome
LDLITEPIQPHKSSSLRSKEFLEYQISDFLNFSLHSLFFGLASNEGPLVDFKIYDKIVIPKPEE
RFPKKESEEGKKLDSFDKRVEEYYSDKLEKKIERKLNTEEKNVIDREKTRIWGEVNKLEEIRSI
IDEINEIKKQKHISEKSKLLGEKWKKVNNIQETLLSQEYVSLISNLSDELTNKKKELLAKKYSK
FDDKIKKIKEDYGLEFDENTIKKEGEKAFLNPDKFSKYQFSSSYLKLIGEIARSLITYKGFLDL
NKYPIIFRKPINKVKKIHNLEPDEWKYYIQFGYEQINNPKLETENILGIDRGLTHILAYSVFEP
RSSKFILNKLEPNPIEGWKWKLRKLRRSIQNLERRWRAQDNVKLPENQMKKNLRSIEDKVENLY
HNLSRKIVDLAKEKNACIVFEKLEGQGMKQHGRKKSDRLRGLNYKLSLFDYGKIAKLIKYKAEI
EGIPIYRIDSAYTSQNCAKCVLESRRFAQPEEISCLDDFKEGDNLDKRILEGTGLVEAKIYKKL
LKEKKEDFEIEEDIAMFDTKKVIKENKEKTVILDYVYTRRKEIIGTNHKKNIKGIAKYTGNTKI
GYCMKHGQVDADLNASRTIALCKNFDINNPEIWK (SEQ ID NO: 23)

>CasZc.6|109|3300010293.a:Ga0116204_1008574_00822 Zr6.cluA_2a.c10_like protein
MSDESLVSSEDKLAIKIKIVPNAEQAKMLDEMFKKWSSICNRISRGKEDIETLRPDEGKELQFN
STQLNSATMDVSDLKKAMARQGERLEAEVSKLRGRYETIDASLRDPSRRHTNPQKPSSFYPSDW
DISGRLTPRFHTARHYSTELRKLKAKEDKMLKTINKIKNGKIVFKPKRITLWPSSVNMAFKGSR
LLLKPFANGFEMELPIVISPQKTADGKSQKASAEYMRNALLGLAGYSINQLLFGMNRSQKMLAN
AKKPEKVEKFLEQMKNKDANFDKKIKALEGKWLLDRKLKESEKSSIAVVRTKFFKSGKVELNED
YLKLLKHMANEILERDGFVNLNKYPILSRKPMKRYKQKNIDNLKPNMWKYYIQFGYEPIFERKA
SGKPKNIMGIDRGLTHLLAVAVFSPDQQKFLFNHLESNPIMHWKWKLRKIRRSIQHMERRIRAE
KNKHIHEAQLKKRLGSIEEKTEQHYHIVSSKIINWAIEYEAAIVLESLSHMKQRGGKKSVRTRA
LNYALSLFDYEKVARLITYKARIRGIPVYDVLPGMTSKTCATCLLNGSQGAYVRGLETTKAAGK
ATKRKNMKIGKCMVCNSSENSMIDADLNAARVIAICKYKNLNDPQPAGSRKVFKRF
(SEQ ID NO: 24)

>CasZc.7|126|3300005573.a:Ga0078972_1001015_00056 Zr6.cluA_2a.c10_like protein
MLALKLKIMPTEKQAEILDAMFWKWASICSRIAKMKKKVSVKENKKELSKKIPSNSDIWFSKTQ
LCQAEVDVGDHKKALKNFEKRQESLLDELKYKVKAINEVINDESKREIDPNNPSKFRIKDSTKK
GNLNSPKFFTLKKWQKILQENEKRIKKKESTIEKLKRGNIFFNPTKISLHEEEYSINFGSSKLL
LNCFYKYNKKSGINSDQLENKFNEFQNGLNIICSPLQPIRGSSKRSFEFIRNSIINFLMYSLYA
KLFGIPRSVKALMKSNKDENKLKLEEKLKKKKSSFNKTVKEFEKMIGRKLSDNESKILNDESKK
FFEIIKSNNKYIPSEEYLKLLKDISEEIYNSNIDFKPYKYSILIRKPLSKFKSKKLYNLKPTDY
KYYLQLSYEPFSKQLIATKTILGIDRGLKHLLAVSVFDPSQNKFVYNKLIKNPVFKWKKRYHDL
KRSIRNRERRIRALTGVHIHENQLIKKLKSMKNKINVLYHNVSKNIVDLAKKYESTIVLERLEN
LKQHGRSKGKRYKKLNYVLSNFDYKKIESLISYKAKKEGVPVSNINPKYTSKTCAKCLLEVNQL
SELKNEYNRDSKNSKIGICNIHGQIDADLNAARVIALCYSKNLNEPHFK
(SEQ ID NO: 25)

FIG. 1 (Cont.)

>CasZd.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_25 IS605 OrfB family transposase id=132383911 partial=0 scaf=CG10_big_fil_rev_8_21_14_0.10_scaffold_4477 bin=CG10_big_fil_rev_8_21_14_0_10_test2 sample=CG10_big_fil_rev_8_21_14_0_10 proj=Crystal Geyser
VINLFGYKFALYPNKTQEELLNKHLGECGWLYNKAIEQNEYYKADSNIEEAQKKFELLPDKNSD
EAKVLRGNISKDNYVYRTLVKKKKSEINVQIRKAVVLRPAETIRNLAKVKKKGLSVGRLKFIPI
REWDVLPFKQSDQIRLEENYLILEPYGRLKFKMHRPLLGKPKTFCIKRTATDRWTISFSTEYDD
SNMRKNDGGQVGIDVGLKTHLRLSNENPDEDPRYPNPKIWKRYDRRLTILQRRISKSKKLGKNR
TRLRLRLSRLWEKIRNSRADLIQNETYEILSENKLIAIEDLNVKGMQEKKDKKGRKGRTRAQEK
GLHRSISDAAFSEFRRVLEYKAKRFGSEVKPVSAIDSSKECHNCGNKKGMPLESRIYECPKCGL
KIDRDLNSAKVILARATGVRPGSNARADTKISATAGASVQTEGTVSEDFRQQMETSDQKPMQGE
GSKEPPMNPEHKSSGRGSKHVNIGCKNKVGLYNEDENSRSTEKQIMDENRSTTEDMVEIGALHS

PVLTT (SEQ ID NO: 26)

>CasZd.2|053|3300001245.a:JGI12048J13642_10201286_01511 Zr6.c9.inclusive
MIASIDYEAVSQALIVFEFKAKGKDSQYQAIDEAIRSYRFIRNSCLRYWMDNKKVGKYDLNKYC
KVLAKQYPFANKLNSQARQSAAECSWSAISRFYDNCKRKVSGKKGFPKFKKHARSVEYKTSGWK
LSENRKAITFTDKNGIGKLKLKGTYDLHFSQLEDMKRVRLVRRADGYYVQFCISVDVKVETEPT
GKAIGLDVGIKYFLADSSGNTIENPQFYRKAEKKLNRANRRKSKKYIRGVKPQSKNYHKARCRY
ARKHLRVSRQRKEYCKRVAYCVIHSNDVVAYEDLNVKGMVKNRHLAKSISDVAWSTFRHWLEYF
AIKYGKLTIPVAPHNTSQNCSNCDKKVPKSLSTRTHICHHCGYSEDRDVNAAKNILKKALSTVG

QTGSLKLGEIEPLLVLEQSCTRKFDL (SEQ ID NO: 27)

>CasZe.1|19ft_2_nophage_noknown_scaffold_0_545 \N id=8269792 partial=0 scaf=19ft_2_nophage_noknown_scaffold_0 bin=19ft_2_UNK sample=19FT_2_THINNED proj=Rifle Sediment CSP1
LAEENTLHLTLAMSLPLNDLPENRTRSELWRRQWLPQKKLSLLLGVNQSVRKAAADCLRWFEPY
QELLWWEPTDPDGKKLLDKEGRPIKRTAGHMRVLRKLEEIAPFRGYQLGSAVKNGLRHKVADLL
LSYAKRKLDPQFTDKTSYPSIGDQFPIVWTGAFVCYEQSITGQLYLYLPLFPRGSHQEDITNNY
DPDRGPALQVFGEKEIARLSRSTSGLLLPLQFDKWGEATFIRGENNPPTWKATHRRSDKKWLSE
VLLREKDFQPKRVELLVRNGRIFVNVACEIPTKPLLEVENFMGVSFGLEHLVTVVVINRDGNVV
HQRQEPARRYEKTYFARLERLRRRGGPFSQELETFHYRQVAQIVEEALRFKSVPAVEQVGNIPK
GRYNPRLNLRLSYWPFGKLADLTSYKAVKEGLPKPYSVYSATAKMLCSTCGAANKEGDQPISLK

GPTVYCGNCGTRHNTGFNTALNLARRAQELFVKGVVAR (SEQ ID NO: 28)

FIG. 1 (Cont.)

>CasZe.2|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_49 Tax=RIFCSPHIGHO2_01_FULL_RIF_OD1_04_46_36_curated id=87353177 partial=0 scaf=RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646 bin=RIFCSPHIGHO2_01_FULL_OD1_Andersenbacteria_46_36_curated sample=2500-curated-genomes-non-redundant proj=2500 Genomes
mSQSLLKWHDMAGRDKDASRSLQKSAVEGVLLHLTASHRVALEMLEKSVSQTVAVTMEAAQQRL
VIVLEDDPTKATSRKRVISADLQFTREEFGSLPNWAQKLASTCPEIATKYADKHINSIRIAWGV
AKESTNGDAVEQKLQWQIRLLDVTMFLQQLVLQLADKALLEQIPSSIRGGIGQEVAQQVTSHIQ
LLDSGTVLKAELPTISDRNSELARKQWEDAIQTVCTYALPFSRERARILDPGKYAAEDPRGDRL
INIDPMWARVLKGPTVKSLPLLFVSGSSIRIVKLTLPRKHAAGHKHTFTATYLVLPVSREWINS
LPGTVQEKVQWWKKPDVLATQELLVGKGALKKSANTLVIPISAGKKRFFNHILPALQRGFPLQW
QRIVGRSYRRPATHRKWFAQLTIGYTNPSSLPEMALGIHFGMKDILWWALADKQGNILKDGSIP
GNSILDFSLQEKGKIERQQKAGKNVAGKKYGKSLLNATYRVVNGVLEFSKGISAEHASQPIGLG
LETIRFVDKASGSSPVNARHSNWNYGQLSGIFANKAGPAGFSVTEITLKKAQRDLSDAEQARVL
AIEATKRFASRIKRLATKRKDDTLFV (SEQ ID NO: 29)

>CasZe.3|rifcsphigho2_01_scaffold_10981_6 Tax=RIFCSPHIGHO2_01_FULL_OD1_47_10b_curated id=55513311 partial=0 scaf=rifcsphigho2_01_scaffold_10981 bin=RIFCSPHIGHO2_01_FULL_OD1_47_10b sample=2500-genomes proj=2500 Genomes
VEPVEKERFYYRTYTFRLDGQPRTQNLTTQSGWGLLTKAVLDNTKHYWEIVHHARIANQPIVFE
NPVIDEQGNPKLNKLGQPRFWKRPISDIVNQLRALFENQNPYQLGSSLIQGTYWDVAENLASWY
ALNKEYLAGTATWGEPSFPEPHPLTEINQWMPLTFSSGKVVRLLKNASGRYFIGLPILGENNPC
YRMRTIEKLIPCDGKGRVTSGSLILFPLVGIYAQQHRRMTDICESIRTEKGKLAWAQVSIDYVR
EVDKRRRMRRTRKSQGWIQGPWQEVFILRLVLAHKAPKLYKPRCFAGISLGPKTLASCVILDQD
ERVVEKQQWSGSELLSLIHQGEERLRSLREQSKPTWNAAYRKQLKSLINTQVFTIVTFLRERGA
AVRLESIARVRKSTPAPPVNFLLSHWAYRQITERLKDLAIRNGMPLTHSNGSYGVRFTCSQCGA
TNQGIKDPTKYKVDIESETFLCSICSHREIAAVNTATNLAKQLLDE (SEQ ID NO: 30)

>CasZe.4b|rifcsplowo2_02_scaffold_57876_2 hypothetical protein Tax=GWA2_OD1_56_11 id=66016044 partial=0 scaf=rifcsplowo2_02_scaffold_57876 bin=RifCSPlowO2_02_full_UNK sample=RIFCSPLOWO2_02_FULL proj=RifleCSP2_LowO2
mNDTETSETLTSHRTVCAHLHVVGETGSLPRLVEAALAELITLNGRATQALLSLAKNGLVLRRD
KEENLIAAELTLPCRKNKYADVAAKAGEPILATRINNKGKLVTKKWYGEGNSYHIVRFTPETGM
FTVRVFDRYAFDEELLHLHSEVVFGSDLPKGIKAKTDSLPANFLQAVFTSFLELPFQGFPDIVV
KPAMKQAAEQLLSYVQLEAGENQQAEYPDTNERDPELRLVEWQKSLHELSVRTEPFEFVRARDI
DYYAETDRRGNRFVNITPEWTKFAESPFARRLPLKIPPEFCILLRRKTEGHAKIPNRIYLGLQI
FDGVTPDSTLGVLATAEDGKLFWWHDHLDEFSNLEGKPEPKLKNKPQLLMVSLEYDREQRFEES
VGGDRKICLVTLKETRNFRRGWNGRILGIHFQHNPVITWALMDHDAEVLEKGFIEGNAFLGKAL
DKQALNEYLQKGGKWVGDRSFGNKLKGITHTLASLIVRLAREKDAWIALEEISWVQKQSADSVA
NHEIVEQPHHSLTR (SEQ ID NO: 31)

FIG. 1 (Cont.)

>CasZe.4|RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495_33 hypothetical protein Tax=GWA2_OD1_56_11 id=88324948 partial=0
scaf=RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495
bin=RIFCSPLOWO2_01_FULL_OD1_Taylorbacteria_45_34b_curated
sample=2500-curated-genomes proj=2500 Genomes
mNDTETSETLTSHRTVCAHLHVVGETGSLPRLVEAALAELITLNGRATQALLSLAKNGLVLRRD
KEENLIAAELTLPCRKNKYADVAAKAGEPILATRINNKGKLVTKKWYGEGNSYHIVRFTPETGM
FTVRVFDRYAFDEELLHLHSEVVFGSDLPKGIKAKTDSLPANFLQAVFTSFLELPFQGFPDIVV
KPAMKQAAEQLLSYVQLEAGENQQAEYPDTNERDPELRLVEWQKSLHELSVRTEPFEFVRARDI
DYYAETDRRGNRFVNITPEWTKFAESPFARRLPLKIPPEFCILLRRKTEGHAKIPNRIYLGLQI
FDGVTPDSTLGVLATAEDGKLFWWHDHLDEFSNLEGKPEPKLKNKPQLLMVSLEYDREQRFEES
VGGDRKICLVTLKETRNFRRGRHGHTRTDRLPAGNTLWRADFATSAEVAAPKWNGRILGIHFQH
NPVITWALMDHDAEVLEKGFIEGNAFLGKALDKQALNEYLQKGGKWVGDRSFGNKLKGITHTLA
SLIVRLAREKDAWIALEEISWVQKQSADSVANRRFSMWNYSRLATLIEWLGTDIATRDCGTAAP
LAHKVSDYLTHFTCPECGACRKAGQKKEIADTVRAGDILTCRKCGFSGPIPDNFIAEFVAKKAL
ERMLKKKPV (SEQ ID NO: 32)

>CasZf.1|rifcsphigho2_01_scaffold_566_121
Tax=RIFCSPHIGHO2_01_FULL_RIF_OD1_10_44_22b_curated id=55167017
partial=0 scaf=rifcsphigho2_01_scaffold_566
bin=RifCSPhighO2_01_full_UNK sample=RIFCSPHIGHO2_01_FULL
proj=RifleCSP2_HighO2
mAKRNFGEKSEALYRAVRFEVRPSKEELSILLAVSEVLRMLFNSALAERQQVFTEFIASLYAEL
KSASVPEEISEIRKKLREAYKEHSISLFDQINALTARRVEDEAFASVTRNWQEETLDALDGAYK
SFLSLRRKGDYDAHSPRSRDSGFFQKIPGRSGFKIGEGRIALSCGAGRKLSFPIPDYQQGRLAE
TTKLKKFELYRDQPNLAKSGRFWISVVYELPKPEATTCQSEQVAFVALGASSIGVVSQRGEEVI
ALWRSDKHWVPKIEAVEERMKRRVKGSRGWLRLLNSGKRRMHMISSRQHVQDEREIVDYLVRNH
GSHFVVTELVVRSKEGKLADSSKPERGGSLGLNWAAQNTGSLSRLVRQLEEKVKEHGGSVRKHK
LTLTEAPPARGAENKLWMARKLRESFLKEV (SEQ ID NO: 33)

>CasZf.2|rifcsplowo2_01_scaffold_81231_1 transposase, IS605 OrfB family, central region id=58737995 partial=0
scaf=rifcsplowo2_01_scaffold_81231 bin=RifCSPlowO2_01_full_UNK
sample=RIFCSPLOWO2_01_FULL proj=RifleCSP2_LowO2
LAKNDEKELLYQSVKFEIYPDESKIRVLTRVSNILVLVWNSALGERRARFELYIAPLYEELKKF
PRKSAESNALRQKIREGYKEHIPTFFDQLKKLLTPMRKEDPALLGSVPRAYQEETLNTLNGSFV
SFMTLRRNNDMDAKPPKGRAEDRFHEISGRSGFKIDGSEFVLSTKEQKLRFPIPNYQLEKLKEA
KQIKKFTLYQSRDRRFWISIAYEIELPDQRPFNPEEVIYIAFGASSIGVISPEGEKVIDFWRPD
KHWKPKIKEVENRMRSCKKGSRAWKKRAAARRKMYAMTQRQQKLNHREIVASLLRLGFHFVVTE
YTVRSKPGKLADGSNPKRGGAPQGFNWSAQNTGSFGEFILWLKQKVKEQGGTVQTFRLVLGQSE
RPEKRGRDNKIEMVRLLREKYLESQTIVV (SEQ ID NO: 34)

FIG. 1 (Cont.)

>CasZf.3|rifcsphigho2_01_scaffold_4702_81 IS605 OrfB family transposase id=55366684 partial=0
scaf=rifcsphigho2_01_scaffold_4702 bin=RifCSPhighO2_01_full_UNK
sample=RIFCSPHIGHO2_01_FULL proj=RifleCSP2_HighO2
mAKGKKKEGKPLYRAVRFEIFPTSDQITLFLRVSKNLQQVWNEAWQERQSCYEQFFGSIYERIG
QAKKRAQEAGFSEVWENEAKKGLNKKLRQQEISMQLVSEKESLLQELSIAFQEHGVTLYDQING
LTARRIIGEFALIPRNWQEETLDSLDGSFKSFLALRKNGDPDAKPPRQRVSENSFYKIPGRSGF
KVSNGQIYLSFGKIGQTLTSVIPEFQLKRLETAIKLKKFELCRDERDMAKPGRFWISVAYEIPK
PEKVPVVSKQITYLAIGASRLGVVSPKGEFCLNLPRSDYHWKPQINALQERLEGVVKGSRKWKK
RMAACTRMFAKLGHQQKQHGQYEVVKKLLRHGVHFVVTELKVRSKPGALADASKSDRKGSPTGP
NWSAQNTGNIARLIQKLTDKASEHGGTVIKRNPPLLSLEERQLPDAQRKIFIAKKLREEFLADQ
K (SEQ ID NO: 35)

>CasZg.1|rifcsp13_1_sub10_scaffold_3_54 \N id=12787801 partial=0
scaf=rifcsp13_1_sub10_scaffold_3 bin=RifCSP13_1_UNK
sample=RIFCSP13_1 proj=Rifle Sediment CSP2
mAKREKKDDVVLRGTKMRIYPTDRQVTLMDMWRRRCISLWNLLLNLETAAYGAKNTRSKLGWRS
IWARVVEENHAKALIVYQHGKCKKDGSFVLKRDGTVKHPPRERFPGDRKILLGLFDALRHTLDK
GAKCKCNVNQPYALTRAWLDETGHGARTADIIAWLKDFKGECDCTAISTAAKYCPAPPTAELLT
KIKRAAPADDLPVDQAILLDLFGALRGGLKQKECDHTHARTVAYFEKHELAGRAEDILAWLIAH
GGTCDCKIVEEAANHCPGPRLFIWEHELAMIMARLKAEPRTEWIGDLPSHAAQTVVKDLVKALQ
TMLKERAKAAAGDESARKTGFPKFKKQAYAAGSVYFPNTTmFFDVAAGRVQLPNGCGSMRCEIP
RQLVAELLERNLKPGLVIGAQLGLLGGRIWRQGDRWYLSCQWERPQPTLLPKTGRTAGVKIAAS
IVFTTYDNRGQTKEYPMPPADKKLTAVHLVAGKQNSRALEAQKEKEKKLKARKERLRLGKLEKG
HDPNALKPLKRPRVRRSKLFYKSAARLAACEAIERDRRDGFLHRVTNEIVHKFDAVSVQKMSVA
PMMRRQKQKEKQIESKKNEAKKEDNGAAKKPRNLKPVRKLLRHVAMARGRQFLEYKYNDLRGPG
SVLIADRLEPEVQECSRCGTKNPQMKDGRRLLRCIGVLPDGTDCDAVLPRNRNAARNAEKRLRK
HREAHNA (SEQ ID NO: 36)

>CasZg.2|033|3300009991.a:Ga0105042_100140_01533
Zr6.c9.inclusive
MNEVLPIPAVGEDAADTIMRGSKMRIYPSVRQAATMDLWRRRCIQLWNLLLELEQAAYSGENRR
TQIGWRSIWATVVEDSHAEAVRVAREGKKRKDGTFRKAPSGKEIPPLDPAMLAKIQRQMNGAVD
VDPKTGEVTPAQPRLFMWEHELQKIMARLKQAPRTHWIDDLPSHAAQSVVKDLIKALQAMLRER
KKRASGIGGRDTGFPKFKKNRYAAGSVYFANTQLRFEAKRGKAGDPDAVRGEFARVKLPNGVGW
MECRMPRHINAAHAYAQATLMGGRIWRQGENWYLSCQWKMPKPAPLPRAGRTAAIKIAAAIPIT
TVDNRGQTREYAMPPIDRERIAAHAAAGRAQSRALEARKRRAKKREAYAKKRHAKKLERGIAAK
PPGRARIKLSPGFYAAAAKLAKLEAEDANAREAWLHEITTQIVRNFDVIAVPRMEVAKLMKKPE
PPEEKEEQVKAPWQGKRRSLKAARVMMRRTAMALIQTTLKYKAVDLRGPQAYEEIAPLDVTAAA
CSGCGVLKPEWKMARAKGREIMRCQEPLPGGKTCNTVLTYTRNSARVIGRELAVRLAERQKA
(SEQ ID NO: 37)

FIG. 1 (Cont.)

>CasZh.1|rifcsp2_19_4_full_scaffold_168_120 transposase
id=102318404 partial=0 scaf=rifcsp2_19_4_full_scaffold_168
bin=RifCSP19_4_full_UNK sample=RIFCSP19_4_FULL proj=Rifle
Sediment CSP1
mTTQKTYNFCFYDQRFFELSKEAGEVYSRSLEEFWKIYDETGVWLSKFDLQKHMRNKLERKLLH
SDSFLGAMQQVHANLASWKQAKKVVPDACPPRKPKFLQAILFKKSQIKYKNGFLRLTLGTEKEF
LYLKWDINIPLPIYGSVTYSKTRGWKINLCLETEVEQKNLSENKYLSIDLGVKRVATIFDGENT
ITLSGKKFMGLMHYRNKLNGKTQSRLSHKKKGSNNYKKIQRAKRKTTDRLLNIQKEMLHKYSSF
IVNYAIRNDIGNIIIGDNSSTHDSPNMRGKTNQKISQNPEQKLKNYIKYKFESISGRVDIVPEP
YTSRKCPHCKNIKKSSPKGRTYKCKKCGFIFDRDGVGAINIYNENVSFGQIISPGRIRSLTEPI
GMKFHNEIYFKSYVAA (SEQ ID NO: 38)

>CasZi.1|RBG_13_scaffold_1401_19
Tax=RBG_13_Planctomycetes_46_10_curated id=16386779 partial=0
scaf=RBG_13_scaffold_1401 bin=RBG_13_Planctomycetes_46_10
sample=2500-genomes proj=2500 Genomes
mSVRSFQARVECDKQTMEHLWRTHKVFNERLPEIIKILFKMKRGECGQNDKQKSLYKSISQSIL
EANAQNADYLLNSVSIKGWKPGTAKKYRNASFTWADDAAKLSSQGIHVYDKKQVLGDLPGMMSQ
MVCRQSVEAISGHIELTKKWEKEHNEWLKEKEKWESEDEHKKYLDLREKFEQFEQSIGGKITKR
RGRWHLYLKWLSDNPDFAAWRGNKAVINPLSEKAQIRINKAKPNKKNSVERDEFFKANPEMKAL
DNLHGYYERNFVRRRKTKKNPDGFDHKPTFTLPHPTIHPRWFVFNKPKTNPEGYRKLILPKKAG
DLGSLEMRLLTGEKNKGNYPDDWISVKFKADPRLSLIRPVKGRRVVRKGKEQGQTKETDSYEFF
DKHLKKWRPAKLSGVKLIFPDKTPKAAYLYFTCDIPDEPLTETAKKIQWLETGDVTKKGKKRKK
KVLPHGLVSCAVDLSMRRGTTGFATLCRYENGKIHILRSRNLWVGYKEGKGCHPYRWTEGPDLG
HIAKHKREIRILRSKRGKPVKGEESHIDLQKHIDYMGEDRFKKAARTIVNFALNTENAASKNGF
YPRADVLLLENLEGLIPDAEKERGINRALAGWNRRHLVERVIEMAKDAGFKRRVFEIPPYGTSQ
VCSKCGALGRRYSIIRENNRREIRFGYVEKLFACPNCGYCANADHNASVNLNRRFLIEDSFKSY
YDWKRLSEKKQKEEIETIESKLMDKLCAMHKISRGSISK (SEQ ID NO: 39)

FIG. 1 (Cont.)

>CasZi.2|026|3300009652.a:Ga0123330_1010394_01528
Zr6.c2c1.inclusive
MHLWRTHCVFNQRLPALLKRLFAMRRGEVGGNEAQRQVYQRVAQFVLARDAKDSVDLLNAVSLR
KRSANSAFKKKATISCNGQAREVTGEEVFAEAVALASKGVFAYDKDDMRAGLPDSLFQPLTRDA
VACMRSHEELVATWKKEYREWRDRKSEWEAEPEHALYLNLRPKFEEGEAARGGRFRKRAERDHA
YLDWLEANPQLAAWRRKAPPAVVPIDEAGKRRIARAKAWKQASVRAEEFWKRNPELHALHKIHV
QYLREFVRPRRTRRNKRREGFKQRPTFTMPDPVRHPRWCLFNAPQTSPQGYRLLRLPQSRRTVG
SVELRLLTGPSDGAGFPDAWVNVRFKADPRLAQLRPVKVPRTVTRGKNKGAKVEADGFRYYDDQ
LLIERDAQVSGVKLLFRDIRMAPFADKPIEDRLLSATPYLVFAVEIKDEARTERAKAIRFDETS
ELTKSGKKRKTLPAGLVSVAVDLDTRGVGFLTRAVIGVPEIQQTHHGVRLLQSRYVAVGQVEAR
ASGEAEWSPGPDLAHIARHKREIRRLRQLRGKPVKGERSHVRLQAHIDRMGEDRFKKAARKIVN
EALRGSNPAAGDPYTRADVLLYESLETLLPDAERERGINRALLRWNRAKLIEHLKRMCDDAGIR
HFPVSPFGTSQVCSKCGALGRRYSLARENGRAVIRFGWVERLFACPNPECPGRRPDRPDRPFTC
NSDHNASVNLHRVFALGDQAVAAFRALAPRDSPARTLAVKRVEDTLRPQLMRVHKLADAGVDSP
F (SEQ ID NO: 40)

>CasZj.1|130|3300012532.a:Ga0137373_10000316_00700 c2c4
MATLVYRYGVRAHGSARQQDAVVSDPAMLEQLRLGHELRNALVGVQHRYEDGKRAVWSGFASVA
AADHRVTTGETAVAELEKQARAEHSADRTAATRQGTAESLKAARAAVKQARADRKAAMAAVAEQ
AKPKIQALGDDRDAEIKDLYRRFCQDGVLLPRCGRCAGDLRSDGDCTDCGAAHEPRKLYWATYN
AIREDHQTAVKLVEAKRKAGQPARLRFRRWTGDGTLTVQLQRMHGPACRCVTCAEKLTRRARKT
DPQAPAVAADPAYPPTDPPRDPALLASGQGKWRNVLQLGTWIPPGEWSAMSRAERRRVGRSHIG
WQLGGGRQLTLPVQLHRQMPADADVAMAQLTRVRVGGRHRMSVALTAKLPDPPQVQGLPPVALH
LGWRQRPDGSLRVATWACPQPLDLPPAVADVVVSHGGRWGEVIMPARWLADAEVPPRLLGRRDK
AMEPVLEALADWLEAHTEACTARMTPALVRRWRSQGRLAGLTNRWRGQPPTGSAEILTYLEAWR
IQDKLLWERESHLRRRLAARRDDAWRRVASWLARHAGVLVVDDADIAELRRRDDPADTDPTMPA
SAAQAARARAALAAPGRLRHLATITATRDGLGVHTVASAGLTRLHRKCGHQAQPDPRYAASAVV
TCPGCGNGYDQDYNAAMLMLDRQQQP (SEQ ID NO: 41)

>CasZk.1|012|3300005602.a:Ga0070762_10001740_01506
Zr6.c9.inclusive
MSRVELHRAYKFRLYPTPAQVAELAEWERQLRRLYNLAHSQRLAAMQRHVRPKSPGVLKSECLS
CGAVAVAEIGTDGKAKKTVKHAVGCSVLECRSCGGSPDAEGRTAHTAACSFVDYYRQGREMTQL
LEEDDQLARVVCSARQETLRDLEKAWQRWHKMPGFGKPHFKKRIDSCRIYFSTPKSWAVDLGYL
SFTGVASSVGRIKIRQDRVWPGDAKFSSCHVVRDVDEWYAVFPLTFTKEIEKPKGGAVGINRGA
VHAIADSTGRVVDSPKFYARSLGVIRHRARLLDRKVPFGRAVKPSPTKYHGLPKADIDAAAARV
NASPGRLVYEARARGSIAAAEAHLAALVLPAPRQTSQLPSEGRNRERARRFLALAHQRVRRQRE
WFLHNESAHYAQSYTKIAIEDWSTKEMTSSEPRDAEEMKRVTRARNRSILDVGWYELGRQIAYK
SEATGAEFAKVDPGLRETETHVPEAIVRERDVDVSGMLRGEAGISGTCSRCGGLLRASASGHAD
AECEVCLHVEVGDVNAAVNVLKRAMFPGAAPPSKEKAKVTIGIKGRKKKRAA
(SEQ ID NO: 42)

FIG. 1 (Cont.)

>CasZk.2|016|3300005921.a:Ga0070766_10011912_01491
Zr6.c9.inclusive
MSRVELHRAYKFRLYPTPVQVAELSEWERQLRRLYNLGHEQRLLTLTRHLRPKSPGVLKGECLS
CDSTQVQEVGADGRPKTTVRHAEQCPTLACRSCGALRDAEGRTAHTVACAFVDYYRQGREMTEL
LAADDQLARVVCSARQEVLRDLDKAWQRWRKMPGFGKPRFKRRTDSCRIYFSTPKAWKLEGGHL
SFTGAATTVGAIKMRQDRNWPASVQFSSCHVVRDVDEWYAVFPLTFVAEVARPKGGAVGINRGA
VHAIADSTGRVVDSPRYYARALGVIRHRARLFDRKVPSGHAVKPSPTKYRGLSAIEVDRVARAT
GFTPGRVVTEALNRGGVAYAECALAAIAVLGHGPERPLTSDGRNREKARKFLALAHQRVRRQRE
WFLHNESAHYARTYSKIAIEDWSTKEMTASEPQGEETRRVTRSRNRSILDVGWYELGRQLAYKT
EATGAEFAQVDPGLKETETNVPKAIADARDVDVSGMLRGEAGISGTCSKCGGLLRAPASGHADA
ECEICLNVEVGDVNAAVNVLKRAMFPGDAPPASGEKPKVSIGIKGRQKKKKAA
(SEQ ID NO: 43)

>CasZk.3|106|3300009698.a:Ga0116216_10000905_01565 c2c9
MEAIATGMSPERRVELGILPGSVELKRAYKFRLYPMKVQQAELSEWERQLRRLYNLAHEQRLAA
LLRYRDWDFQKGACPSCRVAVPGVHTAACDHVDYFRQAREMTQLLEVDAQLSRVICCARQEVLR
DLDKAWQRWRKKLGGRPRFKRRTDSCRIYLSTPKHWEIAGRYLRLSGLASSVGEIRIEQDRAFP
EGALLSSCSIVRDVDEWYACLPLTFTQPIERAPHRSVGLNRGVVHALADSDGRVVDSPKFFERA
LATVQKRSRDLARKVSGSRNAHKARIKLAKAHQRVRRQRAAFLHQESAYYSKGFDLVALEDMSV
RKMTATAGEAPEMGRGAQRDLNRGILDVGWYELARQIDYKRLAHGGELLRVDPGQTTPLACVTE
EQPARGISSACAVCGIPLARPASGNARMRCTACGSSQVGDVNAAENVLTRALSSAPSGPKSPKA
SIKIKGRQKRLGTPANRAGEASGGDPPVRGPVEGGTLAYVVEPVSESQSDT
(SEQ ID NO: 44)

>CasZk.4|072|3300006028.a:Ga0070717_10000077_00263
Zr6.c9.inclusive
MTVRTYKYRAYPTPEQAEALTSWLRFASQLYNAALEHRKNAWGRHDAHGRGFRFWDGDAAPRKK
SDPPGRWVYRGGGGAHISKNDQGKLLTEFRREHAELLPPGMPALVQHEVLARLERSMAAFFQRA
TKGQKAGYPRWRSEHRYDSLTFGLTSPSKERFDPETGESLGRGKTVGAGTYHNGDLRLTGLGEL
RILEHRRIPMGAIPKSVIVRRSGKRWFVSIAMEMPSVEPAASGRPAVGLDMGVVTWGTAFTADT
SAAAALVADLRRMATDPSDCRRLEELEREAAQLSEVLAHCRARGLDPARPRRCPKELTKLYRRS
LHRLGELDRACARIRRRLQAAHDIAEPVPDEAGSAVLIEGSNAGMRHARRVARTQRRVARRTRA
GHAHSNRRKKAVQAYARAKERERSARGDHRHKVSRALVRQFEEISVEALDIKQLTVAPEHNPDP
QPDLPAHVQRRRNRGELDAAWGAFFAALDYKAADAGGRVARKPAPHTTQECARCGTLVPKPISL
RVHRCPACGYTAPRTVNSARNVLQRPLEEPGRAGPSGANGRGVPHAVA (SEQ ID NO: 45)

FIG. 1 (Cont.)

>CasZ1.1|004|3300001256.a:JGI12210J13797_10004690_01983 c2c9
MNCRYRYRIYPTPGQRQSLARLFGCVRVVWNDALFLCRQSEKLPKNSELQKLCITQAKKTEARG
WLGQVSAIPLQQSVADLGVAFKNFFQSRSGKRKGKKVNPPRVKRRNNRQGARFTRGGFKVKTSK
VYLARIGDIKIKWSRPLPSEPSSVTVIKDCAGQYFLSFVVEVKPEIKPPKNPSIGIDLGLKTFA
SCSNGEKIDSPDYSRLYRKLKRCQRRLAKRQRGSKRRERMRVKVAKLNAQIRDKRKDFLHKLST
KVVNENQVIALEDLNVGGMLKNRKLSRAISQAGWYEFRSLCEGKAEKHNRDFRVISRWEPTSQV
CSECGYRWGKIDLSVRSIVCINCGVEHDRDDNASVNIEQAGLKVGVGHTHDSKRTGSACKTSNG

AVCVEPSTHREYVQLTLFDW (SEQ ID NO: 46)

>CasZ1.2|014|3300005660.a:Ga0073904_10021651_01988 c2c9
MKSRWTFRCYPTPEQEQHLARTFGCVRFVWNWALRARTDAFRAGERIGYPATDKALTLLKQQPE
TVWLNEVSSVCLQQALRDLQVAFSNFFDKRAAHPSFKRKEARQSANYTERGFSFDHERRILKLA
KIGAIKVKWSRKAIPHPSSIRLIRTASGKYFVSLVVETQPAPMPETGESVGVDFGVARLATLSN
GERISNPKHGAKWQRRLAFYQKRLARATKGSKRRMRIKRHVARIHEKIGNSRSDTLHKLSTDLV
TRFDLICVEDLNLRGMVKNHSLARSLHDASIGSAIRMIEEKAERYGKNVVKIDRWFPSSKTCSD
CGHIVEQLPLNVREWTCPECGTTHDRDANAAANILAVGQTVSAHGGTVRRSRAKASERKSQRSA

NRQGVNRA (SEQ ID NO: 47)

FIG. 2

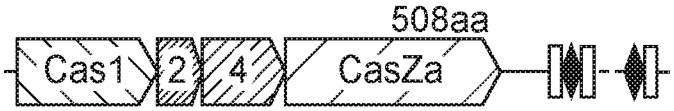
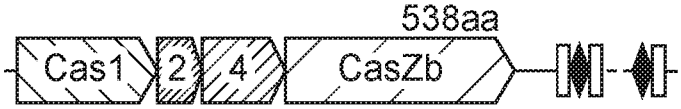
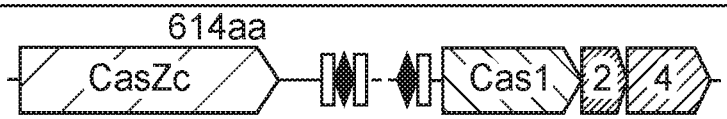
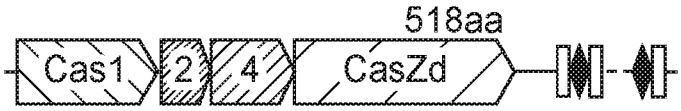
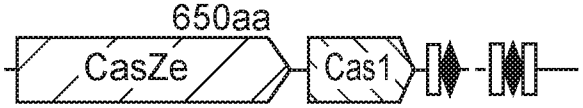
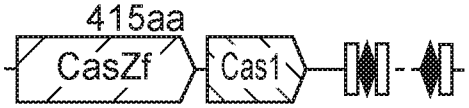
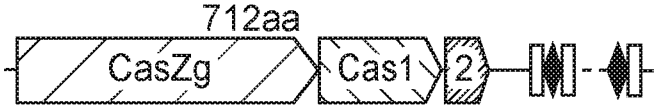
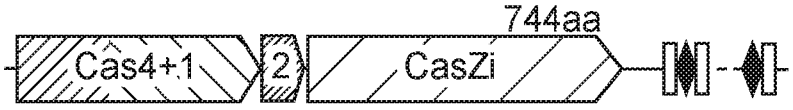
	#	Cas1	Organism	Locus example
Za	5	I-B	DPANN	
Zb	7	I-B	DPANN	
Zc	5	I-B	DPANN	
Zd	1	I-B	DPANN	
Ze	4	III-D*	CPR	
Zf	3	III-D	CPR	
Zg	1	I-E	Phage?	
Zh	1	II-C	MGE	
Zi	1	I-U	Plancto	

FIG. 3

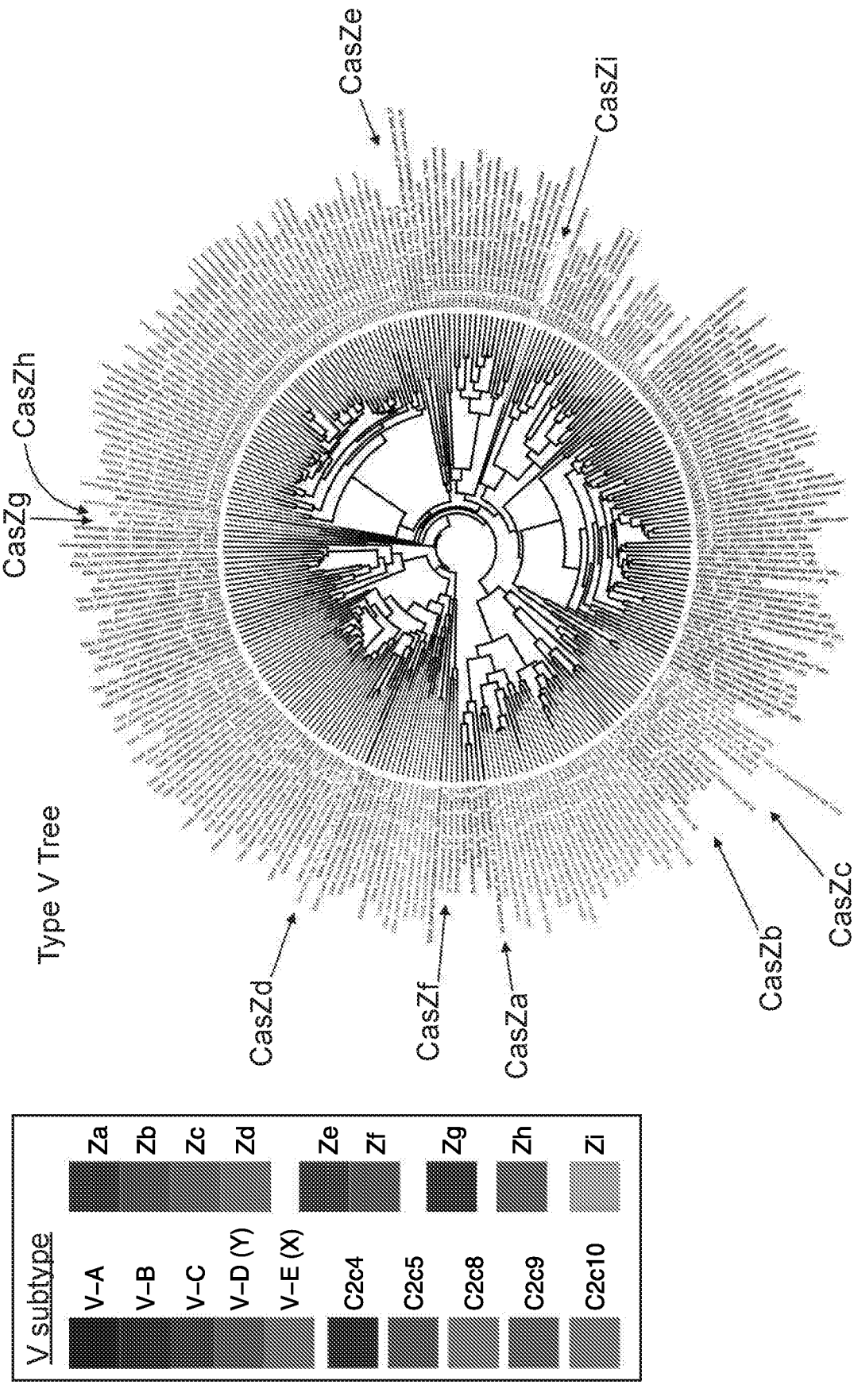

Type V Tree
CasZe
CasZi
CasZh
CasZg
CasZd
CasZf
CasZa
CasZb
CasZc
V subtype
V-A
V-B
V-C
V-D (Y)
V-E (X)
C2c4
C2c5
C2c8
C2c9
C2c10
Za
Zb
Zc
Zd
Ze
Zf
Zg
Zh
Zi

FIG. 4

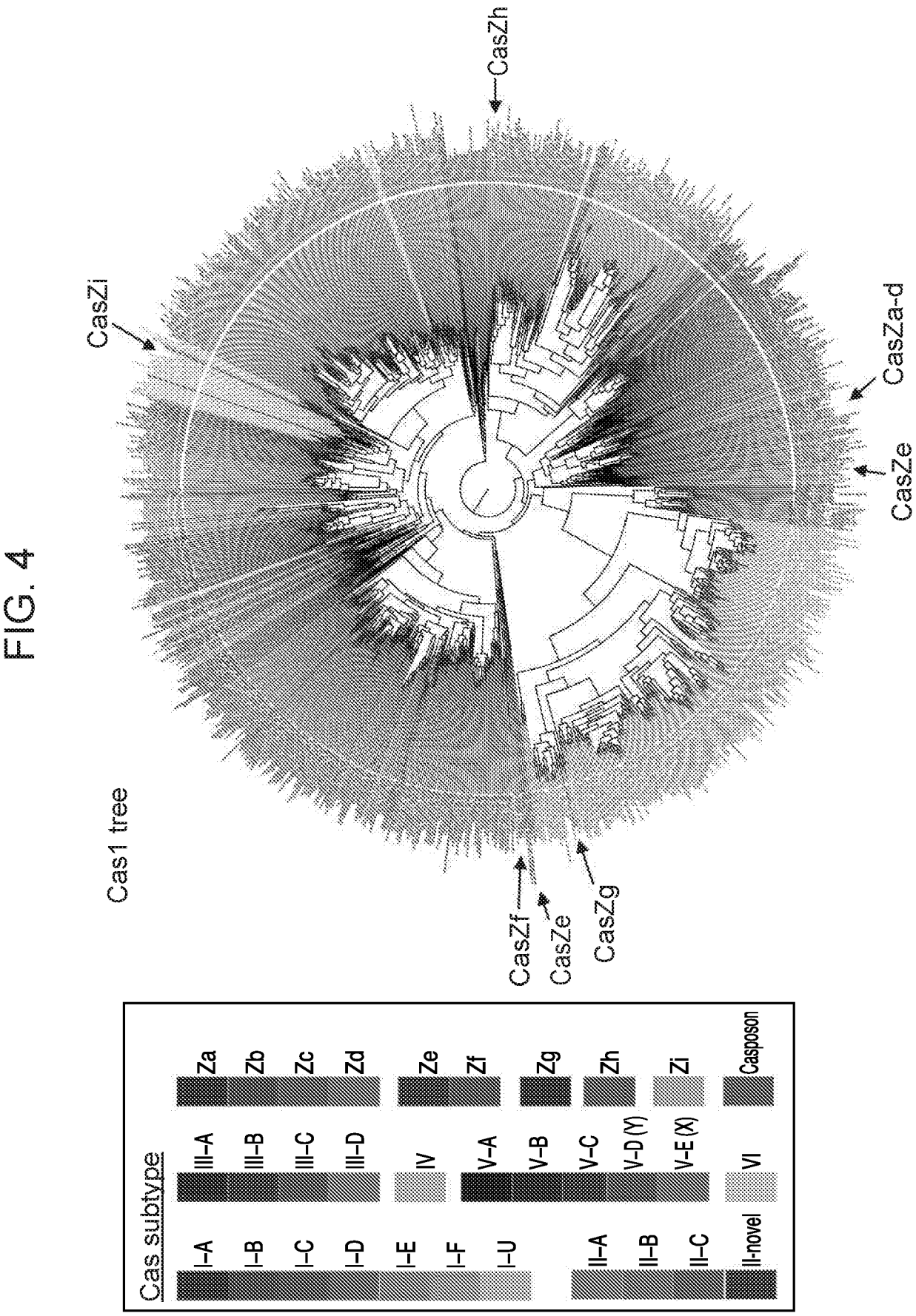

CasZi
CasZh
CasZa-d
CasZe
Cas1 tree
CasZf
CasZe
CasZg
Cas subtype
I–A
I–B
I–C
I–D
I–E
I–F
I–U
III–A
III–B
III–C
III–D
IV
V–A
V–B
V–C
V–D (Y)
V–E (X)
VI
II–A
II–B
II–C
II-novel
Za
Zb
Zc
Zd
Ze
Zf
Zg
Zh
Zi
Casposon

Figure 5

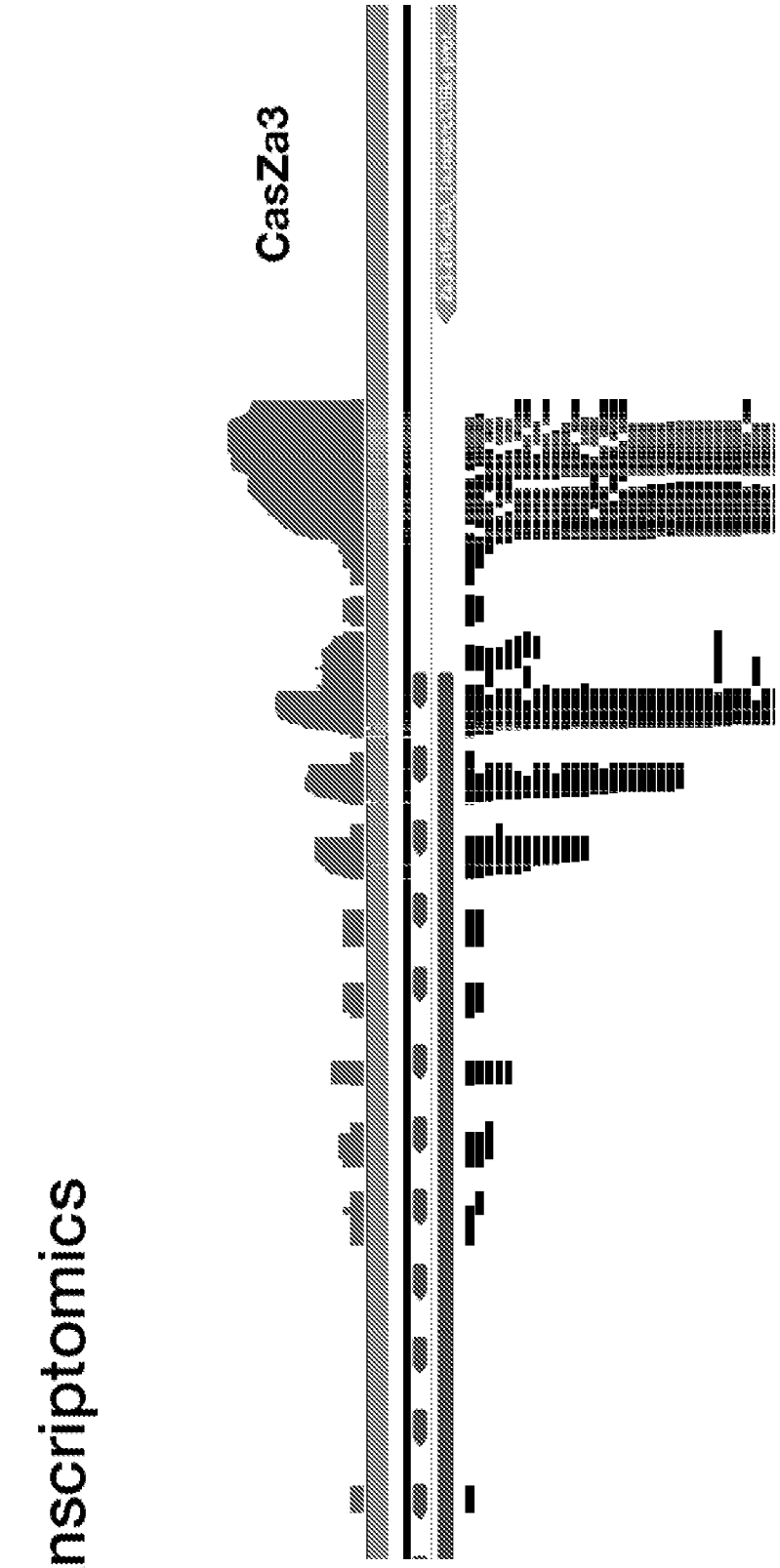


Figure 5 (Cont. 1)

Processing
(Putative guide:
17RP + 18SP)

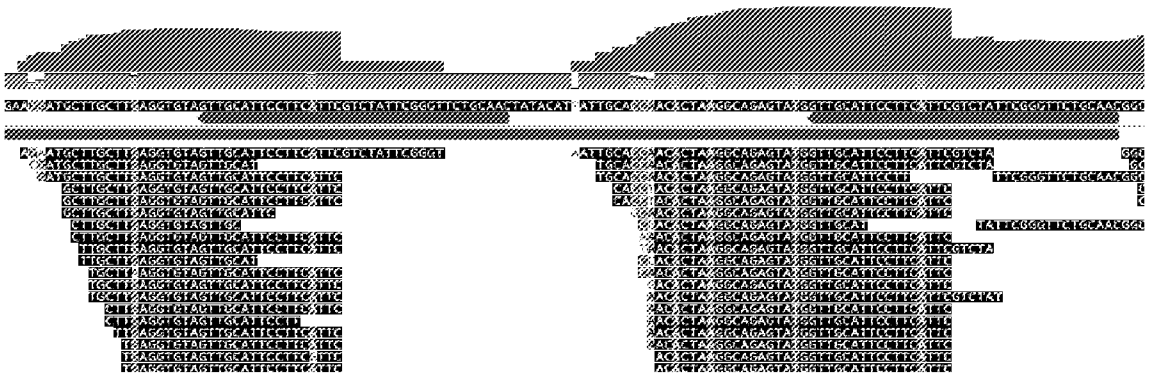


Figure 5 (Cont. 2)

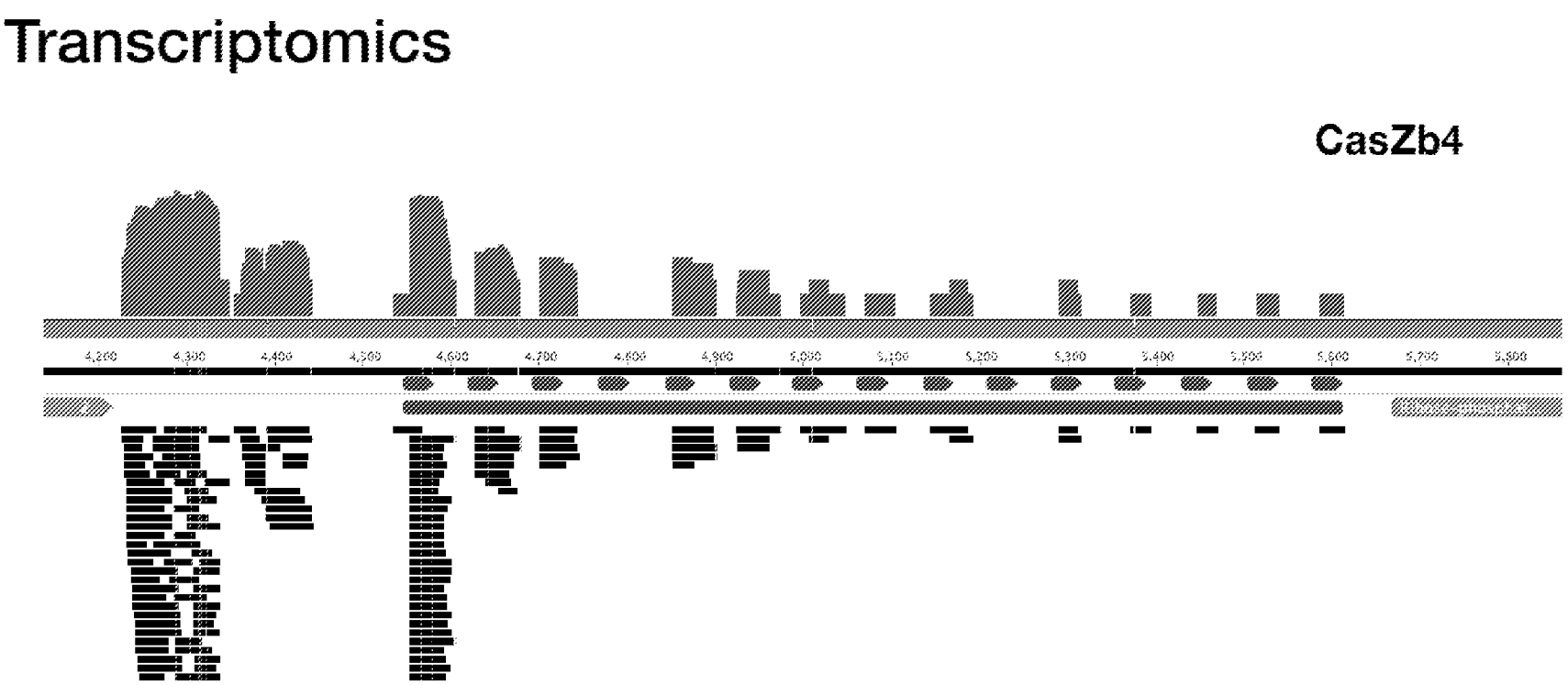


Figure 5 (Cont. 3)

Processing (Putative guide: 29RP + 21SP)

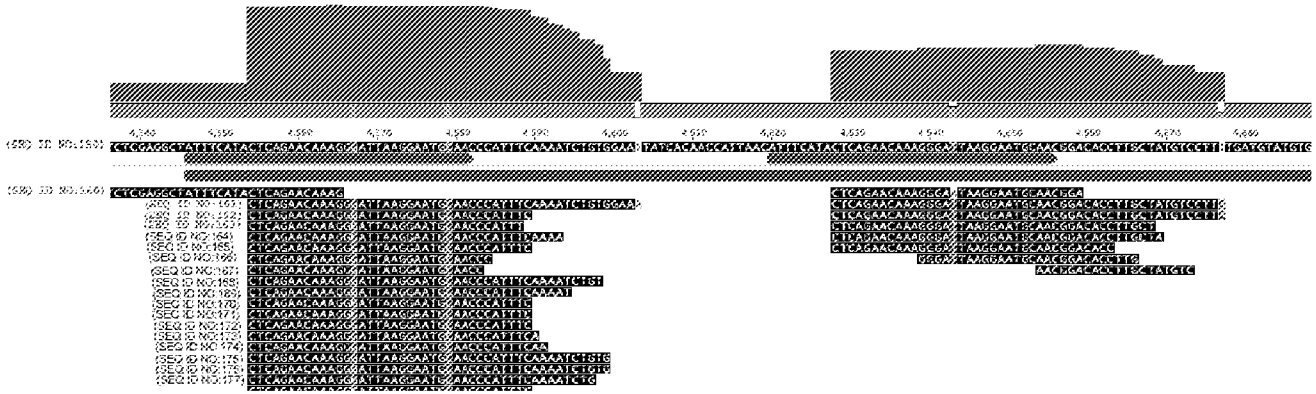
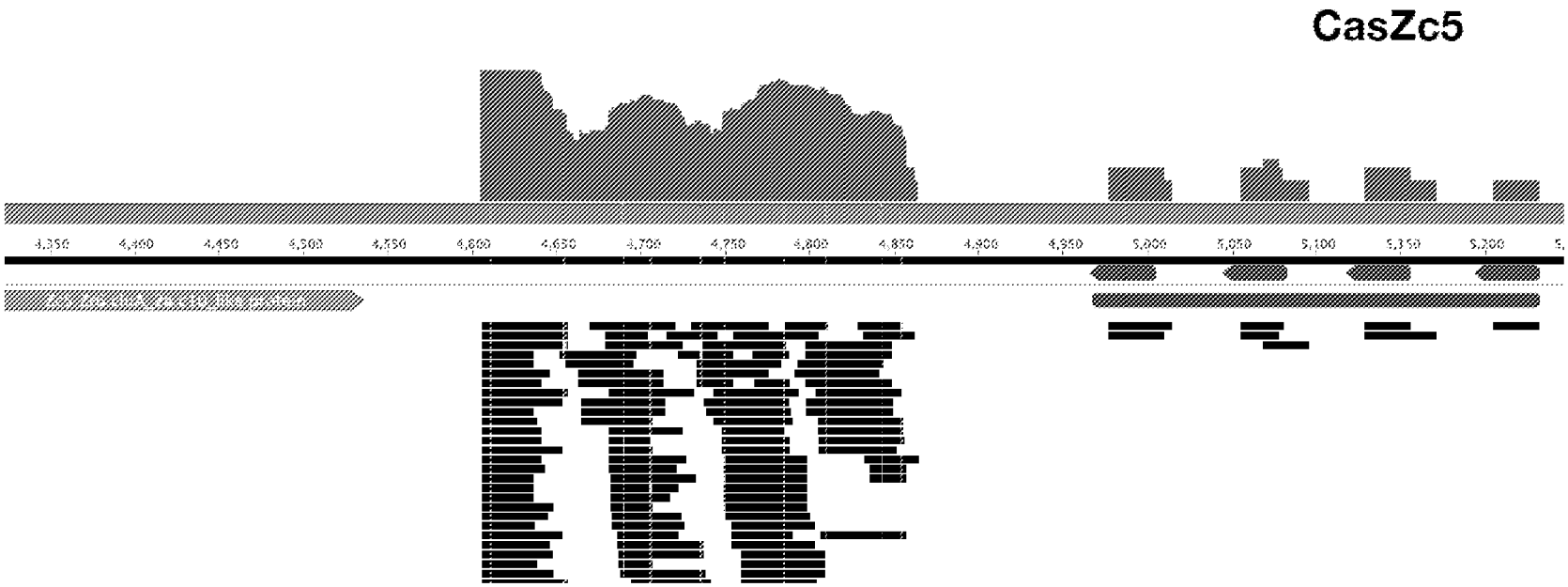


Figure 5 (Cont. 4)

Transcriptomics



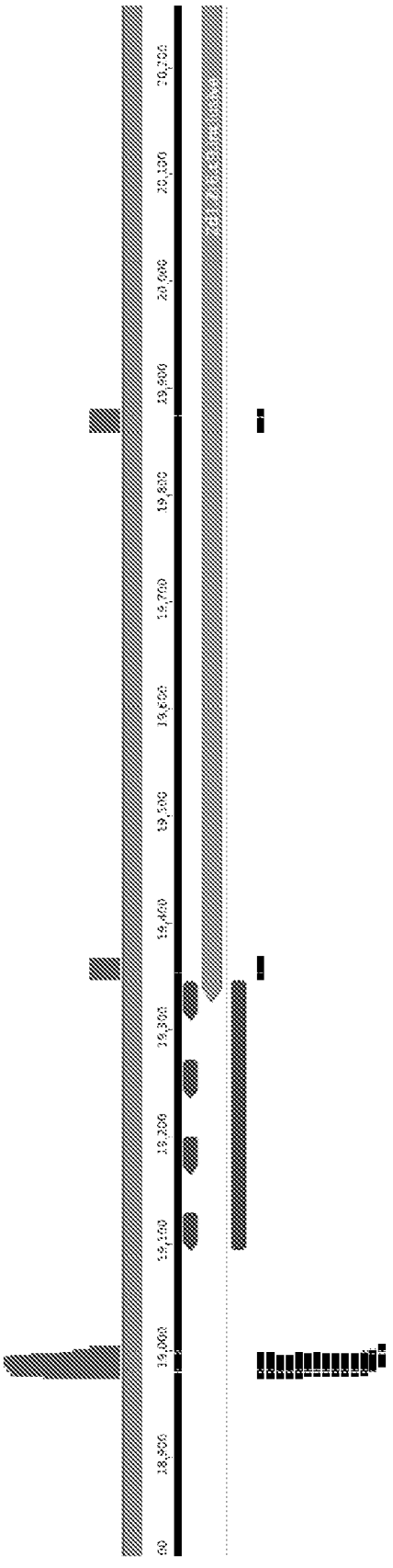


Figure 5 (Cont. 5)

Figure 5 (Cont. 6)

Transcriptomics

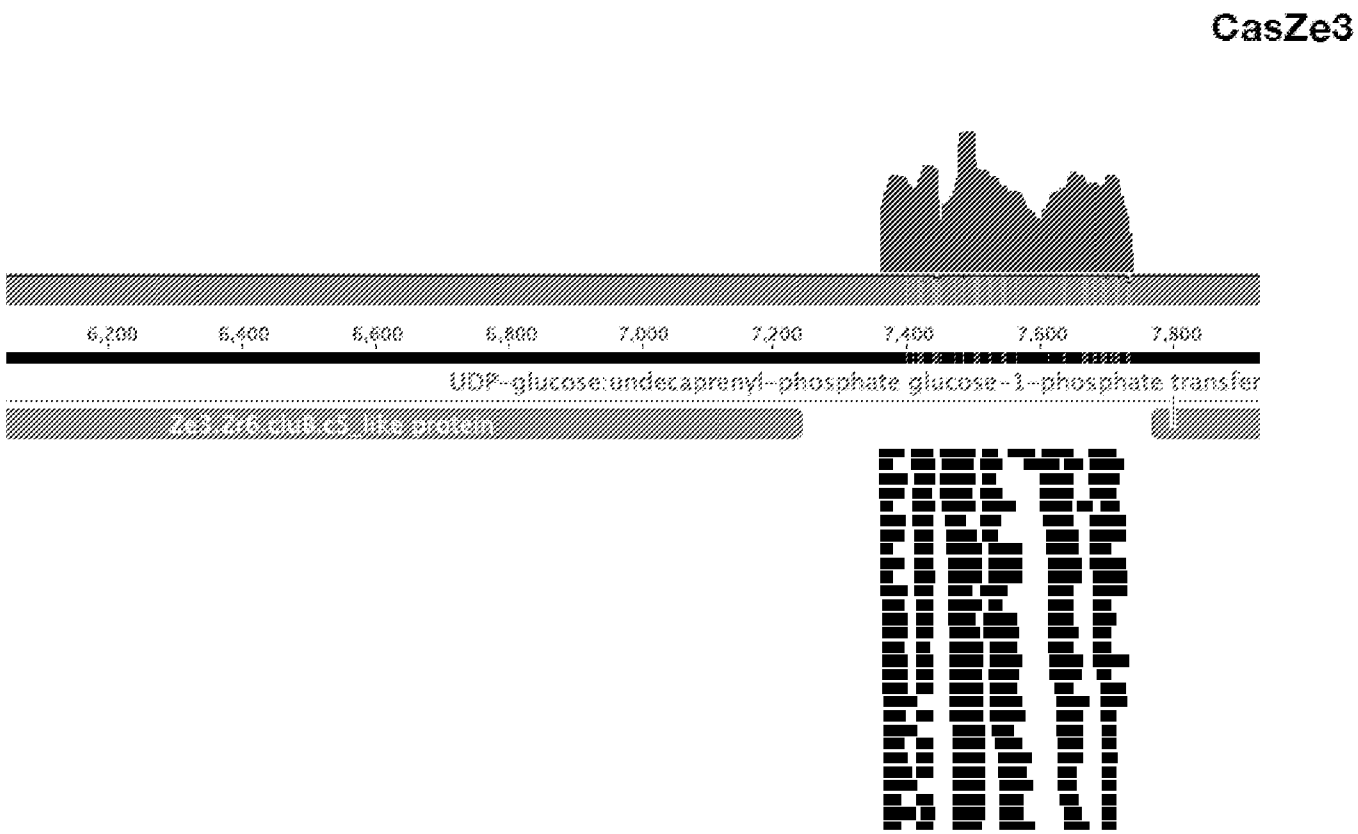


Figure 6

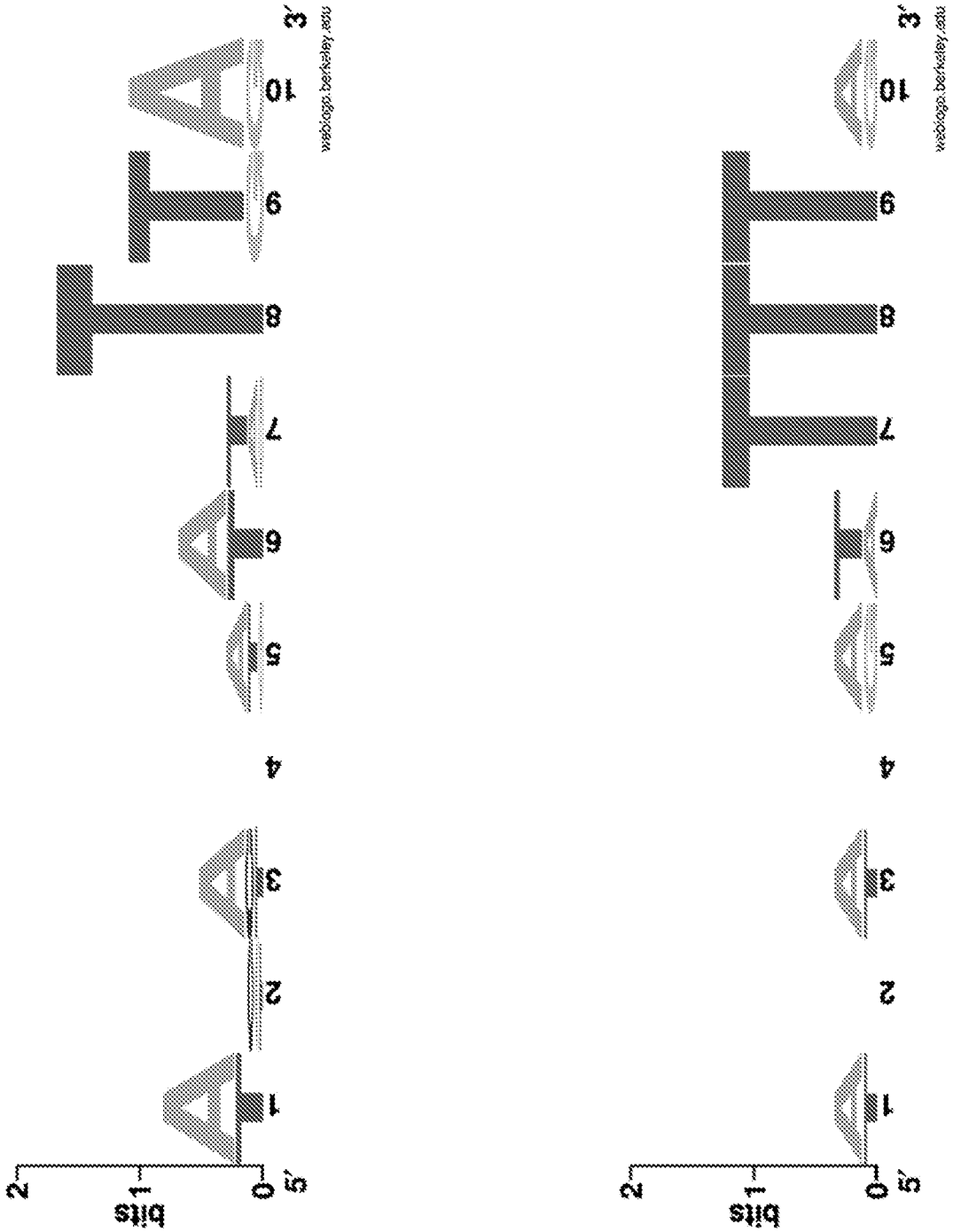

bits
2
1
0
5'
1
2
3
4
5
6
7
8
9
10
3'
weblogo.berkeley.edu
bits
2
1
0
5'
1
2
3
4
5
6
7
8
9
10
3'
weblogo.berkeley.edu

FIG. 7

>Cas14a.1|gwa1_scaffold_1795_curated|25635..27224|revcom
PEEKTSKLKPNSINLAANYDANEKFNCKECKFHPFKNKKRYEFNFYNNLHGCKSCTKST
NNPAVKRIEIGYQKLKFEIKNQMEAYPWFGRLRINFYSDEKRKMSELNEMQVTGVKNKI
FFDAIECAWREILKKRFRESKETLITIPKLKNKAGHGARKHRNKKLLIRRRAFMKKNFH
FLDNDSISYRSFANNIACVLPSKVGVDIGGIISPDVGKDIKPVDISLNLMWASKEGIKS
GRKVEIYSTQYDGNMVKKIEAETGEDKSWGKNRKRRQTSLLLSIPKPSKQVQEFDFKEW
PRYKDIEKKVQWRGFPIKIIFDSNHNSIEFGTYQGGKQKVLPIPFNDSKTTPLGSKMNK
LEKLRFNSKIKSRLGSAIAANKFLEAARTYCVDSLYHEVSSANAIGKGKIFIEYYLEIL
SQNYIEAAQKQLQRFIESIEQWFVADPFQGRLKQYFKDDLKRAKCFLCANREVQTTCYA
AVKLHKSCAEKVKDKNKELAIKERNNKEDAVIKEVEASNYPRVIRLKLTKTITNKAM
(SEQ ID NO:306)

>Cas14a.2|gwa2_scaffold_18027_curated|7105..8628
ELIVNENKDPLNIGKTAKAVFKEIDPTSINRAANYDASIELACKECKFKPFNNTKRHDF
SFYSNWHRCSPNSCLQSTYRAKIRKTEIGYEKLKNEILNQMQYYPWFGRLYQNFFNDQR
DKMTSLDEIQVTGVQNKIFFNTVEKAWREIIKKRFRDNKETMRTIPDLKNKSGHGSRKL
SNKSLLRRRFAFAQKSFKLVDNSDVSYRAFSNNVACVLPSKIGVDIGGIINKDLKREYI
PQEITFNVFWKQHDGLKKGRNIEIHSVQYKGEIVKRIEADTGEDKAWGKNRQRRFTSLI
LKITPKQGGKKIWKFPEKKNASDYEYFPIPIEFILDNGDASIKFGGEEGEVGKQKHLLI
PFNDSKATPLSSKQMLLETSRFNAEVKSTIGLALYANYFVSYARNYVIKSTYHKNSKKG
QIVTEIYLESISQNFVRAIQRQLQSLMLNLKDWGFMQTHKKELKKYFGSDLEGSKGGQK
RREKEEKIEKEIEASYLPRLIRLSLTKSVTKAEEM (SEQ ID NO:305)

>Cas14a.3|rifcsphigho2_02_scaffold_2167_curated|30296..31798|revcom
KEPLNIGKTAKAVFKEIDPTSLNRAANYDASIELNCKECKFKPFKNVKRYEFNFYNNWY
RCNPNSCLQSTYKAQVRKVEIGYEKLKNEILTQMQYYPWFGRLYQNFFHDERDKMTSLD
EIQVIGVQNKVFFNTVEKAWREIIKKRFKDNKETMETIPELKHAAGHGKRKLSNKSLLR
RRFAFVQKSFKFVDNSDVSYRSFSNNIACVLPSRIGVDLGGVISRNPKREYIPQEISFN
AFWKQHEGLKKGRNIEIQSVQYKGETVKRIEADTGEDKAWGKNRQRRFTSLILKLVPKQ
GGKKVWKYPEKRNEGNYEYFPIPIEFILDSGETSIRFGGDEGEAGKQKHLVIPFNDSKA
TPLASQQTLLENSRFNAEVKSCIGLAIYANYFYGYARNYVISSIYHKNSKNGQAITAIY
LESIAHNYVKAIERQLQNLLLNLRDFSFMESHKKELKKYFGGDLEGTGGAQKRREKEEK
IEKEIEQSYLPRLIRLSLTKMVTKQVEM (SEQ ID NO:304)

FIG. 7 (Cont.)

>Cas14a.4|CG10_big_fil_rev_8_21_14_0.10_scaffold_20906_curated|649..2829
SESENKIIEQYYAFLYSFRDKYEKPEFKNRGDIKRKLQNKWEDFLKEQNLKNDKKLSNY
IFSNRNFRRSYDREEENEEGIDEKKSKPKRINCFEKEKNLKDQYDKDAINASANKDGAQ
KWGCFECIFFPMYKIESGDPNKRIIINKTRFKLFDFYLNLKGCKSCLRSTYHPYRSNVY
IESNYDKLKREIGNFLQQKNIFQRMRKAKVSEGKYLTNLDEYRLSCVAMHFKNRWLFFD
SIQKVLRETIKQRLKQMRESYDEQAKTKRSKGHGRAKYEDQVRMIRRRAYSAQAHKLLD
NGYITLFDYDDKEINKVCLTAINQEGFDIGGYLNSDIDNVMPPIEISFHLKWKYNEPIL
NIESPFSKAKISDYLRKIREDLNLERGKEGKARSKKNVRRKVLASKGEDGYKKIFTDFF
SKWKEELEGNAMERVLSQSSGDIQWSKKKRIHYTTLVLNINLLDKKGVGNLKYYEIAEK
TKILSFDKNENKFWPITIQVLLDGYEIGTEYDEIKQLNEKTSKQFTIYDPNTKIIKIPF
TDSKAVPLGMLGINIATLKTVKKTERDIKVSKIFKGGLNSKIVSKIGKGIYAGYFPTVD
KEILEEVEEDTLDNEFSSKSQRNIFLKSIIKNYDKMLKEQLFDFYSFLVRNDLGVRFLT
DRELQNIEDESFNLEKRFFETDRDRIARWFDNTNTDDGKEKFKKLANEIVDSYKPRLIR
LPVVRVIKRIQPVKQREM (SEQ ID NO:178)

>Cas14a.5|rifcsplowo2_01_scaffold_34461_curated|4968..6521
KYSTRDFSELNEIQVTACKQDEFFKVIQNAWREIIKKRFLENRENFIEKKIFKNKKGRG
KRQESDKTIQRNRASVMKNFQLIENEKIILRAPSGHVACVFPVKVGLDIGGFKTDDLEK
NIFPPRTITINVFWKNRDRQRKGRKLEVWGIKARTKLIEKVHKWDKLEEVKKKRLKSLE
QKQEKSLDNWSEVNNDSFYKVQIDELQEKIDKSLKGRTMNKILDNKAKESKEAEGLYIE
WEKDFEGEMLRRIEASTGGEEKWGKRRQRRHTSLLLDIKNNSRGSKEIINFYSYAKQGK
KEKKIEFFPFPLTITLDAEEESPLNIKSIPIEDKNATSKYFSIPFTETRATPLSILGDR
VQKFKTKNISGAIKRNLGSSISSCKIVQNAETSAKSILSLPNVKEDNNMEIFINTMSKN
YFRAMMKQMESFIFEMEPKTLIDPYKEKAIKWFEVAASSRAKRKLKKLSKADIKKSELL
LSNTEEFEKEKQEKLEALEKEIEEFYLPRIVRLQLTKTILETPVM (SEQ ID
NO:179)

>Cas14a.6|3300012359.a|Ga0137385_10000156|41289..42734
KKLQLLGHKILLKEYDPNAVNAAANFETSTAELCGQCKMKPFKNKRRFQYTFGKNYHGC
LSCIQNVYYAKKRIVQIAKEELKHQLTDSIASIPYKYTSLFSNTNSIDELYILKQERAA
FFSNTNSIDELYITGIENNIAFKVISAIWDEIIKKRRQRYAESLTDTGTVKANRGHGGT
AYKSNTRQEKIRALQKQTLHMVTNPYISLARYKNNYIVATLPRTIGMHIGAIKDRDPQK
KLSDYAINFNVFWSDDRQLIELSTVQYTGDMVRKIEAETGENNKWGENMKRTKTSLLLE
ILTKKTTDELTFKDWAFSTKKEIDSVTKKTYQGFPIGIIFEGNESSVKFGSQNYFPLPF
DAKITPPTAEGFRLDWLRKGSFSSQMKTSYGLAIYSNKVTNAIPAYVIKNMFYKIARAE
NGKQIKAKFLKKYLDIAGNNYVPFIIMQHYRVLDTFEEMPISQPKVIRLSLTKTQHIII
KKDKTDSKM (SEQ ID NO:180)

FIG. 7 (Cont.)

>Cas14b.1|rifcsplowo2_01_scaffold_239_curated|54653..56257
NTSNLINLGKKAINISANYDANLEVGCKNCKFLSSNGNFPRQTNVKEGCHSCEKSTYEP
SIYLVKIGERKAKYDVLDSLKKFTFQSLKYQSKKSMKSRNKKPKELKEFVIFANKNKAF
DVIQKSYNHLILQIKKEINRMNSKKRKKNHKRRLFRDREKQLNKLRLIESSNLFLPREN
KGNNHVFTYVAIHSVGRDIGVIGSYDEKLNFETELTYQLYFNDDKRLLYAYKPKQNKII
KIKEKLWNLRKEKEPLDLEYEKPLNKSITFSIKNDNLFKVSKDLMLRRAKFNIQGKEKL
SKEERKINRDLIKIKGLVNSMSYGRFDELKKEKNIWSPHIYREVRQKEIKPCLIKNGDR
IEIFEQLKKKMERLRRFREKRQKKISKDLIFAERIAYNFHTKSIKNTSNKINIDQEAKR
GKASYMRKRIGYETFKNKYCEQCLSKGNVYRNVQKGCSCFENPFDWIKKGDENLLPKKN
EDLRVKGAFRDEALEKQIVKIAFNIAKGYEDFYDNLGESTEKDLKLKFKVGTTINEQES
LKL (SEQ ID NO:181)

>Cas14b.2|rifcsplowo2_01_scaffold_282_curated|77370..78983
TSNPIKLGKKAINISANYDSNLQIGCKNCKFLSYNGNFPRQTNVKEGCHSCEKSTYEPP
VYTVRIGERRSKYDVLDSLKKFIFLSLKYRQSKKMKTRSKGIRGLEEFVISANLKKAMD
VIQKSYRHLILNIKNEIVRMNGKKRNKNHKRLLFRDREKQLNKLRLIEGSSFFKPPTVK
GDNSIFTCVAIHNIGRDIGIAGDYFDKLEPKIELTYQLYYEYNPKKESEINKRLLYAYK
PKQNKIIEIKEKLWNLRKEKSPLDLEYEKPLTKSITFLVKRDGVFRISKDLMLRKAKFI
IQGKEKLSKEERKINRDLIKIKSNIISLTYGRFDELKKDKTIWSPHIFRDVKQGKITPC
IERKGDRMDIFQQLRKKSERLRENRKKRQKKISKDLIFAERIAYNFHTKSIKNTSNLIN
IKHEAKRGKASYMRKRIGNETFRIKYCEQCFPKNNVYKNVQKGCSCFEDPFEYIKKGNE
DLIPNKNQDLKAKGAFRDDALEKQIIKVAFNIAKGYEDFYENLKKTTEKDIRLKFKVGT
IISEEM (SEQ ID NO:182)

>Cas14b.3|rifcsphigho2_01_scaffold_36781_curated|2592..4217
NNSINLSKKAINISANYDANLQVRCKNCKFLSSNGNFPRQTDVKEGCHSCEKSTYEPPV
YDVKIGEIKAKYEVLDSLKKFTFQSLKYQLSKSMKFRSKKIKELKEFVIFAKESKALNV
INRSYKHLILNIKNDINRMNSKKRIKNHKGRLFLDRQKQLSKLKLIEGSSFFVPAKNVG
NKSVFTCVAIHSIGRDIGIAGLYDSFTKPVNEITYQIFFSGERRLLYAYKPKQLKILSI
KENLWSLKNEKKPLDLLYEKPLGKNLNFNVKGGDLFRVSKDLMIRNAKFNVHGRQRLSD
EERLINRNFIKIKGEVVSLSYGRFEELKKDRKLWSPHIFKDVRQNKIKPCLVMQGQRID
IFEQLKRKLELLKKIRKSRQKKLSKDLIFGERIAYNFHTKSIKNTSNKINIDSDAKRGR
ASYMRKRIGNETFKLKYCDVCFPKANVYRRVQNGCSCSENPYNYIKKGDKDLLPKKDEG
LAIKGAFRDEKLNKQIIKVAFNIAKGYEDFYDDLKKRTEKDVDLKFKIGTTVLDQKPME
IFDGIVITWL (SEQ ID NO:183)

FIG. 7 (Cont.)

>Cas14b.4|cg1_0.2_scaffold_785_c_curated|32521..34155
LLTTVVETNNLAKKAINVAANFDANIDRQYYRCTPNLCRFIAQSPRETKEKDAGCSSCT
QSTYDPKVYVIKIGKLLAKYEILKSLKRFLFMNRYFKQKKTERAQQKQKIGTELNEMSI
FAKATNAMEVIKRATKHCTYDIIPETKSLQMLKRRRHRVKVRSLLKILKERRMKIKKIP
NTFIEIPKQAKKNKSDYYVAAALKSCGIDVGLCGAYEKNAEVEAEYTYQLYYEYKGNSS
TKRILYCYNNPQKNIREFWEAFYIQGSKSHVNTPGTIRLKMEKFLSPITIESEALDFRV
WNSDLKIRNGQYGFIKKRSLGKEAREIKKGMGDIKRKIGNLTYGKSPSELKSIHVYRTE
RENPKKPRAARKKEDNFMEIFEMQRKKDYEVNKKRRKEATDAAKIMDFAEEPIRHYHTN
NLKAVRRIDMNEQVERKKTSVFLKRIMQNGYRGNYCRKCIKAPEGSNRDENVLEKNEGC
LDCIGSEFIWKKSSKEKKGLWHTNRLLRRIRLQCFTTAKAYENFYNDLFEKKESSLDII
KLKVSITTKSM (SEQ ID NO:184)

>Cas14b.5|rifcsphigho2_02_scaffold_55589_curated|1904..3598
ASTMNLAKQAINFAANYDSNLEIGCKGCKFMSTWSKKSNPKFYPRQNNQANKCHSCTYS
TGEPEVPIIEIGERAAKYKIFTALKKFVFMSVAYKERRRQRFKSKKPKELKELAICSNR
EKAMEVIQKSVVHCYGDVKQEIPRIRKIKVLKNHKGRLFYKQKRSKIKIAKLEKGSFFK
TFIPKVHNNGCHSCHEASLNKPILVTTALNTIGADIGLINDYSTIAPTETDISWQVYYE
FIPNGDSEAVKKRLLYFYKPKGALIKSIRDKYFKKGHENAVNTGFFKYQGKIVKGPIKF
VNNELDFARKPDLKSMKIKRAGFAIPSAKRLSKEDREINRESIKIKNKIYSLSYGRKKT
LSDKDIIKHLYRPVRQKGVKPLEYRKAPDGFLEFFYSLKRKERRLRKQKEKRQKDMSEI
IDAADEFAWHRHTGSIKKTTNHINFKSEVKRGKVPIMKKRIANDSFNTRHCGKCVKQGN
AINKYYIEKQKNCFDCNSIEFKWEKAALEKKGAFKLNKRLQYIVKACFNVAKAYESFYE
DFRKGEEESLDLKFKIGTTTTLKQYPQNKARAM (SEQ ID NO:185)

>Cas14b.6|CG03_land_8_20_14_0.80_scaffold_2214_curated|6634
..8466|revcom
HSHNLMLTKLGKQAINFAANYDANLEIGCKNCKFLSYSPKQANPKKYPRQTDVHEDGNI
ACHSCMQSTKEPPVYIVPIGERKSKYEILTSLNKFTFLALKYKEKKRQAFRAKKPKELQ
ELAIAFNKEKAIKVIDKSIQHLILNIKPEIARIQRQKRLKNRKGKLLYLHKRYAIKMGL
IKNGKYFKVGSPKKDGKKLLVLCALNTIGRDIGIIGNIEENNRSETEITYQLYFDCLDA
NPNELRIKEIEYNRLKSYERKIKRLVYAYKPKQTKILEIRSKFFSKGHENKVNTGSFNF
ENPLNKSISIKVKNSAFDFKIGAPFIMLRNGKFHIPTKKRLSKEEREINRTLSKIKGRV
FRLTYGRNISEQGSKSLHIYRKERQHPKLSLEIRKQPDSFIDEFEKLRLKQNFISKLKK
QRQKKLADLLQFADRIAYNYHTSSLEKTSNFINYKPEVKRGRTSYIKKRIGNEGFEKLY
CETCIKSNDKENAYAVEKEELCFVCKAKPFTWKKTNKDKLGIFKYPSRIKDFIRAAFTV
AKSYNDFYENLKKKDLKNEIFLKFKIGLILSHEKKNHISIAKSVAEDERISGKSIKNIL
NKSIKLEKNCYSCFFHKEDM (SEQ ID NO:186)

FIG. 7 (Cont.)

>Cas14b.7|3300013125.a|Ga0172369_10000737|994..2652|revcom
SLERVIDKRNLAKKAINIAANFDANINKGFYRCETNQCMFIAQKPRKTNNTGCSSCLQS
TYDPVIYVVKVGEMLAKYEILKSLKRFVFMNRSFKQKKTEKAKQKERIGGELNEMSIFA
NAALAMGVIKRAIRHCHVDIRPEINRLSELKKTKHRVAAKSLVKIVKQRKTKWKGIPNS
FIQIPQKARNKDADFYVASALKSGGIDIGLCGTYDKKPHADPRWTYQLYFDTEDESEKR
LLYCYNDPQAKIRDFWKTFYERGNPSMVNSPGTIEFRMEGFFEKMTPISIESKDFDFRV
WNKDLLIRRGLYEIKKRKNLNRKAREIKKAMGSVKRVLANMTYGKSPTDKKSIPVYRVE
REKPKKPRAVRKEENELADKLENYRREDFLIRNRRKREATEIAKIIDAAEPPIRHYHTN
HLRAVKRIDLSKPVARKNTSVFLKRIMQNGYRGNYCKKCIKGNIDPNKDECRLEDIKKC
ICCEGTQNIWAKKEKLYTGRINVLNKRIKQMKLECFNVAKAYENFYDNLAALKEGDLKV
LKLKVSIPALNPEASDPEEDM (SEQ ID NO:187)

>Cas14b.8|3300013125.a|Ga0172369_10010464|885..2489|revcom
NASINLGKRAINLSANYDSNLVIGCKNCKFLSFNGNFPRQTNVREGCHSCDKSTYAPEV
YIVKIGERKAKYDVLDSLKKFTFQSLKYQIKKSMRERSKKPKELLEFVIFANKDKAFNV
IQKSYEHLILNIKQEINRMNGKKRIKNHKKRLFKDREKQLNKLRLIGSSSLFFPRENKG
DKDLFTYVAIHSVGRDIGVAGSYESHIEPISDLTYQLFINNEKRLLYAYKPKQNKIIEL
KENLWNLKKEKKPLDLEFTKPLEKSITFSVKNDKLFKVSKDLMLRQAKFNIQGKEKLSK
EERQINRDFSKIKSNVISLSYGRFEELKKEKNIWSPHIYREVKQKEIKPCIVRKGDRIE
LFEQLKRKMDKLKKFRKERQKKISKDLNFAERIAYNFHTKSIKNTSNKINIDQEAKRGK
ASYMRKRIGNESFRKKYCEQCFSVGNVYHNVQNGCSCFDNPIELIKKGDEGLIPKGKED
RKYKGALRDDNLQMQIIRVAFNIAKGYEDFYNNLKEKTEKDLKLKFKIGTTISTQESNN
KEM (SEQ ID NO:188)

>Cas14b.9|3300013127.a|Ga0172365_10004421|633..2366|revcom
SNLIKLGKQAINFAANYDANLEVGCKNCKFLSSTNKYPRQTNVHLDNKMACRSCNQSTM
EPAIYIVRIGEKKAKYDIYNSLTKFNFQSLKYKAKRSQRFKPKQPKELQELSIAVRKEK
ALDIIQKSIDHLIQDIRPEIPRIKQQKRYKNHVGKLFYLQKRRKNKLNLIGKGSFFKVF
SPKEKKNELLVICALTNIGRDIGLIGNYNTIINPLFEVTYQLYYDYIPKKNNKNVQRRL
LYAYKSKNEKILKLKEAFFKRGHENAVNLGSFSYEKPLEKSLTLKIKNDKDDFQVSPSL
RIRTGRFFVPSKRNLSRQEREINRRLVKIKSKIKNMTYGKFETARDKQSVHIFRLERQK
EKLPLQFRKDEKEFMEEFQKLKRRTNSLKKLRKSRQKKLADLLQLSEKVVYNNHTGTLK
KTSNFLNFSSSVKRGKTAYIKELLGQEGFETLYCSNCINKGQKTRYNIETKEKCFSCKD
VPFVWKKKSTDKDRKGAFLFPAKLKDVIKATFTVAKAYEDFYDNLKSIDEKKPYIKFKI
GLILAHVRHEHKARAKEEAGQKNIYNKPIKIDKNCKECFFFKEEAM (SEQ ID NO:189
)

FIG. 7 (Cont.)

>Cas14b.10|CG08_land_8_20_14_0.20_scaffold_1609_curated|6134..7975
NTTRKKFRKRTGFPQSDNIKLAYCSAIVRAANLDADIQKKHNQCNPNLCVGIKSNEQSRKYEHSDRQALLCYACNQSTGAPKVDYIQIGEIGAKYKILQMVNAYDFLSLAYNLTKLRNGKSRGHQRMSQLDEVVIVADYEKATEVIKRSINHLLDDIRGQLSKLKKRTQNEHITEHKQSKIRRKLRKLSRLLKRRRWKWGTIPNPYLKNWVFTKKDPELVTVALLHKLGRDIGLVNRSKRRSKQKLLPKVGFQLYYKWESPSLNNIKKSKAKKLPKRLLIPYKNVKLFDNKQKLENAIKSLLESYQKTIKVEFDQFFQNRTEEIIAEEQQTLERGLLKQLEKKKNEFASQKKALKEEKKKIKEPRKAKLLMEESRSLGFLMANVSYALFNTTIEDLYKKSNVVSGCIPQEPVVVFPADIQNKGSLAKILFAPKDGFRIKFSGQHLTIRTAKFKIRGKEIKILTKTKREILKNIEKLRRVWYREQHYKLKLFGKEVSAKPRFLDKRKTSIERRDPNKLADQTDDRQAELRNKEYELRHKQHKMAERLDNIDTNAQNLQTLSFWVGEADKPPKLDEKDARGFGVRTCISAWKWFMEDLLKKQEEDPLLKLKLSIM (SEQ ID NO:190)

>Cas14b.11|CG_4_10_14_0.8_um_filter_scaffold_20762_curated|1372..3219
PKKPKFQKRTGFPQPDNLRKEYCLAIVRAANLDADFEKKCTKCEGIKTNKKGNIVKGRTYNSADKDNLLCYACNISTGAPAVDYVFVGALEAKYKILQMVKAYDFHSLAYNLAKLWKGRGRGHQRMGGLNEVVIVSNNEKALDVIEKSLNHFHDEIRGELSRLKAKFQNEHLHVHKESKLRRKLRKISRLLKRRRWKWDVIPNSYLRNFTFTKTRPDFISVALLHRVGRDIGLVTKTKIPKPTDLLPQFGFQIYYTWDEPKLNKLKKSRLRSEPKRLLVPYKKIELYKNKSVLEEAIRHLAEVYTEDLTICFKDFFETQKRKFVSKEKESLKRELLKELTKLKKDFSERKTALKRDRKEIKEPKKAKLLMEESRSLGFLAANTSYALFNLIAADLYTKSKKACSTKLPRQLSTILPLEIKEHKSTTSLAIKPEEGFKIRFSNTHLSIRTPKFKMKGADIKALTKRKREILKNATKLEKSWYGLKHYKLKLYGKEVAAKPRFLDKRNPSIDRRDPKELMEQIENRRNEVKDLEYEIRKGQHQMAKRLDNVDTNAQNLQTKSFWVGEADKPPELDSMEAKKLGLRTCISAWKWFMKDLVLLQEKSPNLKLKLSLTEM (SEQ ID NO:191)

>Cas14b.12|CG22_combo_CG10-13_8_21_14_all_scaffold_2003_curated|553..2880|revcom
KFSKRQEGFLIPDNIDLYKCLAIVRSANLDADVQGHKSCYGVKKNGTYRVKQNGKKGVKEKGRKYVFDLIAFKGNIEKIPHEAIEEKDQGRVIVLGKFNYKLILNIEKNHNDRASLEIKNKIKKLVQISSLETGEFLSDLLSGKIGIDEVYGIIEPDVFSGKELVCKACQQSTYAPLVEYMPVGELDAKYKILSAIKGYDFLSLAYNLSRNRANKKRGHQKLGGGELSEVVISANYDKALNVIKRSINHYHVEIKPEISKLKKKMQNEPLKVMKQARIRRELHQLSRKVKRLKWKWGMIPNPELQNIIFEKKEKDFVSYALLHTLGRDIGLFKDTSMLQVPNISDYGFQIYYSWEDPKLNSIKKIKDLPKRLLIPYKRLDFYIDTILVAKVIKNLIELYRKSYVYETFGEEYGYAKKAEDILFDWDSINLSEGIEQKIQKIKDEFSDLLYEARESKRQNFVESFENILGLYDKNFASDRNSYQEKIQSMIIKKQQENIEQKLKREFKEVIERGFEGMDQNKKYYKVLSPNIKGGLLYTDTNNLGFFRSHLAFMLLSKISDDLYRKNNLVSKGGNKGILDQTPETMLTLEFGKSNLPNISIKRKFFNIKYNSSWIGIRKPKFSIKGAVIREITKKVRDEQRLIKSLEGVWHKSTHFKRWGKPRFNLPRHPDREKNNDDNLMESITSRREQIQLLLREKQKQQEKMAGRLDKIDKEIQNLQTANFQIKQIDKKPALTEKSEGKQSVRNALSAWKWFMEDLIKYQKRTPILQLKLAKM (SEQ ID NO: 192)

FIG. 7 (Cont.)

>Cas14b.13|rifcsphigho2_01_scaffold_82367_curated|1523..3856|revcom
KFSKRQEGFVIPENIGLYKCLAIVRSANLDADVQGHVSCYGVKKNGTYVLKQNGKKSIR
EKGRKYASDLVAFKGDIEKIPFEVIEEKKKEQSIVLGKFNYKLVLDVMKGEKDRASLTM
KNKSKKLVQVSSLGTDEFLLTLLNEKFGIEEIYGIIEPEVFSGKKLVCKACQQSTYAPL
VEYMPVGELDSKYKILSAIKGYDFLSLAYNLARHRSNKKRGHQKLGGGELSEVVISANN
AKALNVIKRSLNHYYSEIKPEISKLRKKMQNEPLKVGKQARMRRELHQLSRKVKRLKWK
WGKIPNLELQNITFKESDRDFISYALLHTLGRDIGMFNKTEIKMPSNILGYGFQIYYDW
EEPKLNTIKKSKNTPKRILIPYKKLDFYNDSILVARAIKELVGLFQESYEWEIFGNEYN
YAKEAEVELIKLDEESINGNVEKKLQRIKENFSNLLEKAREKKRQNFIESFESIARLYD
ESFTADRNEYQREIQSFIIEKQKQSIEKKLKNEFKKIVEKKFNEQEQGKKHYRVLNPTI
INEFLPKDKNNLGFLRSKIAFILLSKISDDLYKKSNAVSKGGEKGIIKQQPETILDLEF
SKSKLPSINIKKKLFNIKYTSSWLGIRKPKFNIKGAKIREITRRVRDVQRTLKSAESSW
YASTHFRRWGFPRFNQPRHPDKEKKSDDRLIESITLLREQIQILLREKQKGQKEMAGRL
DDVDKKIQNLQTANFQIKQTGDKPALTEKSAGKQSFRNALSAWKWFMENLLKYQNKTPD
LKLKIARTVM (SEQ ID NO: 193)

>Cas14b.14|gwc1_scaffold_8732_curated|2705..4537
KWIEPNNIDFNKCLAITRSANLDADVQGHKMCYGIKTNGTYKAIGKINKKHNTGIIEKR
RTYVYDLIVTKEKNEKIVKKTDFMAIDEEIEFDEKKEKLLKKYIKAEVLGTGELIRKDL
NDGEKFDDLCSIEEPQAFRRSELVCKACNQSTYASDIRYIPIGEIEAKYKILKAIKGYD
FLSLKYNLGRLRDSKKRGHQKMGQGELKEFVICANKEKALDVIKRSLNHYLNEVKDEIS
RLNKKMQNEPLKVNDQARWRRELNQISRRLKRLKWKWGEIPNPELKNLIFKSSRPEFVS
YALIHTLGRDIGLINETELKPNNIQEYGFQIYYKWEDPELNHIKKVKNIPKRFIIPYKN
LDLFGKYTILSRAIEGILKLYSSSSFQYKSFKDPNLFAKEGEKKITNEDFELGYDEKIKK
IKDDFKSYKKALLEKKKNTLEDSLNSILSVYEQSLLTEQINNVKKWKEGLLKSKESIHK
QKKIENIEDIISRIEELKNVEGWIRTKERDIVNKEETNLKREIKKELKDSYYEEVRKDF
SDLKKGEESEKKPFREEPKPIVIKDYIKFDVLPGENSALGFFLSHLSFNLFDSIQYELF
EKSRLSSSKHPQIPETILDL (SEQ ID NO: 194)

>Cas14b.15|3300010293.a|Ga0116204_1008574|2134..4032
FRKFVKRSGAPQPDNLNKYKCIAIVRAANLDADIMSNESSNCVMCKGIKMNKRKTAKGA
AKTTELGRVYAGQSGNLLCTACTKSTMGPLVDYVPIGRIRAKYTILRAVKEYDFLSLAY
NLARTRVSKKGGRQKMHSLSELVIAAEYEIAWNIIKSSVIHYHQETKEEISGLRKKLQA
EHIHKNKEARIRREMHQISRRIKRLKWKWHMIPNSELHNFLFKQQDPSFVAVALLHTLG
RDIGMINKPKGSAKREFIPEYGFQIYYKWMNPKLNDINKQKYRKMPKRSLIPYKNLNVF
GDRELIENAMHKLLKLYDENLEVKGSKFFKTRVVAISSKESEKLKRDLLWKGELAKIKK
DFNADKNKMQELFKEVKEPKKANALMKQSRNMGFLLQNISYGALGLLANRMYEASAKQS
KGDATKQPSIVIPLEMEFGNAFPKLLLRSGKFAMNVSSPWLTIRKPKFVIKGNKIKNIT
KLMKDEKAKLKRLETSYHRATHFRPTLRGSIDWDSPYFSSPKQPNTHRRSPDRLSADIT
EYRGRLKSVEAELREGQRAMAKKLDSVDMTASNLQTSNFQLEKGEDPRLTEIDEKGRSI
RNCISSWKKFMEDLMKAQEANPVIKIKIALKDESSVLSEDSM (SEQ ID NO: 195)

FIG. 7 (Cont.)

>Cas14b.16|3300005573.a|Ga0078972_1001015a|33750..35627
KFHPENLNKSYCLAIVRAANLDADIQGHINCIGIKSNKSDRNYENKLESLQNVELLCKA
CTKSTYKPNINSVPVGEKKAKYSILSEIKKYDFNSLVYNLKKYRKGKSRGHQKLNELRE
LVITSEYKKALDVINKSVNHYLVNIKNKMSKLKKILQNEHIHVGTLARIRRERNRISRK
LDHYRKKWKFVPNKILKNYVFKNQSPDFVSVALLHKLGRDIGLITKTAILQKSFPEYSL
QLYYKYDTPKLNYLKKSKFKSLPKRILISYKYPKFDINSNYIEESIDKLLKLYEESPIY
KNNSKIIEFFKKSEDNLIKSENDSLKRGIMKEFEKVTKNFSSKKKKLKEELKLKNEDKN
SKMLAKVSRPIGFLKAYLSYMLFNIISNRIFEFSRKSSGRIPQLPSCIINLGNQFENFK
NELQDSNIGSKKNYKYFCNLLLKSSGFNISYEEEHLSIKTPNFFINGRKLKEITSEKKK
IRKENEQLIKQWKKLTFFKPSNLNGKKTSDKIRFKSPNNPDIERKSEDNIVENIAKVKY
KLEDLLSEQRKEFNKLAKKHDGVDVEAQCLQTKSFWIDSNSPIKKSLEKKNEKVSVKKK
MKAIRSCISAWKWFMADLIEAQKETPMIKLKLALM (SEQ ID NO: 196)

>Cas14c.1|CG10_big_fil_rev_8_21_14_0.10_scaffold_4477_curat
ed|19327..20880|revcom
TTLVPSHLAGIEVMDETTSRNEDMIQKETSRSNEDENYLGVKNKCGINVHKSGRGSSKH
EPNMPPEKSGEGQMPKQDSTEMQQRFDESVTGETQVSAGATASIKTDARANSGPRVGTA
RALIVKASNLDRDIKLGCKPCEYIRSELPMGKKNGCNHCEKSSDIASVPKVESGFRKAK
YELVRRFESFAADSISRHLGKEQARTRGKRGKKDKKEQMGKVNLDEIAILKNESLIEYT
ENQILDARSNRIKEWLRSLRLRLRTRNKGLKKSKSIRRQLITLRRDYRKWIKPNPYRPD
EDPNENSLRLHTKLGVDIGVQGGDNKRMNSDDYETSFSITWRDTATRKICFTKPKGLLP
RHMKFKLRGYPELILYNEELRIQDSQKFPLVDWERIPIFKLRGVSLGKKKVKALNRITE
APRLVVAKRIQVNIESKKKKVLTRYVYNDKSINGRLVKAEDSNKDPLLEFKKQAEEINS
DAKYYENQEIAKNYLWGCEGLHKNLLEEQTKNPYLAFKYGFLNIV (SEQ ID
NO:197)

>Cas14c.2|3300001245.a|JGI12048J13642_10201286|4257..5489|r
evcom
LDFKRTCSQELVLLPEIEGLKLSGTQGVTSLAKKLINKAANVDRDESYGCHHCIHTRTS
LSKPVKKDCNSCNQSTNHPAVPITLKGYKIAFYELWHRFTSWAVDSISKALHRNKVMGK
VNLDEYAVVDNSHIVCYAVRKCYEKRQRSVRLHKRAYRCRAKHYNKSQPKVGRIYKKSK
RRNARNLKKEAKRYFQPNEITNGSSDALFYKIGVDLGIAKGTPETEVKVDVSICFQVYY
GDARRVLRVRKMDELQSFHLDYTGKLKLKGIGNKDTFTIAKRNESLKWGSTKYEVSRAH
KKFKPFGKKGSVKRKCNDYFRSIASWSCEAASQRAQSNLKNAFPYQKALVKCYKNLDYK
GVKKNDMWYRLCSNRIFRYSRIAEDIAQYQSDKGKAKFEFVILAQSVAEYDISAIM
(SEQ ID NO: 198)

FIG. 7 (Cont.)

>Cas14d.1|RIFCSPHIGHO2_01_FULL_CPR_46_36_rifcsphigho2_01_scaffold_646_curated|49808..51616|revcom
VFLTDDKRKTALRKIRSAFRKTAEIALVRAQEADSLDRQAKKLTIETVSFGAPGAKNAF
IGSLQGYNWNSHRANVPSSGSAKDVFRITELGLGIPQSAHEASIGKSFELVGNVVRYTA
NLLSKGYKKGAVNKGAKQQREIKGKEQLSFDLISNGPISGDKLINGQKDALAWWLIDKM
GFHIGLAMEPLSSPNTYGITLQAFWKRHTAPRRYSRGVIRQWQLPFGRQLAPLIHNFFR
KKGASIPIVLTNASKKLAGKGVLLEQTALVDPKKWWQVKEQVTGPLSNIWERSVPLVLY
TATFTHKHGAAHKRPLTLKVIRISSGSVFLLPLSKVTPGKLVRAWMPDINILRDGRPDE
AAYKGPDLIRARERSFPLAYTCVTQIADEWQKRALESNRDSITPLEAKLVTGSDLLQIH
STVQQAVEQGIGGRISSPIQELLAKDALQLVLQQLFMTVDLLRIQWQLKQEVADGNTSE
KAVGWAIRISNIHKDAYKTAIEPCTSALKQAWNPLSGFEERTFQLDASIVRKRSTAKTP
DDELVIVLRQQAAEMTVAVTQSVSKELMELAVRHSATLHLLVGEVASKQLSRSADKDRG
AMDHWKLLSQSM (SEQ ID NO: 199)

>Cas14d.2|rifcsphigho2_01_scaffold_10981_curated|5762..7246|revcom
EDLLQKALNTATNVAAIERHSCISCLFTESEIDVKYKTPDKIGQNTAGCQSCTFRVGYS
GNSHTLPMGNRIALDKLRETIQRYAWHSLLFNVPPAPTSKRVRAISELRVAAGRERLFT
VITFVQTNILSKLQKRYAANWTPKSQERLSRLREEGQHILSLLESGSWQQKEVVREDQD
LIVCSALTKPGLSIGAFCRPKYLKPAKHALVLRLIFVEQWPGQIWGQSKRTRRMRRRKD
VERVYDISVQAWALKGKETRISECIDTMRRHQQAYIGVLPFLILSGSTVRGKGDCPILK
EITRMRYCPNNEGLIPLGIFYRGSANKLLRVVKGSSFTLPMWQNIETLPHPEPFSPEGW
TATGALYEKNLAYWSALNEAVDWYTGQILSSGLQYPNQNEFLARLQNVIDSIPRKWFRP
QGLKNLKPNGQEDIVPNEFVIPQNAIRAHHVIEWYHKTNDLVAKTLLGWGSQTTLNQTR
PQGDLRFTYTRYYFREKEVPEV (SEQ ID NO: 200)

>Cas14d.3|RIFCSPLOWO2_01_FULL_OD1_45_34b_rifcsplowo2_01_scaffold_3495_curated|25656..27605|revcom
VPKKKLMRELAKKAVFEAIFNDPIPGSFGCKRCTLIDGARVTDAIEKKQGAKRCAGCEP
CTFHTLYDSVKHALPAATGCDRTAIDTGLWEILTALRSYNWMSFRRNAVSDASQKQVWS
IEELAIWADKERALRVILSALTHTIGKLKNGFSRDGVWKGGKQLYENLAQKDLAKGLFA
NGEIFGKELVEADHDMLAWTIVPNHQFHIGLIRGNWKPAAVEASTAFDARWLTNGAPLR
DTRTHGHRGRRFNRTEKLTVLCIKRDGGVSEEFRQERDYELSVMLLQPKNKLKPEPKGE
LNSFEDLHDHWWFLKGDEATALVGLTSDPTVGDFIQLGLYIRNPIKAHGETKRRLLICF
EPPIKLPLRRAFPSEAFKTWEPTINVFRNGRRDTEAYYDIDRARVFEFPETRVSLEHLS
KQWEVLRLEPDRENTDPYEAQQNEGAELQVYSLLQEAAQKMAPKVVIDPFGQFPLELFS
TFVAQLFNAPLSDTKAKIGKPLDSGFVVESHLHLLEEDFAYRDFVRVTFMGTEPTFRVI
HYSNGEGYWKKTVLKGKNNIRTALIPEGAKAAVDAYKNKRCPLTLEAAILNEEKDRRLV
LGNKALSLLAQTARGNLTILEALAAEVLRPLSGTEGVVHLHACVTRHSTLTESTETDNM
(SEQ ID NO: 201)

FIG. 7 (Cont.)

>Cas14e.1|rifcsphigho2_01_scaffold_566_curated|113069..114313
VEKLFSERLKRAMWLKNEAGRAPPAETLTLKHKRVSGGHEKVKEELQRVLRSLSGTNQAAWNLGLSGGREPKSSDALKGEKSRVVLETVVFHSGHNRVLYDVIEREDQVHQRSSIMHMRRKGSNLLRLWGRSGKVRRKMREEVAEIKPVWHKDSRWLAIVEEGRQSVVGISSAGLAVFAVQESQCTTAEPKPLEYVVSIWFRGSKALNPQDRYLEFKKLKTTEALRGQQYDPIPFSLKRGAGCSLAIRGEGIKFGSRGPIKQFFGSDRSRPSHADYDGKRRLSLFSKYAGDLADLTEEQWNRTVSAFAEDEVRRATLANIQDFLSISHEKYAERLKKRIESIEEPVSASKLEAYLSAIFETFVQQREALASNFLMRLVESVALLISLEEKSPRVEFRVARYLAESKEGFNRKAM (SEQ ID NO: 202)

>Cas14e.2|rifcsplowo2_01_scaffold_81231_curated|976..2217
VVITQSELYKERLLRVMEIKNDRGRKEPRESQGLVLRFTQVTGGQEKVKQKLWLIFEGFSGTNQASWNFGQPAGGRKPNSGDALKGPKSRVTYETVVFHFGLRLLSAVIERHNLKQQRQTMAYMKRRAAARKKWARSGKKCSRMRNEVEKIKPKWHKDPRWFDIVKEGEPSIVGISSAGFAIYIVEEPNFPRQDPLEIEYAISIWFRRDRSQYLTFKKIQKAEKLKELQYNPIPFRLKQEKTSLVFESGDIKFGSRGSIEHFRDEARGKPPKADMDNNRRLTMFSVFSGNLTNLTEEQYARPVSGLLAPDEKRMPTLLKKLQDFFTPIHEKYGERIKQRLANSEASKRPFKKLEEYLPAIYLEFRARREGLASNWVLVLINSVRTLVRIKSEDPYIEFKVSQYLLEKEDNKAL (SEQ ID NO: 203)

>Cas14e.3|rifcsphigho2_01_scaffold_4702_curated|82881..84230|revcom
KQDALFEERLKKAIFIKRQADPLQREELSLLPPNRKIVTGGHESAKDTLKQILRAINGTNQASWNPGTPSGKRDSKSADALAGPKSRVKLETVVFHVGHRLLKKVVEYQGHQKQQHGLKAFMRTCAAMRKKWKRSGKVVGELREQLANIQPKWHYDSRPLNLCFEGKPSVVGLRSAGIALYTIQKSVVPVKEPKPIEYAVSIWFRGPKAMDREDRCLEFKKLKIATELRKLQFEPIVSTLTQGIKGFSLYIQGNSVKFGSRGPIKYFSNESVRQRPPKADPDGNKRLALFSKFSGDLSDLTEEQWNRPILAFEGIIRRATLGNIQDYLTVGHEQFAISLEQLLSEKESVLQMSIEQQRLKKNLGKKAENEWVESFGAEQARKKAQGIREYISGFFQEYCSQREQWAENWVQQLNKSVRLFLTIQDSTPFIEFRVARYLPKGEKKKGKAM (SEQ ID NO: 204)

>Cas14f.1|rifcsp13_1_sub10_scaffold_3_curated|38906..41041
ANHAERHKRLRKEANRAANRNRPLVADCDTGDPLVGICRLLRRGDKMQPNKTGCRSCEQVEPELRDAILVSGPGRLDNYKYELFQRGRAMAVHRLLKRVPKLNRPKKAAGNDEKKAENKKSEIQKEKQKQRRMMPAVSMKQVSVADFKHVIENTVRHLFGDRRDREIAECAALRAASKYFLKSRRVRPRKLPKLANPDHGKELKGLRLREKRAKLKKEKEKQAELARSNQKGAVLHVATLKKDAPPMPYEKTQGRNDYTTFVISAAIKVGATRGTKPLLTPQPREWQCSLYWRDGQRWIRGGLLGLQAGIVLGPKLNRELLEAVLQRPIECRMSGCGNPLQVRGAAVDFFMTTNPFYVSGAAYAQKKFKPFGTKRASEDGAAAKAREKLMTQLAKVLDKVVTQAAHSPLDGIWETRPEAKLRAMIMALEHEWIFLRPGPCHNAAEEVIKCDCTGGHAILWALIDEARGALEHKEFYAVTRAHTHDCEKQKLGGRLAGFLDLLIAQDVPLDDAPAARKIKTLLEATPPAPCYKAATSIATCDCEGKFDKLWAIIDATRAGHGTEDLWARTLAYPQNVNCKCKAGKDLTHRLADFLGLLIKRDGPFRERPPHKVTGDRKLVFSGDKKCKGHQYVILAKAHNEEVVRAWISRWGLKSRTNKAGYAATELNLLLNWLSICRRRWMDMLTVQRDTPYIRMKTGRLVVDDKKERKAM (SEQ ID NO: 205)

FIG. 7 (Cont.)

>Cas14f.2|3300009991.a|Ga0105042_100140|1624..3348
AKQREALRVALERGIVRASNRTYTLVTNCTKGGPLPEQCRMIERGKARAMKWEPKLVGC
GSCAAATVDLPAIEEYAQPGRLDVAKYKLTTQILAMATRRMMVRAAKLSRRKGQWPAKV
QEEKEEPPEPKKMLKAVEMRPVAIVDFNRVIQTTIEHLWAERANADEAELKALKAAAAY
FGPSLKIRARGPPKAAIGRELKKAHRKKAYAERKKARRKRAELARSQARGAAAHAAIRE
RDIPPMAYERTQGRNDVTTIPIAAAIKIAATRGARPLPAPKPMKWQCSLYWNEGQRWIR
GGMLTAQAYAHAANIHRPMRCEMWGVGNPLKVRAFEGRVADPDGAKGRKAEFRLQTNAF
YVSGAAYRNKKFKPFGTDRGGIGSARKKRERLMAQLAKILDKVVSQAAHSPLDDIWHTR
PAQKLRAMIKQLEHEWMFLRPQAPTVEGTKPDVDVAGNMQRQIKALMAPDLPPIEKGSP
AKRFTGDKRKKGERAVRVAEAHSDEVVTAWISRWGIQTRRNEGSYAAQELELLLNWLQI
CRRRWLDMTAAQRVSPYIRMKSGRMITDAADEGVAPIPLVENM (SEQ ID NO: 206
)

>Cas14g.1|RBG_13_scaffold_1401_curated|15949..18180
KSISGRSIKHMACLKDMLKSEITEIEEKQKKESLRKWDYYSKFSDEILFRRNLNVSANH
DANACYGCNPCAFLKEVYGFRIERRNNERIISYRRGLAGCKSCVQSTGYPPIEFVRRKF
GADKAMEIVREVLHRRNWGALARNIGREKEADPILGELNELLLVDARPYFGNKSAANET
NLAFNVITRAAKKFRDEGMYDIHKQLDIHSEEGKVPKGRKSRLIRIERKHKAIHGLDPG
ETWRYPHCGKGEKYGVWLNRSRLIHIKGNEYRCLTAFGTTGRRMSLDVACSVLGHPLVK
KKRKKGKKTVDGTELWQIKKATETLPEDPIDCTFYLYAAKPTKDPFILKVGSLKAPRWK
KLHKDFFEYSDTEKTQGQEKGKRVVRRGKVPRILSLRPDAKFKVSIWDDPYNGKNKEGT
LLRMELSGLDGAKKPLILKRYGEPNTKPKNFVFWRPHITPHPLTFTPKHDFGDPNKKTK
RRRVFNREYYGHLNDLAKMEPNAKFFEDREVSNKKNPKAKNIRIQAKESLPNIVAKNGR
WAAFDPNDSLWKLYLHWRGRRKTIKGGISQEFQEFKERLDLYKKHEDESEWKEKEKLWE
NHEKEWKKTLEIHGSIAEVSQRCVMQSMMGPLDGLVQKKDYVHIGQSSLKAADDAWTFS
ANRYKKATGPKWGKISVSNLLYDANQANAELISQSISKYLSKQKDNQGCEGRKMKFLIK
IIEPLRENFVKHTRWLHEMTQKDCEVRAQFSRVSM (SEQ ID NO:207)

>Cas14g.2|3300009652.a|Ga0123330_1010394|2814..5123
FPSDVGADALKHVRMLQPRLTDEVRKVALTRAPSDRPALARFAAVAQDGLAFVRHLNVS
ANHDSNCTFPRDPRDPRRGPCEPNPCAFLREVWGFRIVARGNERALSYRRGLAGCKSCV
QSTGFPSVPFHRIGADDCMRKLHEILKARNWRLLARNIGREREADPLLTELSEYLLVDA
RTYPDGAAPNSGRLAENVIKRAAKKFRDEGMRDIHAQLRVHSREGKVPKGRLQRLRRIE
RKHRAIHALDPGPSWEAEGSARAEVQGVAVYRSQLLRVGHHTQQIEPVGIVARTLFGVG
RTDLDVAVSVLGAPLTKRKKGSKTLESTEDFRIAKARETRAEDKIEVAFVLYPTASLLR
DEIPKDAFPAMRIDRFLLKVGSVQADREILLQDDYYRFGDAEVKAGKNKGRTVTRPVKV
PRLQALRPDAKFRVNVWADPFGAGDSPGTLLRLEVSGVTRRSQPLRLLRYGQPSTQPAN
FLCWRPHRVPDPMTFTPRQKFGERRKNRRTRRPRVFERLYQVHIKHLAHLEPNRKWFEE
ARVSAQKWAKARAIRRKGAEDIPVVAPPAKRRWAALQPNAELWDLYAHDREARKRFRGG
RAAEGEEFKPRLNLYLAHEPEAEWESKRDRWERYEKKWTAVLEEHSRMCAVADRTLPQF
LSDPLGARMDDKDYAFVGKSALAVAEAFVEEGTVERAQGNCSITAKKKFASNASRKRLS
VANLLDVSDKADRALVFQAVRQYVQRQAENGGVEGRRMAFLRKLLAPLRQNFVCHTRWL
HM (SEQ ID NO: 208)

FIG. 7 (Cont.)

>Cas14h.1|3300005602.a|Ga0070762_10001740|7377..9071|revcom
AARKKKRGKIGITVKAKEKSPPAAGPFMARKLVNVAANVDGVEVHLCVECEADAHGSAS
ARLLGGCRSCTGSIGAEGRLMGSVDVDRERVIAEPVHTETERLGPDVKAFEAGTAESKY
AIQRGLEYWGVDLISRNRARTVRKMEEADRPESSTMEKTSWDEIAIKTYSQAYHASENH
LFWERQRRVRQHALALFRRARERNRGESPLQSTQRPAPLVLAALHAEAAAISGRARAEY
VLRGPSANVRAAAADIDAKPLGHYKTPSPKVARGFPVKRDLLRARHRIVGLSRAYFKPS
DVVRGTSDAIAHVAGRNIGVAGGKPKEIEKTFTLPFVAYWEDVDRVVHCSSFKADGPWV
RDQRIKIRGVSSAVGTFSLYGLDVAWSKPTSFYIRCSDIRKKFHPKGFGPMKHWRQWAK
ELDRLTEQRASCVVRALQDDEELLQTMERGQRYYDVFSCAATHATRGEADPSGGCSRCE
LVSCGVAHKVTKKAKGDTGIEAVAVAGCSLCESKLVGPSKPRVHRQMAALRQSHALNYL
RRLQREWEALEAVQAPTPYLRFKYARHLEVRSM (SEQ ID NO: 209)

>Cas14h.2|3300005921.a|Ga0070766_10011912|384..2081
AAKKKKQRGKIGISVKPKEGSAPPADGPFMARKLVNVAANVDGVEVNLCIECEADAHGS
APARLLGGCKSCTGSIGAEGRLMGSVDVDRADAIAKPVNTETEKLGPDVQAFEAGTAET
KYALQRGLEYWGVDLISRNRSRTVRRTEEGQPESATMEKTSWDEIAIKSYTRAYHASEN
HLFWERQRRVRQHALALFKRAKERNRGDSTLPREPGHGLVAIAALACEAYAVGGRNLAE
TVVRGPTFGTARAVRDVEIASLGRYKTPSPKVAHGSPVKRDFLRARHRIVGLARAYYRP
SDVVRGTSDAIAHVAGRNIGVAGGKPRAVEAVFTLPFVAYWEDVDRVVHCSSFQVSAPW
NRDQRMKIAGVTTAAGTFSLHGGELKWAKPTSFYIRCSDTRRKFRPKGFGPMKRWRQWA
KDLDRLVEQRASCVVRALQDDAALLETMERGQRYYDVFACAVTHATRGEADRLAGCSRC
ALTPCQEAHRVTTKPRGDAGVEQVQTSDCSLCEGKLVGPSKPRLHRTLTLLRQEHGLNY
LRRLQREWESLEAVQVPTPYLRFKYARHLEVRSM (SEQ ID NO: 210)

>Cas14h.3|3300009698.a|Ga0116216_10000905|8005..9504
TDSQSESVPEVVYALTGGEVPGRVPPDGGSAEGARNAPTGLRKQRGKIKISAKPSKPGS
PASSLARTLVNEAANVDGVQSSGCATCRMRANGSAPRALPIGCVACASSIGRAPQEETV
CALPTTQGPDVRLLEGGHALRKYDIQRALEYWGVDLIGRNLDRQAGRGMEPAEGATATM
KRVSMDELAVLDFGKSYYASEQHLFAARQRRVRQHAKALKIRAKHANRSGSVKRALDRS
RKQVTALAREFFKPSDVVRGDSDALAHVVGRNLGVSRHPAREIPQTFTLPLCAYWEDVD
RVISCSSLLAGEPFARDQEIRIEGVSSALGSLRLYRGAIEWHKPTSLYIRCSDTRRKFR
PRGGLKKRWRQWAKDLDRLVEQRACCIVRSLQADVELLQTMERAQRFYDVHDCAATHVG
PVAVRCSPCAGKQFDWDRYRLLAALRQEHALNYLRRLQREWESLEAQQVKMPYLRFKYA
RKLEVSGPLIGLEVRREPSMGTAIAEM (SEQ ID NO: 211)

>Cas14u.1|3300009029.a|Ga0066793_10010091|37..1113|revcom
AGTAGRRHGSLGARRSINIAGVTDRHGRWGCESCVYTRDQAGNRARCAPCDQSTYAPDV
QEVTIGQRQAKYTIFLTLQSFSWTNTMRNNKRAAAGRSKRTTGKRIGQLAEIKITGVGL
AHAHNVIQRSLQHNITKMWRAEKGKSKRVARLKKAKQLTKRRAYFRRRMSRQSRGNGFF
RTGKGGIHAVAPVKIGLDVGMIASGSSEPADEQTVTLDAIWKGRKKKIRLIGAKGELAV
AACRFREQQTKGDKCIPLILQDGEVRWNQNNWQCHPKKLVPLCGLEVSRKFVSQADRLA
QNKVASPLAARFDKTSVKGTLVESDFAAVLVNVTSIYQQCHAMLLRSQEPTPSLRVQRT
ITSM (SEQ ID NO: 212)

FIG. 7 (Cont.)

>Cas14u.2|3300002172.a|JGI24730J26740_1002785|496..1605|rev
com
GVRFSPAQSQVFFRTVIPQSVEARFAINMAAIHDAAGAFGCSVCRFEDRTPRNAKAVHG
CSPCTRSTNRPDVFVLPVGAIKAKYDVFMRLLGFNWTHLNRRQAKRVTVRDRIGQLDEL
AISMLTGKAKAVLKKSICHNVDKSFKAMRGSLKKLHRKASKTGKSQLRAKLSDLRERTN
TTQEGSHVEGDSDVALNKIGLDVGLVGKPDYPSEESVEVVVCLYFVGKVLILDAQGRIR
DMRAKQYDGFKIPIIQRGQLTVLSVKDLGKWSLVRQDYVLAGDLRFEPKISKDRKYAEC
VKRIALITLQASLGFKERIPYYVTKQVEIKNASHIAFVTEAIQNCAENFREMTEYLMKY
QEKSPDLKVLLTQLM (SEQ ID NO: 213)

>Cas14u.3|19ft_2_nophage_noknown_scaffold_0_curated|508188.
.509648
RAVVGKVFLEQARRALNLATNFGTNHRTGCNGCYVTPGKLSIPQDGEKNAAGCTSCLMK
ATASYVSYPKPLGEKVAKYSTLDALKGFPWYSLRLNLRPNYRGKPINGVQEVAPVSKFR
LAEEVIQAVQRYHFTELEQSFPGGRRRLRELRAFYTKEYRRAPEQRQHVVNGDRNIVVV
TVLHELGFSVGMFNEVELLPKTPIECAVNVFIRGNRVLLEVRKPQFDKERLLVESLWKK
DSRRHTAKWTPPNNEGRIFTAEGWKDFQLPLLLGSTSRSLRAIEKEGFVQLAPGRDPDY
NNTIDEQHSGRPFLPLYLYLQGTISQEYCVFAGTWVIPFQDGISPYSTKDTFQPDLKRK
AYSLLLDAVKHRLGNKVASGLQYGRFPAIEELKRLVRMHGATRKIPRGEKDLLKKGDPD
TPEWWLLEQYPEFWRLCDAAAKRVSQNVGLLLSLKKQPLWQRRWLESRTRNEPLDNLPL
SMALTLHLTNEEAL (SEQ ID NO: 214)

>Cas14u.4|rifcsp2_19_4_full_scaffold_168_curated|84455..856
57
AAVYSKFYIENHFKMGIPETLSRIRGPSIIQGFSVNENYINIAGVGDRDFIFGCKKCKY
TRGKPSSKKINKCHPCKRSTYPEPVIDVRGSISEFKYKIYNKLKQEPNQSIKQNTKGRM
NPSDHTSSNDGIIINGIDNRIAYNVIFSSYKHLMEKQINLLRDTTKRKARQIKKYNNSG
KKKHSLRSQTKGNLKNRYHMLGMFKKGSLTITNEGDFITAVRKVGLDISLYKNESLNKQ
EVETELCLNIKWGRTKSYTVSGYIPLPINIDWKLYLFEKETGLTLRLFGNKYKIQSKKF
LIAQLFKPKRPPCADPVVKKAQKWSALNAHVQQMAGLFSDSHLLKRELKNRMHKQLDFK
SLWVGTEDYIKWFEELSRSYVEGAEKSLEFFRQDYFCFNYTKQTTM (SEQ ID NO: 215
)

>Cas14u.5|3300012532.a|Ga0137373_10000316|3286..5286
PQQQRDLMLMAANYDQDYGNGCGPCTVVASAAYRPDPQAQHGCKRHLRTLGASAVTHVG
LGDRTATITALHRLRGPAALAARARAAQAASAPMTPDTDAPDDRRRLEAIDADDVVLVG
AHRALWSAVRRWADDRRAALRRRLHSEREWLLKDQIRWAELYTLIEASGTPPQGRWRNT
LGALRGQSRWRRVLAPTMRATCAETHAELWDALAELVPEMAKDRRGLLRPPVEADALWR
APMIVEGWRGGHSVVVDAVAPPLDLPQPCAWTAVRLSGDPRQRWGLHLAVPPLGQVQPP
DPLKATLAVSMRHRGGVRVRTLQAMAVDADAPMQRHLQVPLTLQRGGGLQWGIHSRGVR
RREARSMASWEGPPIWTGLQLVNRWKGQGSALLAPDRPPDTPPYAPDAAVAPAQPDTKR
ARRTLKEACTVCRCAPGHMRQLQVTLTGDGTWRRFRLRAPQGAKRKAEVLKVATQHDER
IANYTAWYLKRPEHAAGCDTCDGDSRLDGACRGCRPLLVGDQCFRRYLDKIEADRDDGL
AQIKPKAQEAVAAMAAKRDARAQKVAARAAKLSEATGQRTAATRDASHEARAQKELEAV
ATEGTTVRHDAAAVSAFGSWVARKGDEYRHQVGVLANRLEHGLRLQELMAPDSVVADQQ
RASGHARVGYRYVLTAM (SEQ ID NO:216)

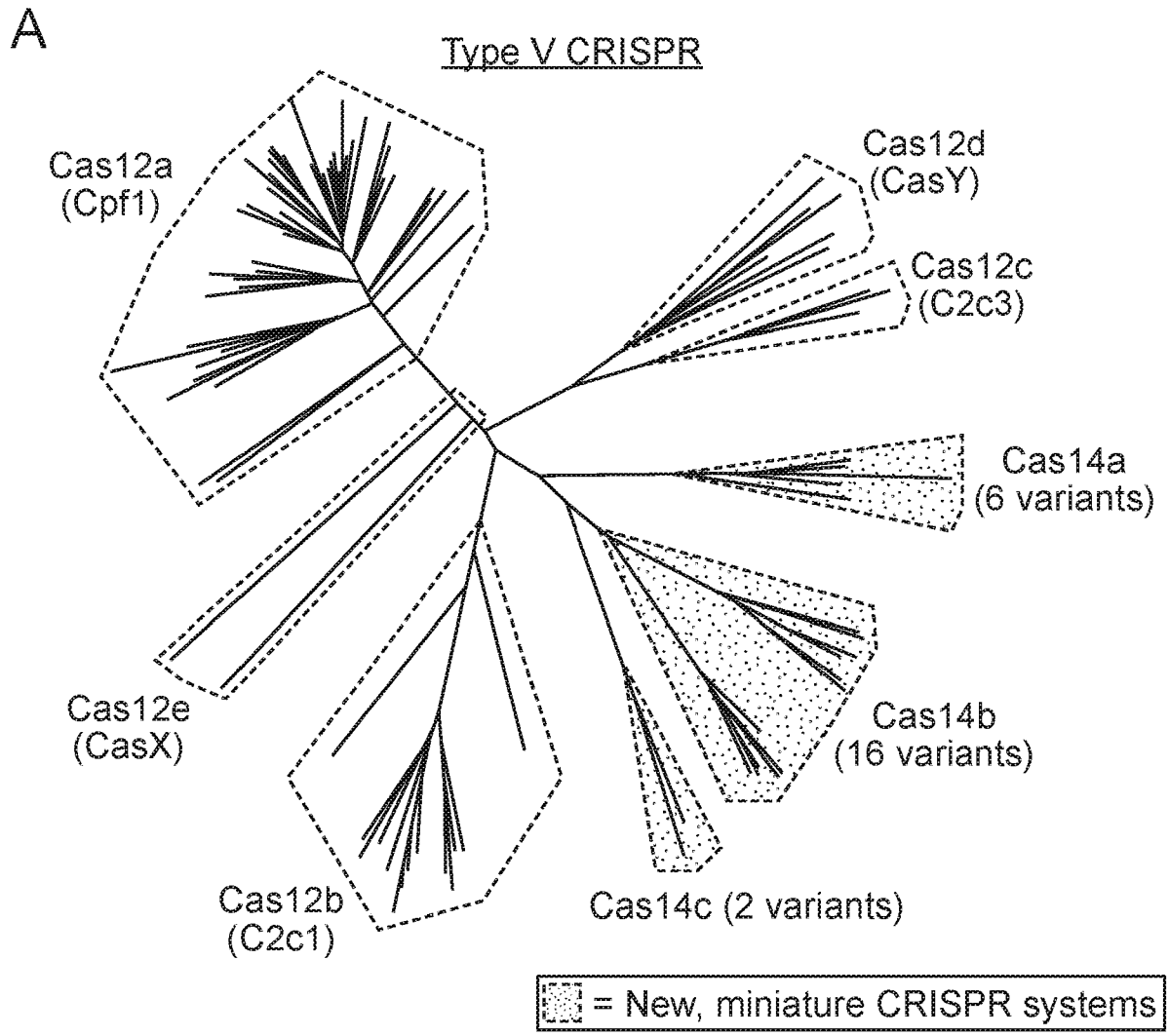
FIG. 7 (Cont.)

>Cas14u.6|3300006028.a|Ga0070717_10000077|54519..56201|revcom
AVAHPVGRGNAGSPGARGPEELPRQLVNRASNVTRPATYGCAPCRHVRLSIPKPVLTGC
RACEQTTHPAPKRAVRGGADAAKYDLAAFFAGWAADLEGRNRRRQVHAPLDPQPDPNHE
PAVTLQKIDLAEVSIEEFQRVLARSVKHRHDGRASREREKARAYAQVAKKRRNSHAHGA
RTRRAVRRQTRAVRRAHRMGANSGEILVASGAEDPVPEAIDHAAQLRRRIRACARDLEG
LRHLSRRYLKTLEKPCRRPRAPDLGRARCHALVESLQAAERELEELRRCDSPDTAMRRL
DAVLAAAAASTDATFATGWTVVGMDLGVAPRGSAAPEVSPMEMAISVFWRKGSRRVIVSK
PIAGMPIRRHELIRLEGLGTLRLDGNHYTGAGVTKGRGLSEGTEPDFREKSPSTLGFTL
SDYRHESRWRPYGAKQGKTARQFFAAMSRELRALVEHQVLAPMGPPLLEAHERRFETLL
KGQDNKSIHAGGGGRYVWRGPPDSKKRPAADGDWFRFGRGHADHRGWANKRHELAANYL
QSAFRLWSTLAEAQEPTPYARYKYTRVTM (SEQ ID NO:217)

>Cas14u.7|3300001256.a|JGI12210J13797_10004690|5792..7006
WDFLTLQVYERHTSPEVCVAGNSTKCASGTRKSDHTHGVGVKLGAQEINVSANDDRDHE
VGCNICVISRVSLDIKGWRYGCESCVQSTPEWRSIVRFDRNHKEAKGECLSRFEYWGAQ
SIARSLKRNKLMGGVNLDELAIVQNENVVKTSLKHLFDKRKDRIQANLKAVKVRMRERR
KSGRQRKALRRQCRKLKRYLRSYDPSDIKEGNSCSAFTKLGLDIGISPNKPPKIEPKVE
VVFSLFYQGACDKIVTVSSPESPLPRSWKIKIDGIRALYVKSTKVKFGGRTFRAGQRNN
RRKVRPPNVKKGKRKGSRSQFFNKFAVGLDAVSQQLPIASVQGLWGRAETKKAQTICLK
QLESNKPLKESQRCLFLADNWVVRVCGFLRALSQRQGPTPYIRYRYRCNM (SEQ ID
NO:218)

>Cas14u.8|3300005660.a|Ga0073904_10021651|765..1943
ARNVGQRNASRQSKRESAKARSRRVTGGHASVTQGVALINAAANADRDHTTGCEPCTWE
RVNLPLQEVIHGCDSCTKSSPFWRDIKVVNKGYREAKEEIMRIASGISADHLSRALSHN
KVMGRLNLDEVCILDFRTVLDTSLKHLTDSRSNGIKEHIRAVHRKIRMRRKSGKTARAL
RKQYFALRRQWKAGHKPNSIREGNSLTALRAVGFDVGVSEGTEPMPAPQTEVVLSVFYK
GSATRILRISSPHPIAKRSWKVKIAGIKALKLIRREHDFSFGRETYNASQRAEKRKFSP
HAARKDFFNSFAVQLDRLAQQLCVSSVENLWVTEPQQKLLTLAKDTAPYGIREGARFAD
TRARLAWNWVFRVCGFTRALHQEQEPTPYCRFTWRSKM (SEQ ID NO:219)

FIG. 8


A
Type V CRISPR
Cas12a
(Cpf1)
Cas12d
(CasY)
Cas12c
(C2c3)
Cas14a
(6 variants)
Cas12e
(CasX)
Cas14b
(16 variants)
Cas12b
(C2c1)
Cas14c (2 variants)
= New, miniature CRISPR systems

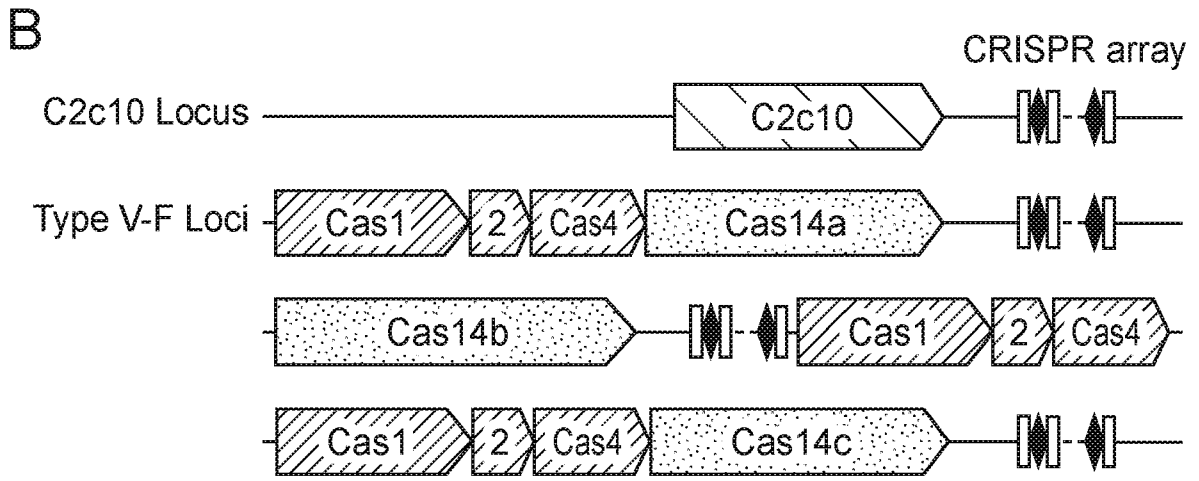
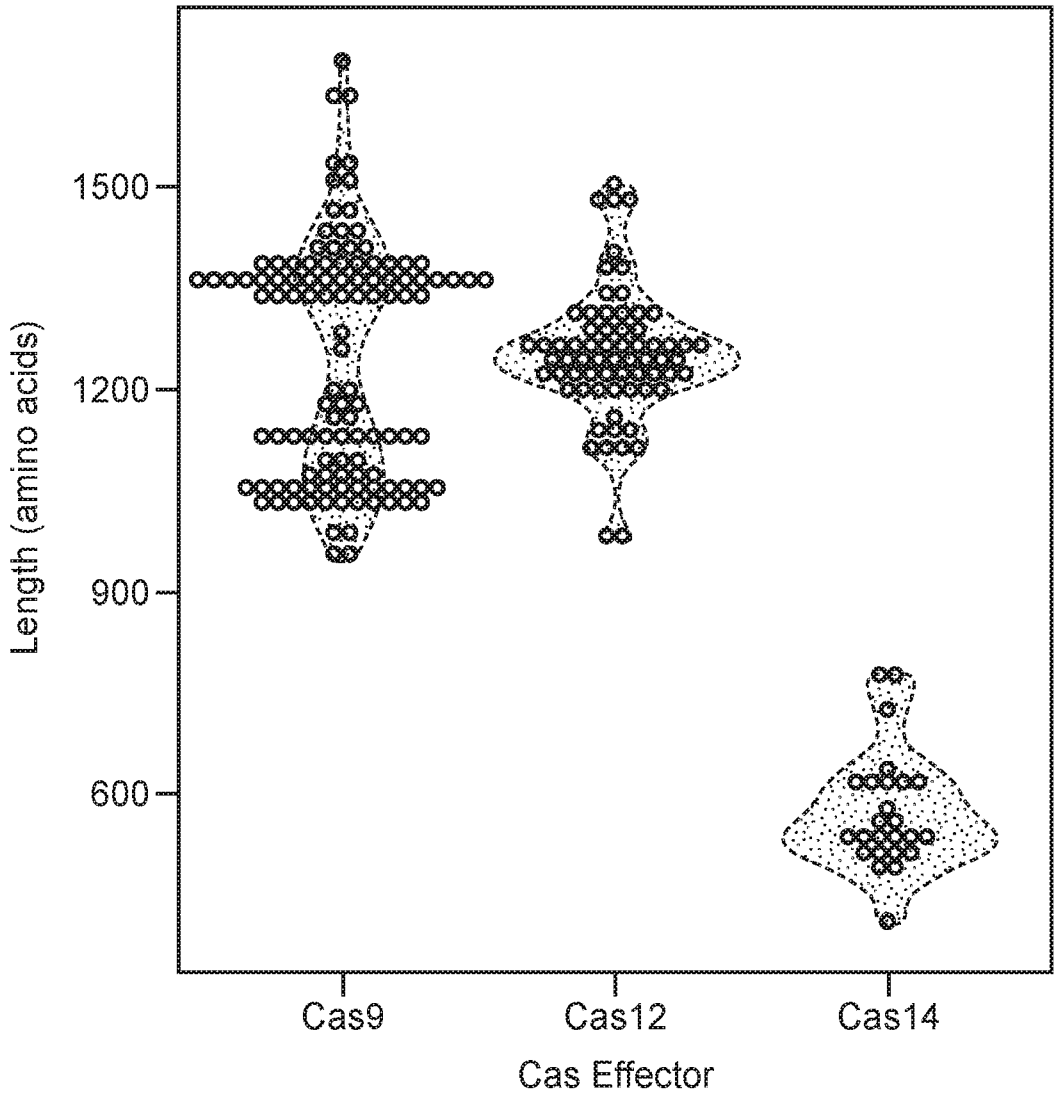

B
CRISPR array
C2c10 Locus
C2c10
Type V-F Loci
Cas1
2
Cas4
Cas14a
Cas14b
Cas1
2
Cas4
Cas1
2
Cas4
Cas14c

FIG. 8 (Cont.)

C


1500
1200
900
600
Length (amino acids)
Cas9
Cas12
Cas14
Cas Effector

D

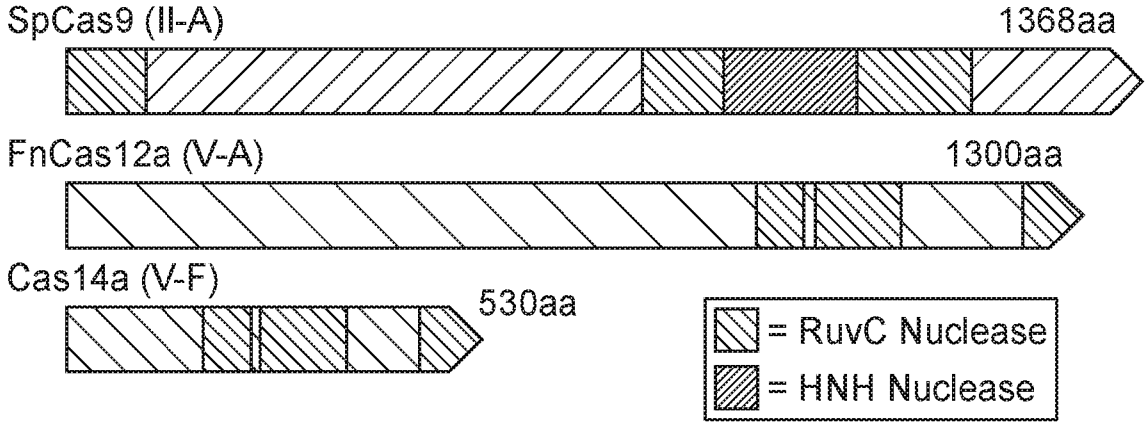

SpCas9 (II-A)
1368aa
FnCas12a (V-A)
1300aa
Cas14a (V-F)
530aa
= RuvC Nuclease
= HNH Nuclease

FIG. 9

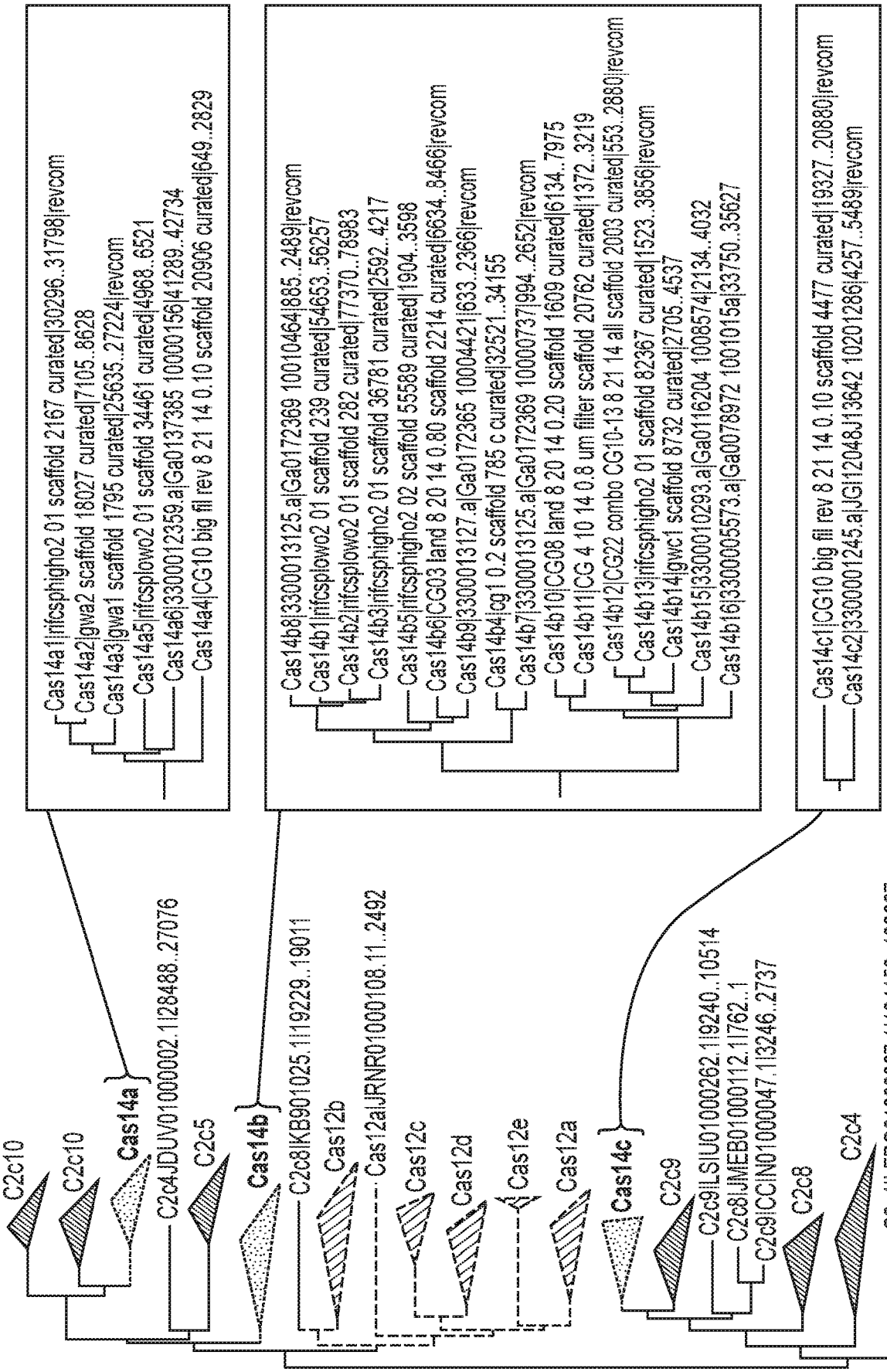

Cas14a1|nifcsphigho2 01 scaffold 2167 curated|30296..31798|revcom
Cas14a2|gwa2 scaffold 18027 curated|7105..8628
Cas14a3|gwa1 scaffold 1795 curated|25635..27224|revcom
Cas14a5|nifcsplowo2 01 scaffold 34461 curated|4968..6521
Cas14a6|3300012359.a|Ga0137385 100001561|41289..42734
Cas14a4|CG10 big fil rev 8 21 14 0.10 scaffold 20906 curated|649..2829
Cas14b8|3300013125.a|Ga0172369 100104641|885..2489|revcom
Cas14b1|nifcsplowo2 01 scaffold 239 curated|54653..56257
Cas14b2|nifcsplowo2 01 scaffold 282 curated|77370..78983
Cas14b3|nifcsphigho2 01 scaffold 36781 curated|2592..4217
Cas14b5|nifcsphigho2 02 scaffold 55589 curated|1904..3598
Cas14b6|CG03 land 8 20 14 0.80 scaffold 2214 curated|6634..8466|revcom
Cas14b9|3300013127.a|Ga0172365 100044211|633..2366|revcom
Cas14b4|cg1 0.2 scaffold 785 c curated|32521..34155
Cas14b7|3300013125.a|Ga0172369 100007371|994..2652|revcom
Cas14b10|CG08 land 8 20 14 0.20 scaffold 1609 curated|6134..7975
Cas14b11|CG 4 10 14 0.8 um filter scaffold 20762 curated|1372..3219
Cas14b12|CG22 combo CG10-13 8 21 14 all scaffold 2003 curated|553..2880|revcom
Cas14b13|nifcsphigho2 01 scaffold 82367 curated|1523..3856|revcom
Cas14b14|gwc1 scaffold 8732 curated|2705..4537
Cas14b15|3300010293.a|Ga0116204 100857412|134..4032
Cas14b16|3300005573.a|Ga0078972 100101 5a|33750..35627
Cas14c1|CG10 big fil rev 8 21 14 0.10 scaffold 4477 curated|19327..20880|revcom
Cas14c2|3300001245.a|JGI12048J13642 10201286|4257..5489|revcom
C2c10
C2c10
Cas14a
C2c4JDUV01000002.1|28488..27076
C2c5
Cas14b
C2c8|KB901025.1|19229..19011
Cas12b
Cas12a|JRNR01000108.11..2492
Cas12c
Cas12d
Cas12e
Cas12a
Cas14c
C2c9
C2c9|LSIU01000262.1|9240..10514
C2c8|JMEB01000112.1|762..1
C2c9|CCIN01000047.1|3246..2737
C2c8
C2c4
C2c4|LFPO01000007.1|134156..132627

FIG. 10

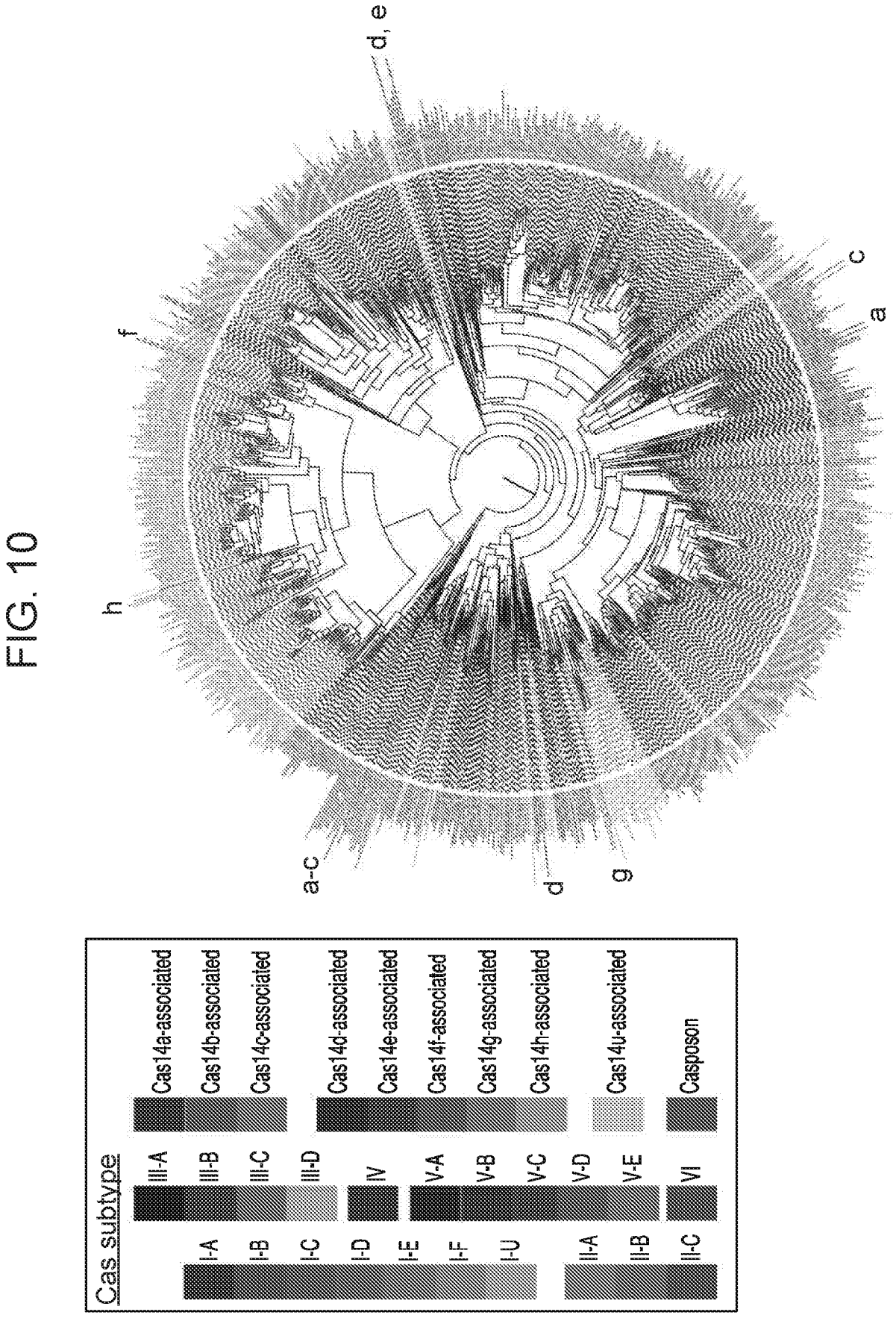

h
f
d, e
c
a
a-c
d
g
Cas subtype
I-A
I-B
I-C
I-D
I-E
I-F
I-U
II-A
II-B
II-C
III-A
III-B
III-C
III-D
IV
V-A
V-B
V-C
V-D
V-E
VI
Cas14a-associated
Cas14b-associated
Cas14c-associated
Cas14d-associated
Cas14e-associated
Cas14f-associated
Cas14g-associated
Cas14h-associated
Cas14u-associated
Casposon

Figure 11, Panels A-B

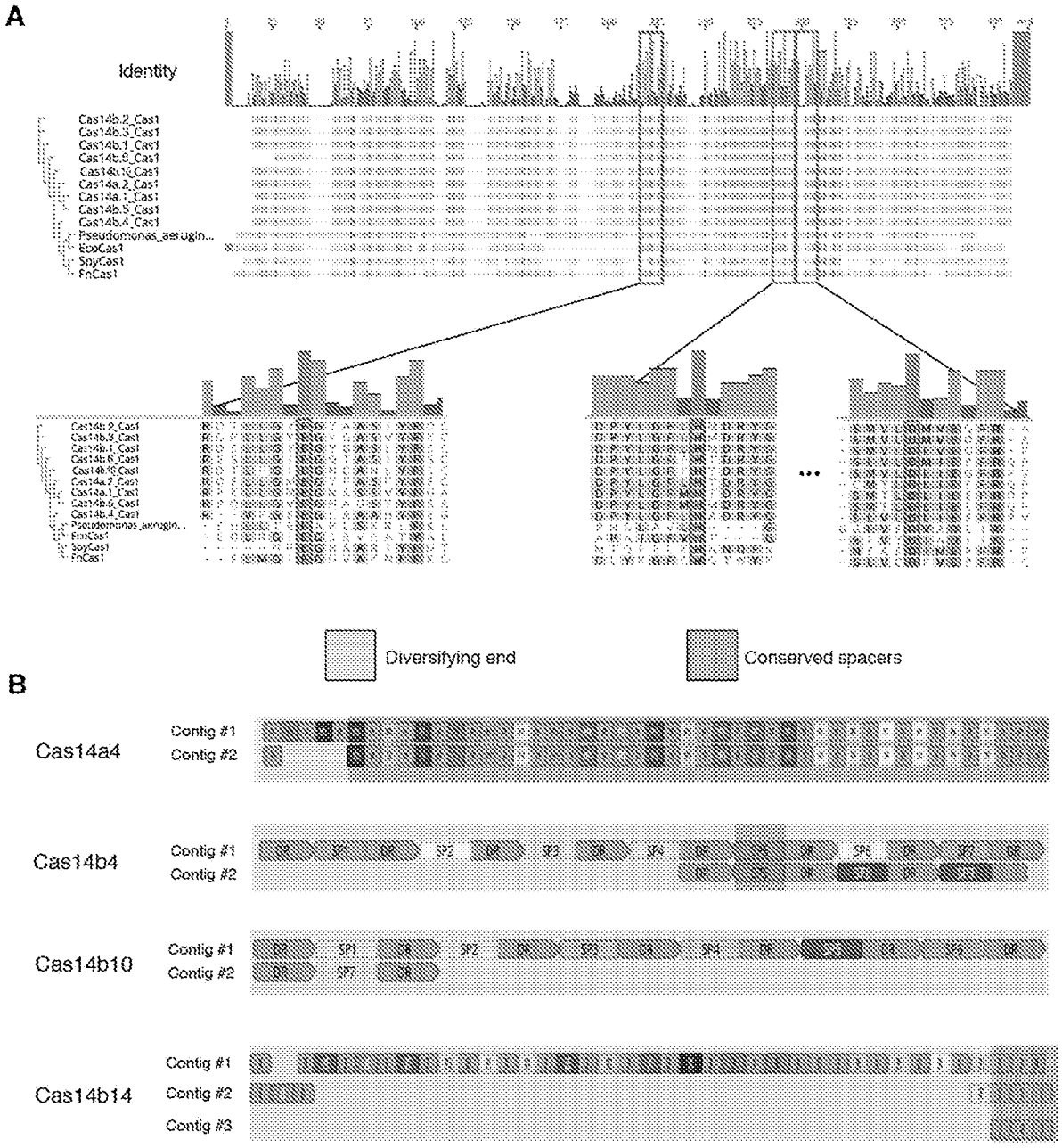


Figure 12, Panels A-D

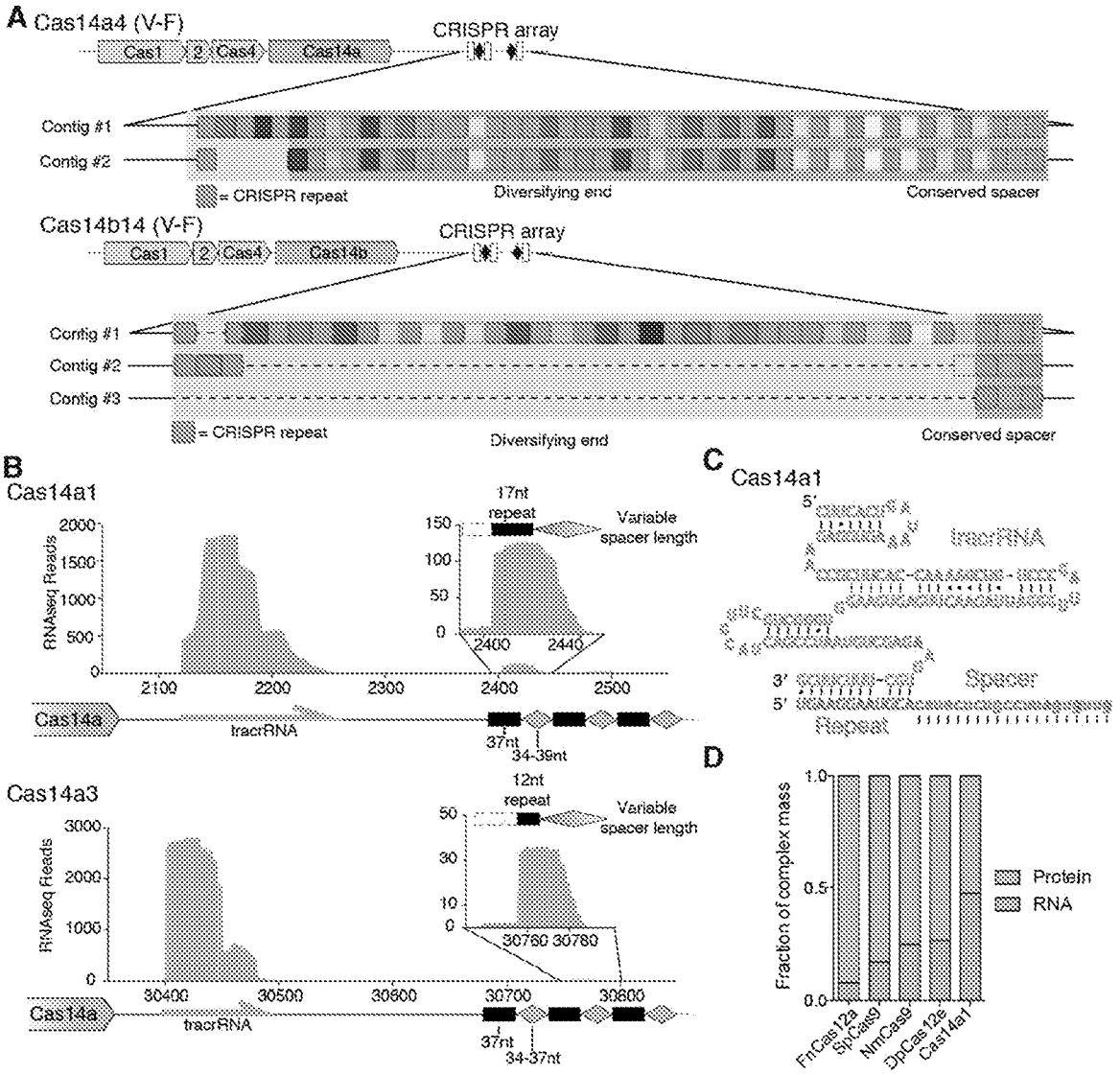


FIG. 13

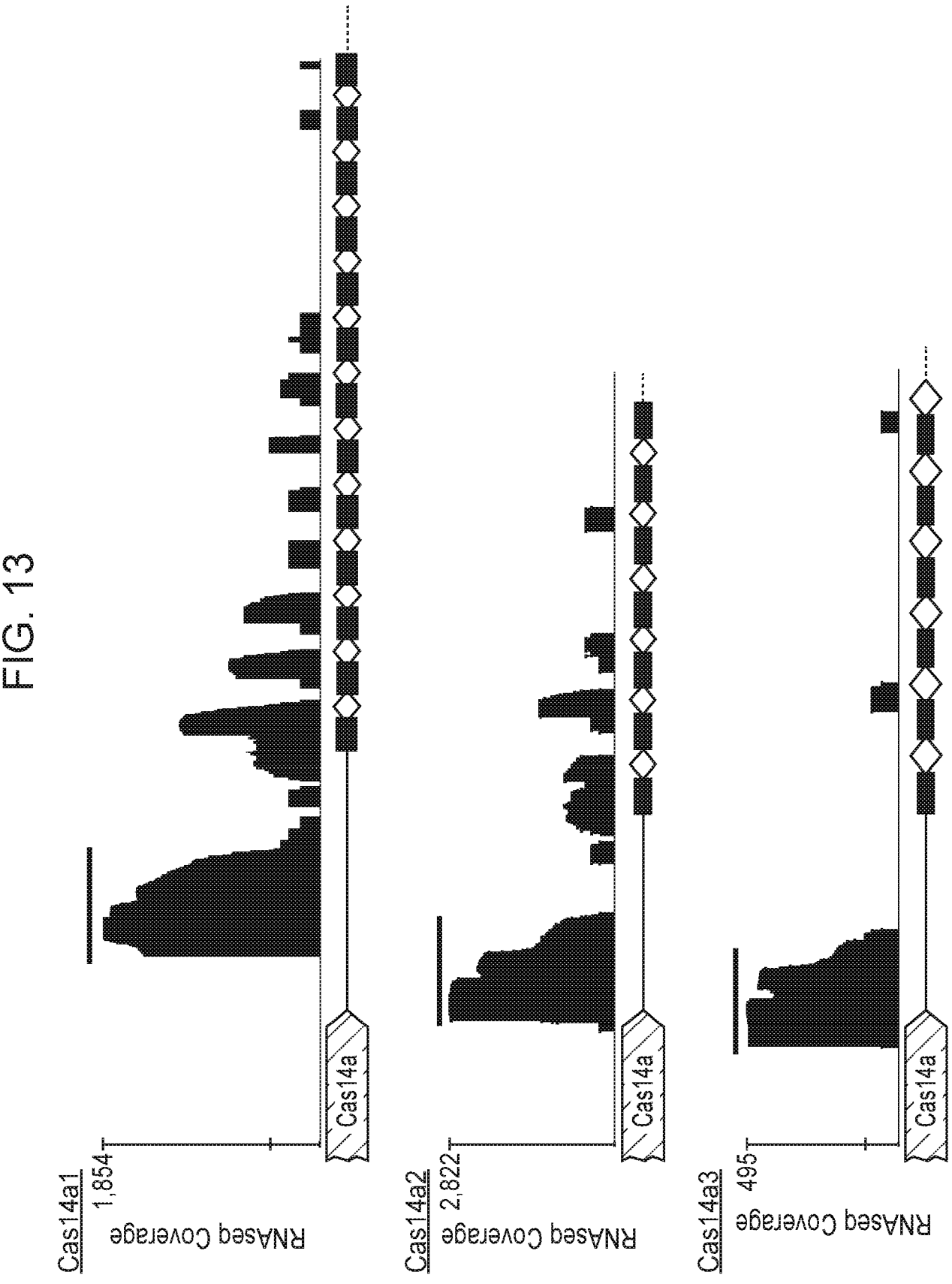
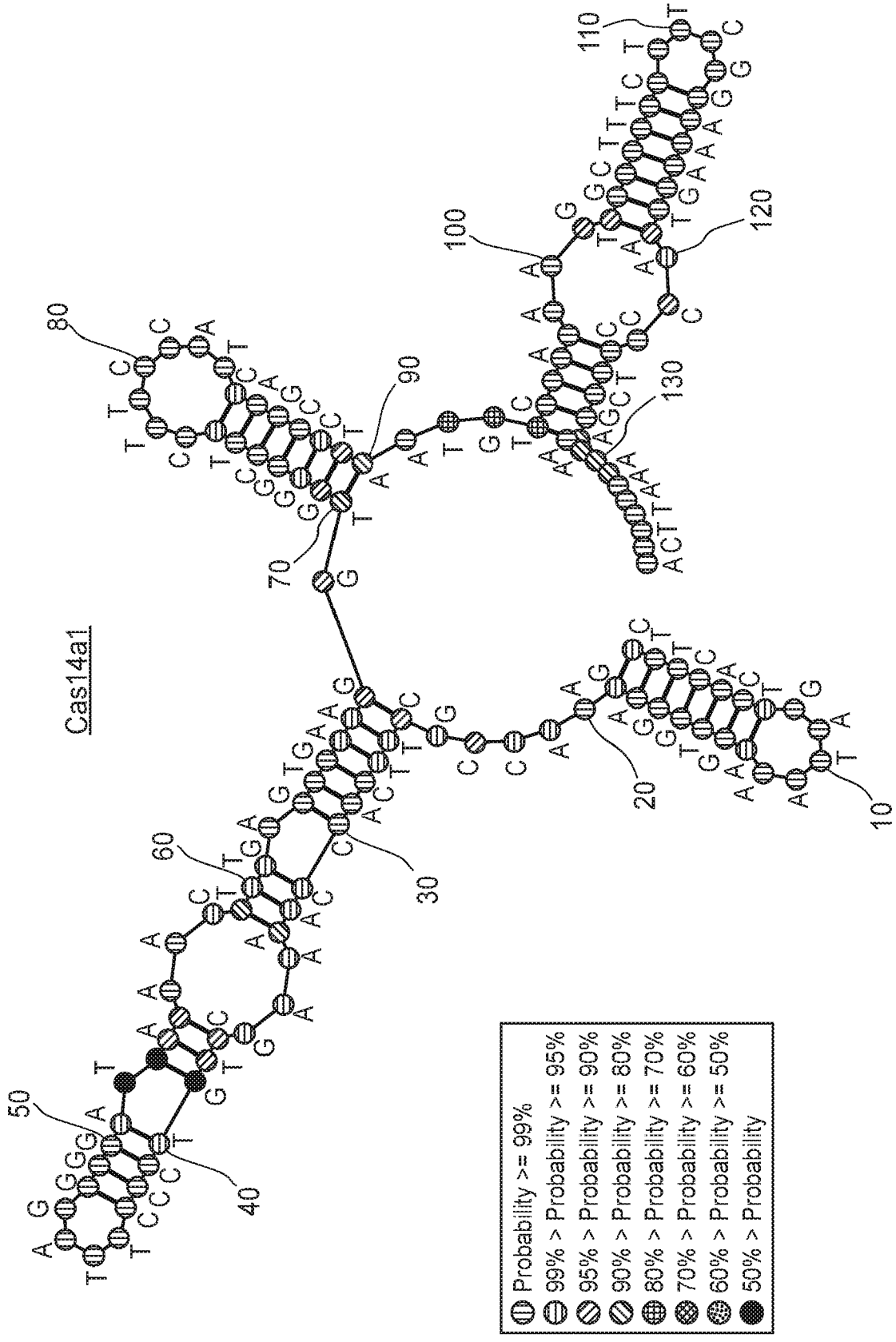

A
Cas14a1
1,854
RNAseq Coverage
Cas14a
Cas14a2
2,822
RNAseq Coverage
Cas14a
Cas14a3
495
RNAseq Coverage
Cas14a

FIG. 13 (Cont.)

A (Cont.)



A (Cont.)

FIG. 13 (Cont.)

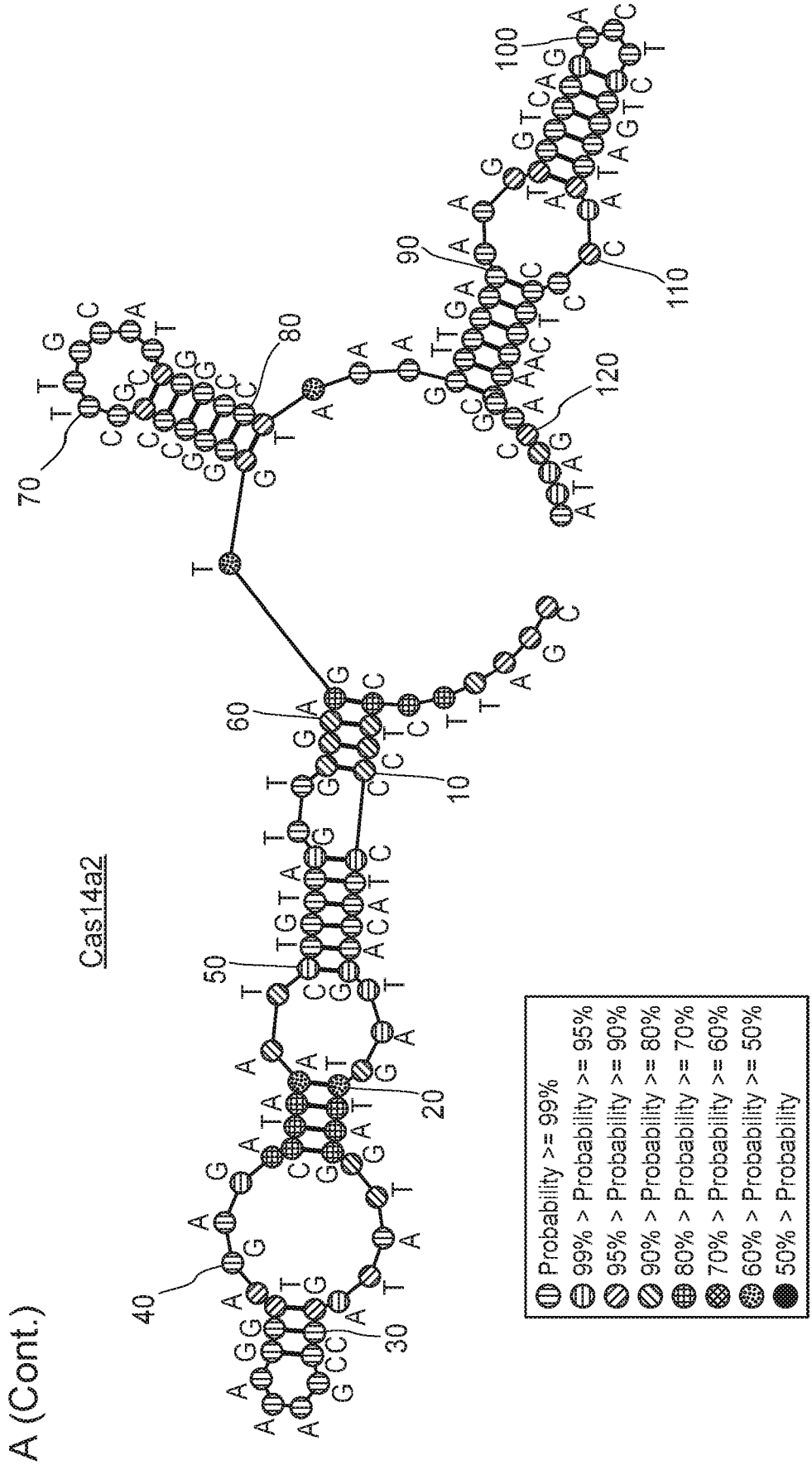


FIG. 13 (Cont.)

A (Cont.)

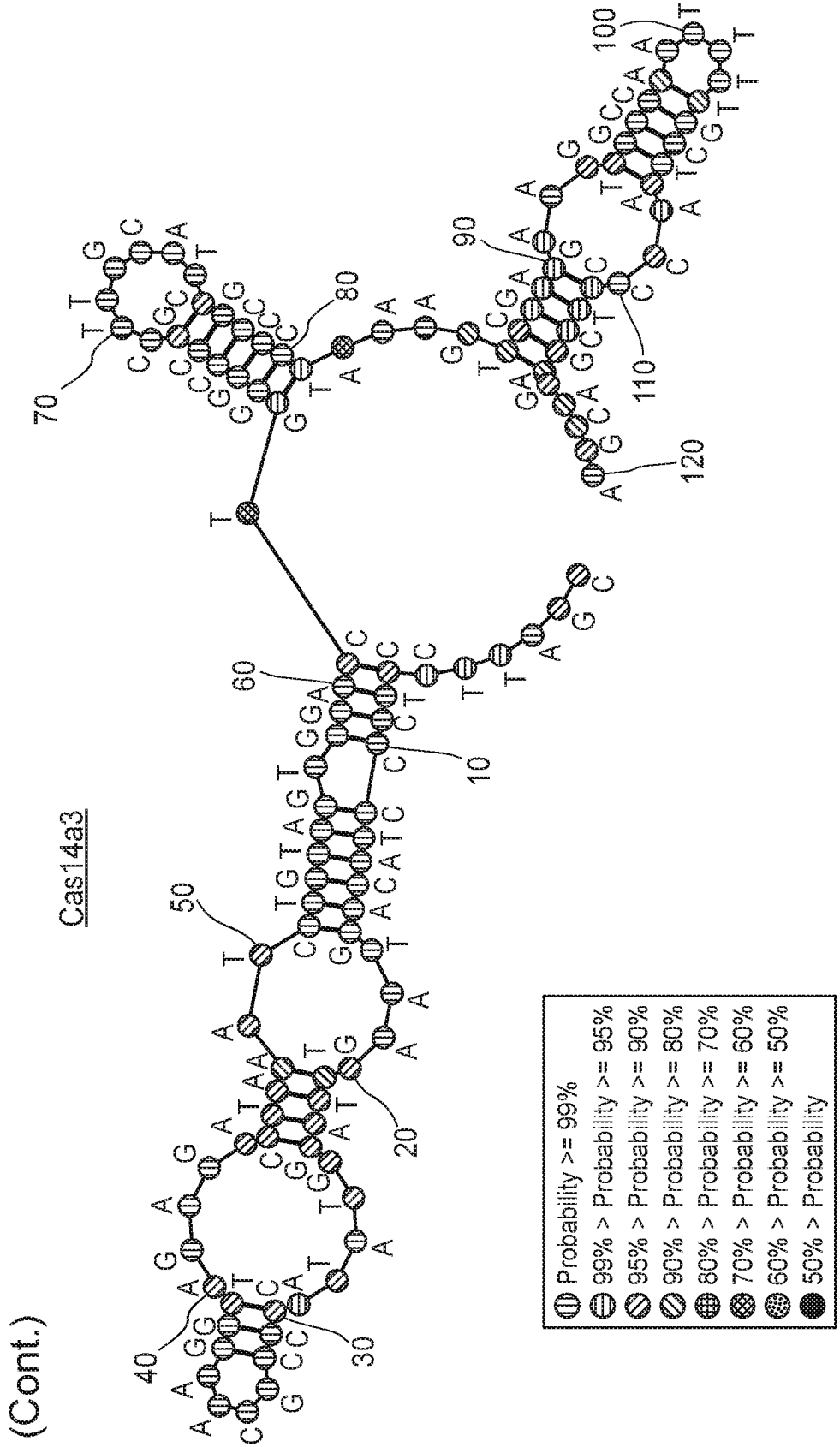


FIG. 13 (Cont.)

B

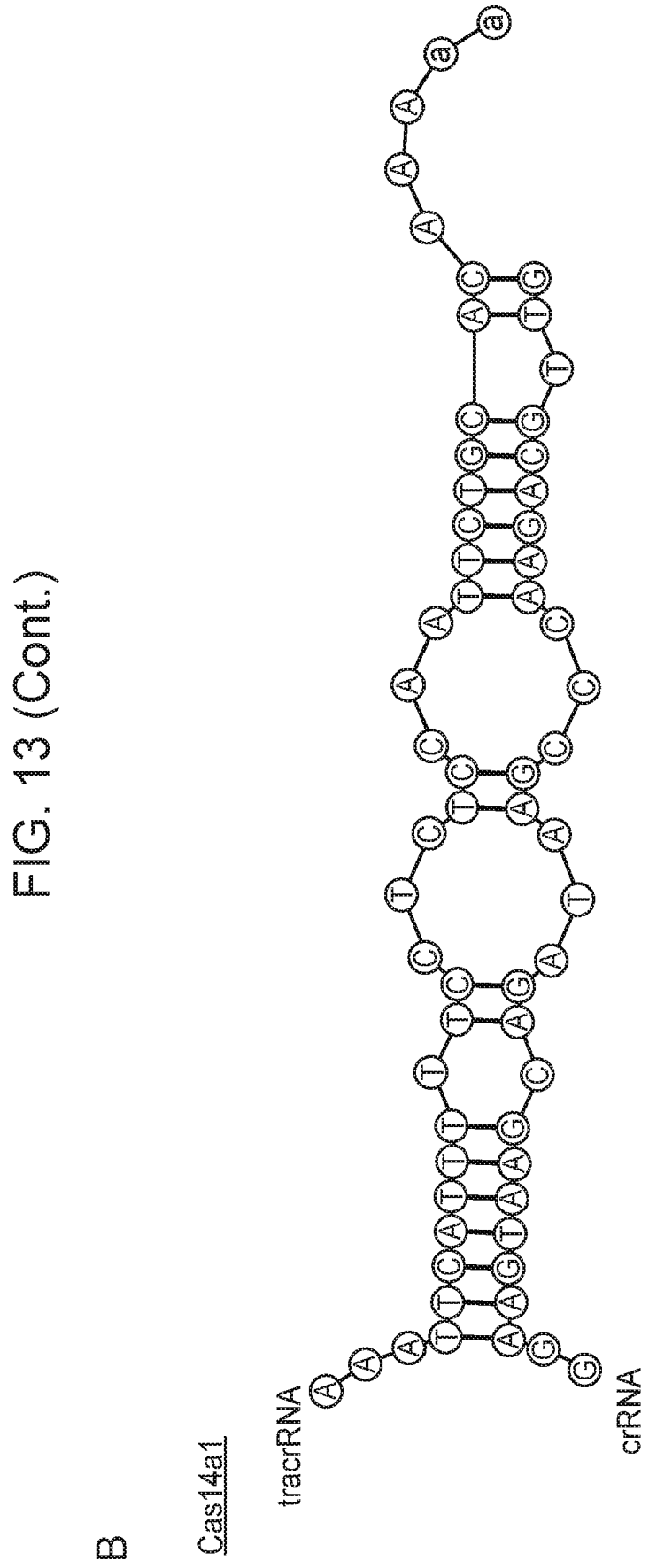


Figure 14, Panels A-B

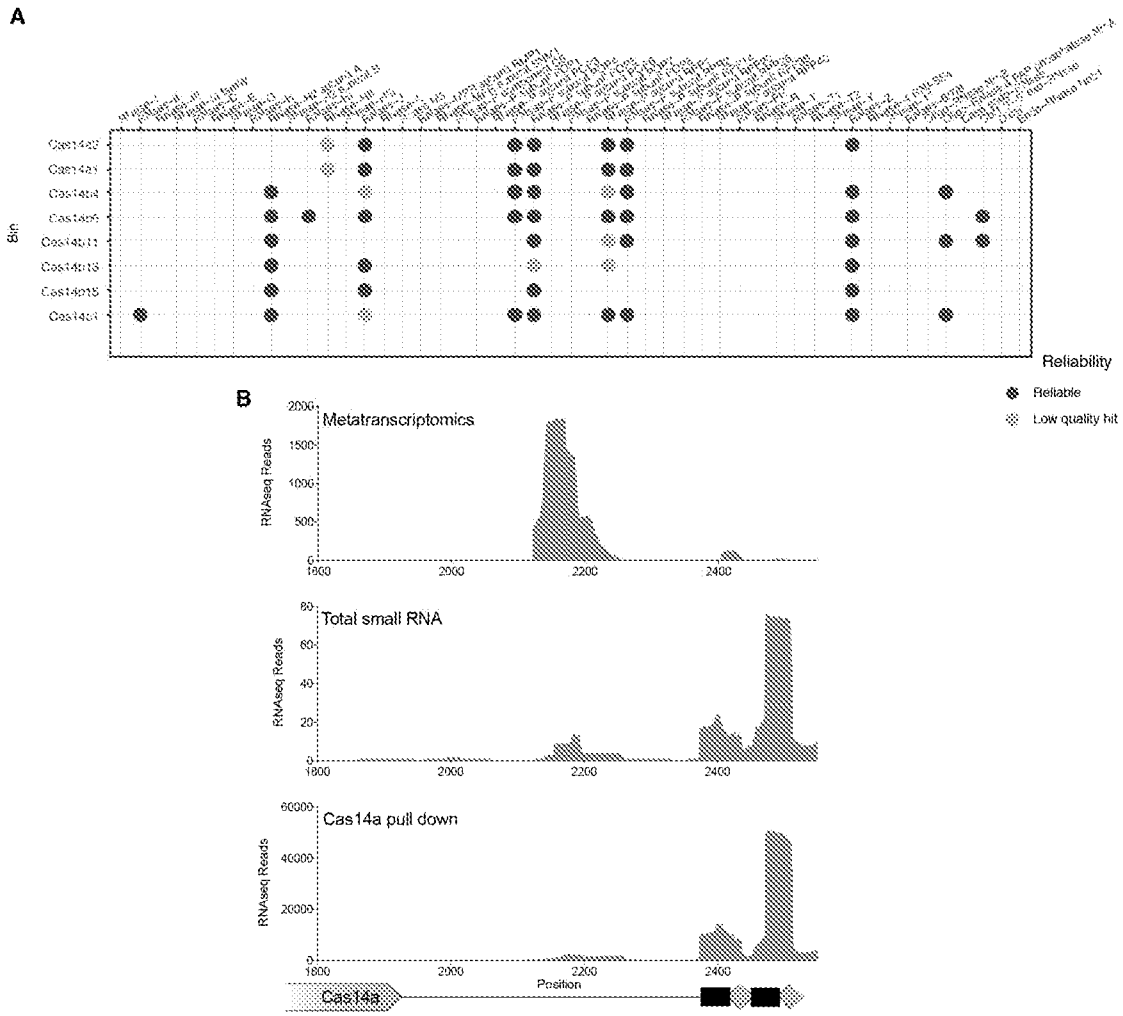


FIG. 15

A

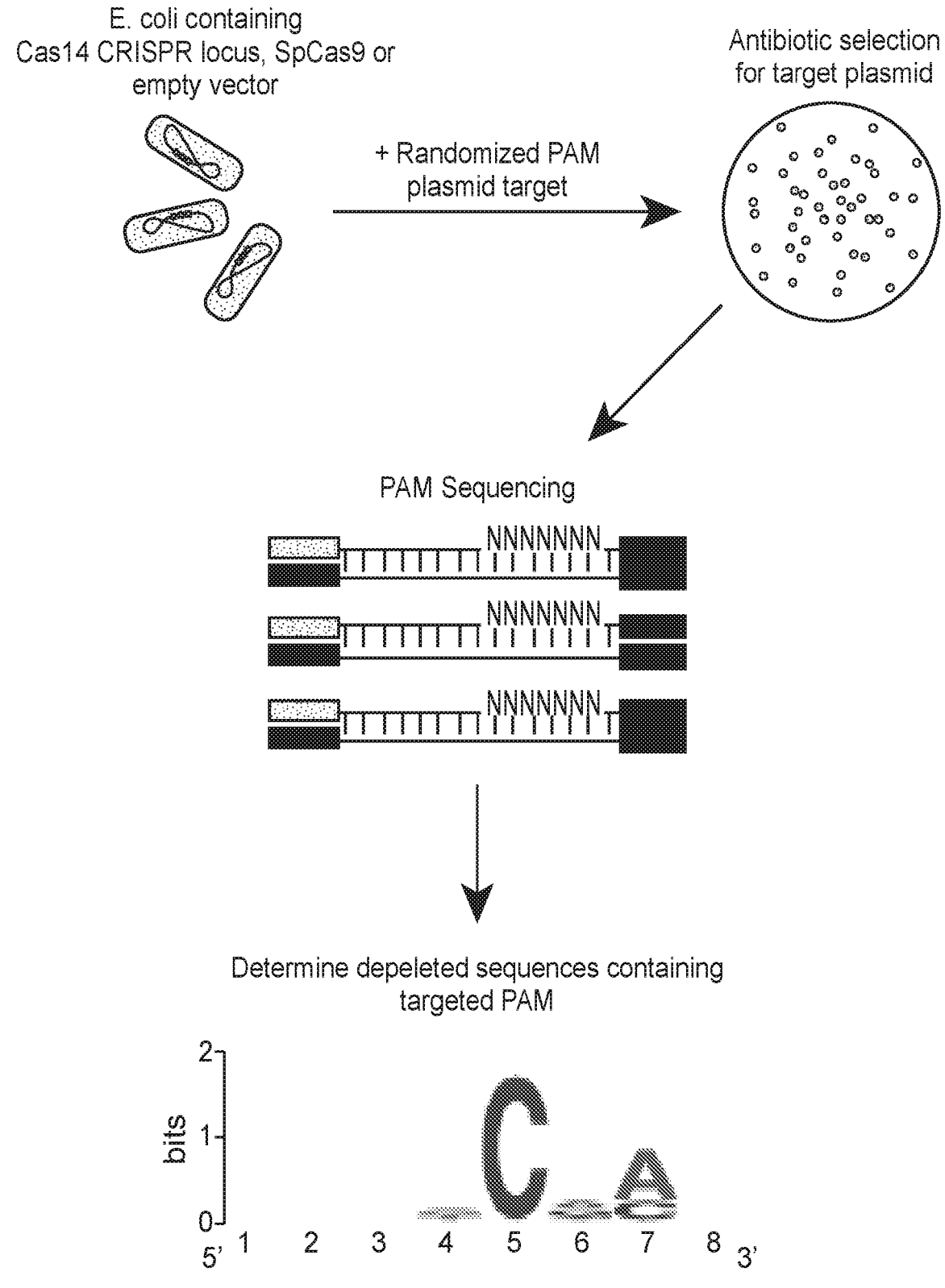
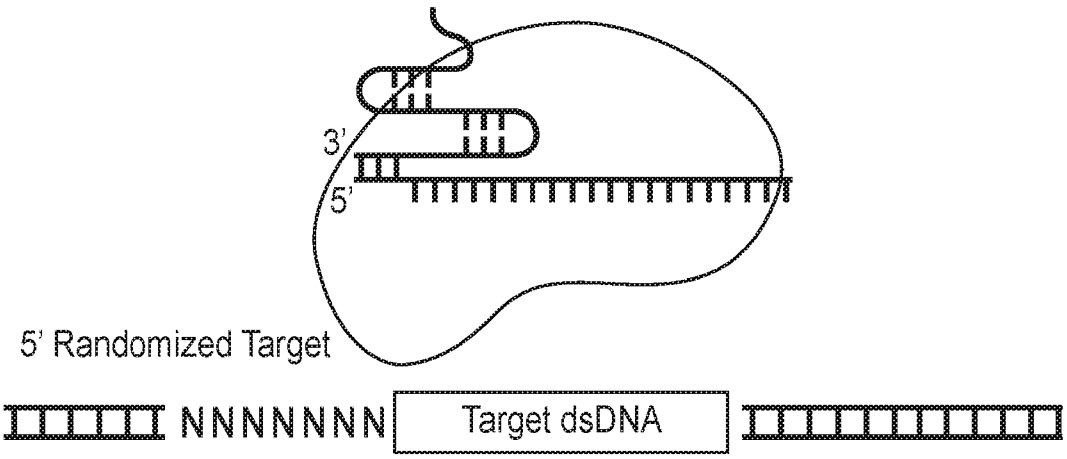
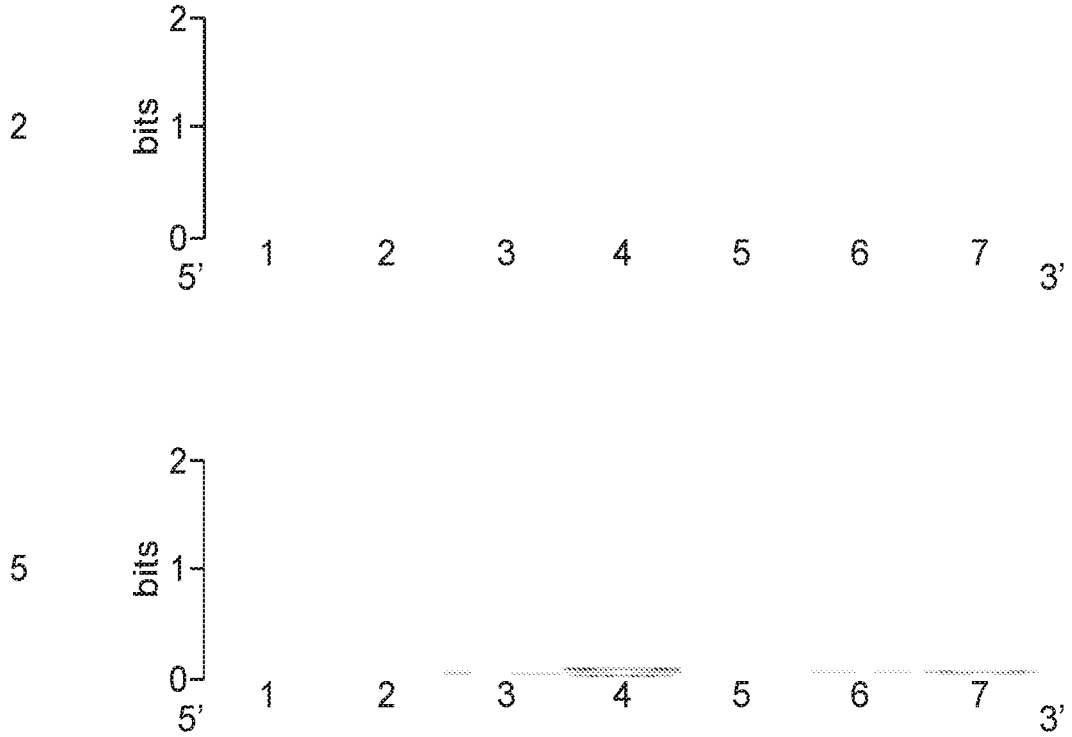

E. coli containing
Cas14 CRISPR locus, SpCas9 or
empty vector
+ Randomized PAM
plasmid target
Antibiotic selection
for target plasmid
PAM Sequencing
NNNNNNN
NNNNNNN
NNNNNNN
Determine depeleted sequences containing
targeted PAM
bits
2
1
0
5'
1
2
3
4
5
6
7
8
3'

FIG. 15 (Cont.)

B



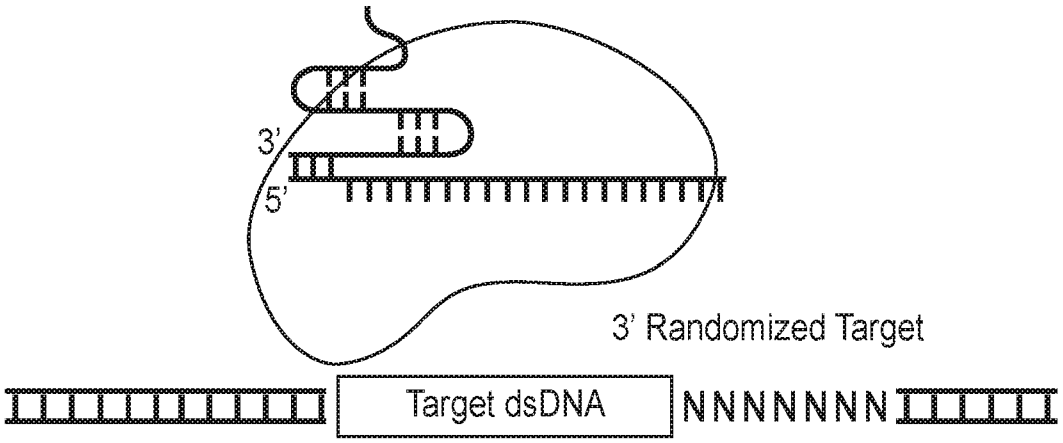
PDVT



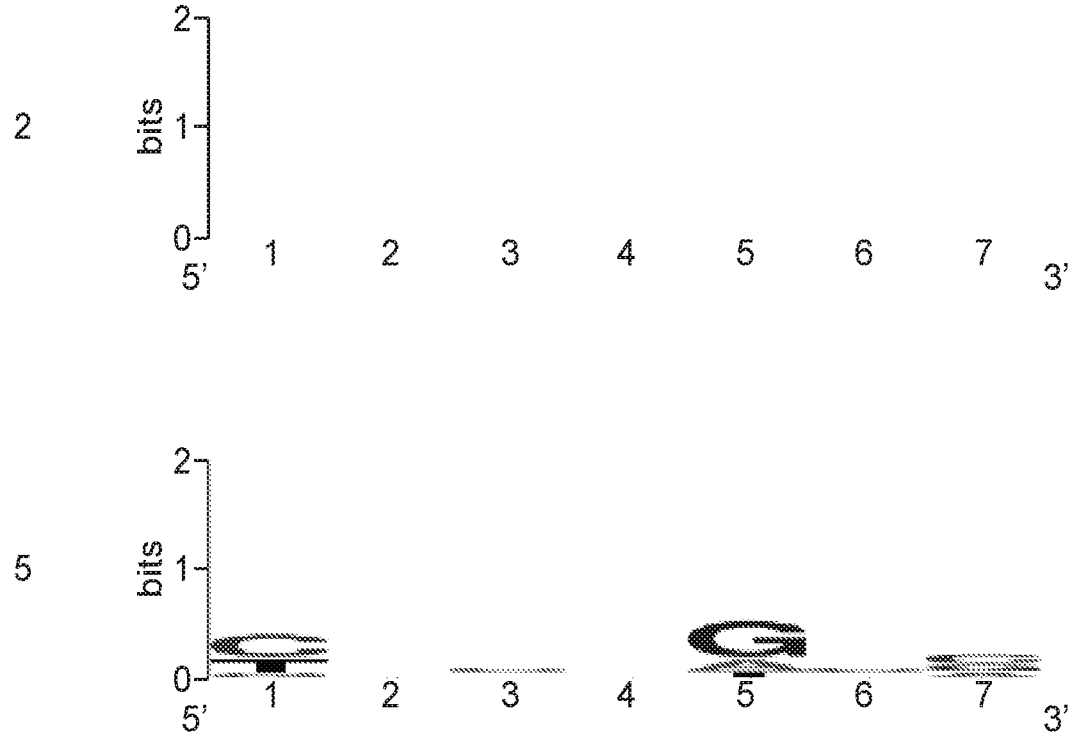
20 No Sequences

FIG. 15 (Cont.)

C



PDVT



20 No Sequences

FIG. 15 (Cont.)

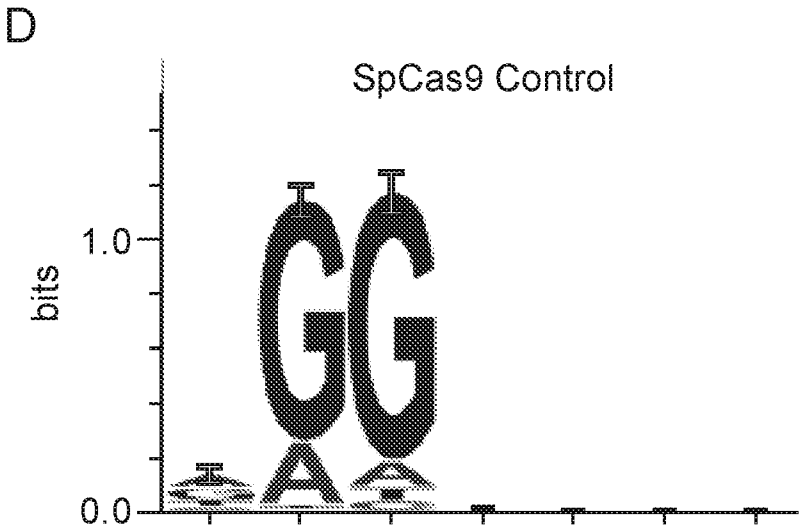


FIG. 16

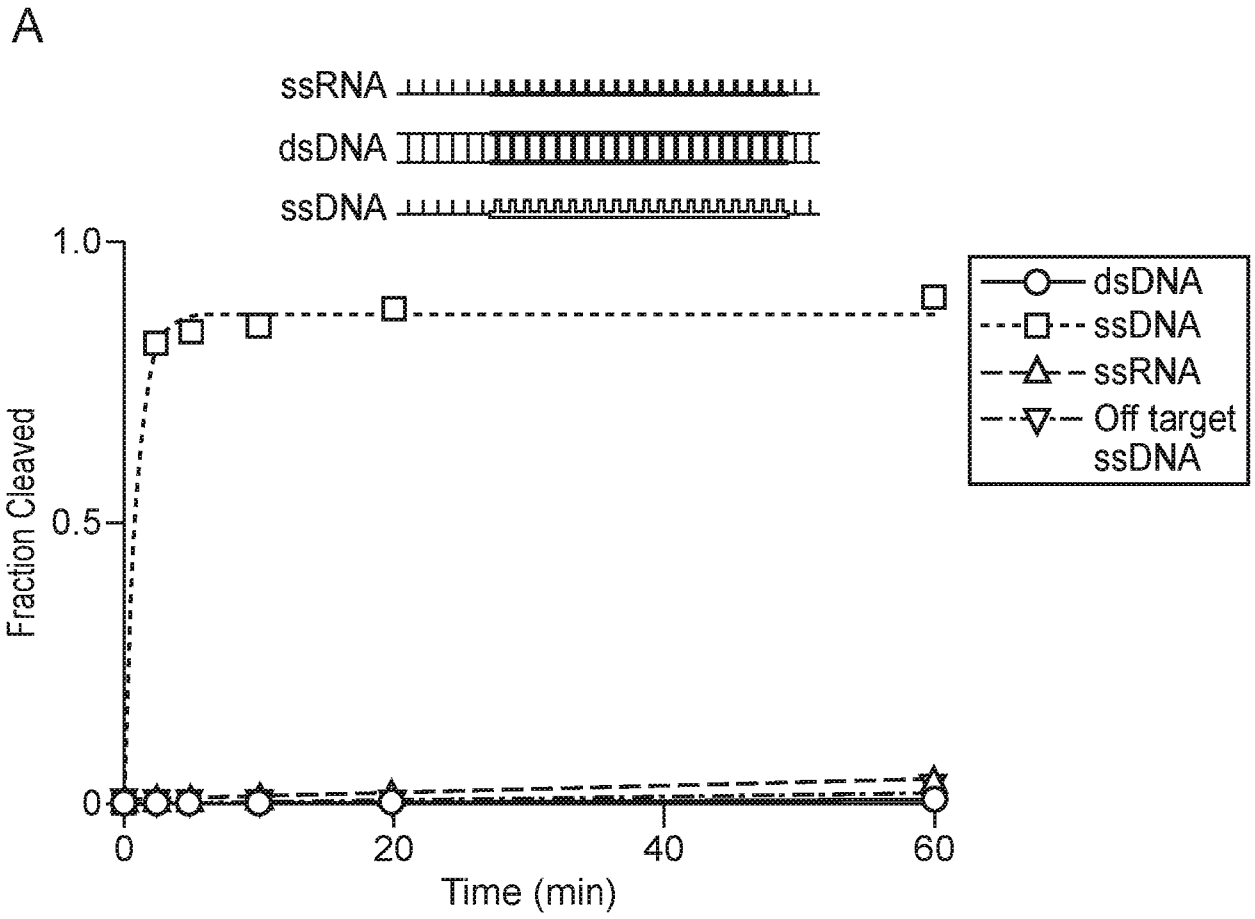
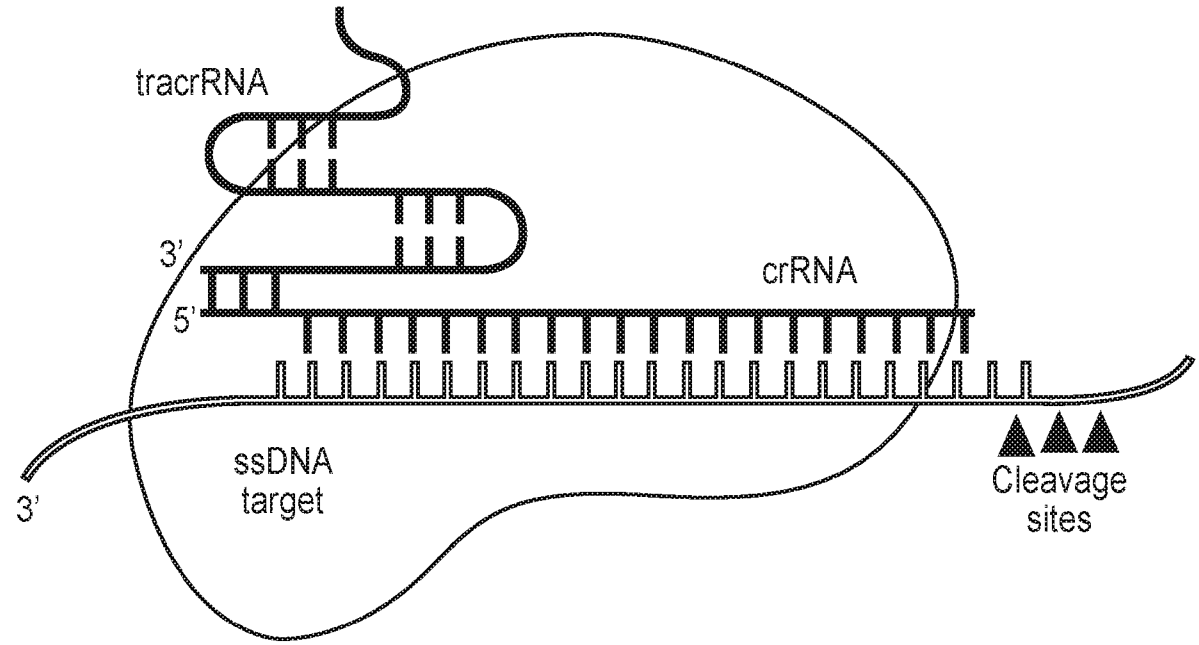


FIG. 16 (Cont.)

B


tracrRNA
3'
5'
crRNA
ssDNA
target
3'
Cleavage
sites

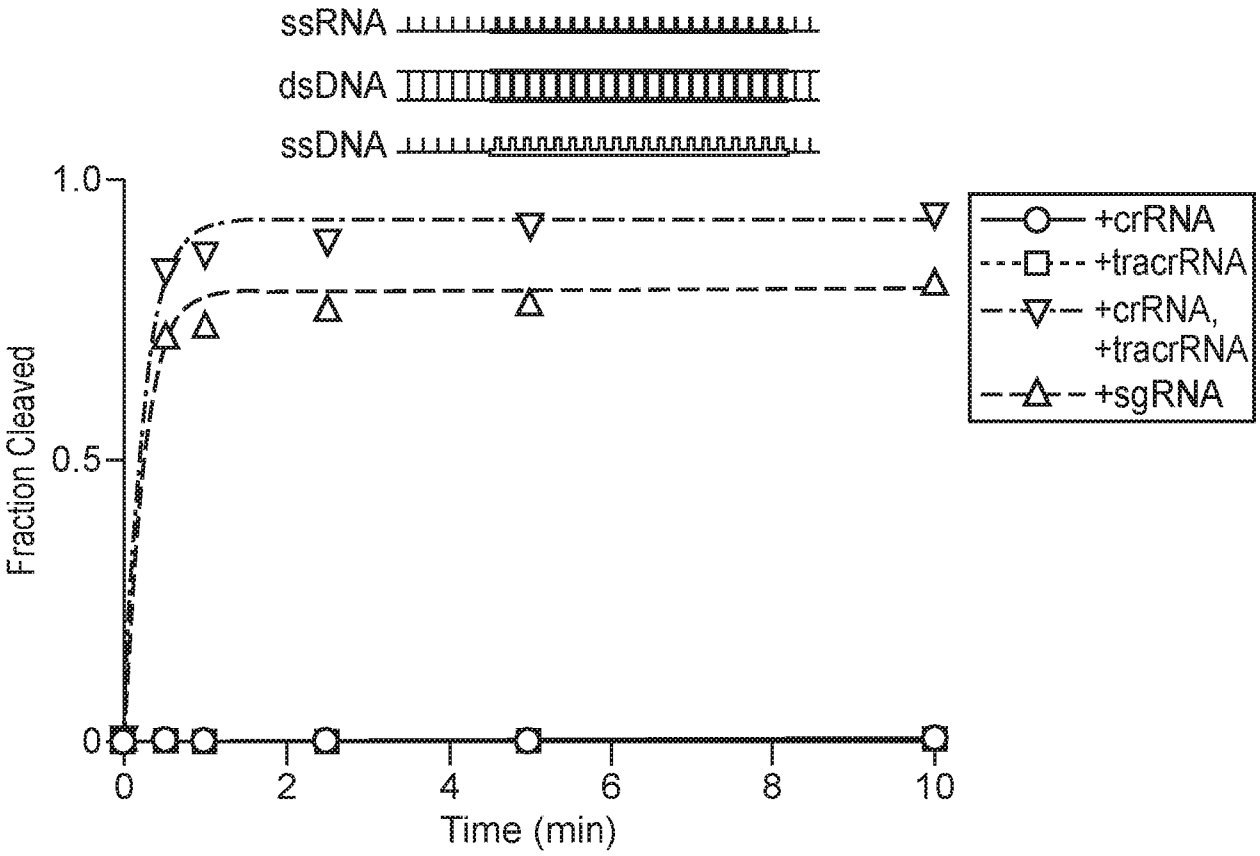
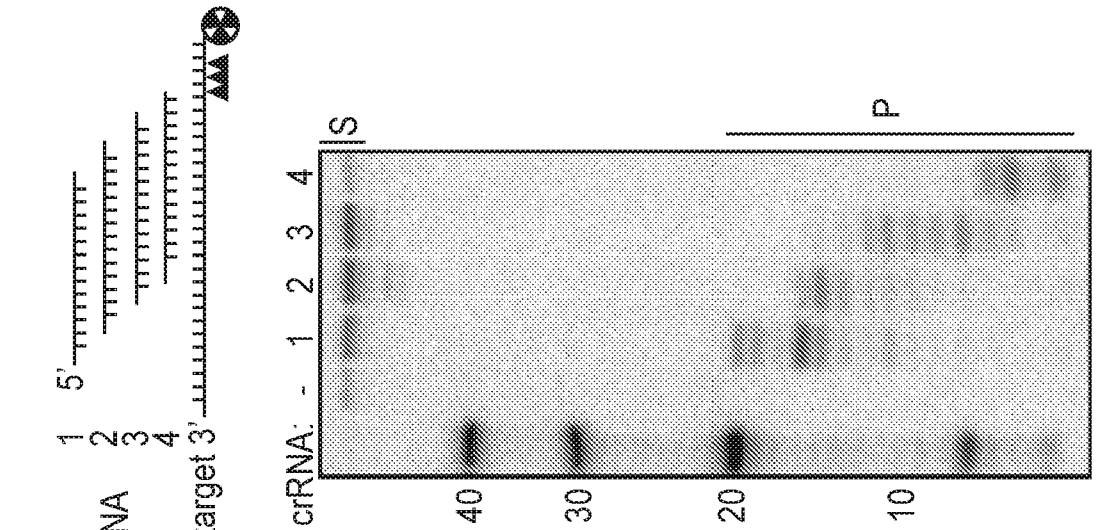

ssRNA
dsDNA
ssDNA
1.0
0.5
0
Fraction Cleaved
0
2
4
6
8
10
Time (min)
+crRNA
+tracrRNA
+crRNA,
+tracrRNA
+sgRNA

FIG. 16 (Cont.)


C
crRNA
1
2
3
4
5'
ssDNA target 3'
crRNA: - 1 2 3 4
S
P
40
30
20
10

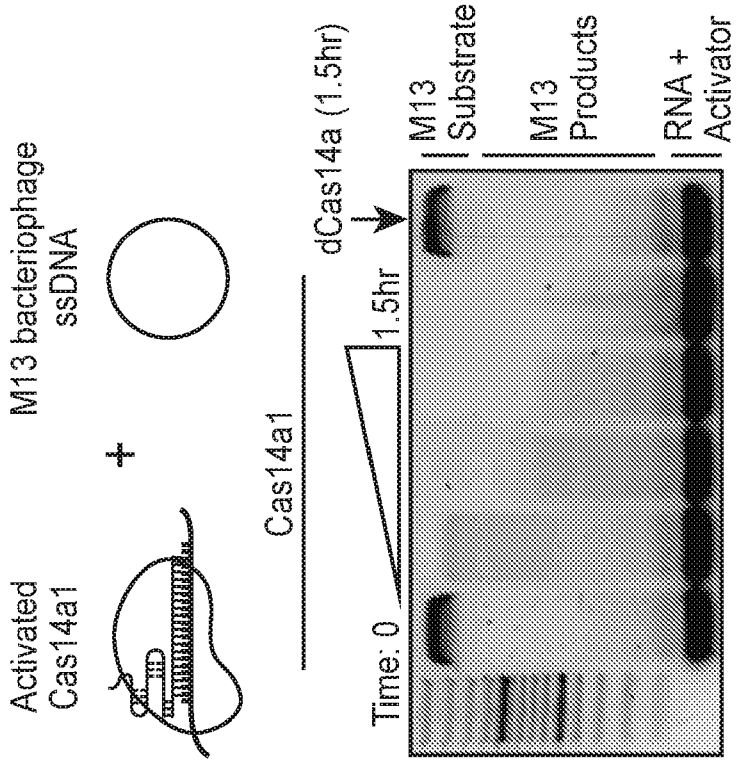

D
Activated Cas14a1
M13 bacteriophage ssDNA
Cas14a1
dCas14a (1.5hr)
Time: 0
1.5hr
M13 Substrate
M13 Products
RNA + Activator
3kb
1kb
0.5kb
100bp

FIG. 17

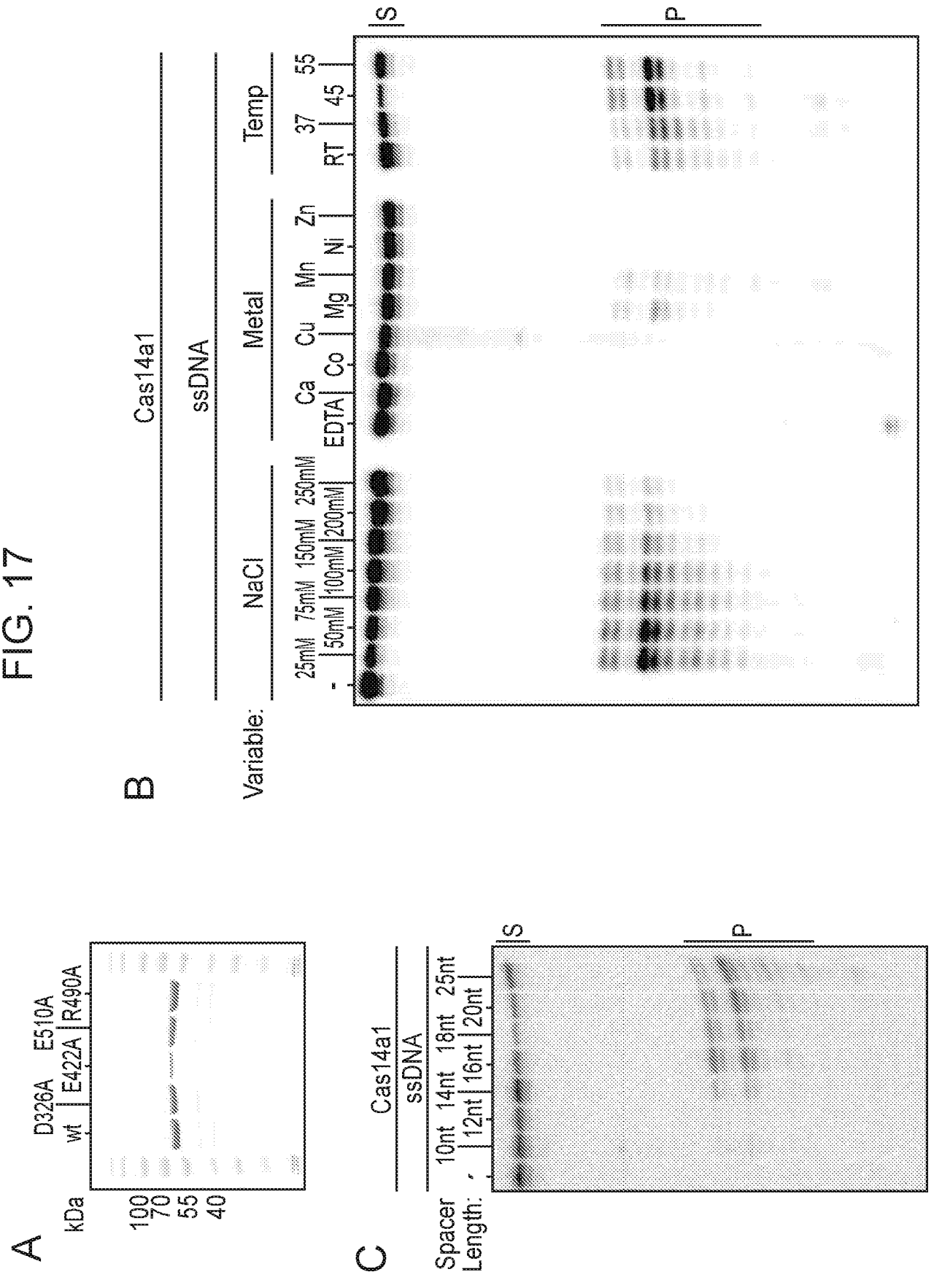

A
kDa
100
70
55
40
wt
D326A
E422A
E510A
R490A
B
Cas14a1
ssDNA
Variable:
NaCl
Metal
Temp
-
25mM
50mM
75mM
100mM
150mM
200mM
250mM
EDTA
Ca
Co
Cu
Mg
Mn
Ni
Zn
RT
37
45
55
S
P
C
Cas14a1
ssDNA
Spacer Length:
-
10nt
12nt
14nt
16nt
18nt
20nt
25nt
S
P

FIG. 17 (Cont.)

D

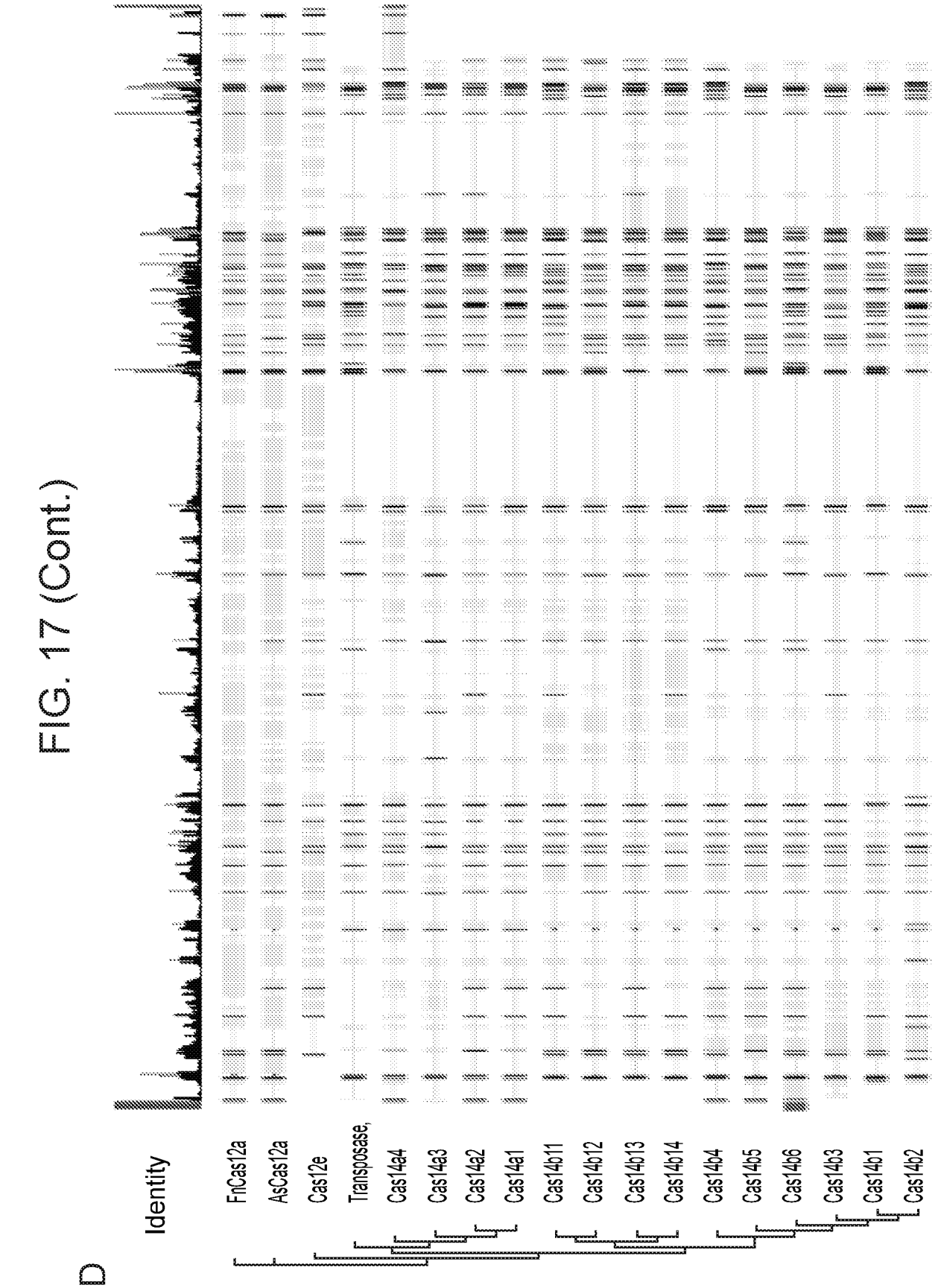

Identity
FnCas12a
AsCas12a
Cas12e
Transposase,
Cas14a4
Cas14a3
Cas14a2
Cas14a1
Cas14b11
Cas14b12
Cas14b13
Cas14b14
Cas14b4
Cas14b5
Cas14b6
Cas14b3
Cas14b1
Cas14b2

FIG. 17 (Cont.)

D (Cont.)

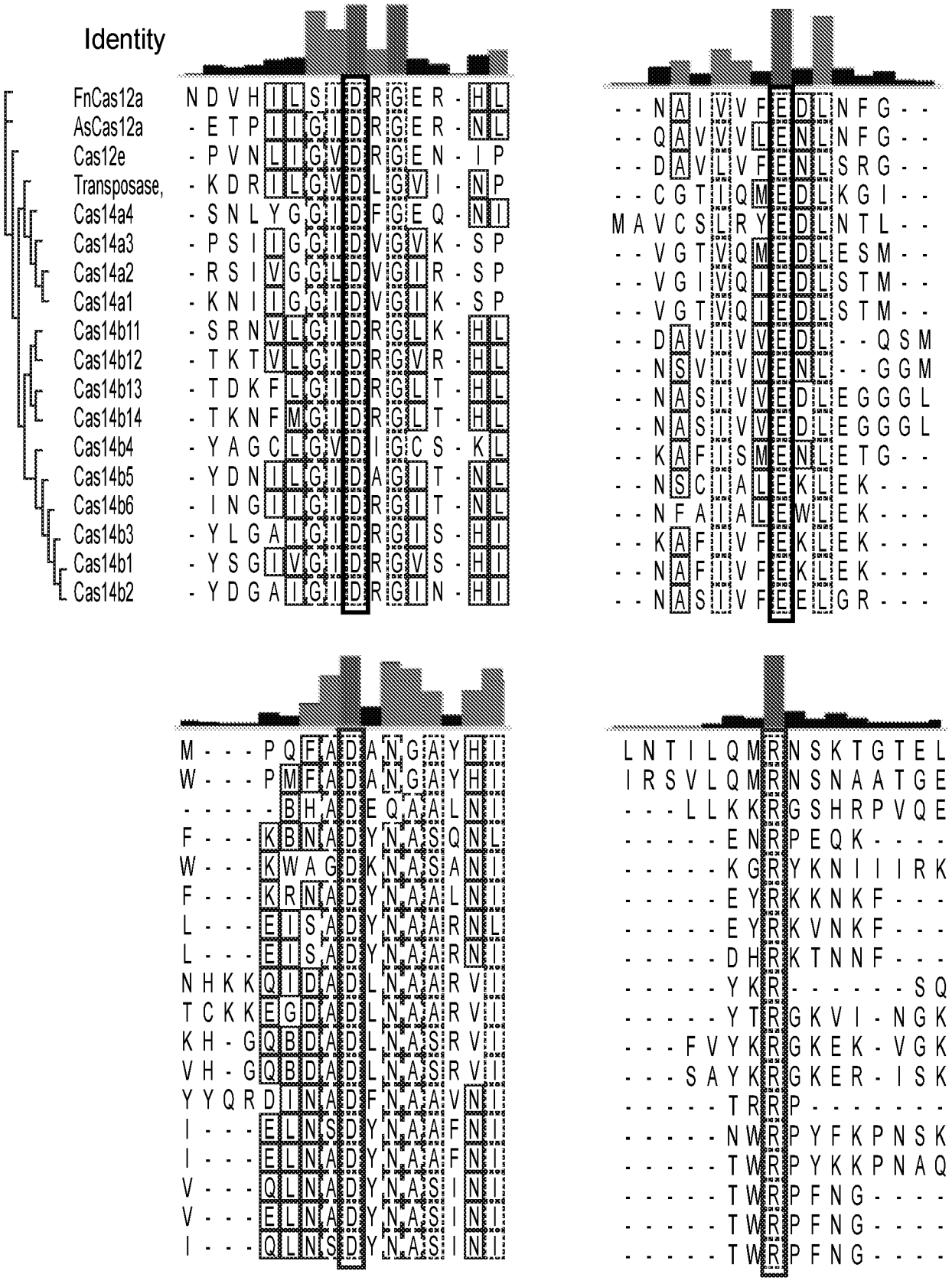


FIG. 17 (Cont.)

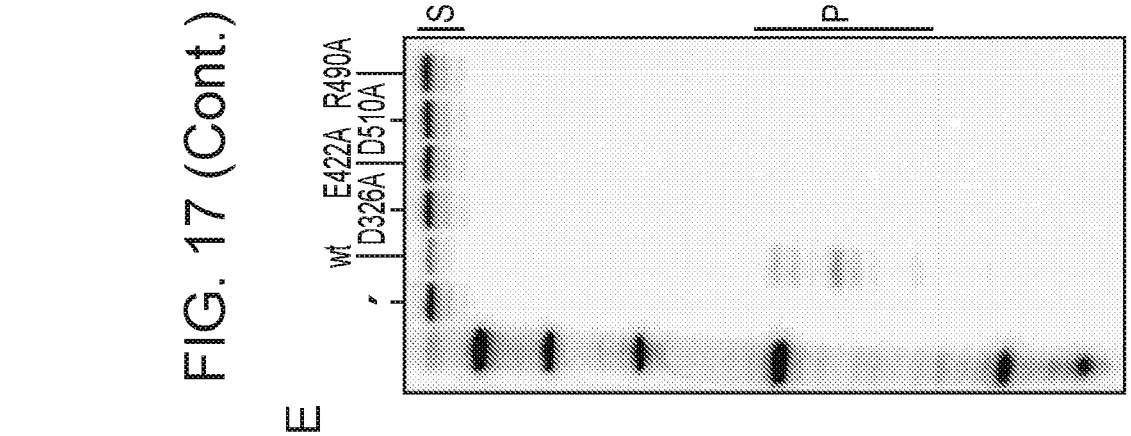


FIG. 18

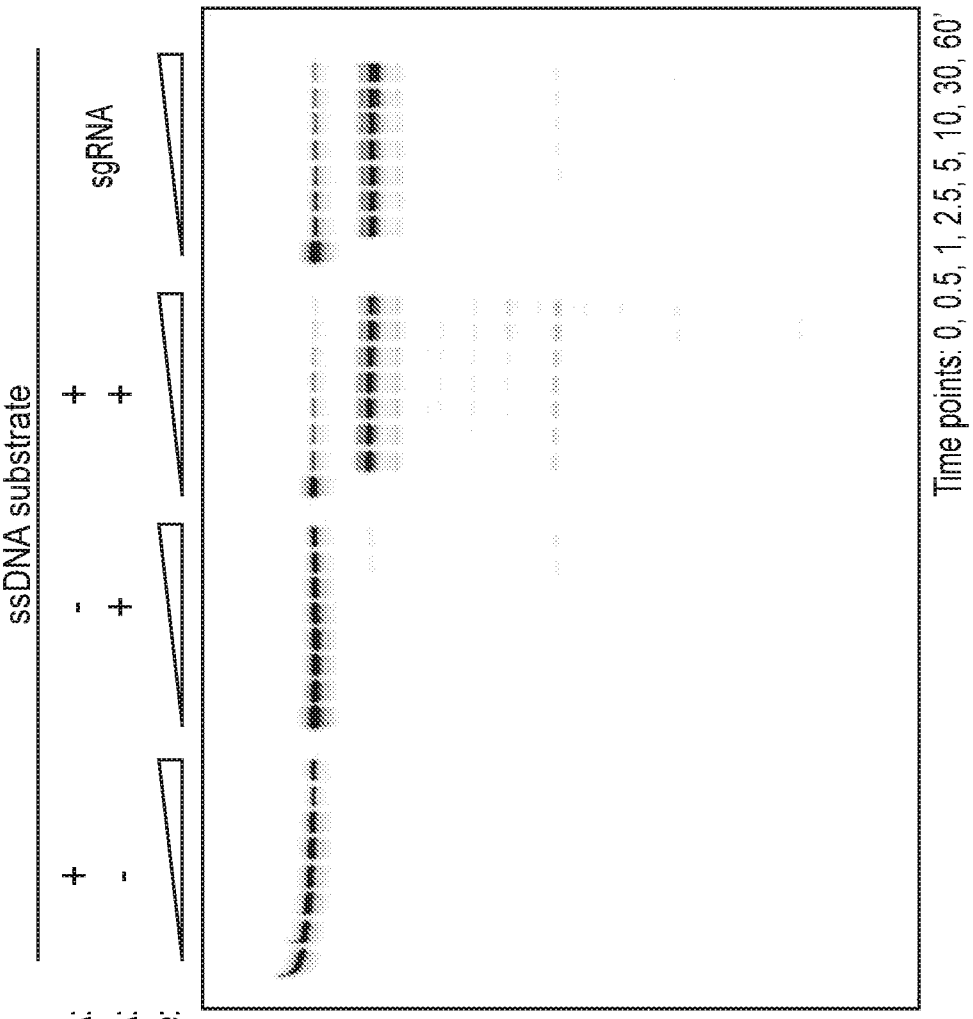


FIG. 19

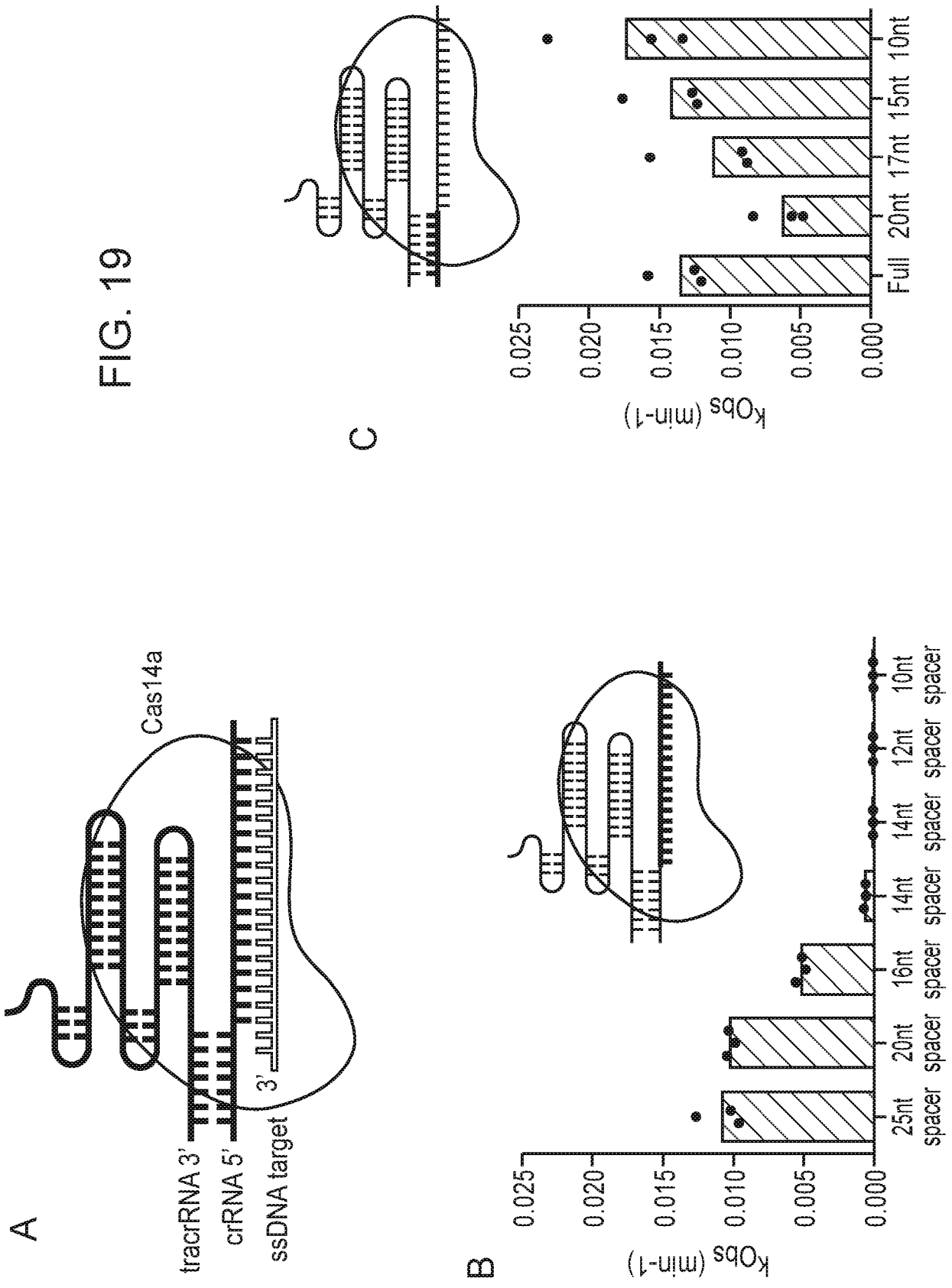

A
Cas14a
tracrRNA 3'
crRNA 5'
3'
ssDNA target
B
kObs (min-1)
0.025
0.020
0.015
0.010
0.005
0.000
25nt spacer
20nt spacer
16nt spacer
14nt spacer
14nt spacer
12nt spacer
10nt spacer
C
kObs (min-1)
0.025
0.020
0.015
0.010
0.005
0.000
Full
20nt
17nt
15nt
10nt

FIG. 19 (Cont.)

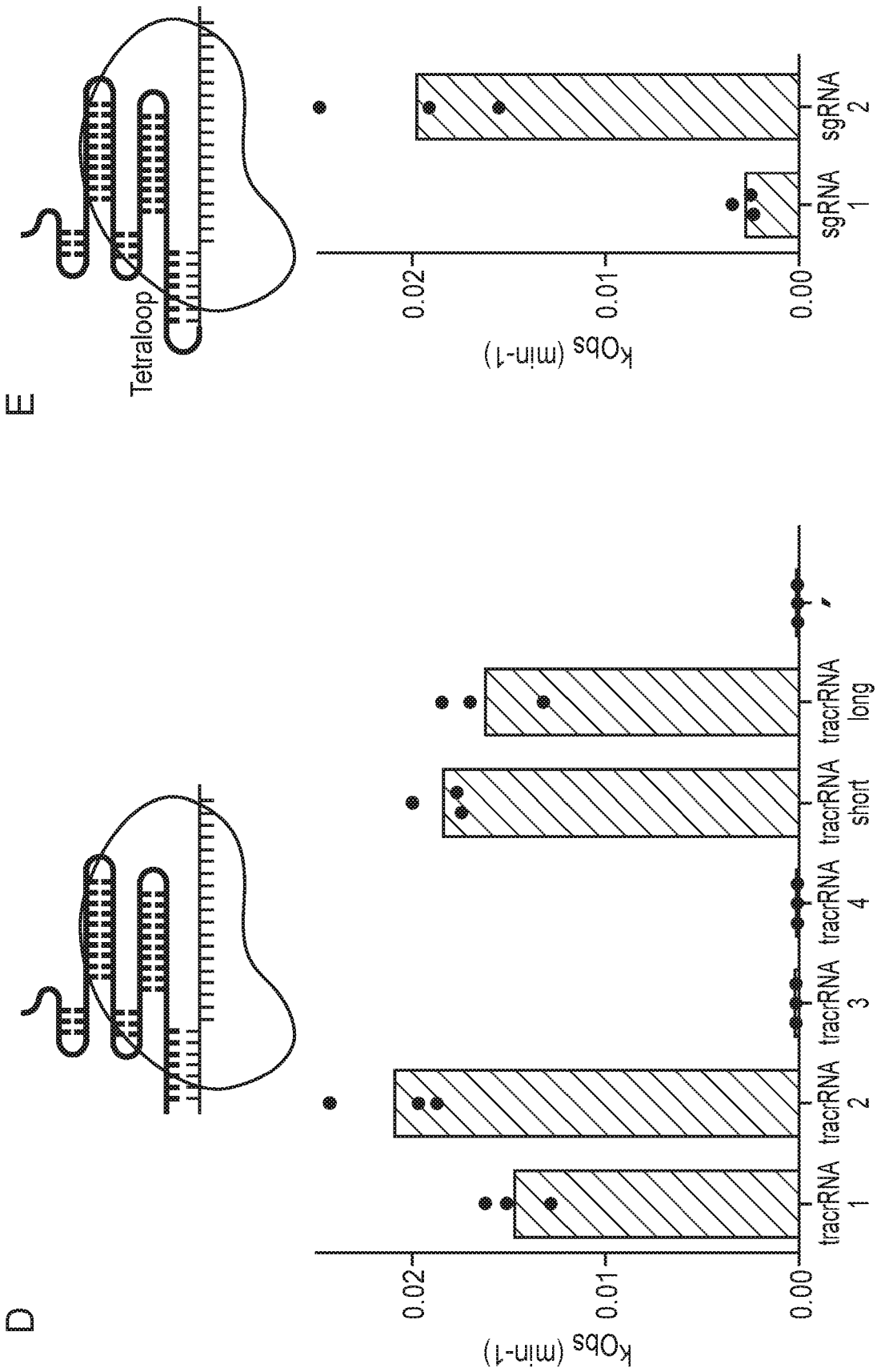

D
E
Tetraloop
kObs (min-1)
0.00
0.01
0.02
tracrRNA 1
tracrRNA 2
tracrRNA 3
tracrRNA 4
tracrRNA short
tracrRNA long
sgRNA 1
sgRNA 2

FIG. 19 (Cont.)

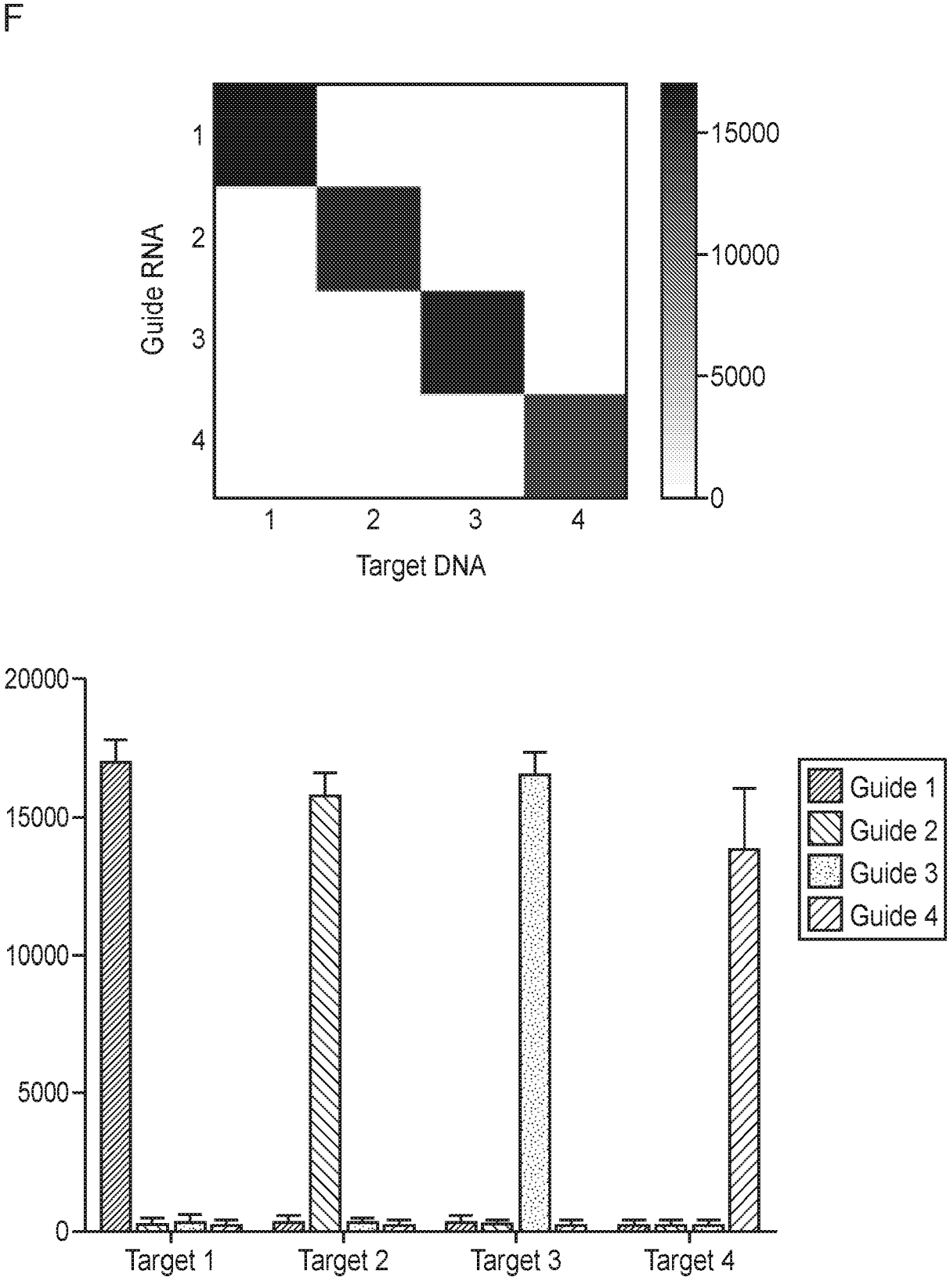

F
Guide RNA
1
2
3
4
Target DNA
15000
10000
5000
0
Fluorescence (AU)
20000
15000
10000
5000
0
Target 1
Target 2
Target 3
Target 4
Guide 1
Guide 2
Guide 3
Guide 4

FIG. 20

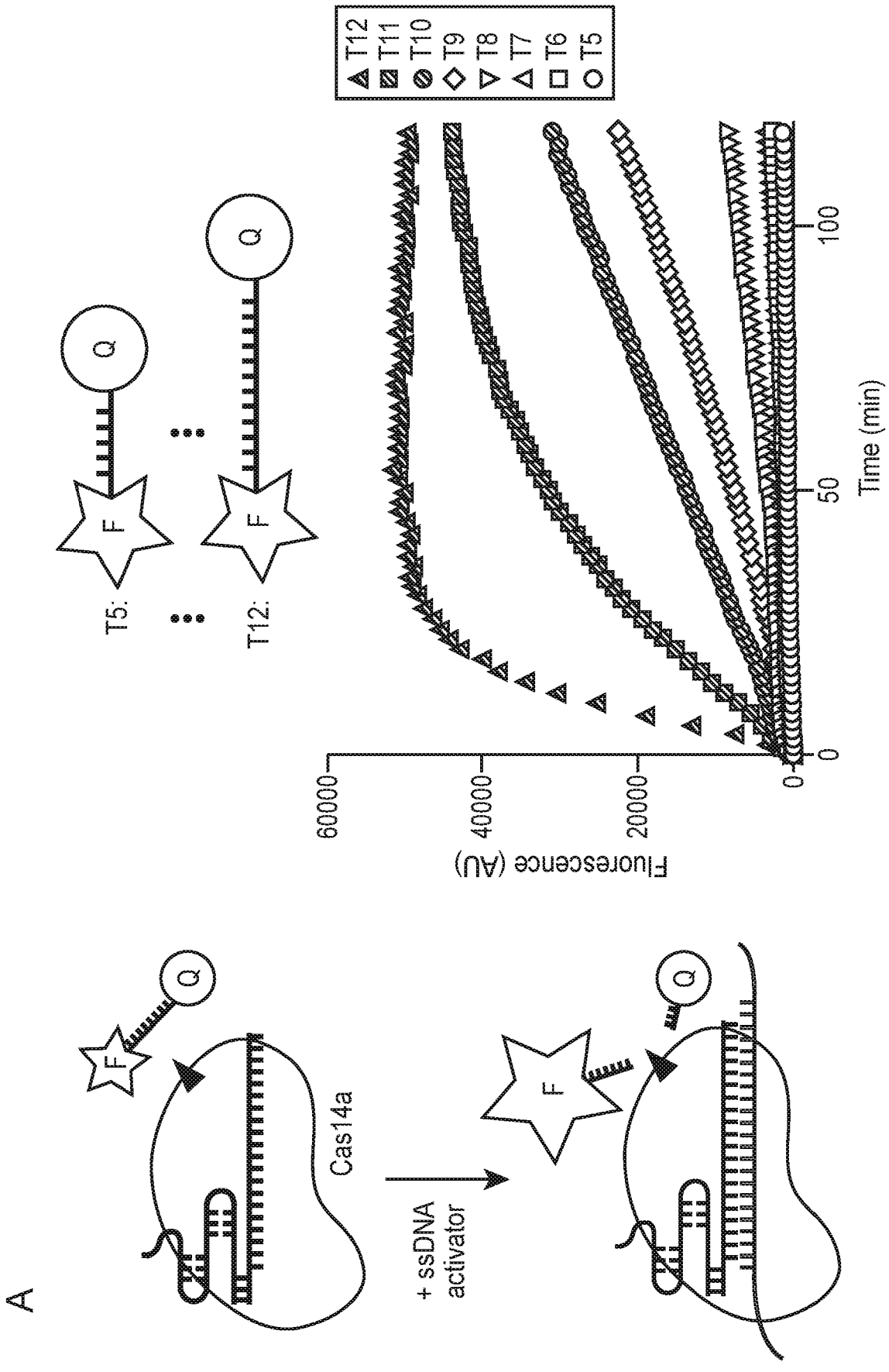

A
Cas14a
+ ssDNA activator
F
Q
T5:
T12:
T12
T11
T10
T9
T8
T7
T6
T5
Fluorescence (AU)
60000
40000
20000
0
0
50
100
Time (min)

FIG. 20 (Cont.)

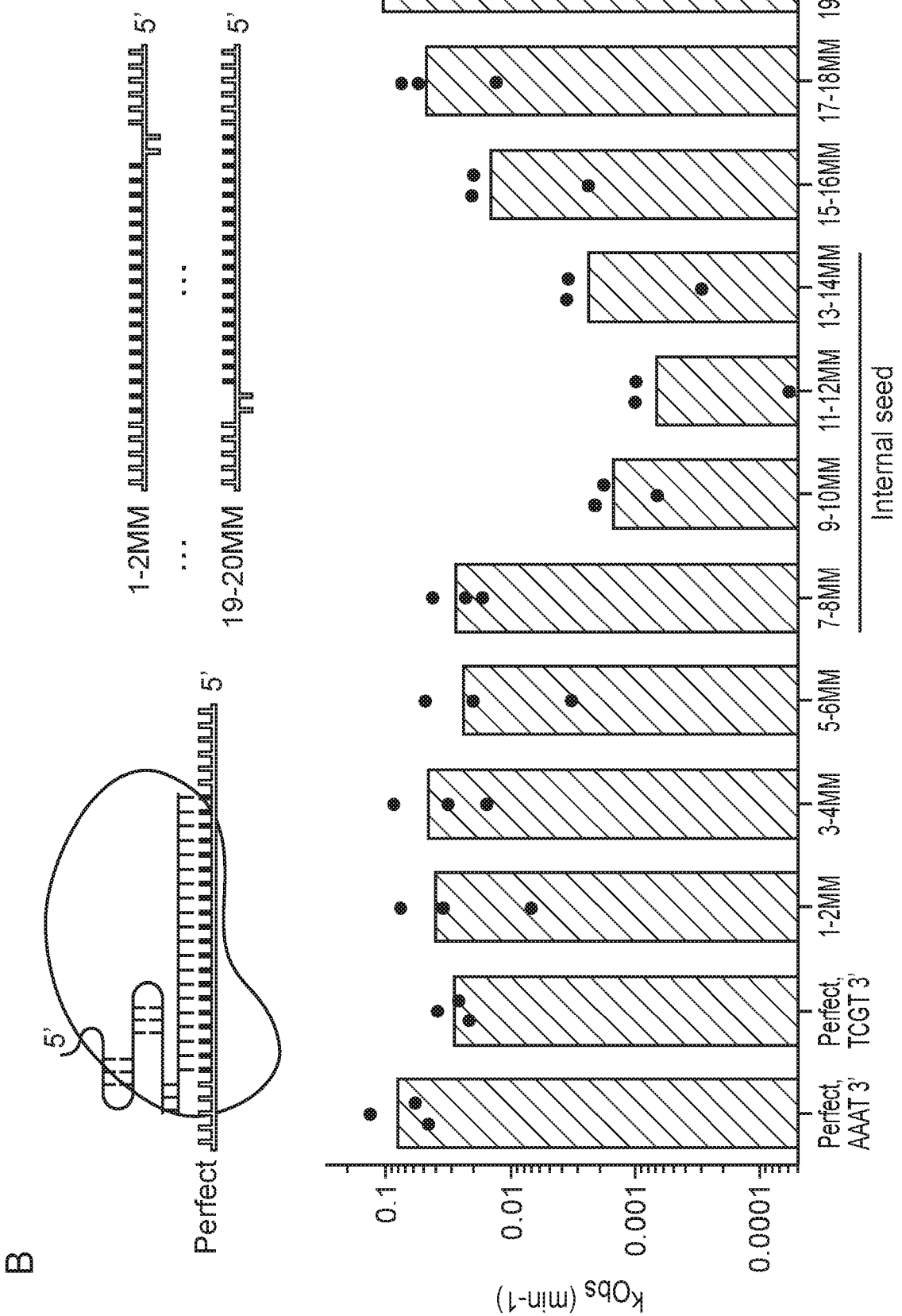

B
5'
Perfect
5'
1-2MM
19-20MM
5'
5'
K_{Obs} (min-1)
0.1
0.01
0.001
0.0001
Perfect, AAAT 3'
Perfect, TCGT 3'
1-2MM
3-4MM
5-6MM
7-8MM
9-10MM
11-12MM
13-14MM
15-16MM
17-18MM
19-20MM
Internal seed

FIG. 20 (Cont.)

C

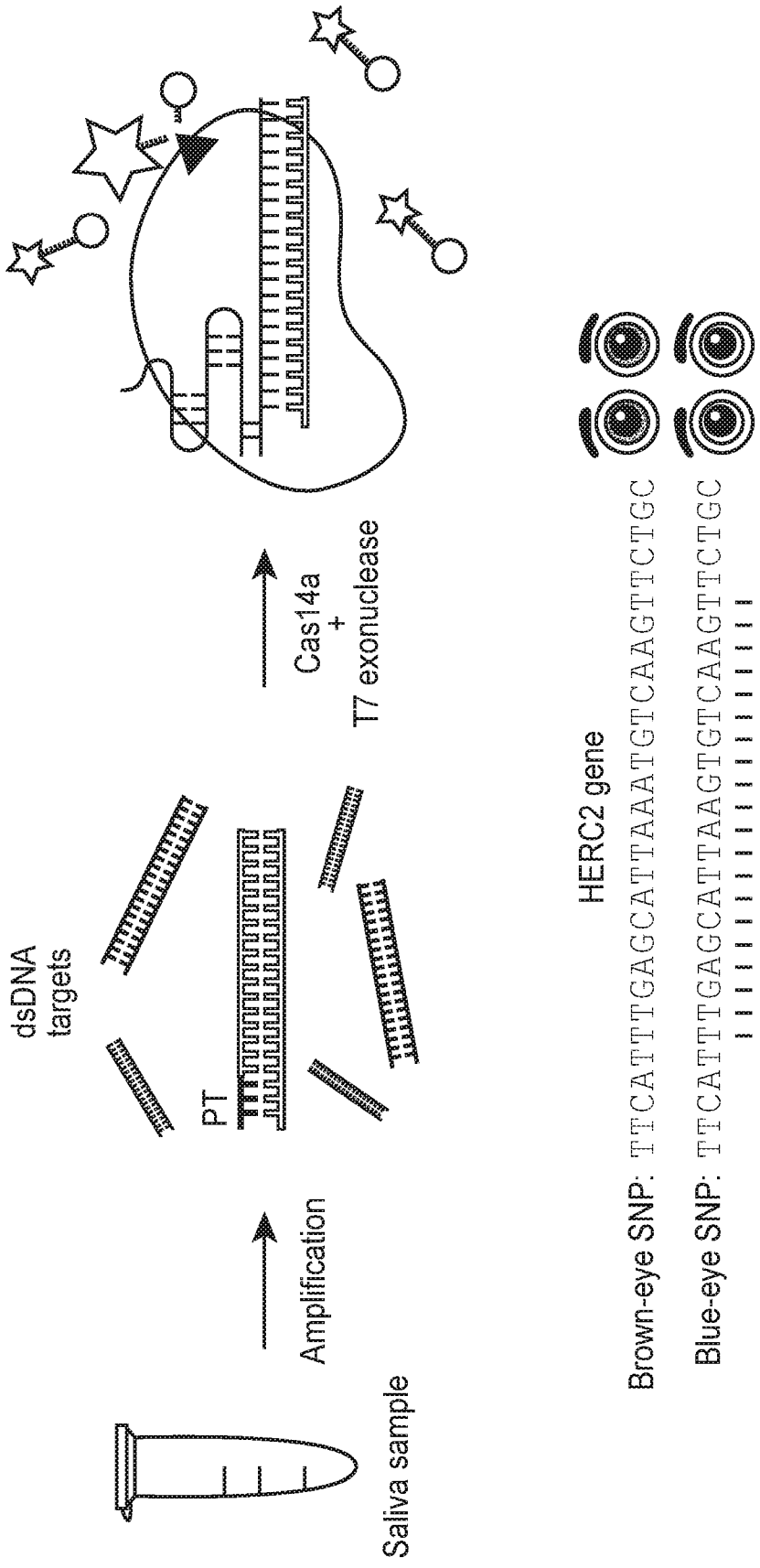

Saliva sample
Amplification
PT
dsDNA
targets
Cas14a
+
T7 exonuclease
HERC2 gene
Brown-eye SNP: TTCATTTGAGCATTAAATGTCAAGTTCTGC
Blue-eye SNP: TTCATTTGAGCATTAAGTGTCAAGTTCTGC
crRNA: AACTCGTAATTCACAGTTCA

FIG. 20 (Cont.)

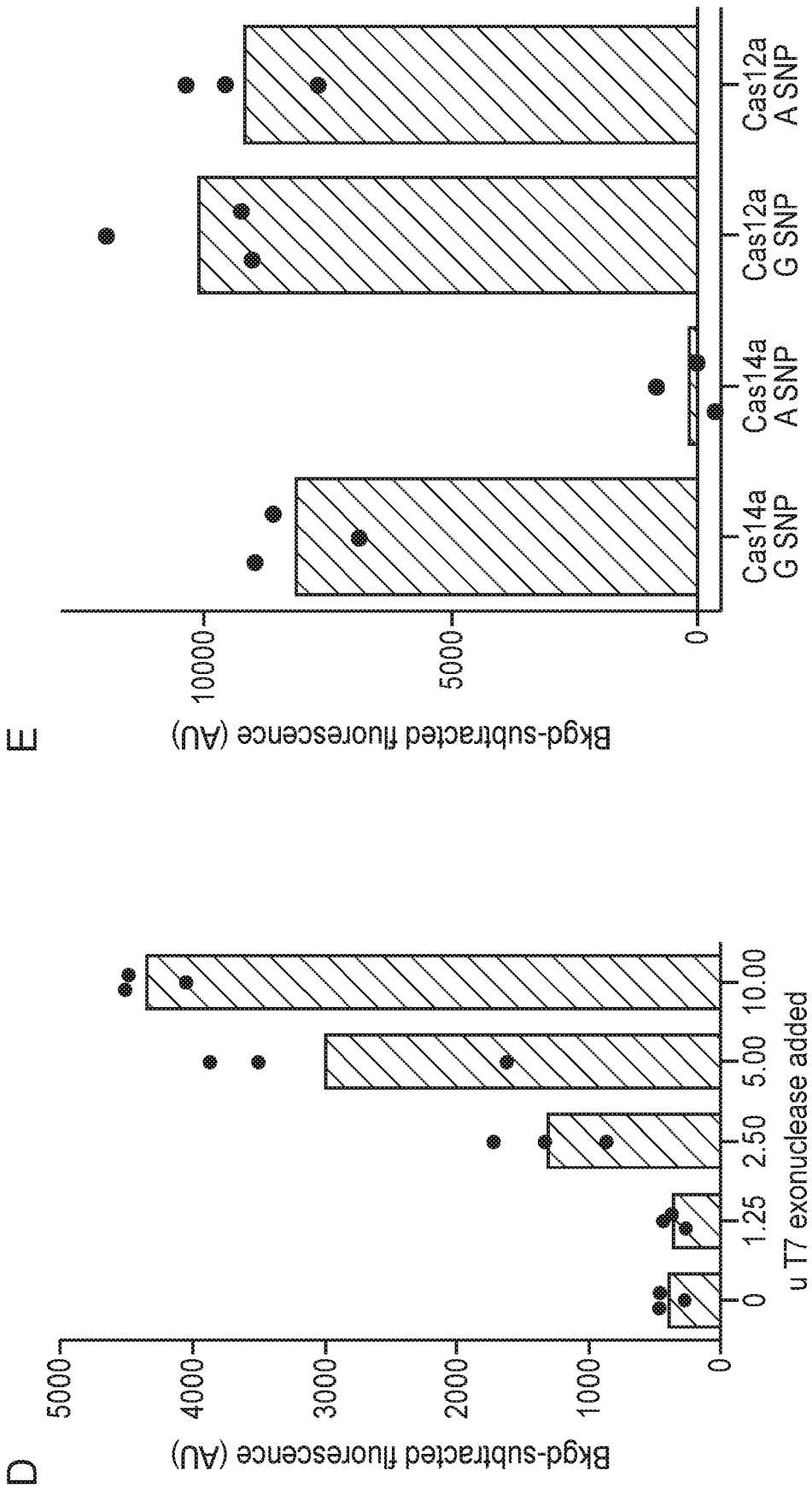

D
Bkgd-subtracted fluorescence (AU)
5000
4000
3000
2000
1000
0
0
1.25
2.50
5.00
10.00
u T7 exonuclease added
E
Bkgd-subtracted fluorescence (AU)
10000
5000
0
Cas14a G SNP
Cas14a A SNP
Cas12a G SNP
Cas12a A SNP

FIG. 21

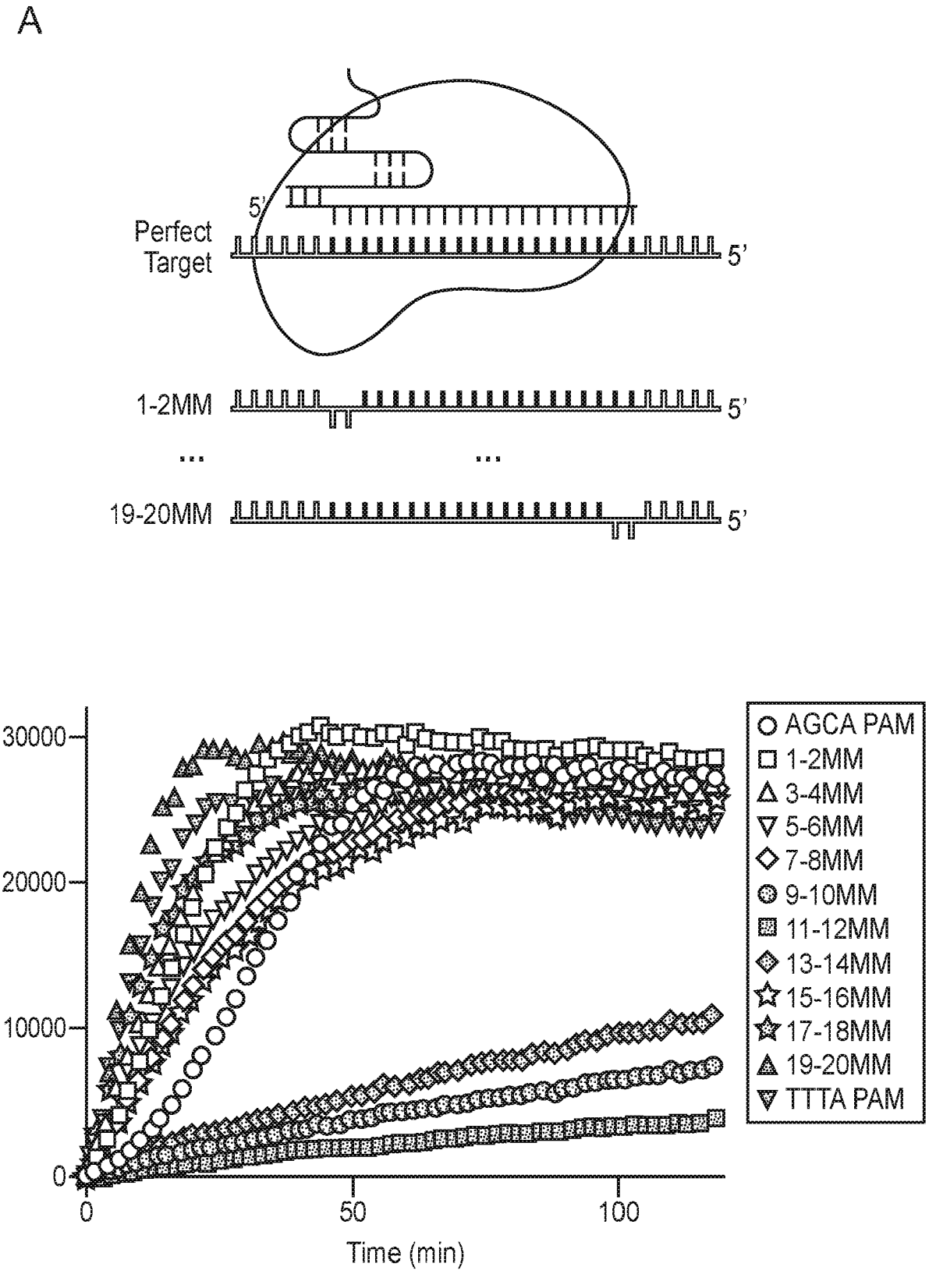

A
5'
Perfect Target
5'
1-2MM
5'
...
...
19-20MM
5'
30000
20000
10000
0
Fluorescence (AU)
0
50
100
Time (min)
AGCA PAM
1-2MM
3-4MM
5-6MM
7-8MM
9-10MM
11-12MM
13-14MM
15-16MM
17-18MM
19-20MM
TTTA PAM

FIG. 21 (Cont.)

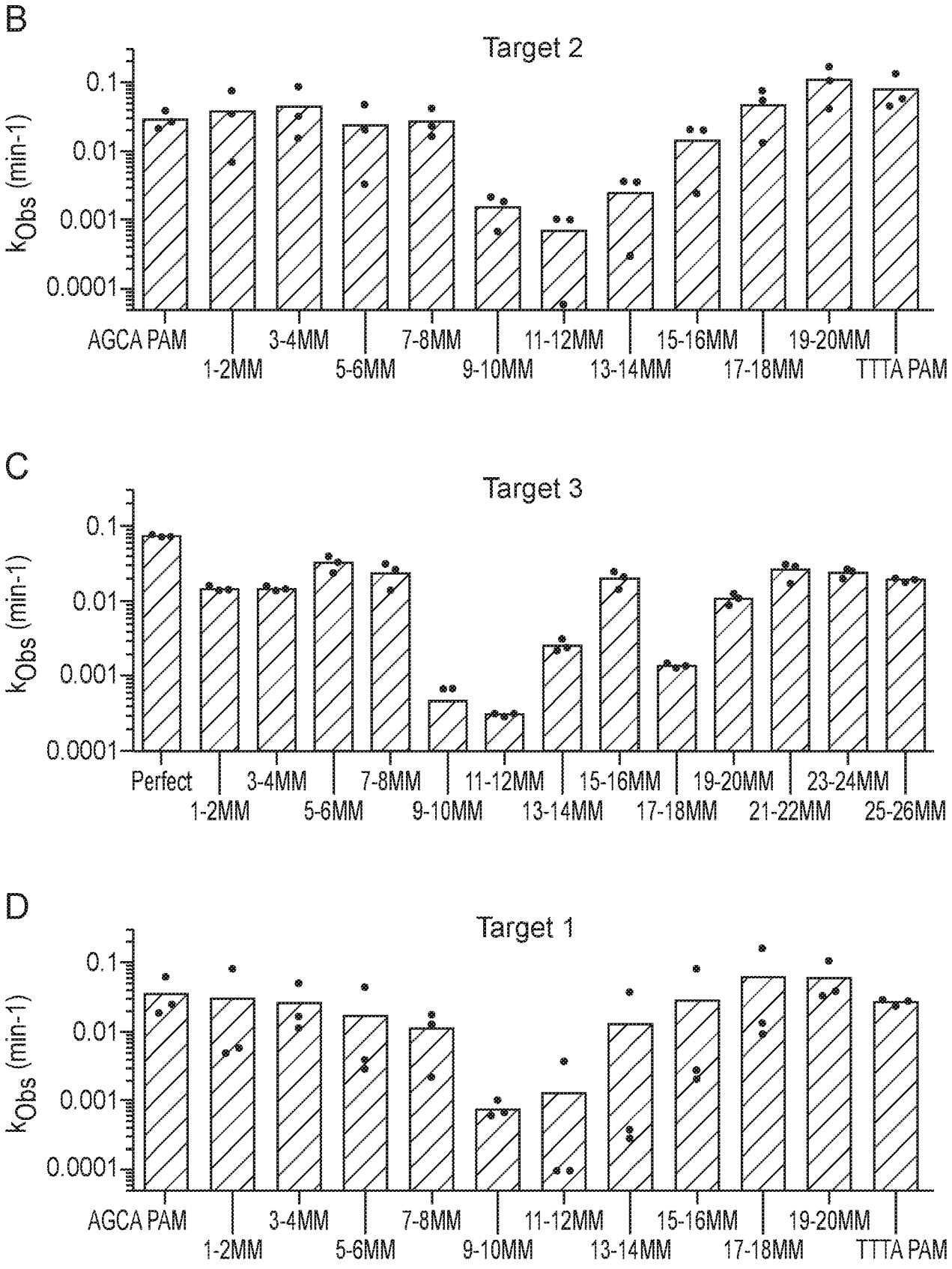

B
Target 2
k_{Obs} (min-1)
0.1
0.01
0.001
0.0001
AGCA PAM
1-2MM
3-4MM
5-6MM
7-8MM
9-10MM
11-12MM
13-14MM
15-16MM
17-18MM
19-20MM
TTTA PAM
C
Target 3
k_{Obs} (min-1)
0.1
0.01
0.001
0.0001
Perfect
1-2MM
3-4MM
5-6MM
7-8MM
9-10MM
11-12MM
13-14MM
15-16MM
17-18MM
19-20MM
21-22MM
23-24MM
25-26MM
D
Target 1
k_{Obs} (min-1)
0.1
0.01
0.001
0.0001
AGCA PAM
1-2MM
3-4MM
5-6MM
7-8MM
9-10MM
11-12MM
13-14MM
15-16MM
17-18MM
19-20MM
TTTA PAM

FIG. 21 (Cont.)

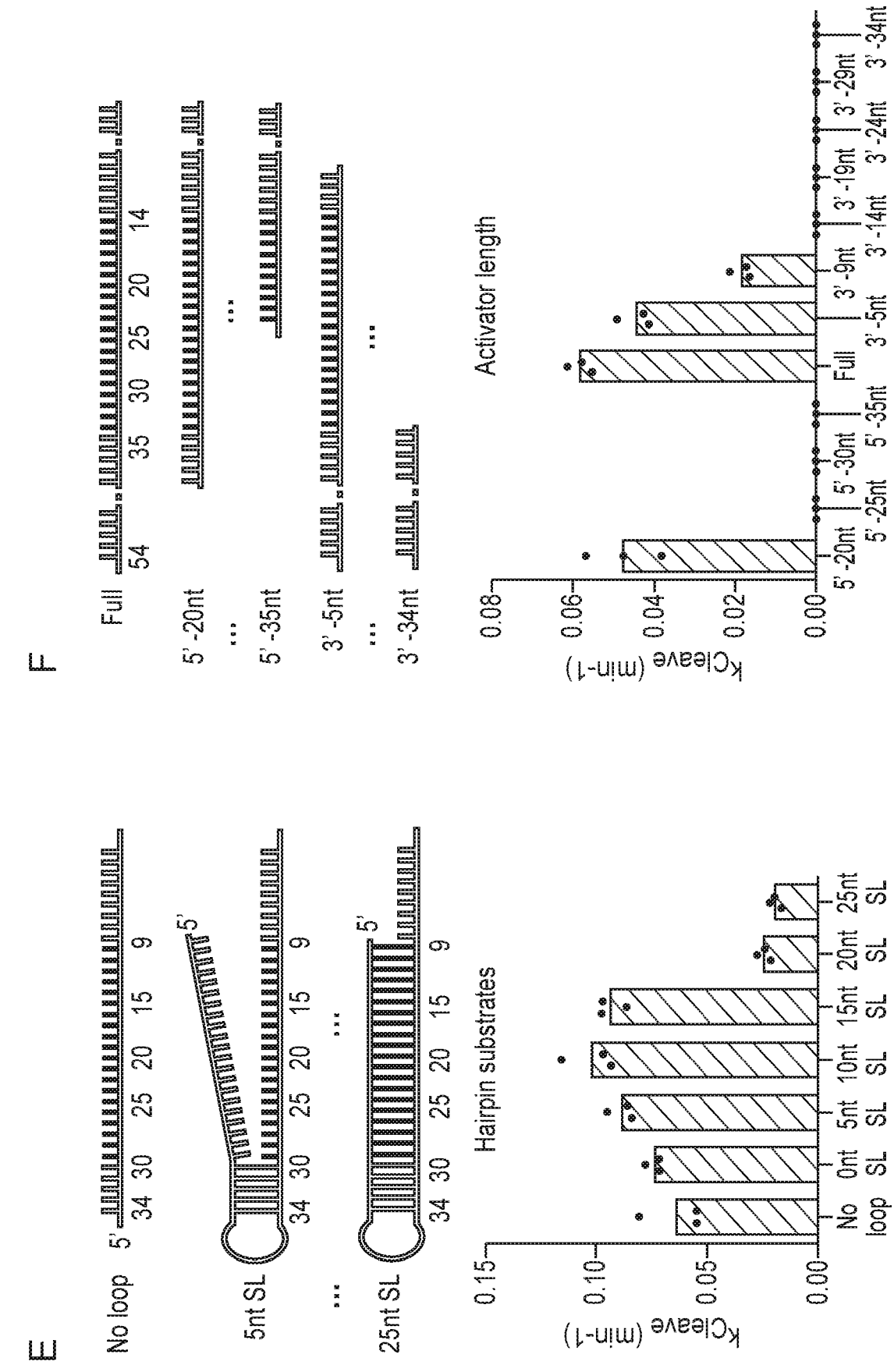

E
No loop
5'
34 30 25 20 15 9
5nt SL
5'
34 30 25 20 15 9
25nt SL
5'
34 30 25 20 15 9
Hairpin substrates
kCleave (min-1)
0.15
0.10
0.05
0.00
No loop
0nt SL
5nt SL
10nt SL
15nt SL
20nt SL
25nt SL
F
Full
54 35 30 25 20 14
5' -20nt
5' -35nt
3' -5nt
3' -34nt
Activator length
kCleave (min-1)
0.08
0.06
0.04
0.02
0.00
5' -20nt
5' -25nt
5' -30nt
5' -35nt
Full
3' -5nt
3' -9nt
3' -14nt
3' -19nt
3' -24nt
3' -29nt
3' -34nt

FIG. 22

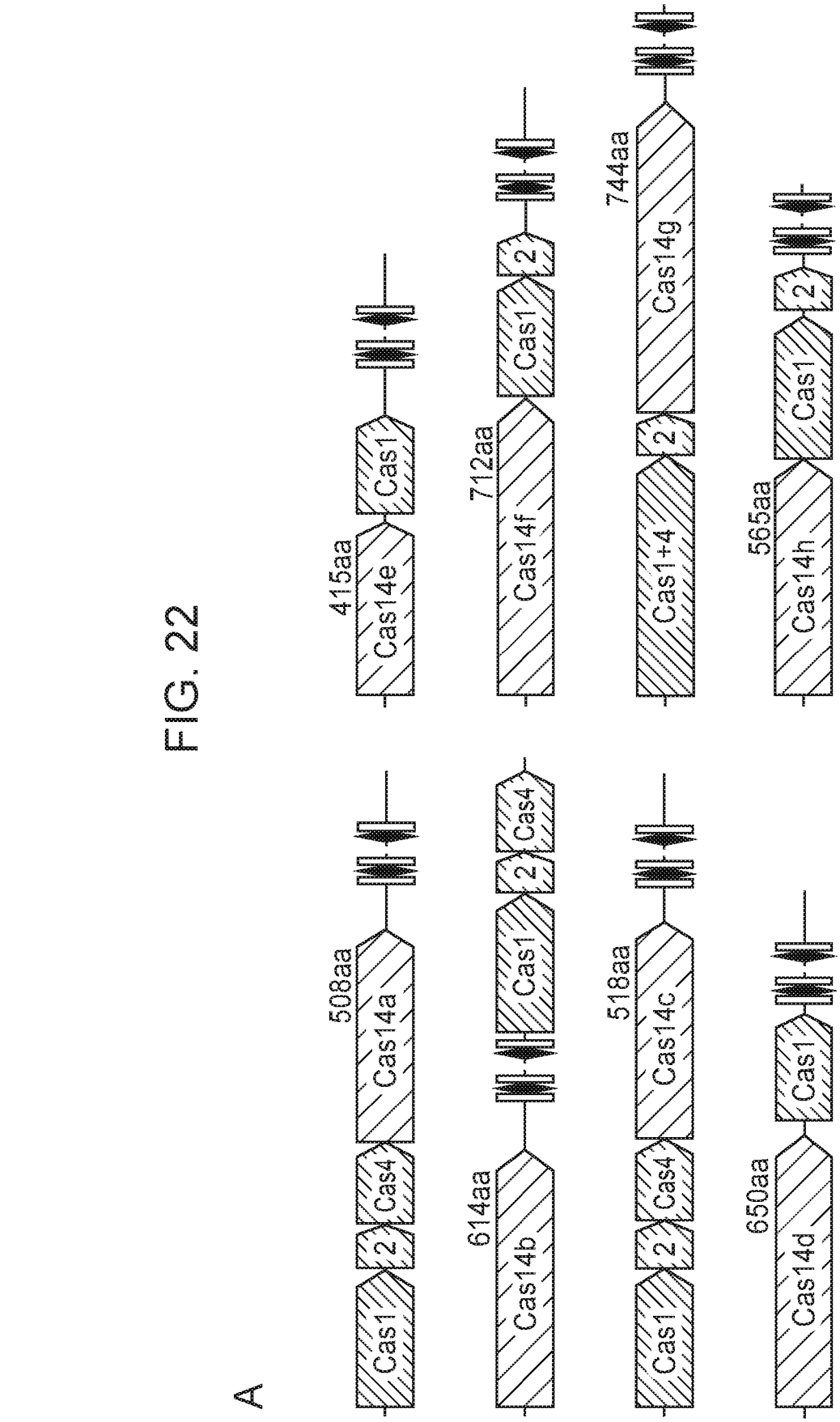

A
Cas1
2
Cas4
Cas14a
508aa
Cas14e
415aa
Cas1
Cas14b
614aa
Cas1
2
Cas4
Cas14f
712aa
Cas1
2
Cas1
2
Cas4
Cas14c
518aa
Cas1
Cas14d
650aa
Cas1
2
Cas1+4
Cas14g
744aa
Cas1
2
Cas14h
565aa

FIG. 22 (Cont.)

B

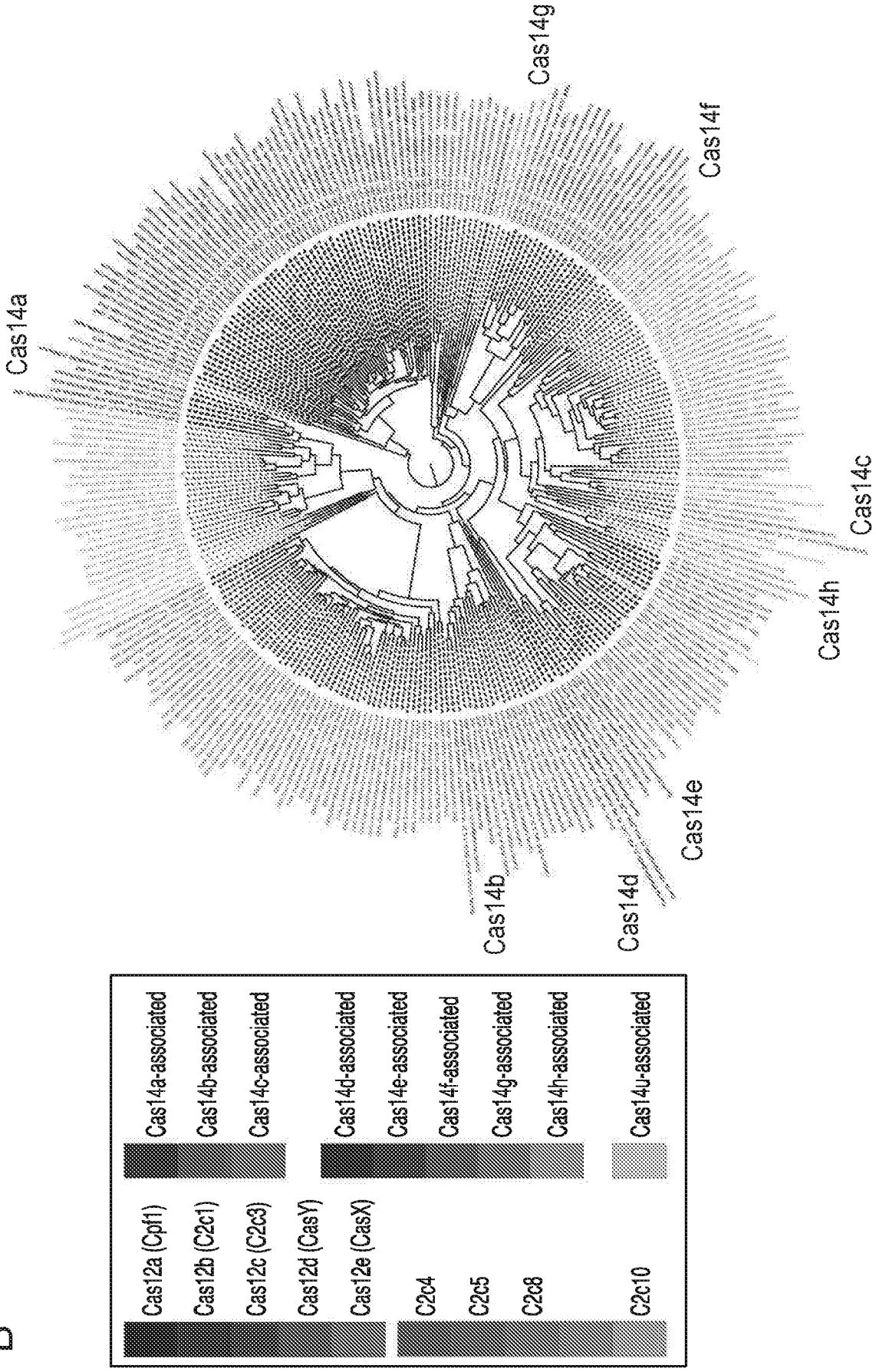

Cas14a
Cas14g
Cas14f
Cas14c
Cas14h
Cas14e
Cas14d
Cas14b
Cas14a-associated
Cas14b-associated
Cas14c-associated
Cas14d-associated
Cas14e-associated
Cas14f-associated
Cas14g-associated
Cas14h-associated
Cas14u-associated
Cas12a (Cpf1)
Cas12b (C2c1)
Cas12c (C2c3)
Cas12d (CasY)
Cas12e (CasX)
C2c4
C2c5
C2c8
C2c10

FIG. 23

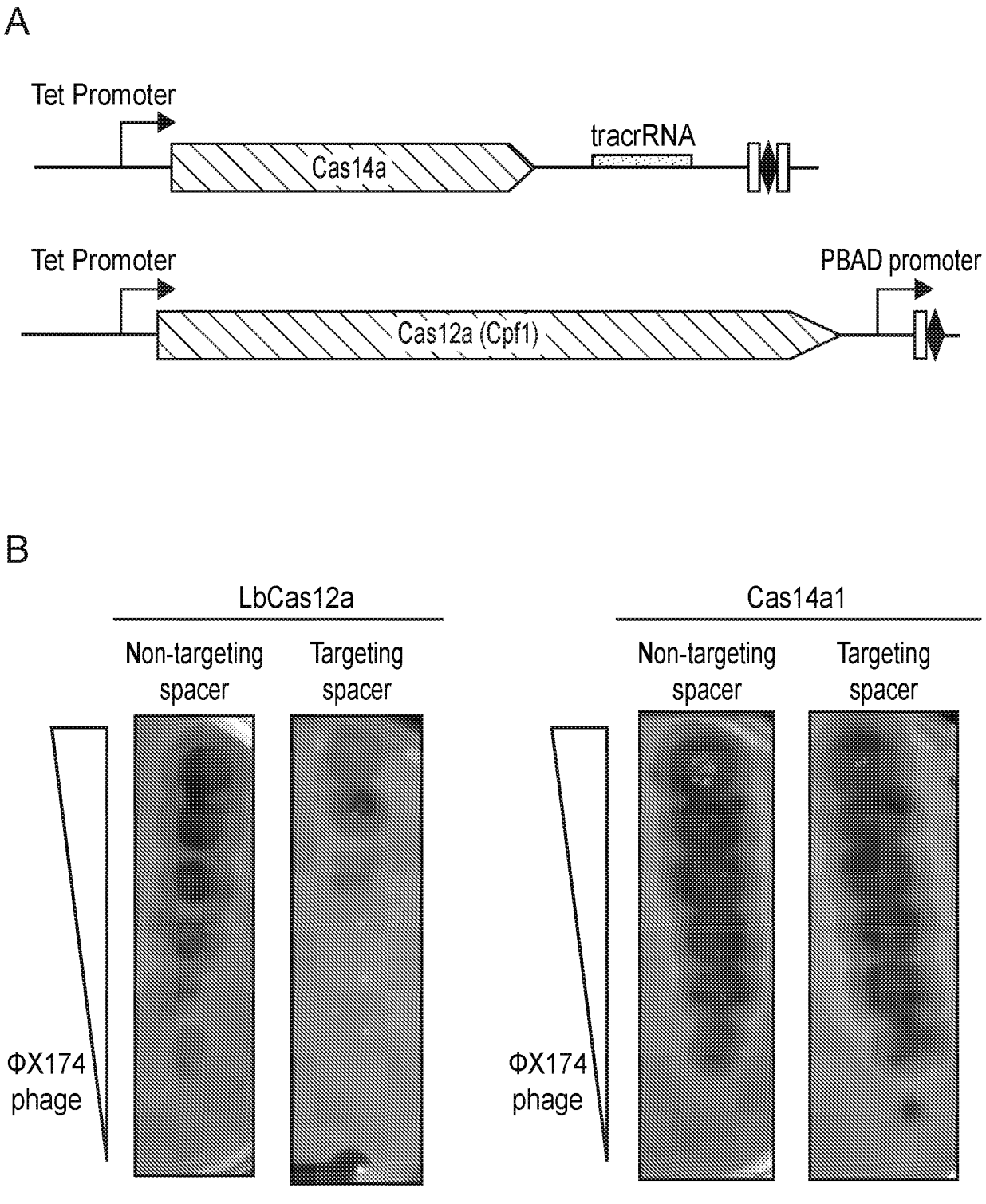

A
Tet Promoter
Cas14a
tracrRNA
Tet Promoter
PBAD promoter
Cas12a (Cpf1)
B
LbCas12a
Non-targeting spacer
Targeting spacer
ΦX174 phage
Cas14a1
Non-targeting spacer
Targeting spacer
ΦX174 phage

FIG. 23 (Cont.)

C

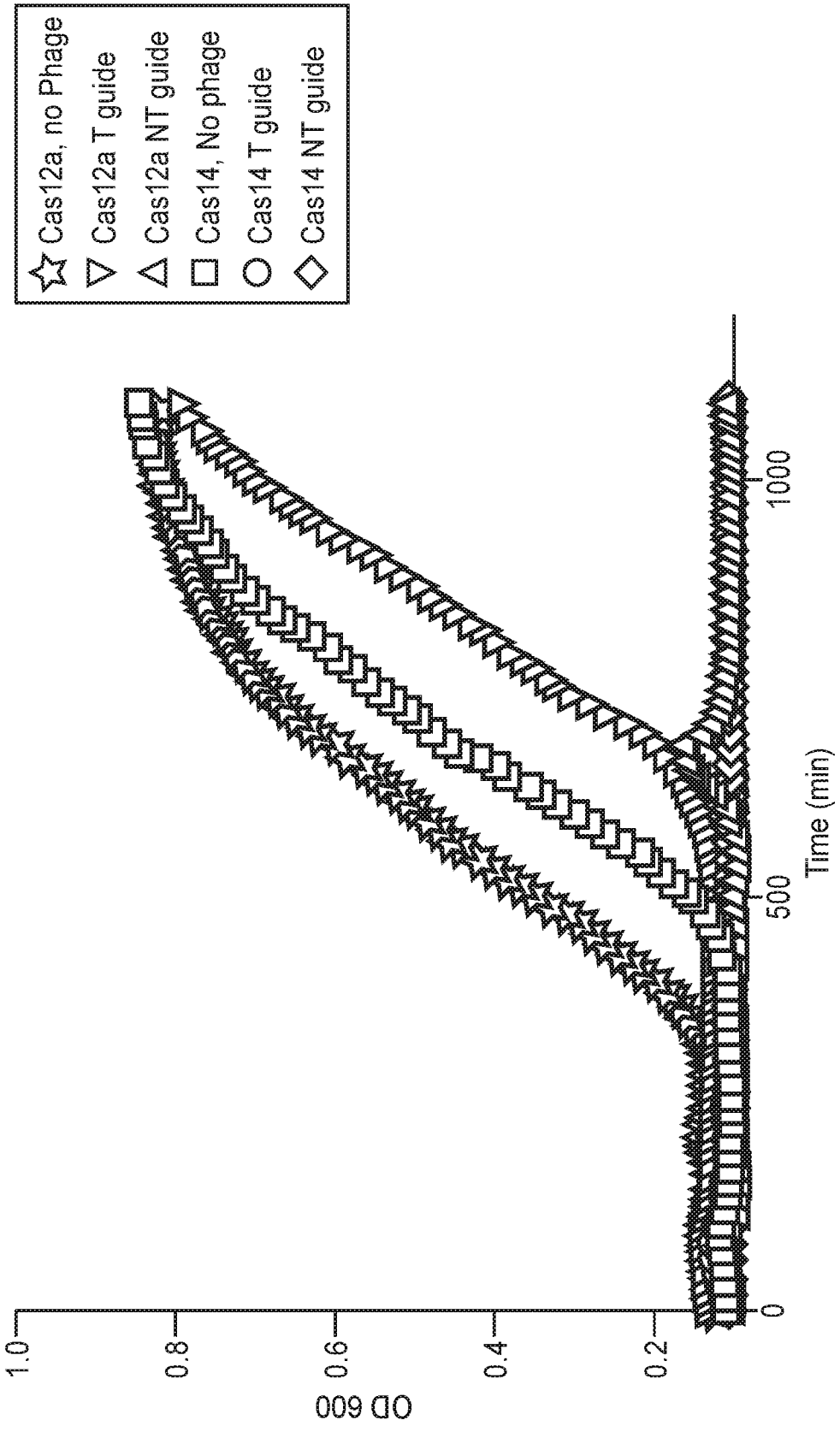

Cas12a, no Phage
Cas12a T guide
Cas12a NT guide
Cas14, No phage
Cas14 T guide
Cas14 NT guide
OD 600
1.0
0.8
0.6
0.4
0.2
0
500
1000
Time (min)

FIG. 24

Plasmid pLBH533_MBP-Cas14b2 expression (Addgene Plasmid Number 112506)

GCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCTCGATCC
CGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGTGCCGGCTCCGGAGAGCTC
TTTAATTAAGCGGCCGCCCTGCAGGACTCGAGTTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATA
TACATATGAAATCTTCTCACCATCACCATCACCATCACCATCACCATGGTTCTTCTATGAAAATCGAAGA
AGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTC
GAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAGGTTG
CGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGCTACGCTCAATCTGG
CCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACCTGGGATGCCGTA
CGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTTATAACAAAGATC
TGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAA
GAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTAT
GCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAGCGG
GTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTACTCCATCGCAGA
AGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCCAACATCGACACC
AGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAACCGTTCGTTGGCG
TGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCT
GCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTAC
GAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGGAAAACGCCCAGAAAGGTGAAATCATGC
CGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCCGCCAGCGGTCG
TCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAACAACAATAACAAT
AACAACAACCTCGGGATCGAGGAAAACCTGTACTTCCAATCCAATGCAATGGAAGAAAGCATTATTACCG
GTGTGAAATTCAAACTGCGCATCGATAAAGAAACCACCAAAAAACTGAACGAGTACTTCGATGAATATGG
CAAAGCAATTAACTTCGCCGTGAAGATCATTCAGAAAGAACTGGCAGATGATCGTTTTGCAGGTAAAGCA
AAACTGGACCAGAATAAAAACCCGATCCTGGATGAAAACGGCAAAAAAATCTATGAATTCCCGGATGAAT
TTTGCAGCTGTGGTAAACAGGTTAACAAGTACGTTAACAACAAACCGTTTTGCCAAGAGTGCTATAAAAT
CCGCTTTACCGAAAATGGTATTCGCAAACGTATGTATAGCGCCAAAGGTCGTAAAGCCGAACATAAAATC
AATATCCTGAACAGCACCAACAAGATCAGCAAAACCCATTTTAACTATGCCATTCGCGAAGCCTTCATTC
TGGATAAAAGCATCAAAAAGCAGCGCAAAAAACGTAATGAACGTCTGCGTGAAAGTAAAAAACGTCTGCA
GCAGTTTATCGATATGCGTGATGGTAAACGTGAAATTTGCCCGACCATTAAAGGTCAGAAAGTGGATCGT
TTTATTCATCCGAGCTGGATCACCAAAGATAAAAAGCTGGAAGATTTTCGCGGTTATACCCTGAGCATTA

SUBSTITUTE SHEET (RULE 26)

FIG. 24 (Cont.)

TCAACAGCAAAATTAAGATTCTGGATCGCAACATCAAACGCGAAGAAAAAAGCCTGAAAGAAAAAGGCCA
GATCATCTTTAAAGCCAAACGTCTGATGCTGGATAAATCCATTCGTTTTGTTGGTGATCGCAAAGTGCTG
TTTACAATTAGTAAAACCCTGCCGAAAGAGTATGAACTGGATCTGCCGAGCAAAGAAAAACGGCTGAATT
GGCTGAAAGAGAAGATCGAGATTATCAAGAACCAGAAACCGAAATATGCCTATCTGCTGCGCAAAAACAT
TGAGAGCGAAAAAAAACCGAACTATGAGTACTATCTGCAGTACACCCTGGAAATTAAACCGGAACTGAAA
GATTTTTATGATGGTGCCATTGGTATTGACCGTGGCATTAATCATATTGCCGTTTGCACCTTTATTAGCA
ACGATGGTAAAGTTACCCCTCCGAAATTTTTCAGCAGCGGTGAAATTCTGCGTCTGAAAAATCTGCAGAA
AGAGCGTGATCGCTTTCTGCTGCGTAAACACAACAAAAATCGCAAAAAAGGCAACATGCGCGTGATCGAA
AACAAAATCAATCTGATCCTGCACCGTTATAGCAAGCAGATTGTTGATATGGCCAAAAAGCTGAATGCCA
GCATTGTTTTTGAAGAACTGGGTCGTATTGGTAAAAGCCGCACCAAAATGAAAAAAAGCCAGCGTTATAA
ACTGAGCCTGTTCATCTTCAAGAAACTGAGCGATCTGGTTGATTACAAAAGCCGTCGTGAAGGTATTCGT
GTTACCTATGTTCCGCCTGAATATACCAGCAAAGAATGTAGCCATTGCGGTGAAAAAGTTAATACCCAGC
GTCCGTTTAATGGCAACTATAGCCTGTTTAAATGCAACAAATGTGGCATCCAGCTGAACAGCGATTATAA
TGCAAGCATCAACATTGCGAAAAAGGGCCTGAAAATTCCGAATAGCACCTAATAACATTGGAAGTGGATA
ACGGATCCGCGATCGCGGCGCGCCACCTGGTGGCCGGCCGGTACCACGCGTGCGCGCTGATCCGGCTGCT
AACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGG
CCTCTAAACGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCACAGGACGGGTG
TGGTCGCCATGATCGCGTAGTCGATAGTGGCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAA
GCGGTCGGACAGTGCTCCGAGAACGGGTGCGCATAGAAATTGCATCAACGCATATAGCGCTAGCAGCACG
CCATAGTGACTGGCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCA
AGCCTATGCCTACAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCATACA
CGGTGCCTGACTGCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGT
CAAACATGAGAATTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATA
ATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTT
TCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAACATTGAAA
AAGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCT
GTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTT
ACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGAT
GAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGT
CGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATG
GCATGACAGTAAGAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCT
GACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTT

SUBSTITUTE SHEET (RULE 26)

FIG. 24 (Cont.)

GATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAA
TGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGA
CTGGATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCT
GATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCT
CCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGA
GATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGAT
TTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCC
CTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC
TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCG
GATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCC
TTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCT
AATCCTGTTACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAG
TTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGA
CCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGC
GGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCC
TGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAG
GGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTT
TGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCT
GATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGA
TGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCAATGGTGCACTCTCAGTACAATCTGC
TCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCG
ACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCT
GTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGC
GGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTT
GAGTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGT
TTGGTCACTGATGCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGA
GGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACT
GGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGAT
GTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTG
ACTTCCGCGTTTCCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGA
CGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAA
CCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGC

SUBSTITUTE SHEET (RULE 26)

FIG. 24 (Cont.)

TGCCCGAGATGCGCCCGCGTGCTGCTGGAGATGGCGCGGACGCGATGGATATGTTCTGCCAAGGGTTGGT
TTGCGCATTCACAGTTCTCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCGAGG
TGCCCGCCGGCTTCCATTCAGGTCGAGGTGGCCCGGCTCCATGCACCGCGACGCAACGCGGGGAGGCAGAC
AAGGTATAGGGCGGCGCCTACAATCCATGCCAACCCGTTCCATGTGCTCGTCGCCGAGGCGGCATAAATCGCC
GTGACGATCAGCGGTCCAATGATCGAAGTTAGGCTGGTAAGAGCCGCGAGCGATCCTTGAAGCTGTCCCT
GATGGTCGTCATCTACCTGCCTGGACAGCATGGCCTGCAACGCGGGCATCCCGATGCCGCCGGAAGCGAG
AAGAATCATAATGGGGAAGGCCATCCAGCCTCGCGTCGCGAACGCCAGCAAGACGTAGCCCAGCGCGTCG
GCCGCCATGCCCGGCGATAATGGCCTGCTTCTCGCCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTT
GAGCGAGGGCGTGCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCG
GTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTC
ATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGG
GCATCGGTCGACGCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTT
GAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCT
GCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGA
TGTCGGCGATATAGGCGCCA (SEQ ID NO: 493)

FIG. 24 (Cont.)

Plasmid pLBH547_Tet-Cas14b2Locus (AddGene Plasmid Number: 112504)

ATCGTTGATAGAGTTATTTTACCACTCCCTATCAGTGATAGAGAAAAGAATTCAAAAGATCTAAAGAGGA
GAAAGGATCTATGGAAGAAAGCATTATTACCGGTGTGAAATTCAAACTGCGCATCGATAAAGAAACCACC
AAAAAACTGAACGAGTACTTCGATGAATATGGCAAAGCAATTAACTTCGCCGTGAAGATCATTCAGAAAG
AACTGGCAGATGATCGTTTTGCAGGTAAAGCAAAACTGGACCAGAATAAAAACCCGATCCTGGATGAAAA
CGGCAAAAAAATCTATGAATTCCCGGATGAATTTTGCAGCTGTGGTAAACAGGTTAACAAGTACGTTAAC
AACAAACCGTTTTGCCAAGAGTGCTATAAAATCCGCTTTACCGAAAATGGTATTCGCAAACGTATGTATA
GCGCCAAAGGTCGTAAAGCCGAACATAAAATCAATATCCTGAACAGCACCAACAAGATCAGCAAAACCCA
TTTTAACTATGCCATTCGCGAAGCCTTCATTCTGGATAAAAGCATCAAAAAGCAGCGCAAAAAACGTAAT
GAACGTCTGCGTGAAAGTAAAAAACGTCTGCAGCAGTTTATCGATATGCGTGATGGTAAACGTGAAATTT
GCCCGACCATTAAAGGTCAGAAAGTGGATCGTTTTATTCATCCGAGCTGGATCACCAAAGATAAAAAGCT
GGAAGATTTTCGCGGTTATACCCTGAGCATTATCAACAGCAAAATTAAGATTCTGGATCGCAACATCAAA
CGCGAAGAAAAAAGCCTGAAAGAAAAAGGCCAGATCATCTTTAAAGCCAAACGTCTGATGCTGGATAAAT
CCATTCGTTTTGTTGGTGATCGCAAAGTGCTGTTTACAATTAGTAAAACCCTGCCGAAAGAGTATGAACT
GGATCTGCCGAGCAAAGAAAAACGGCTGAATTGGCTGAAAGAGAAGATCGAGATTATCAAGAACCAGAAA
CCGAAATATGCCTATCTGCTGCGCAAAAACATTGAGAGCGAAAAAAAACCGAACTATGAGTACTATCTGC
AGTACACCCTGGAAATTAAACCGGAACTGAAAGATTTTTATGATGGTGCCATTGGTATTGACCGTGGCAT
TAATCATATTGCCGTTTGCACCTTTATTAGCAACGATGGTAAAGTTACCCCTCCGAAATTTTTCAGCAGC
GGTGAAATTCTGCGTCTGAAAAATCTGCAGAAAGAGCGTGATCGCTTTCTGCTGCGTAAACACAACAAAA
ATCGCAAAAAAGGCAACATGCGCGTGATCGAAAACAAAATCAATCTGATCCTGCACCGTTATAGCAAGCA
GATTGTTGATATGGCCAAAAAGCTGAATGCCAGCATTGTTTTTGAAGAACTGGGTCGTATTGGTAAAAGC
CGCACCAAAATGAAAAAAAGCCAGCGTTATAAACTGAGCCTGTTCATCTTCAAGAAACTGAGCGATCTGG
TTGATTACAAAAGCCGTCGTGAAGGTATTCGTGTTACCTATGTTCCGCCTGAATATACCAGCAAAGAATG
TAGCCATTGCGGTGAAAAAGTTAATACCCAGCGTCCGTTTAATGGCAACTATAGCCTGTTTAAATGCAAC
AAATGTGGCATCCAGCTGAACAGCGATTATAATGCAAGCATCAACATTGCGAAAAAGGGCCTGAAAATTC
CGAATAGCACCTAATAATGTTGGTTAAGCCACAATATGGAATATTGTTCTTATGGACAGTATTGACTTAA
ATTAATAATCTTCGAAGGCTATATGCGGAAGATTTGGCGTTGTTGTAACGCAATAAGGGGTAACCCTGAA
AAGGTTTGAAATCATATAAACCTAGTTTTATTTGAGTTTAGGCTCAGATAAAATGAACAGACCAATCTTT
AATTCCGTTCTGATTTAAAAAATCAGAATCTCTTTATAAATAGTATTACAAAAAGTGTACATTCCAAAAT

FIG. 24 (Cont.)

CCGAAAGCAGAATTGACCTTTTTAAGCCTAAAAAAGCCAAATTTCAAGGCTCTTTCATACTCAGAACAAA
GGGATTAAGGAATGCAACTACCTTACACCGCTTGCGAAAGTTGTCAGAAGATAATCTTTCATACTCAGAA
CAAAGGGATTAAGGAATGCAACTATCTTATCCATTTCTTGACATCAAATTTTCTTGCAGCATCTGAATTG
CTTAATTGCTTTCCTTGCTTCAGCAGGAAATAGCCAAGATTTTCCAGTTCTGCTGGCGTTATTGCAAGCA
CGGGGATTTTGTGCTTGCTGTAGATTTTGTTCTTCAGGACCTTCTTTTTCCATGCTAGAGTCACACTGGC
TCACCTTCGGGTGGGCCTTTCTGCGTTTATACCTAGGGATATATTCCGCTTCCTCGCTCACTGACTCGCT
ACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCC
AGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCCTGA
CAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCG
TTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGT
TATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACT
GTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCC
GGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTC
ATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGG
TTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGC
AAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAATATTTCTAGAT
TTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTTACTAGTG
CTTGGATTCTCACCAATAAAAAACGCCCGGCGGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTG
AGGTCATTACTGGATCTATCAACAGGAGTCCAAGCGAGCTCGATATCAAATTACGCCCCGCCCTGCCACT
CATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAAC
CTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCG
AAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGA
AAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGA
ATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGC
TCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATAC
GAAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATT
TTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACT
GACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTT
TTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCT
TATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAGATATCGACG
TCTTAAGACCCACTTTCACATTTAAGTTGTTTTTCTAATCCGCATATGATCAATTCAAGGCCGAATAAGA
AGGCTGGCTCTGCACCTTGGTGATCAAATAATTCGATAGCTTGTCGTAATAATGGCGGCATACTATCAGT

SUBSTITUTE SHEET (RULE 26)

FIG. 24 (Cont.)

AGTAGGTGTTTCCCTTTCTTCTTTAGCGACTTGATGCTCTTGATCTTCCAATACGCAACCTAAAGTAAAA
TGCCCCCACAGCGCTGAGTGCATATAATGCATTCTCTAGTGAAAAAACCTTGTTGGCATAAAAAGGCTAATT
GATTTTCGAGAGTTTCATACTGTTTTTCTGTAGGCCGTGTACCTAAATGTACTTTTGCTCCATCGCGATG
ACTTAGTAAAGCACATCTAAAACTTTTAGCGTTATTACGTAAAAAATCTTGCCAGCTTTCCCCTTCTAAA
GGGCAAAAGTGAGTATGGTGCCTATCTAACATCTCAATGGCTAAGGCGTCGAGCAAAGCCCGCTTATTTT
TTACATGCCAATACAATGTAGGCTGCTCTACACCTAGCTTCTGGGCGAGTTTACGGGTTGTTAAACCTTC
GATTCCGACCTCATTAAGCAGCTCTAATGCGCTGTTAATCACTTTACTTTTATCTAATCTAGACATCATT
AATTCCTAATTTTTTGTTGACACTCT (SEQ ID NO: 494)

FIG. 24 (Cont.)

Plasmid pLBH559_Tet-HisCas14a1Locus (AddGene Plasmid Number: 112502)

TAATTCCTAATTTTTGTTGACACTCTATCGTTGATAGAGTTATTTTACCACTCCCTATCAGTGATAGAGA
AAAGAATTCAAAAGATCTAAAGAGGAAAGGATCTATGAAATCTCACCATCACCATCACCATGAAAACC
TGTACTTCCAATCCAATATTGGAAGTGGAATGGCCAAAAACACCATTACCAAAACACTGAAACTGCGTAT
TGTGCGTCCGTATAATAGCGCAGAAGTGGAAAAAATTGTTGCCGACGAAAAAACACAACCGCGAAAAAATC
GCACTGGAAAAGAACAAAGACAAAGTGAAAGAAGCCTGCAGCAAACATCTGAAAGTTGCAGCATATTGTA
CCACACAGGTTGAACGTAATGCATGCCTGTTTTGTAAAGCACGTAAACTGGATGACAAATTCTACCAAAA
ACTGCGTGGTCAGTTTCCGGATGCAGTTTTTTGGCAAGAAATCAGCGAAATTTTTCGCCAGCTGCAGAAA
CAGGCAGCAGAAATCTATAATCAGAGCCTGATCGAACTGTACTACGAGATTTTTATCAAAGGCAAAGGTA
TTGCAAATGCCAGCAGCGTTGAACATTATCTGAGTGATGTTTGTTATACCCGTGCAGCAGAACTGTTTAA
AAACGCAGCAATTGCAAGCGGTCTGCGTAGCAAAAATCAAAAGGCAATTTTCGTCTGAAAGAACTGAAAAC
ATGAAAAGTGGTCTGCCGACCACCAAAAGCGATAATTTTCCGATTCCGCTGGTTAAACAGAAAGGTGGTC
AGTATACCGGTTTTGAAATTAGCAATCATAATAGCGACTTCATCAAGATTCCGTTTGGTCGTTGGCA
GGTCAAAAAAGAGATTGATAAATATCGTCCGTGGGAGAAATTTGACTTTGAACAGGTTCAGAAAAGCCCG
AAACCGATTAGCCTGCTGCTGAGCACCCAGCGTCGTAAACGTAATAAGGGTTGGAGCAAAGATGAAGGCA
CCGAAGCCGAAATCAAAAAAGTTATGAATGGCGATTATCAGACCAGCTACATTGAAGTTAAACGTGGCAG
CAAAATCGGTGAAAAAAAGCGCATGGATGCTTGAATCTGAAGCATTGATGTTCCGAAAATTGATAAAGGTGTG
GATCCGAGCATTATTGGTGGTATTGATGTTGGTGTTAAATCACCGCTGGTTTGCGCAATTAACAATGCAT
TTAGCCGTTATAGCATCAGCGATAACGACCTGTTTCACTTCAACAAGAAAATGTTTGCACGTCGTCGTAT
CCTGCTGAAAAAAAACCGTCATAAACGTGCAGGTCATGGTGCAAAAAACACAAACTGAAACCGATCACCATT
CTGACCGAAAAAAGTGAACGTTTTCGCAAAAAGCTGATTGAACGTTGGGCATGTGAAATCGCGGATTTCT
TCATTAAAAACAAAGTTGGCACCGTGCAGATGGAAAATCTGGAAAGCATGAAACGTAAAGAGGACAGCTA
TTTTAACATTCGCCTGCGTGGCTTTTGGCCGTATGCAGAAATGCAGAACAAAATCGAATTCAAACTGAAG
CAGTATGGCATCGAAATTCGTAAAGTTGCACCGAATAATACCAGCAAAACCTGTAGCAAATGTGGCCATC
TGAACAACTATTTCAACTTCGAGTACCGCAAGAAAAACACAAATTCCCGCACTTTAAATGCGAAAAATGCAA
CTTCAAAGAAAACGCCGATTATAATGCAGCCCTGAATATTTCAAACCCGAAACTGAAAAGCACCAAAGAG
GAACCGTAAATATTTATACTTTATCCTTCATTGACAAAAATGAGAATGTTATCCCAGATAACATTTG
ATGTACACAGATTCACACTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAAGCTGTCCCTTAGGGATT
AGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCC
TCGAAACAAATTCATTTTTCCTCTCCAATTCTGCACAAAAAAGGTGAGTCCTTATAAACCGGCGTGCAG

FIG. 24 (Cont.)

AACGCCGGCTCACCTTTTTCTTCATTCGATTTTATGCTTAAAAGCCGTAAAAACGCGGAATTCGGCGCC
GTTGCAGAACCCGAATAGACGGAATGCAACTACCTTACACCGCTTGCGAAAGTTGTCAGAAGA
TAATCTTGCAGAACCCGAATAGACGGAATGCAATCTTGACAGAGCCCGATTGCGTTATCTCCA
GGAGAAACATATAAAAGCATCAACCGCTGATCGGACTAGAGTCACACTGGCTCACCTTCGGGTGGGCCTT
TCTGCGTTTATACCTAGGGATATATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTG
CGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTAACAGGG
AAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGA
CGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGCGGCTCCC
TCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCA
TTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAACCCCCCGT
TCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCA
CCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAA
ACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTC
AGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACC
AAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTCAGTGCAATTTATCTCT
TCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTTACTAGTGCTTGGATTCTCACCAATAA
AAAACGCCCGGCGGCAACCGAGCGTTCTGAACAAATCCAGATGGAGTTCTGAGGTCATTACTGGATCTAT
CAACAGGAGTCCAAGCGAGCTCGATATCAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAA
TTCATTAAGCATTCTGCCGACATGGAAGCCATCACAAACGGCATGATGAACCTGAATCGCCAGCGGCATC
AGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTCCATATTGG
CCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAA
CCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGC
CGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAAC
AAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACGAAATTCCGGATGAGCATT
CATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAA
AAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATT
GTTCTTTACCGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTC
CTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAG
TTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAGATATCGACGTCTTAAGACCCACTTTCAC
ATTTAAGTTGTTTTTCTAATCCGCATATGATCAATTCAAGGCCGAATAAGAAGGCTGGCTCTGCACCTTG
GTGATCAAATAATTCGATAGCTTGTCGTAATAATGGCGGCATACTATCAGTAGTAGGTGTTTCCCTTTCT
TCTTTAGCGACTTGATGCTCTTGATCTTCCAATACGCAACCTAAAGTAAAATGCCCCACAGCGCTGAGTG

FIG. 24 (Cont.)

CATATAATGCATTCTCTAGTGAAAAACCTTGTTGGCATAAAAAGGCTAATTGATTTTCGAGAGTTTCATA
CTGTTTTTCTGTAGGCCGTGTACCTAAATGTACTTTTGATCCATCGCGATGACTTAGTAAAGCACATCTA
AAACTTTTAGCGTTATTACGTAAAAAATCTTGCCAGCTTTCCCCTTCTAAAGGGCAAAAGTGAGTATGGT
GCCTATCTAACATCTCAATGGCTAAGGCGTCGAGCAAAGCCCGCTTATTTTTTACATGCCAATACAATGT
AGGCTGCTCTACACCTAGCTTCTGGGCGAGTTTACGGGTTGTTAAACCTTCGATTCCGACCTCATTAAGC
AGCTCTAATGCGCTGTTAATCACTTTACTTTTATCTAATCTAGACATCAT (SEQ ID NO: 495)

SUBSTITUTE SHEET (RULE 26)

FIG. 24 (Cont.)

Plasmid pLBH545_Tet-Cas14a1_Locus (Addgene Plasmid Number: 112501)

TTACTTTTATCTAATCTAGACATCATTAATTCCTAATTTTTGTTGACACTCTATCGTTGATAGAGTTATT
TTACCACTCCCTATCAGTGATAGAGAAAAGAATTCAAAAGATCTAAAGAGGAGAAAGGATCTATGGCCAA
AAACACCATTACCAAAACACTGAAACTGCGTATTGTGCGTCCGTATAATAGCGCAGAAGTGGAAAAAATT
GTTGCCGACGAAAAAAACAACCGCGAAAAAATCGCACTGGAAAAGAACAAAGACAAAGTGAAAGAAGCCT
GCAGCAAACATCTGAAAGTTGCAGCATATTGTACCACACAGGTTGAACGTAATGCATGCCTGTTTTGTAA
AGCACGTAAACTGGATGACAAATTCTACCAAAAACTGCGTGGTCAGTTTCCGGATGCAGTTTTTTGGCAA
GAAATCAGCGAAATTTTTCGCCAGCTGCAGAAACAGGCAGCAGAAATCTATAATCAGAGCCTGATCGAAC
TGTACTACGAGATTTTTATCAAAGGCAAAGGTATTGCAAATGCCAGCAGCGTTGAACATTATCTGAGTGA
TGTTTGTTATACCCGTGCAGCAGAACTGTTTAAAAACGCAGCAATTGCAAGCGGTCTGCGTAGCAAAATC
AAAAGCAATTTTCGTCTGAAAGAACTGAAAAACATGAAAAGTGGTCTGCCGACCACCAAAAGCGATAATT
TTCCGATTCCGCTGGTTAAACAGAAAGGTGGTCAGTATACCGGTTTTGAAATTAGCAATCATAATAGCGA
CTTCATCATCAAGATTCCGTTTGGTCGTTGGCAGGTCAAAAAAGAGATTGATAAATATCGTCCGTGGGAG
AAATTTGACTTTGAACAGGTTCAGAAAAGCCCGAAACCGATTAGCCTGCTGCTGAGCACCCAGCGTCGTA
AACGTAATAAAGGTTGGAGCAAAGATGAAGGCACCGAAGCCGAAATCAAAAAAGTTATGAATGGCGATTA
TCAGACCAGCTACATTGAAGTTAAACGTGGCAGCAAAATCGGTGAAAAAAGCGCATGGATGCTGAATCTG
AGCATTGATGTTCCGAAAATTGATAAAGGTGTGGATCCGAGCATTATTGGTGGTATTGATGTTGGTGTTA
AATCACCGCTGGTTTGCGCAATTAACAATGCATTTAGCCGTTATAGCATCAGCGATAACGACCTGTTTCA
CTTCAACAAGAAAATGTTTGCACGTCGTCGTATCCTGCTGAAAAAAAACCGTCATAAACGTGCAGGTCAT
GGTGCAAAAAACAAACTGAAACCGATCACCATTCTGACCGAAAAAAGTGAACGTTTTCGCAAAAAGCTGA
TTGAACGTTGGGCATGTGAAATCGCGGATTTCTTCATTAAAAACAAAGTTGGCACCGTGCAGATGGAAAA
TCTGGAAAGCATGAAACGTAAAGAGGACAGCTATTTTAACATTCGCCTGCGTGGCTTTTGGCCGTATGCA
GAAATGCAGAACAAAATCGAATTCAAACTGAAGCAGTATGGCATCGAAATTCGTAAAGTTGCACCGAATA
ATACCAGCAAAACCTGTAGCAAATGTGGCCATCTGAACAACTATTTCAACTTCGAGTACCGCAAGAAAAA
CAAATTCCCGCACTTTAAATGCGAAAAATGCAACTTCAAAGAAAACGCCGATTATAATGCAGCCCTGAAT
ATTTCAAACCCGAAACTGAAAAGCACCAAAGAGGAACCGTAAATATTTATACTTTATTATCCTTCATTGA
CAAAAATGAGAATGTTATCCCAGATAACATTTGATGTACACAGATTCACACTTCACTGATAAAGTGGAGA
ACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTA
ATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTTTTCCTCTCCAATTCTGCACA

SUBSTITUTE SHEET (RULE 26)

FIG. 24 (Cont.)

AAAAAAGGTGAGTCCTTATAAACCGGCGTGCAGAACGCCGGCTCACCTTTTTTCTTCATTCGATTTTATG
CTTAAAAGCCGTAAAAACGCGGAATTCGGCGCCGTTGCAGAACCCGAATAGACGAATGAAGGAATGCAAC
TACCTTACACCGCTTGCGAAAGTTGTCAGAAGATAATCTTGCAGAACCCGAATAGACGAATGAAGGAATG
CAATCTTGACAGAGCCCGATTGCGTTATCTCCAGGAGAAACATATAAAAGCATCAACCGCTGATCGGACT
AGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCTGCGTTTATACCTAGGGATATATTCCGCTTCCTCG
CTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATT
TCCTGGAAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGG
CTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTAT
AAAGATACCAGGCGTTTCCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGG
TGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCG
CTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGT
CTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAG
TTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCCTCCAA
GCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTT
TTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATA
AAATATTTCTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATA
AGTTGTTACTAGTGCTTGGATTCTCACCAATAAAAAACGCCCGGCGGCAACCGAGCGTTCTGAACAAATC
CAGATGGAGTTCTGAGGTCATTACTGGATCTATCAACAGGAGTCCAAGCGAGCTCGATATCAAATTACGC
CCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAA
ACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGG
TGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGG
ATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACAC
GCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAA
ACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTC
TTTCATTGCCATACGAAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAA
AACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGG
TACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGT
ATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACG
CCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTC
GCCAGATATCGACGTCTTAAGACCCACTTTCACATTTAAGTTGTTTTTCTAATCCGCATATGATCAATTC
AAGGCCGAATAAGAAGGCTGGCTCTGCACCTTGGTGATCAAATAATTCGATAGCTTGTCGTAATAATGGC
GGCATACTATCAGTAGTAGGTGTTTCCCTTTCTTCTTTAGCGACTTGATGCTCTTGATCTTCCAATACGC

SUBSTITUTE SHEET (RULE 26)

FIG. 24 (Cont.)

AACCTAAAGTAAAATGCCCCACAGCGCTGAGTGCATATAATGCATTCTCTAGTGAAAAACCTTGTTGGCA
TAAAAAGGCTAATTGATTTTCGAGAGTTTCATACTGTTTTTCTGTAGGCCGTGTACCTAAATGTACTTTT
GCTCCATCGCGATGACTTAGTAAAGCACATCTAAAACTTTTAGCGTTATTACGTAAAAAATCTTGCCAGC
TTTCCCCTTCTAAAGGGCAAAAGTGAGTATGGTGCCTATCTAACATCTCAATGGCTAAGGCGTCGAGCAA
AGCCCGCTTATTTTTTACATGCCAATACAATGTAGGCTGCTCTACACCTAGCTTCTGGGCGAGTTTACGG
GTTGTTAAACCTTCGATTCCGACCTCATTAAGCAGCTCTAATGCGCTGTTAATCACT (SEQ ID NO:496)

SUBSTITUTE SHEET (RULE 26)

FIG. 24 (Cont.)

Plasmid pLBH531_MBP-Cas14a1 expression (Addgene Plasmid Number: 112500)

ACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCATCGGTGATGTCGGCG
ATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCG
AGATCTCGATCCCGCGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAGTGCCGGC
TCCGGAGAGCTCTTTAATTAAGCGGCCGCCCTGCAGGACTCGAGTTCTAGAAATAATTTTGTTTAACTTT
AAGAAGGAGATATACATATGAAATCTTCTCACCATCACCATCACCATCACCATCACCATGGTTCTTCTAT
GAAAATCGAAGAAGGTAAACTGGTAATCTGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTC
GGTAAGAAATTCGAGAAAGATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAAT
TCCCACAGGTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGCTA
CGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTGTATCCGTTTACC
TGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTGTTGAAGCGTTATCGCTGATTT
ATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGAAGAGATCCCGGCGCTGGATAAAGAACTGAA
AGCGAAAGGTAAGAGCGCGCTGATGTTCAACCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCT
GACGGGGGTTATGCGTTCAAGTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTG
GCGCGAAAGCGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTA
CTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTGGGCATGGTCC
AACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTTCAAGGGTCAACCATCCAAAC
CGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAGTCCGAACAAAGAGCTGGCAAAAGAGTTCCT
CGAAAACTATCTGCTGACTGATGAAGGTCTGGAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCG
CTGAAGTCTTACGAGGAAGAGTTGGCGAAAGATCCACGTATTGCCGCCACTATGGAAAACGCCCAGAAAG
GTGAAATCATGCCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGC
CGCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTCGAACAACAAC
AACAATAACAATAACAACAACCTCGGGATCGAGGAAAACCTGTACTTCCAATCCAATGCAATGGCCAAAA
ACACCATTACCAAAACACTGAAACTGCGTATTGTGCGTCCGTATAATAGCGCAGAAGTGGAAAAAATTGT
TGCCGACGAAAAAAACAACCGCGAAAAAATCGCACTGGAAAAGAACAAAGACAAAGTGAAAGAAGCCTGC
AGCAAACATCTGAAAGTTGCAGCATATTGTACCACACAGGTTGAACGTAATGCATGCCTGTTTTGTAAAG
CACGTAAACTGGATGACAAATTCTACCAAAAACTGCGTGGTCAGTTTCCGGATGCAGTTTTTTGGCAAGA
AATCAGCGAAATTTTTCGCCAGCTGCAGAAACAGGCAGCAGAAATCTATAATCAGAGCCTGATCGAACTG

SUBSTITUTE SHEET (RULE 26)

FIG. 24 (Cont.)

TACTACGAGATTTTTATCAAAGGCAAAGGTATTGCAAATGCCAGCAGCGTTGAACATTATCTGAGTGATG
TTTGTTATACCCGTGCAGCAGAACTGTTTAAAAACGCAGCAATTGCAAGCGGTCTGCGTAGCAAAATCAA
AAGCAATTTTCGTCTGAAAGAACTGAAAAACATGAAAAGTGGTCTGCCGACCACCAAAAGCGATAATTTT
CCGATTCCGCTGGTTAAACAGAAAGGTGGTCAGTATACCGGTTTTGAAATTAGCAATCATAATAGCGACT
TCATCATCAAGATTCCGTTTGGTCGTTGGCAGGTCAAAAAAGAGATTGATAAATATCGTCCGTGGGAGAA
ATTTGACTTTGAACAGGTTCAGAAAAGCCCGAAACCGATTAGCCTGCTGCTGAGCACCCAGCGTCGTAAA
CGTAATAAAGGTTGGAGCAAAGATGAAGGCACCGAAGCCGAAATCAAAAAAGTTATGAATGGCGATTATC
AGACCAGCTACATTGAAGTTAAACGTGGCAGCAAAATCTGTGAAAAAAGCGCATGGATGCTGAATCTGAG
CATTGATGTTCCGAAAATTGATAAAGGTGTGGATCCGAGCATTATTGGTGGTATTGATGTTGGTGTTAAA
TCACCGCTGGTTTGCGCAATTAACAATGCATTTAGCCGTTATAGCATCAGCGATAACGACCTGTTTCACT
TCAACAAGAAAATGTTTGCACGTCGTCGTATCCTGCTGAAAAAAAACCGTCATAAACGTGCAGGTCATGG
TGCAAAAAACAAACTGAAACCGATCACCATTCTGACCGAAAAAAGTGAACGTTTTCGCAAAAAGCTGATT
GAACGTTGGGCATGTGAAATCGCGGATTTCTTCATTAAAAACAAAGTTGGCACCGTGCAGATGGAAAATC
TGGAAAGCATGAAACGTAAAGAGGACAGCTATTTTAACATTCGCCTGCGTGGCTTTTGGCCGTATGCAGA
AATGCAGAACAAAATCGAATTCAAACTGAAGCAGTATGGCATCGAAATTCGTAAAGTTGCACCGAATAAT
ACCAGCAAAACCTGTAGCAAATGTGGCCATCTGAACAACTATTTCAACTTCGAGTACCGCAAGAAAAACA
AATTCCCGCACTTTAAATGCGAAAAATGCAACTTCAAAGAAAACGCCGATTATAATGCAGCCCTGAATAT
TTCAAACCCGAAACTGAAAAGCACCAAAGAGGAACCGTAATAACATTGGAAGTGGATAACGGATCCGCGA
TCGCGGCGCGCCACCTGGTGGCCGGCCGGTACCACGCGTGCGCGCTGATCCGGCTGCTAACAAAGCCCGA
AAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGG
TCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGATATCCACAGGACGGGTGTGGTCGCCATGA
TCGCGTAGTCGATAGTGGCTCCAAGTAGCGAAGCGAGCAGGACTGGGCGGCGGCCAAAGCGGTCGGACAG
TGCTCCGAGAACGGGTGCGCATAGAAATTGCATCAACGCATATAGCGCTAGCAGCACGCCATAGTGACTG
GCGATGCTGTCGGAATGGACGATATCCCGCAAGAGGCCCGGCAGTACCGGCATAACCAAGCCTATGCCTA
CAGCATCCAGGGTGACGGTGCCGAGGATGACGATGAGCGCATTGTTAGATTTCATACACGGTGCCTGACT
GCGTTAGCAATTTAACTGTGATAAACTACCGCATTAAAGCTTATCGATGATAAGCTGTCAAACATGAGAA
TTCTTGAAGACGAAAGGGCCTCGTGATACGCCTATTTTTATAGGTTAATGTCATGATAATAATGGTTTCT
TAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATT
CAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAACATTGAAAAAGGAAGAGTAT
GAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCAC
CCAGAAACGCTGGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGG

SUBSTITUTE SHEET (RULE 26)

FIG. 24 (Cont.)

ATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAA
AGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACAC
TATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAA
GAGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGG
AGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAA
CCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGCAGCAATGGCAACAACGT
TGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGC
GGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGA
GCCGGTGAGCGTGGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAG
TTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTC
ACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCAT
TTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGT
TTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG
CGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTA
CCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGC
CGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACC
AGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAG
GCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAAC
TGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCC
GGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTAT
AGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCC
TATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTT
CTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGC
CGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTC
TCCTTACGCATCTGTGCGGTATTTCACACCGCAATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCA
TAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAAC
ACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCC
GGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCTGCGGTAAAGCTCAT
CAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAGTTTCTCCAG
AAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGAT
GCCTCCGTGTAAGGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGA
TACGGGTTACTGATGATGAACATGCCCGGTTACTGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGAT

SUBSTITUTE SHEET (RULE 26)

FIG. 24 (Cont.)

GCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAATACAGATGTAGGTGTTCCA
CAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTT
CCAGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCA
GCAGTCGCTTCACGTTCGCTCGCGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCT
AGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGTGGCCAGGACCCAACGCTGCCCGAGATGC
GCCGCGTGCGGCTGCTGGAGATGGCGGACGCGATGGATATGTTCTGCCAAGGGTTGGTTTGCGCATTCAC
AGTTCTCCGCAAGAATTGATTGGCTCCAATTCTTGGAGTGGTGAATCCGTTAGCGAGGTGCCGCCGGCTT
CCATTCAGGTCGAGGTGGCCCGGCTCCATGCACCGCGACGCAACGCGGGGAGGCAGACAAGGTATAGGGC
GGCGCCTACAATCCATGCCAACCCGTTCCATGTGCTCGCCGAGGCGGCATAAATCGCCGTGACGATCAGC
GGTCCAATGATCGAAGTTAGGCTGGTAAGAGCCGCGAGCGATCCTTGAAGCTGTCCCTGATGGTCGTCAT
CTACCTGCCTGGACAGCATGGCCTGCAACGCGGGCATCCCGATGCCGCCGGAAGCGAGAAGAATCATAAT
GGGGAAGGCCATCCAGCCTCGCGTCGCGAACGCCAGCAAGACGTAGCCCAGCGCGTCGGCCGCCATGCCG
GCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGT
GCAAGATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAA
AATGACCCAGAGCGCTGCCGGCACCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCG
ACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGGTTGAAGGCTCTCAAGGGCATCGGTCGAC
GCTCTCCCTTATGCGACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGC
CGCAAGGAATGGTGCATGCAAGGAGATGGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCAT

(SEQ ID NO: 497)

SUBSTITUTE SHEET (RULE 26)

FIG. 25

A

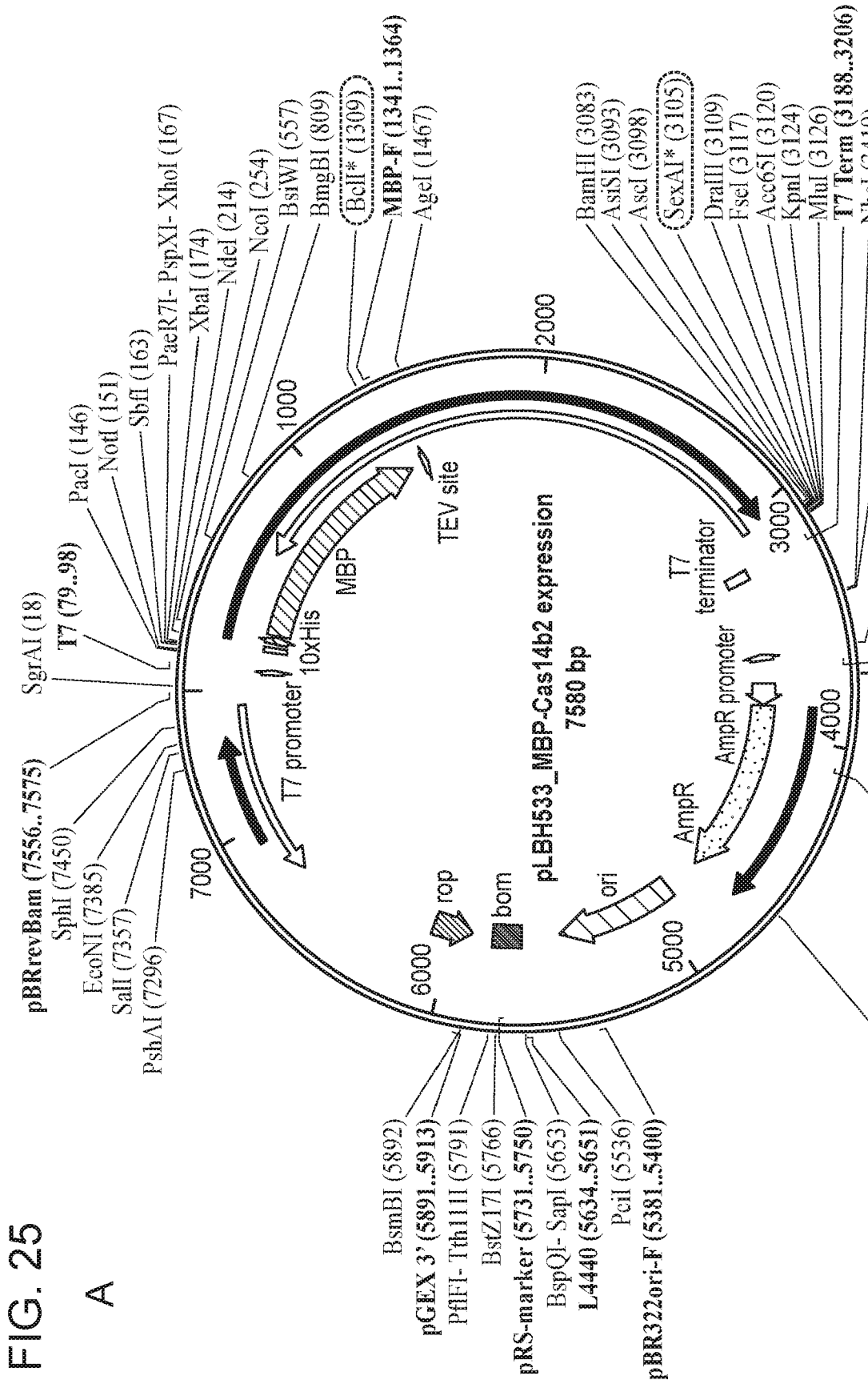

pLBH533_MBP-Cas14b2 expression
7580 bp
1000
2000
3000
4000
5000
6000
7000
T7 promoter
10xHis
MBP
TEV site
T7 terminator
AmpR promoter
AmpR
ori
bom
rop
pBRrevBam (7556..7575)
SphI (7450)
EcoNI (7385)
SalI (7357)
PshAI (7296)
SgrAI (18)
T7 (79..98)
PacI (146)
NotI (151)
SbfI (163)
PacR7I- PspXI- XhoI (167)
XbaI (174)
NdeI (214)
NcoI (254)
BsiWI (557)
BmgBI (809)
BclI* (1309)
MBP-F (1341..1364)
AgeI (1467)
BamHI (3083)
AsiSI (3093)
AscI (3098)
SexAI* (3105)
DraIII (3109)
FseI (3117)
Acc65I (3120)
KpnI (3124)
MluI (3126)
T7 Term (3188..3206)
NheI (3419)
BmtI (3423)
HindIII (3619)
pBRfoEco (3670..3688)
ZraI (3727)
AatII (3729)
Amp-R (4079..4098)
AhdI (4648)
BsmBI (5892)
pGEX 3' (5891..5913)
PflFI- Tth111I (5791)
BstZ17I (5766)
pRS-marker (5731..5750)
BspQI- SapI (5653)
L4440 (5634..5651)
PciI (5536)
pBR322ori-F (5381..5400)

FIG. 25 (Cont.)

B

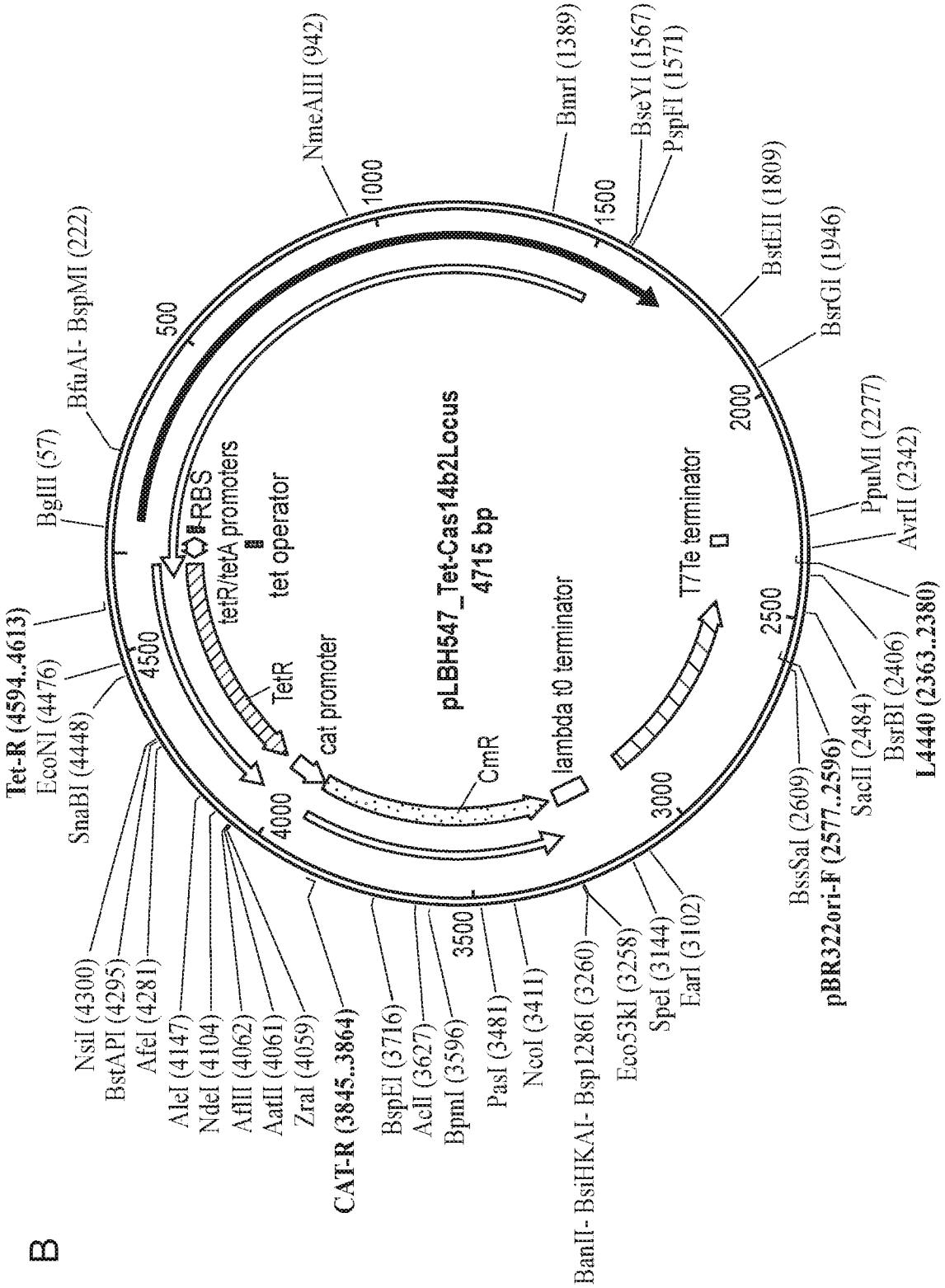

pLBH547_Tet-Cas14b2Locus
4715 bp
Tet-R (4594..4613)
EcoNI (4476)
SnaBI (4448)
NsiI (4300)
BstAPI (4295)
AfeI (4281)
AleI (4147)
NdeI (4104)
AflII (4062)
AatII (4061)
ZraI (4059)
CAT-R (3845..3864)
BspEI (3716)
AclI (3627)
BpmI (3596)
PasI (3481)
NcoI (3411)
BanII- BsiHKAI- Bsp1286I (3260)
Eco53kI (3258)
SpeI (3144)
EarI (3102)
BssSaI (2609)
pBR322ori-F (2577..2596)
SacII (2484)
BsrBI (2406)
L4440 (2363..2380)
PpuMI (2277)
AvrII (2342)
BsrGI (1946)
BstEII (1809)
PspFI (1571)
BseYI (1567)
BmrI (1389)
NmeAIII (942)
BfuAI- BspMI (222)
BglII (57)
RBS
tetR/tetA promoters
tet operator
TetR
cat promoter
CmR
lambda t0 terminator
T7Te terminator
500
1000
1500
2000
2500
3000
3500
4000
4500

FIG. 25 (Cont.)

C

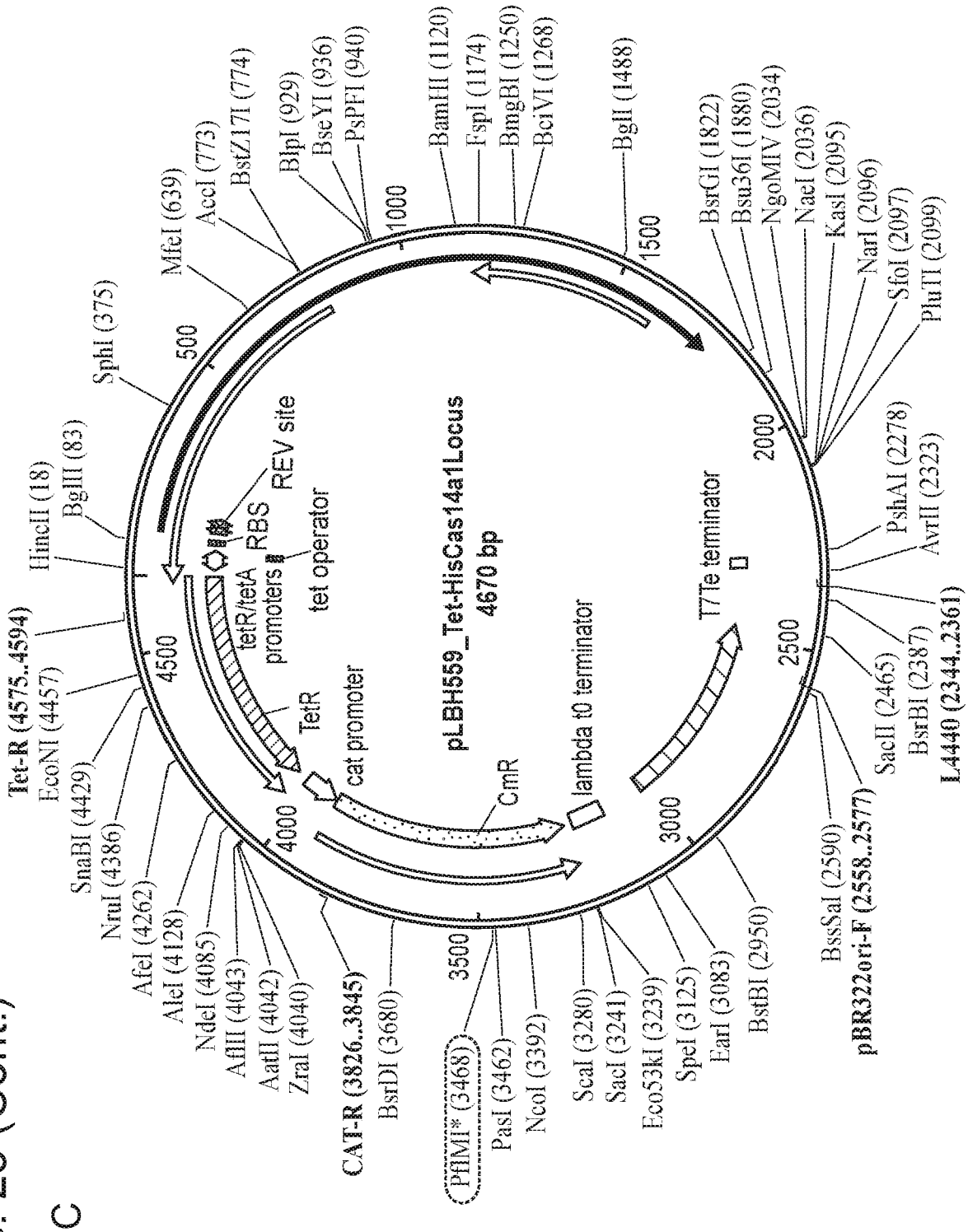

pLBH559_Tet-HisCas14a1Locus
4670 bp
Tet-R (4575..4594)
HincII (18)
BglII (83)
SphI (375)
MfeI (639)
AccI (773)
BstZ17I (774)
BlpI (929)
BseYI (936)
PspFI (940)
BamHI (1120)
FspI (1174)
BmgBI (1250)
BciVI (1268)
BglII (1488)
BsrGI (1822)
Bsu36I (1880)
NgoMIV (2034)
NaeI (2036)
KasI (2095)
NarI (2096)
SfoI (2097)
PluTI (2099)
PshAI (2278)
AvrII (2323)
L4440 (2344..2361)
SacII (2465)
BsrBI (2387)
pBR322ori-F (2558..2577)
BssSaI (2590)
BstBI (2950)
EarI (3083)
SpeI (3125)
Eco53kI (3239)
SacI (3241)
ScaI (3280)
NcoI (3392)
PasI (3462)
PflMI* (3468)
BsrDI (3680)
CAT-R (3826..3845)
ZraI (4040)
AatII (4042)
AflII (4043)
NdeI (4085)
AleI (4128)
AfeI (4262)
NruI (4386)
SnaBI (4429)
EcoNI (4457)
tetR/tetA promoters
RBS
REV site
tet operator
TetR
cat promoter
CmR
lambda t0 terminator
T7Te terminator
500
1000
1500
2000
2500
3000
3500
4000
4500

FIG. 25 (Cont.)

D

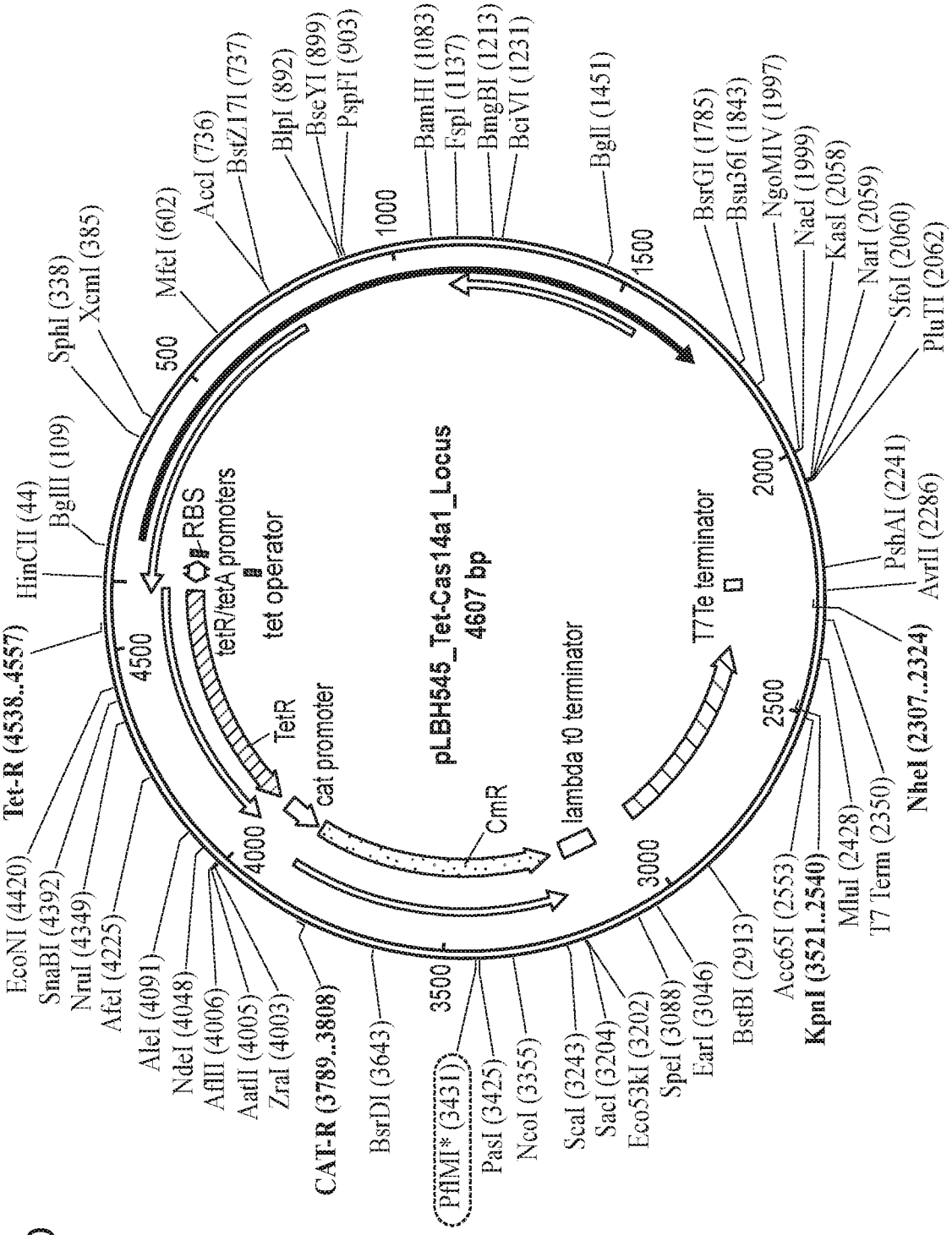
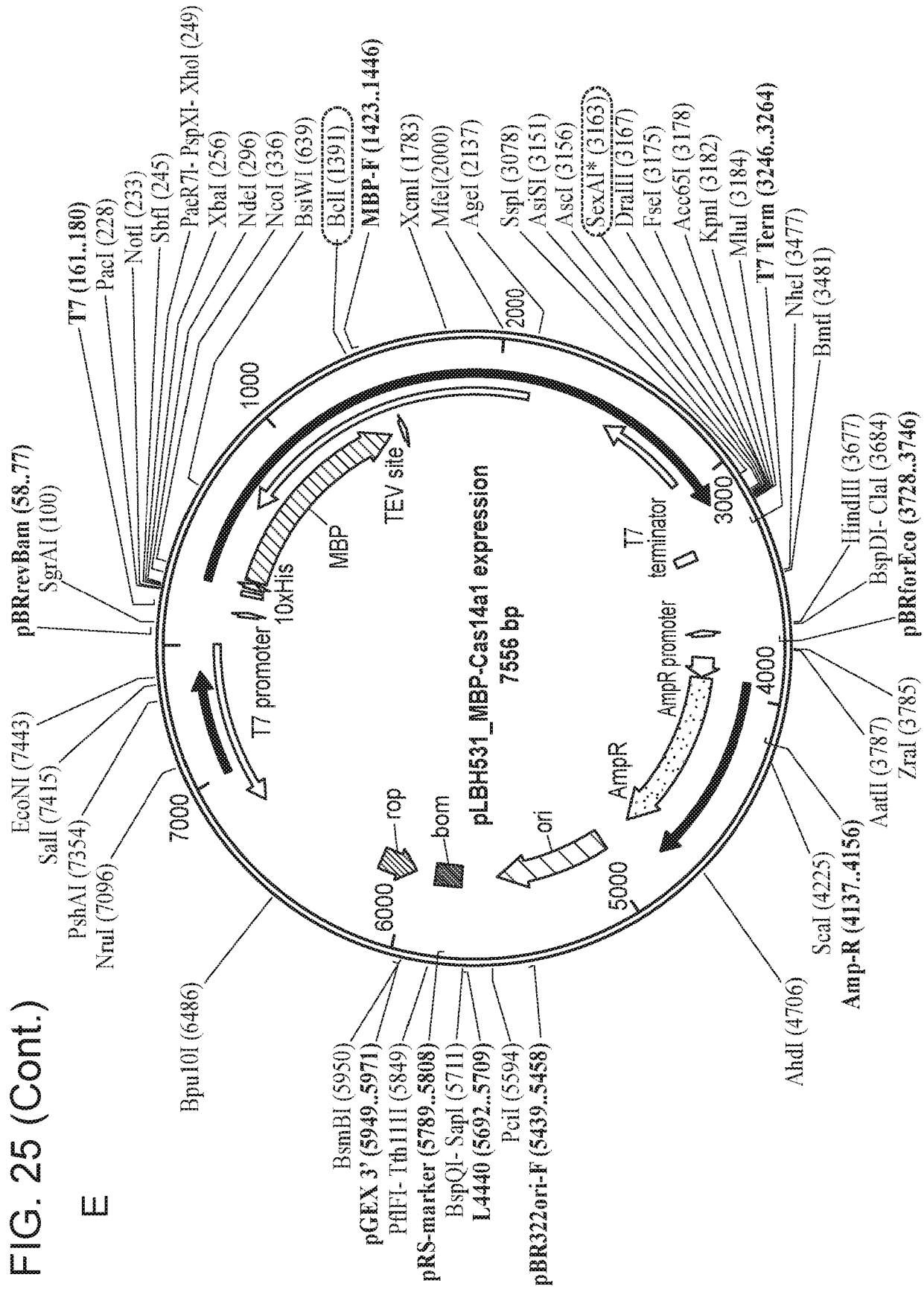

pLBH545_Tet-Cas14a1_Locus
4607 bp
Tet-R (4538..4557)
RBS
tetR/tetA promoters
tet operator
TetR
cat promoter
CmR
lambda t0 terminator
T7Te terminator
CAT-R (3789..3808)
KpnI (3521..2540)
NheI (2307..2324)
HinCII (44)
BglII (109)
SphI (338)
XcmI (385)
MfeI (602)
AccI (736)
BstZ17I (737)
BlpI (892)
BseYI (899)
PspFI (903)
BamHI (1083)
FspI (1137)
BmgBI (1213)
BciVI (1231)
BglI (1451)
BsrGI (1785)
Bsu36I (1843)
NgoMIV (1997)
NaeI (1999)
KasI (2058)
NarI (2059)
SfoI (2060)
PluTI (2062)
PshAI (2241)
AvrII (2286)
EcoNI (4420)
SnaBI (4392)
NruI (4349)
AfeI (4225)
AleI (4091)
NdeI (4048)
AflII (4006)
AatII (4005)
ZraI (4003)
BsrDI (3643)
PHMI* (3431)
PasI (3425)
NcoI (3355)
ScaI (3243)
SacI (3204)
Eco53kI (3202)
SpeI (3088)
EarI (3046)
BstBI (2913)
Acc65I (2553)
MluI (2428)
T7 Term (2350)
500
1000
1500
2000
2500
3000
3500
4000
4500

FIG. 25 (Cont.)

E


pLBH531_MBP-Cas14a1 expression
7556 bp
T7 promoter
10xHis
MBP
TEV site
T7 terminator
AmpR promoter
AmpR
ori
rop
bom
1000
2000
3000
4000
5000
6000
7000
pBRrevBam (58..77)
SgrAI (100)
T7 (161..180)
PacI (228)
NotI (233)
SbfI (245)
PaeR7I- PspXI- XhoI (249)
XbaI (256)
NdeI (296)
NcoI (336)
BsiWI (639)
BclI (1391)
MBP-F (1423..1446)
XcmI (1783)
MfeI(2000)
AgeI (2137)
SspI (3078)
AsiSI (3151)
AscI (3156)
SexAI* (3163)
DraIII (3167)
FseI (3175)
Acc65I (3178)
KpnI (3182)
MluI (3184)
T7 Term (3246..3264)
NheI (3477)
BmtI (3481)
HindIII (3677)
BspDI- ClaI (3684)
pBRforEco (3728..3746)
ZraI (3785)
AatII (3787)
Amp-R (4137..4156)
ScaI (4225)
AhdI (4706)
Bpu10I (6486)
BsmBI (5950)
pGEX 3' (5949..5971)
PflFI- Tth111I (5849)
pRS-marker (5789..5808)
BspQI- SapI (5711)
L4440 (5692..5709)
PciI (5594)
pBR322ori-F (5439..5458)
NruI (7096)
PshAI (7354)
SalI (7415)
EcoNI (7443)

CASZ COMPOSITIONS AND METHODS OF USE

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 62/580,395, filed November 1, 2017, which application is incorporated herein by reference in its entirety.

INCORPORATION BY REFERENCE OF SEQUENCE LISTING PROVIDED AS A TEXT FILE

[0002] A Sequence Listing is provided herewith as a text file, "BERK-374WO_SEQ_LISTING_ST25.txt" created on October 30, 2018 and having a size of 536 KB. The contents of the text file are incorporated by reference herein in their entirety.

INTRODUCTION

[0003] The CRISPR-Cas system, an example of a pathway that was unknown to science prior to the DNA sequencing era, is now understood to confer bacteria and archaea with acquired immunity against phage and viruses. Intensive research has uncovered the biochemistry of this system. CRISPR-Cas systems consist of Cas proteins, which are involved in acquisition, targeting and cleavage of foreign DNA or RNA, and a CRISPR array, which includes direct repeats flanking short spacer sequences that guide Cas proteins to their targets. Class 2 CRISPR-Cas are streamlined versions in which a single Cas protein bound to RNA is responsible for binding to and cleavage of a targeted sequence. The programmable nature of these minimal systems has facilitated their use as a versatile technology that is revolutionizing the field of genome manipulation.

SUMMARY

[0004] The present disclosure provides compositions and methods that include one or more of: (1) a "CasZ" protein (also referred to as a CasZ polypeptide), a nucleic acid encoding the CasZ protein, and/or a modified host cell comprising the CasZ protein (and/or a nucleic acid encoding the same); (2) a CasZ guide RNA that binds to and provides sequence specificity to the CasZ protein, a nucleic acid encoding the CasZ guide RNA, and/or a modified host cell comprising the CasZ guide RNA (and/or a nucleic acid encoding the same); and (3) a CasZ transactivating noncoding RNA (trancRNA) (referred to herein as a "CasZ trancRNA"), a nucleic acid encoding the CasZ trancRNA, and/or a modified host cell comprising the CasZ trancRNA (and/or a nucleic acid encoding the same).

BRIEF DESCRIPTION OF THE DRAWINGS

[0005] **Figure 1** depicts examples of naturally occurring CasZ protein sequences.

[0006] **Figure 2** depicts schematic representations of CasZ loci, which include a Cas1 protein in addition to the CasZ protein.

[0007] **Figure 3** depicts a phylogenetic tree of CasZ sequences in relation to other Class 2 CRISPR/Cas effector protein sequences.

[0008] **Figure 4** depicts a phylogenetic tree of Cas1 sequences from CasZ loci in relation to Cas1 sequences from other Class 2 CRISPR/Cas loci.

[0009] **Figure 5** depicts transcriptomic RNA mapping data demonstrating expression of trancRNA from CasZ loci. The trancRNAs are adjacent to the CasZ repeat array, but do not include the repeat sequence and are not complementary to the repeat sequence. Shown are RNA mapping data for the following loci: CasZa3, CasZb4, CasZc5, CasZd1, and CasZe3. Small repeating aligned arrows represent the repeats of the CRISPR array (indicating the presence of guide RNA-encoding sequence); The peaks outside and adjacent to the repeat arrays represent highly transcribed trancRNAs. Figure 5 (Cont. 1) nucleotide sequences (Top to Bottom: SEQ ID NOs.:312-331). Figure 5 (Cont. 3) nucleotide sequences (Top to Bottom: SEQ ID Nos.:161-177).

[0010] **Figure 6** depicts results for PAM preferences as assayed using PAM depletion assays for CasZc (top) and CasZb (bottom).

[0011] **Figure 7** depict the sequences of Cas14 proteins described herein.

[0012] **Figure 8, Panels A-D** depict the architecture and phylogeny of CRISPR-Cas14 genomic loci.

[0013] **Figure 9** depicts a phylogenetic analysis of Cas14 orthologs.

[0014] **Figure 10** depicts a maximum likelihood tree for Cas1 from known CRISPR systems.

[0015] **Figure 11, Panels A-B** depict the acquisition of new spacers by CRISPR-Cas14 systems.

[0016] **Figure 12, Panels A-D** depict that CRISPR-Cas14a actively adapts and encodes a tracrRNA.

[0017] **Figure 13, Panels A-B** depict metatranscriptomics for CRISPR-Cas14 loci.

[0018] **Figure 14, Panels A-B** depict RNA processing and heterologous expression by CRISPR-Cas14.

[0019] **Figure 15, Panels A-D** depict plasmid depletion by Cas14a1 and SpCas9.

[0020] **Figure 16, Panels A-D** depict CRISPR-Cas14a is an RNA-guided DNA-endonuclease.

[0021] **Figure 17, Panels A-E** depict degradation of ssDNA by Cas14a1.

[0022] **Figure 18** depicts kinetics of Cas14a1 cleavage of ssDNA with various guide RNA components.

[0023] **Figure 19, Panels A-F** depict optimization of Cas14a1 guide RNA components.

[0024] **Figure 20, Panels A-E** depict high fidelity ssDNA DNP detection by CRISPR-Cas14a. Figure 20, panel C provides nucleotide sequences (Top to Bottom: SEQ ID NOs:367-370)

[0025] **Figure 21, Panels A-F** depict the impact of various activators on Cas14a1 cleavage rate.

[0026] **Figure 22, Panels A-B** depict diversity of CRISPR-Cas14 systems.

[0027] **Figure 23, Panels A-C** depict a test of Cas14a1 mediated interference in a heterologous host. Diagram of Cas14a1 and LbCas12a constructs to test interference in *E. coli*.

[0028] **Figure 24** depicts Cas14 nucleotide sequences of plasmids used in the present invention.

[0029] **Figure 25, Panels A-E** depict a sequence map of each of the plasmids disclosed in Figure 24.

DEFINITIONS

[0030] "Heterologous," as used herein, means a nucleotide or polypeptide sequence that is not found in the native nucleic acid or protein, respectively. For example, relative to a CasZ polypeptide, a heterologous polypeptide comprises an amino acid sequence from a protein other than the CasZ polypeptide. In some cases, a portion of a CasZ protein from one species is fused to a portion of a CasZ protein from a different species. The CasZ sequence from each species could therefore be considered heterologous relative to one another. As another example, a CasZ protein (e.g., a dCasZ protein) can be fused to an active domain from a non-CasZ protein (e.g., a histone deacetylase), and the sequence of the active domain could be considered a heterologous polypeptide (it is heterologous to the CasZ protein).

[0031] The terms "polynucleotide" and "nucleic acid," used interchangeably herein, refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxynucleotides. Thus, this term includes, but is not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. The terms "polynucleotide" and "nucleic acid" should be understood to include, as applicable to the embodiment being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides.

[0032] The terms "polypeptide," "peptide," and "protein", are used interchangeably herein, refer to a polymeric form of amino acids of any length, which can include genetically coded and non-genetically coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones. The term includes fusion proteins, including, but not limited to, fusion proteins with a heterologous amino acid sequence, fusions with heterologous and

homologous leader sequences, with or without N-terminal methionine residues; immunologically tagged proteins; and the like.

[0033] The term "naturally-occurring" as used herein as applied to a nucleic acid, a protein, a cell, or an organism, refers to a nucleic acid, cell, protein, or organism that is found in nature.

[0034] As used herein the term "isolated" is meant to describe a polynucleotide, a polypeptide, or a cell that is in an environment different from that in which the polynucleotide, the polypeptide, or the cell naturally occurs. An isolated genetically modified host cell may be present in a mixed population of genetically modified host cells.

[0035] As used herein, the term "exogenous nucleic acid" refers to a nucleic acid that is not normally or naturally found in and/or produced by a given bacterium, organism, or cell in nature. As used herein, the term "endogenous nucleic acid" refers to a nucleic acid that is normally found in and/or produced by a given bacterium, organism, or cell in nature. An "endogenous nucleic acid" is also referred to as a "native nucleic acid" or a nucleic acid that is "native" to a given bacterium, organism, or cell.

[0036] "Recombinant," as used herein, means that a particular nucleic acid (DNA or RNA) is the product of various combinations of cloning, restriction, and/or ligation steps resulting in a construct having a structural coding or non-coding sequence distinguishable from endogenous nucleic acids found in natural systems. Generally, DNA sequences encoding the structural coding sequence can be assembled from cDNA fragments and short oligonucleotide linkers, or from a series of synthetic oligonucleotides, to provide a synthetic nucleic acid which is capable of being expressed from a recombinant transcriptional unit contained in a cell or in a cell-free transcription and translation system. Such sequences can be provided in the form of an open reading frame uninterrupted by internal non-translated sequences, or introns, which are typically present in eukaryotic genes. Genomic DNA comprising the relevant sequences can also be used in the formation of a recombinant gene or transcriptional unit. Sequences of non-translated DNA may be present 5' or 3' from the open reading frame, where such sequences do not interfere with manipulation or expression of the coding regions, and may indeed act to modulate production of a desired product by various mechanisms (see "DNA regulatory sequences", below).

[0037] Thus, e.g., the term "recombinant" polynucleotide or "recombinant" nucleic acid refers to one which is not naturally occurring, e.g., is made by the artificial combination of two otherwise separated segments of sequence through human intervention. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. Such is usually done to replace a codon with a redundant codon encoding the same or a conservative amino acid, while typically introducing or

removing a sequence recognition site. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques.

[0038] Similarly, the term "recombinant" polypeptide refers to a polypeptide which is not naturally occurring, e.g., is made by the artificial combination of two otherwise separated segments of amino sequence through human intervention. Thus, e.g., a polypeptide that comprises a heterologous amino acid sequence is recombinant.

[0039] By "construct" or "vector" is meant a recombinant nucleic acid, generally recombinant DNA, which has been generated for the purpose of the expression and/or propagation of a specific nucleotide sequence(s), or is to be used in the construction of other recombinant nucleotide sequences.

[0040] The terms "DNA regulatory sequences," "control elements," and "regulatory elements," used interchangeably herein, refer to transcriptional and translational control sequences, such as promoters, enhancers, polyadenylation signals, terminators, protein degradation signals, and the like, that provide for and/or regulate expression of a coding sequence and/or production of an encoded polypeptide in a host cell.

[0041] The term "transformation" is used interchangeably herein with "genetic modification" and refers to a permanent or transient genetic change induced in a cell following introduction of new nucleic acid (e.g., DNA exogenous to the cell) into the cell. Genetic change ("modification") can be accomplished either by incorporation of the new nucleic acid into the genome of the host cell, or by transient or stable maintenance of the new nucleic acid as an episomal element. Where the cell is a eukaryotic cell, a permanent genetic change is generally achieved by introduction of new DNA into the genome of the cell. In prokaryotic cells, permanent changes can be introduced into the chromosome or via extrachromosomal elements such as plasmids and expression vectors, which may contain one or more selectable markers to aid in their maintenance in the recombinant host cell. Suitable methods of genetic modification include viral infection, transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate precipitation, direct microinjection, and the like. The choice of method is generally dependent on the type of cell being transformed and the circumstances under which the transformation is taking place (i.e. *in vitro, ex vivo,* or *in vivo*). A general discussion of these methods can be found in Ausubel, et al, Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995.

[0042] "Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression. As used herein, the

terms "heterologous promoter" and "heterologous control regions" refer to promoters and other control regions that are not normally associated with a particular nucleic acid in nature. For example, a "transcriptional control region heterologous to a coding region" is a transcriptional control region that is not normally associated with the coding region in nature.

[0043] A "host cell," as used herein, denotes an *in vivo* or *in vitro* eukaryotic cell, a prokaryotic cell, or a cell from a multicellular organism (e.g., a cell line) cultured as a unicellular entity, which eukaryotic or prokaryotic cells can be, or have been, used as recipients for a nucleic acid (e.g., an expression vector), and include the progeny of the original cell which has been genetically modified by the nucleic acid. It is understood that the progeny of a single cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation. A "recombinant host cell" (also referred to as a "genetically modified host cell") is a host cell into which has been introduced a heterologous nucleic acid, e.g., an expression vector. For example, a subject prokaryotic host cell is a genetically modified prokaryotic host cell (e.g., a bacterium), by virtue of introduction into a suitable prokaryotic host cell of a heterologous nucleic acid, e.g., an exogenous nucleic acid that is foreign to (not normally found in nature in) the prokaryotic host cell, or a recombinant nucleic acid that is not normally found in the prokaryotic host cell; and a subject eukaryotic host cell is a genetically modified eukaryotic host cell, by virtue of introduction into a suitable eukaryotic host cell of a heterologous nucleic acid, e.g., an exogenous nucleic acid that is foreign to the eukaryotic host cell, or a recombinant nucleic acid that is not normally found in the eukaryotic host cell.

[0044] The term "conservative amino acid substitution" refers to the interchangeability in proteins of amino acid residues having similar side chains. For example, a group of amino acids having aliphatic side chains consists of glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains consists of serine and threonine; a group of amino acids having amide-containing side chains consists of asparagine and glutamine; a group of amino acids having aromatic side chains consists of phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains consists of lysine, arginine, and histidine; and a group of amino acids having sulfur-containing side chains consists of cysteine and methionine. Exemplary conservative amino acid substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

[0045] A polynucleotide or polypeptide has a certain percent "sequence identity" to another polynucleotide or polypeptide, meaning that, when aligned, that percentage of bases or amino acids are the same, and in the same relative position, when comparing the two sequences. Sequence similarity can be determined in a number of different manners. To determine sequence identity, sequences can be aligned using the methods and computer programs, including BLAST, available over the world wide web at ncbi.nlm.nih.gov/BLAST. See, e.g., Altschul *et al.* (1990), *J. Mol. Biol.* 215:403-10. Another

alignment algorithm is FASTA, available in the Genetics Computing Group (GCG) package, from Madison, Wisconsin, USA, a wholly owned subsidiary of Oxford Molecular Group, Inc. Other techniques for alignment are described in Methods in Enzymology, vol. 266: Computer Methods for Macromolecular Sequence Analysis (1996), ed. Doolittle, Academic Press, Inc., a division of Harcourt Brace & Co., San Diego, California, USA. Of particular interest are alignment programs that permit gaps in the sequence. The Smith-Waterman is one type of algorithm that permits gaps in sequence alignments. See *Meth. Mol. Biol.* 70: 173-187 (1997). Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences. See *J. Mol. Biol.* 48: 443-453 (1970).

[0046] As used herein, the terms "treatment," "treating," and the like, refer to obtaining a desired pharmacologic and/or physiologic effect. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or may be therapeutic in terms of a partial or complete cure for a disease and/or adverse effect attributable to the disease. "Treatment," as used herein, covers any treatment of a disease in a mammal, e.g., in a human, and includes: (a) preventing the disease from occurring in a subject which may be predisposed to the disease but has not yet been diagnosed as having it; (b) inhibiting the disease, i.e., arresting its development; and (c) relieving the disease, i.e., causing regression of the disease.

[0047] The terms "individual," "subject," "host," and "patient," used interchangeably herein, refer to an individual organism, e.g., a mammal, including, but not limited to, murines, simians, non-human primates, humans, mammalian farm animals, mammalian sport animals, and mammalian pets.

[0048] Before the present invention is further described, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[0049] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[0050] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present invention, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[0051] It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a CasZ polypeptide" includes a plurality of such polypeptides and reference to "the guide RNA" includes reference to one or more guide RNAs and equivalents thereof known to those skilled in the art, and so forth. It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as "solely," "only" and the like in connection with the recitation of claim elements, or use of a "negative" limitation.

[0052] It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of a single embodiment, may also be provided separately or in any suitable sub-combination. All combinations of the embodiments pertaining to the invention are specifically embraced by the present invention and are disclosed herein just as if each and every combination was individually and explicitly disclosed. In addition, all sub-combinations of the various embodiments and elements thereof are also specifically embraced by the present invention and are disclosed herein just as if each and every such sub-combination was individually and explicitly disclosed herein.

[0053] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

DETAILED DESCRIPTION

[0054] The present disclosure provides compositions and methods that include one or more of: (1) a "CasZ" protein (also referred to as a CasZ polypeptide), a nucleic acid encoding the CasZ protein, and/or a modified host cell comprising the CasZ protein (and/or a nucleic acid encoding the same); (2) a CasZ guide RNA that binds to and provides sequence specificity to the CasZ protein, a nucleic acid encoding the CasZ guide RNA, and/or a modified host cell comprising the CasZ guide RNA (and/or a nucleic acid encoding the same); and (3) a CasZ transactivating noncoding RNA (trancRNA) (referred to

herein as a "CasZ trancRNA"), a nucleic acid encoding the CasZ trancRNA, and/or a modified host cell comprising the CasZ trancRNA (and/or a nucleic acid encoding the same).

COMPOSITIONS

CRISPR/CASZ PROTEINS, GUIDE RNAS, AND TRANCRNAS

[0055] Class 2 CRISPR–Cas systems are characterized by effector modules that include a single multidomain protein. In the CasZ system, a CRISPR/Cas endonuclease (e.g., a CasZ protein) interacts with (binds to) a corresponding guide RNA (e.g., a CasZ guide RNA) to form a ribonucleoprotein (RNP) complex that is targeted to a particular site in a target nucleic acid via base pairing between the guide RNA and a target sequence within the target nucleic acid molecule. A guide RNA includes a nucleotide sequence (a guide sequence) that is complementary to a sequence (the target site) of a target nucleic acid. Thus, a CasZ protein forms a complex with a CasZ guide RNA and the guide RNA provides sequence specificity to the RNP complex via the guide sequence. The CasZ protein of the complex provides the site-specific activity. In other words, the CasZ protein is guided to a target site (e.g., stabilized at a target site) within a target nucleic acid (e.g. a target nucleotide sequence within a target chromosomal nucleic acid; or a target nucleotide sequence within a target extrachromosomal nucleic acid , e.g., an episomal nucleic acid, a minicircle nucleic acid, a mitochondrial nucleic acid, a chloroplast nucleic acid, etc.) by virtue of its association with the guide RNA.

[0056] The present disclosure provides compositions comprising a CasZ polypeptide (and/or a nucleic acid encoding the CasZ polypeptide) (e.g., where the CasZ polypeptide can be a naturally existing CasZ protein, a nickase CasZ protein, a dCasZ protein, a chimeric CasZ protein, etc.)(a CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK, or CasZl protein). The present disclosure provides compositions comprising a CasZ guide RNA (and/or a nucleic acid encoding the CasZ guide RNA). For example, the present disclosure provides compositions comprising (a) a CasZ polypeptide (and/or a nucleic acid encoding the CasZ polypeptide) and (b) a CasZ guide RNA (and/or a nucleic acid encoding the CasZ guide RNA). The present disclosure provides a nucleic acid/protein complex (RNP complex) comprising: (a) a CasZ polypeptide; and (b) a CasZ guide RNA. The present disclosure provides compositions comprising a CasZ trancRNA. The present disclosure provides compositions comprising a CasZ trancRNA and one or more of: (a) a CasZ protein, and (b) a CasZ guide RNA (e.g., comprising a CasZ trancRNA and a CasZ protein, a CasZ trancRNA and a CasZ guide RNA, or a CasZ trancRNA and a CasZ protein and a CasZ guide RNA. The present disclosure provides a nucleic acid/protein complex (RNP complex) comprising: (a) a CasZ polypeptide; (b) a CasZ guide RNA; and (c) a CasZ trancRNA. The present disclosure provides compositions comprising a CasZ protein and one or more of: (a) a CasZ trancRNA, and (b) a CasZ guide RNA.

CasZ protein

[0057] A CasZ polypeptide (this term is used interchangeably with the term "CasZ protein", "Cas14", "Cas14 polypeptide", or "Cas14 protein") can bind and/or modify (e.g., cleave, nick, methylate, demethylate, etc.) a target nucleic acid and/or a polypeptide associated with target nucleic acid (e.g., methylation or acetylation of a histone tail) (e.g., in some cases the CasZ protein includes a fusion partner with an activity, and in some cases the CasZ protein provides nuclease activity). In some cases, the CasZ protein is a naturally-occurring protein (e.g., naturally occurs in prokaryotic cells). In other cases, the CasZ protein is not a naturally-occurring polypeptide (e.g., the CasZ protein is a variant CasZ protein, a chimeric protein, and the like). A CasZ protein includes 3 partial RuvC domains (RuvC-I, RuvC-II, and RuvC-III, also referred to herein as subdomains) that are not contiguous with respect to the primary amino acid sequence of the CasZ protein, but form a RuvC domain once the protein is produced and folds. A naturally occurring CasZ protein functions as an endonuclease that catalyzes cleavage at a specific sequence in a targeted nucleic acid (e.g., a double stranded DNA (dsDNA)). The sequence specificity is provided by the associated guide RNA, which hybridizes to a target sequence within the target DNA. The naturally occurring CasZ guide RNA is a crRNA, where the crRNA includes (i) a guide sequence that hybridizes to a target sequence in the target DNA and (ii) a protein binding segment that binds to the CasZ protein.

[0058] In some embodiments, the CasZ protein of the subject methods and/or compositions is (or is derived from) a naturally occurring (wild type) protein. Examples of naturally occurring CasZ proteins (e.g., CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZk, CasZl) are depicted in Figure 1. In some cases, a subject CasZ protein is a CasZa protein. In some cases, a subject CasZ protein is a CasZb protein. In some cases, a subject CasZ protein is a CasZc protein. In some cases, a subject CasZ protein is a CasZd protein. In some cases, a subject CasZ protein is a CasZe protein. In some cases, a subject CasZ protein is a CasZf protein. In some cases, a subject CasZ protein is a CasZg protein. In some cases, a subject CasZ protein is a CasZh protein. In some cases, a subject CasZ protein is a CasZi protein. In some cases, a subject CasZ protein is a CasZj protein. In some cases, a subject CasZ protein is a CasZk protein. In some cases, a subject CasZ protein is a CasZl protein. In some cases, a subject CasZ protein is a CasZe, CasZf, CasZg, or CasZh protein. In some cases, a subject CasZ protein is a CasZj, CasZk, or CasZl protein.

[0059] It is important to note that this newly discovered protein (CasZ) is short compared to previously identified CRISPR-Cas endonucleases, and thus use of this protein as an alternative provides the advantage that the nucleotide sequence encoding the protein is relatively short. This is useful, for example, in cases where a nucleic acid encoding the CasZ protein is desirable, e.g., in situations that employ a viral vector (e.g., an AAV vector), for delivery to a cell such as a eukaryotic cell (e.g.., mammalian cell, human cell, mouse cell, in vitro, ex vivo, in vivo) for research and/or clinical

applications. In addition, in their natural context, the CasZ-encoding DNA sequences are present in loci that also have a Cas1 protein.

[0060] In some cases, a subject CasZ protein has a length of 900 amino acids or less (e.g., 850 amino acids or less, 800 amino acids or less, 750 amino acids or less, or 700 amino acids or less). In some cases, a subject CasZ protein has a length of 850 amino acids or less (e.g., 850 amino acids or less). In some cases, a subject CasZ protein length of 800 amino acids or less (e.g., 750 amino acids or less). In some cases, a subject CasZ protein has a length of 700 amino acids or less. In some cases, a subject CasZ protein has a length of 650 amino acids or less.

[0061] In some cases, a subject CasZ protein has a length in a range of from 350-900 amino acids (e.g., 350-850, 350-800, 350-750, 350-700, 400-900, 400-850, 400-800, 400-750, or 400-700 amino acids).

[0062] In some cases, a subject CasZ protein (e.g., CasZa) has a length in a range of from 350-750 amino acids (e.g., 350-700, 350-550, 450-550, 450-750, 450-650, or 450-550 amino acids). In some cases, a subject CasZ protein (e.g., CasZa) has a length in a range of from 450-750 amino acids (e.g., 500-700 amino acids). In some cases, a subject CasZ protein (e.g., CasZa) has a length in a range of from 350-700 amino acids (e.g., 350-650, 350-600, or 350-550 amino acids). In some cases, a subject CasZ protein (e.g., CasZa) has a length in a range of from 500-700 amino acids. In some cases, a subject CasZ protein (e.g., CasZa) has a length in a range of from 450-550 amino acids. In some cases, a subject CasZ protein (e.g., CasZa) has a length in a range of from 350-550 amino acids.

[0063] In some cases, a subject CasZ protein (e.g., CasZb) has a length in a range of from 350-700 amino acids (e.g., 350-650, or 350-620 amino acids). In some cases, a subject CasZ protein (e.g., CasZb) has a length in a range of from 450-700 amino acids (e.g., 450-650, 500-650 or 500-620 amino acids). In some cases, a subject CasZ protein (e.g., CasZb) has a length in a range of from 500-650 amino acids (e.g., 500-620 amino acids). In some cases, a subject CasZ protein (e.g., CasZb) has a length in a range of from 500-620 amino acids.

[0064] In some cases, a subject CasZ protein (e.g., CasZc) has a length in a range of from 600-800 amino acids (e.g., 600-650 or 700-800 amino acids). In some cases, a subject CasZ protein (e.g., CasZc) has a length in a range of from 600-650 amino acids. In some cases, a subject CasZ protein (e.g., CasZc) has a length in a range of from 700-800 amino acids.

[0065] In some cases, a subject CasZ protein (e.g., CasZd) has a length in a range of from 400-650 amino acids (e.g., 400-600, 400-550, 500-650, 500-600 or 500-550 amino acids). In some cases, a subject CasZ protein (e.g., CasZd) has a length in a range of from 500-600 amino acids. In some cases, a subject CasZ protein (e.g., CasZd) has a length in a range of from 500-550 amino acids. In some cases, a subject CasZ protein (e.g., CasZd) has a length in a range of from 400-550 amino acids.

[0066] In some cases, a subject CasZ protein (e.g., CasZe) has a length in a range of from 450-700 amino acids (e.g., 450--650, 450-615, 475-700, 475-650, or 475-615 amino acids). In some cases, a subject CasZ protein (e.g., CasZe) has a length in a range of from 450-675 amino acids. In some cases, a subject CasZ protein (e.g., CasZe) has a length in a range of from 475-675 amino acids.

[0067] In some cases, a subject CasZ protein (e.g., CasZf) has a length in a range of from 400-550 amino acids (e.g., 400-520, 400-500, 400-475, 415-550, 415-520, 415-500, or 415-475 amino acids). In some cases, a subject CasZ protein (e.g., CasZf) has a length in a range of from 400-475 amino acids (e.g., 400-450 amino acids).

[0068] In some cases, a subject CasZ protein (e.g., CasZg) has a length in a range of from 500-750 amino acids (e.g., 550-750 or 500-700 amino acids). In some cases, a subject CasZ protein (e.g., CasZg) has a length in a range of from 700-750 amino acids. In some cases, a subject CasZ protein (e.g., CasZg) has a length in a range of from 550-600 amino acids.

[0069] In some cases, a subject CasZ protein (e.g., CasZh) has a length in a range of from 380-450 amino acids (e.g., 380-420, 400-450, or 400-420 amino acids). In some cases, a subject CasZ protein (e.g., CasZh) has a length in a range of from 400-420 amino acids.

[0070] In some cases, a subject CasZ protein (e.g., CasZi) has a length in a range of from 700-800 amino acids (e.g., 700-750, 720-800, or 720-750 amino acids). In some cases, a subject CasZ protein (e.g., CasZi) has a length in a range of from 720-780 amino acids.

[0071] In some cases, a subject CasZ protein (e.g., CasZj) has a length in a range of from 600-750 amino acids (e.g., 600-700 or 650-700 amino acids). In some cases, a subject CasZ protein (e.g., CasZj) has a length in a range of from 400-420 amino acids.

[0072] In some cases, a subject CasZ protein (e.g., CasZk) has a length in a range of from 450-600 amino acids (e.g., 450-580, 480-600, 480-580, or 500-600 amino acids). In some cases, a subject CasZ protein (e.g., CasZk) has a length in a range of from 480-580 amino acids.

[0073] In some cases, a subject CasZ protein (e.g., CasZl) has a length in a range of from 350-500 amino acids (e.g., 350-450, 380-450, 350-420, or 380-420 amino acids). In some cases a subject CasZ protein (e.g., CasZl) has a length in a range of from 380-420 amino acids.

[0074] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZa protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity)

with a CasZa protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZa protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZa protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZa amino acid sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZa amino acid sequence of Figure 1 or Figure 7, with the exception that the sequence includes one or more amino acid substitutions (e.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions)(e.g., in some cases such that the CasZ protein is a dCasZ). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZa protein of Figure 1 or Figure 7 and has a length in a range of from 350-800 amino acids (e.g., 350-800, 350-750, 350-700, 350-550, 450-550, 450-750, 450-650, or 450-550 amino acids).

[0075] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZb protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZb protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZb protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZb protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZb amino acid sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZb amino acid sequence of Figure 1 or Figure 7, with the exception that the sequence includes an amino acid substitution (e.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions)(e.g., in some cases such that the CasZ protein is a dCasZ). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more,

99% or more, or 100% sequence identity) with a CasZb protein of Figure 1 or Figure 7 and has a length in a range of from 350-700 amino acids (e.g., 350-650, or 350-620 amino acids).

[0076] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZc protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZc protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZc protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZc protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZc amino acid sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZc amino acid sequence of Figure 1 or Figure 7, with the exception that the sequence includes an amino acid substitution (c.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions)(e.g., in some cases such that the CasZ protein is a dCasZ). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZc protein of Figure 1 or Figure 7 and has a length in a range of from 600-800 amino acids (e.g., 600-650 or 700-800 amino acids).

[0077] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZd protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZd protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZd protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence

identity) with a CasZd protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZd amino acid sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZd amino acid sequence of Figure 1 or Figure 7, with the exception that the sequence includes an amino acid substitution (e.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions)(e.g., in some cases such that the CasZ protein is a dCasZ). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZd protein of Figure 1 or Figure 7 and has a length in a range of from 400-650 amino acids (e.g., 400-600, 400-550, 500-650, 500-600 or 500-550 amino acids).

[0078] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZe protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZc protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZe protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZe protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZe amino acid sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZe amino acid sequence of Figure 1 or Figure 7, with the exception that the sequence includes an amino acid substitution (e.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions)(e.g., in some cases such that the CasZ protein is a dCasZ). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZe protein of Figure 1 or Figure 7 and has a length in a range of from 450-700 amino acids (e.g., 450--650, 450-615, 475-700, 475-650, or 475-615 amino acids).

[0079] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or

more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZf protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZf protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZf protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZf protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZf amino acid sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZf amino acid sequence of Figure 1 or Figure 7, with the exception that the sequence includes an amino acid substitution (e.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions)(e.g., in some cases such that the CasZ protein is a dCasZ). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZf protein of Figure 1 or Figure 7 and has a length in a range of from 400-750 amino acids (e.g., 400-700, 700-650, 400-620, 400-600, 400-550, 400-520, 400-500, 400-475, 415-550, 415-520, 415-500, or 415-475 amino acids).

[0080] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZg protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZg protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZg protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZg protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZg amino acid sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZg amino acid sequence of Figure 1 or Figure 7, with the exception that the sequence includes an amino acid substitution (e.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring

catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions)(e.g., in some cases such that the CasZ protein is a dCasZ). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZg protein of Figure 1 or Figure 7 and has a length in a range of from 500-750 amino acids (e.g., 500-750 amino acids (e.g., 550-750 amino acids)).

[0081] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZh protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZh protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZh protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZh protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZh amino acid sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZh amino acid sequence of Figure 1 or Figure 7, with the exception that the sequence includes an amino acid substitution (e.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions)(e.g., in some cases such that the CasZ protein is a dCasZ). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZh protein of Figure 1 or Figure 7 and has a length in a range of from 380-450 amino acids (e.g., 380-420, 400-450, or 400-420 amino acids).

[0082] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZi protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZi protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more,

97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZi protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZi protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZi amino acid sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZi amino acid sequence of Figure 1 or Figure 7, with the exception that the sequence includes an amino acid substitution (e.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions)(e.g., in some cases such that the CasZ protein is a dCasZ). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZi protein of Figure 1 or Figure 7 and has a length in a range of from 700-800 amino acids (e.g., 700-750, 720-800, or 720-750 amino acids).

[0083] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZj protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZj protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZj protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZj protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZj amino acid sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZj amino acid sequence of Figure 1 or Figure 7, with the exception that the sequence includes an amino acid substitution (e.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions)(e.g., in some cases such that the CasZ protein is a dCasZ). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZj protein of Figure 1 or Figure 7 and has a length in a range of from 600-750 amino acids (e.g., 600-700 or 650-700 amino acids).

[0084] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or

more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZk protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZk protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZk protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZk protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZk amino acid sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZk amino acid sequence of Figure 1 or Figure 7, with the exception that the sequence includes an amino acid substitution (e.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions)(e.g., in some cases such that the CasZ protein is a dCasZ). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZk protein of Figure 1 or Figure 7 and has a length in a range of from 450-600 amino acids (e.g., 450-580, 480-600, 480-580, or 500-600 amino acids).

[0085] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZl protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZl protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZl protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZl protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZl amino acid sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes a CasZl amino acid sequence of Figure 1 or Figure 7, with the exception that the sequence includes an amino acid substitution (e.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring

catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions)(e.g., in some cases such that the CasZ protein is a dCasZ). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZl protein of Figure 1 or Figure 7 and has a length in a range of from 450-600 amino acids (e.g., 450-580, 480-600, 480-580, or 500-600 amino acids).

[0086] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZe, CasZf, CasZg, or CasZh protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZe, CasZf, CasZg, or CasZh protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZe, CasZf, CasZg, or CasZh protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZe, CasZf, CasZg, or CasZh protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having a CasZe, CasZf, CasZg, or CasZh protein sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having a CasZe, CasZf, CasZg, or CasZh protein sequence of Figure 1 or Figure 7, with the exception that the sequence includes an amino acid substitution (e.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZe, CasZf, CasZg, or CasZh protein of Figure 1 or Figure 7 and has a length in a range of from 350-900 amino acids (e.g., 350-850, 350-800, 400-900, 400-850, or 400-800 amino acids).

[0087] In some cases, a subject CasZ protein (of the subject compositions and/or methods) includes an amino acid sequence having 20% or more sequence identity (e.g., 30% or more, 40% or more, 50% or more, 60% or more, 70% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK, or CasZl protein of Figure 1 or Figure 7. For example, in some cases, a subject CasZ protein includes an amino acid sequence having 50% or more sequence identity (e.g., 60% or more, 70% or more, 80% or more, 85% or more, 90% or

more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK, or CasZl protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 80% or more sequence identity (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK, or CasZl protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK, or CasZl protein of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having a CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK, or CasZl protein sequence of Figure 1 or Figure 7. In some cases, a subject CasZ protein includes an amino acid sequence having a CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK, or CasZl protein sequence of Figure 1 or Figure 7, with the exception that the sequence includes an amino acid substitution (e.g., 1, 2, or 3 amino acid substitutions) that reduces the naturally occurring catalytic activity of the protein (e.g., such as at one or more catalytic amino acid positions). In some cases, a subject CasZ protein includes an amino acid sequence having 90% or more sequence identity (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100% sequence identity) with a CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, CasZi, CasZj, CasZK, or CasZl protein of Figure 1 or Figure 7 and has a length in a range of from 350-900 amino acids (e.g., 350-850, 350-800, 400-900, 400-850, or 400-800 amino acids).

CasZ Variants

[0088] A variant CasZ protein has an amino acid sequence that is different by at least one amino acid (e.g., has a deletion, insertion, substitution, fusion) when compared to the amino acid sequence of the corresponding wild type CasZ protein. A CasZ protein that cleaves one strand but not the other of a double stranded target nucleic acid is referred to herein as a "nickase" (e.g., a "nickase CasZ"). A CasZ protein that has substantially no nuclease activity is referred to herein as a dead CasZ protein ("dCasZ") (with the caveat that nuclease activity can be provided by a heterologous polypeptide – a fusion partner – in the case of a chimeric CasZ protein, which is described in more detail below). For any of the CasZ variant proteins described herein (e.g., nickase CasZ, dCasZ, chimeric CasZ), the CasZ variant can include a CasZ protein sequence with the same parameters described above (e.g., domains that are present, percent identity, length, and the like).

Variants – catalytic activity

[0089] In some cases, the CasZ protein is a variant CasZ protein, e.g., mutated relative to the naturally occurring catalytically active sequence, and exhibits reduced cleavage activity (e.g., exhibits 90%, or less, 80% or less, 70% or less, 60% or less, 50% or less, 40% or less, or 30% or less cleavage

activity) when compared to the corresponding naturally occurring sequence. In some cases, such a variant CasZ protein is a catalytically 'dead' protein (has substantially no cleavage activity) and can be referred to as a 'dCasZ.' In some cases, the variant CasZ protein is a nickase (cleaves only one strand of a double stranded target nucleic acid, e.g., a double stranded target DNA). As described in more detail herein, in some cases, a CasZ protein (in some case a CasZ protein with wild type cleavage activity and in some cases a variant CasZ with reduced cleavage activity, e.g., a dCasZ or a nickase CasZ) is fused (conjugated) to a heterologous polypeptide that has an activity of interest (e.g., a catalytic activity of interest) to form a fusion protein (a chimeric CasZ protein).

[0090] Catalytic residues of CasZ include D405, E586 and D684 when numbered according to CasZi.1 (e.g., see Figure 1). Thus, in some cases, the CasZ protein has reduced activity and one or more of the above described amino acids (or one or more corresponding amino acids of any CasZ protein) are mutated (e.g., substituted with an alanine). In some cases, the variant CasZ protein is a catalytically 'dead' protein (is catalytically inactive) and is referred to as 'dCasZ.' A dCasZ protein can be fused to a fusion partner that provides an activity, and in some cases, the dCasZ (e.g., one without a fusion partner that provides catalytic activity – but which can have an NLS when expressed in a eukaryotic cell) can bind to target DNA and can be used for imaging (e.g., the protein can be tagged/labeled) and/or can block RNA polymerase from transcribing from a target DNA. In some cases, the variant CasZ protein is a nickase (cleaves only one strand of a double stranded target nucleic acid, e.g., a double stranded target DNA).

Variants – chimeric CasZ (i.e., fusion proteins)

[0091] As noted above, in some cases, a CasZ protein (in some cases a CasZ protein with wild type cleavage activity and in some cases a variant CasZ with reduced cleavage activity, e.g., a dCasZ or a nickase CasZ) is fused (conjugated) to a heterologous polypeptide that has an activity of interest (e.g., a catalytic activity of interest) to form a fusion protein (a chimeric CasZ protein). A heterologous polypeptide to which a CasZ protein can be fused is referred to herein as a 'fusion partner.'

[0092] In some cases, the fusion partner can modulate transcription (e.g., inhibit transcription, increase transcription) of a target DNA. For example, in some cases the fusion partner is a protein (or a domain from a protein) that inhibits transcription (e.g., a transcriptional repressor, a protein that functions via recruitment of transcription inhibitor proteins, modification of target DNA such as methylation, recruitment of a DNA modifier, modulation of histones associated with target DNA, recruitment of a histone modifier such as those that modify acetylation and/or methylation of histones, and the like). In some cases the fusion partner is a protein (or a domain from a protein) that increases transcription (e.g., a transcription activator, a protein that acts via recruitment of transcription activator proteins, modification of target DNA such as demethylation, recruitment of a DNA modifier, modulation

of histones associated with target DNA, recruitment of a histone modifier such as those that modify acetylation and/or methylation of histones, and the like).

[0093] In some cases, a chimeric CasZ protein includes a heterologous polypeptide that has enzymatic activity that modifies a target nucleic acid (e.g., nuclease activity such as FokI nuclease activity, methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity).

[0094] In some cases, a chimeric CasZ protein includes a heterologous polypeptide that has enzymatic activity that modifies a polypeptide (e.g., a histone) associated with a target nucleic acid (e.g., methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity or demyristoylation activity).

[0095] Examples of proteins (or fragments thereof) that can be used in increase transcription include but are not limited to: transcriptional activators such as VP16, VP64, VP48, VP160, p65 subdomain (e.g., from NFkB), and activation domain of EDLL and/or TAL acitvation domain (e.g., for activity in plants); histone lysine methyltransferases such as SET1A, SET1B, MLL1 to 5, ASH1, SYMD2, NSD1, and the like; histone lysine demethylases such as JHDM2a/b, UTX, JMJD3, and the like; histone acetyltransferases such as GCN5, PCAF, CBP, p300, TAF1, TIP60/PLIP, MOZ/MYST3, MORF/MYST4, SRC1, ACTR, P160, CLOCK, and the like; and DNA demethylases such as Ten-Eleven Translocation (TET) dioxygenase 1 (TET1CD), TET1, DME, DML1, DML2, ROS1, and the like.

[0096] Examples of proteins (or fragments thereof) that can be used in decrease transcription include but are not limited to: transcriptional repressors such as the Krüppel associated box (KRAB or SKD); KOX1 repression domain; the Mad mSIN3 interaction domain (SID); the ERF repressor domain (ERD), the SRDX repression domain (e.g., for repression in plants), and the like; histone lysine methyltransferases such as Pr-SET7/8, SUV4-20H1, RIZ1, and the like; histone lysine demethylases such as JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASC1, JMJD2D, JARID1A/RBP2, JARID1B/PLU-1, JARID1C/SMCX, JARID1D/SMCY, and the like; histone lysine deacetylases such as HDAC1, HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRT1, SIRT2, HDAC11, and the like; DNA methylases such as HhaI DNA m5c-methyltransferase (M.HhaI), DNA methyltransferase 1 (DNMT1), DNA methyltransferase 3a (DNMT3a), DNA methyltransferase 3b (DNMT3b), METI, DRM3 (plants), ZMET2, CMT1, CMT2 (plants), and the like; and periphery recruitment elements such as Lamin A, Lamin B, and the like.

[0097] In some cases, the fusion partner has enzymatic activity that modifies the target nucleic acid (e.g., ssRNA, dsRNA, ssDNA, dsDNA). Examples of enzymatic activity that can be provided by the fusion partner include but are not limited to: nuclease activity such as that provided by a restriction enzyme (e.g., FokI nuclease), methyltransferase activity such as that provided by a methyltransferase (e.g., HhaI DNA m5c-methyltransferase (M.HhaI), DNA methyltransferase 1 (DNMT1), DNA methyltransferase 3a (DNMT3a), DNA methyltransferase 3b (DNMT3b), METI, DRM3 (plants), ZMET2, CMT1, CMT2 (plants), and the like); demethylase activity such as that provided by a demethylase (e.g., Ten-Eleven Translocation (TET) dioxygenase 1 (TET1CD), TET1, DME, DML1, DML2, ROS1, and the like) , DNA repair activity, DNA damage activity, deamination activity such as that provided by a deaminase (e.g., a cytosine deaminase enzyme such as rat APOBEC1), dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity such as that provided by an integrase and/or resolvase (e.g., Gin invertase such as the hyperactive mutant of the Gin invertase, GinH106Y; human immunodeficiency virus type 1 integrase (IN); Tn3 resolvase; and the like), transposase activity, recombinase activity such as that provided by a recombinase (e.g., catalytic domain of Gin recombinase), polymerase activity, ligase activity, helicase activity, photolyase activity, and glycosylase activity).

[0098] In some cases, the fusion partner has enzymatic activity that modifies a protein associated with the target nucleic acid (e.g., ssRNA, dsRNA, ssDNA, dsDNA) (e.g., a histone, an RNA binding protein, a DNA binding protein, and the like). Examples of enzymatic activity (that modifyies a protein associated with a target nucleic acid) that can be provided by the fusion partner include but are not limited to: methyltransferase activity such as that provided by a histone methyltransferase (HMT) (e.g., suppressor of variegation 3-9 homolog 1 (SUV39H1, also known as KMT1A), euchromatic histone lysine methyltransferase 2 (G9A, also known as KMT1C and EHMT2), SUV39H2, ESET/SETDB1, and the like, SET1A, SET1B, MLL1 to 5, ASH1, SYMD2, NSD1, DOT1L, Pr-SET7/8, SUV4-20H1, EZH2, RIZ1), demethylase activity such as that provided by a histone demethylase (e.g., Lysine Demethylase 1A (KDM1A also known as LSD1), JHDM2a/b, JMJD2A/JHDM3A, JMJD2B, JMJD2C/GASC1, JMJD2D, JARID1A/RBP2, JARID1B/PLU-1, JARID1C/SMCX, JARID1D/SMCY, UTX, JMJD3, and the like), acetyltransferase activity such as that provided by a histone acetylase transferase (e.g., catalytic core/fragement of the human acetyltransferase p300, GCN5, PCAF, CBP, TAF1, TIP60/PLIP, MOZ/MYST3, MORF/MYST4, HBO1/MYST2, HMOF/MYST1, SRC1, ACTR, P160, CLOCK, and the like), deacetylase activity such as that provided by a histone deacetylase (e.g., HDAC1, HDAC2, HDAC3, HDAC8, HDAC4, HDAC5, HDAC7, HDAC9, SIRT1, SIRT2, HDAC11, and the like), kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, and demyristoylation activity.

[0099] Additional examples of a suitable fusion partners are dihydrofolate reductase (DHFR) destabilization domain (e.g., to generate a chemically controllable chimeric CasZ protein), and a chloroplast transit peptide. Suitable chloroplast transit peptides include, but are not limited to:

[00100] MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGR VKCMQVWPPIGKKKFETLSYLPPLTRDSRA (SEQ ID NO: 101);
MASMISSSAVTTVSRASRGQSAAMAPFGGLKSMTGFPVRKVNTDITSITSNGGRVKS (SEQ ID NO: 102);
MASSMLSSATMVASPAQATMVAPFNGLKSSAAFPATRKANNDITSITSNGGRVNCMQVWPPIE KKKFETLSYLPDLTDSGGRVNC (SEQ ID NO: 103);
MAQVSRICNGVQNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGSELRPL KVMSSVSTAC (SEQ ID NO: 104);
MAQVSRICNGVWNPSLISNLSKSSQRKSPLSVSLKTQQHPRAYPISSSWGLKKSGMTLIGSELRP LKVMSSVSTAC (SEQ ID NO: 105);
MAQINNMAQGIQTLNPNSNFHKPQVPKSSSFLVFGSKKLKNSANSMLVLKKDSIFMQLFCSFRIS ASVATAC (SEQ ID NO: 106);
MAALVTSQLATSGTVLSVTDRFRRPGFQGLRPRNPADAALGMRTVGASAAPKQSRKPHRFDRR CLSMVV (SEQ ID NO: 107);
MAALTTSQLATSATGFGIADRSAPSSLLRHGFQGLKPRSPAGGDATSLSVTTSARATPKQQRSV QRGSRRFPSVVVC (SEQ ID NO: 108);
MASSVLSSAAVATRSNVAQANMVAPFTGLKSAASFPVSRKQNLDITSIASNGGRVQC (SEQ ID NO: 109);
MESLAATSVFAPSRVAVPAARALVRAGTVVPTRRTSSTSGTSGVKCSAAVTPQASPVISRSAAA A (SEQ ID NO: 110); and
MGAAATSMQSLKFSNRLVPPSRRLSPVPNNVTCNNLPKSAAPVRTVKCCASSWNSTINGAAAT TNGASAASS (SEQ ID NO: 111).

[00101] In some case, a CasZ fusion polypeptide of the present disclosure comprises: a) a CasZ polypeptide of the present disclosure; and b) a chloroplast transit peptide. Thus, for example, a CRISPR-CasZ complex can be targeted to the chloroplast. In some cases, this targeting may be achieved by the presence of an N-terminal extension, called a chloroplast transit peptide (CTP) or plastid transit peptide. Chromosomal transgenes from bacterial sources must have a sequence encoding a CTP sequence fused to a sequence encoding an expressed polypeptide if the expressed polypeptide is to be compartmentalized in the plant plastid (e.g. chloroplast). Accordingly, localization of an exogenous polypeptide to a chloroplast is often 1 accomplished by means of operably linking a polynucleotide sequence encoding a CTP sequence to the 5' region of a polynucleotide encoding the exogenous polypeptide. The CTP is removed in a processing step during translocation into the plastid. Processing efficiency may, however,

be affected by the amino acid sequence of the CTP and nearby sequences at the NH 2 terminus of the peptide. Other options for targeting to the chloroplast which have been described are the maize cab-m7 signal sequence (U.S. Pat. No. 7,022,896, WO 97/41228) a pea glutathione reductase signal sequence (WO 97/41228) and the CTP described in US2009029861.

[00102] In some cases, a CasZ fusion polypeptide of the present disclosure can comprise: a) a CasZ polypeptide of the present disclosure; and b) an endosomal escape peptide. In some cases, an endosomal escape polypeptide comprises the amino acid sequence GLFXALLXLLXSLWXLLLXA (SEQ ID NO: 112), wherein each X is independently selected from lysine, histidine, and arginine. In some cases, an endosomal escape polypeptide comprises the amino acid sequence GLFHALLHLLHSLWHLLLHA (SEQ ID NO: 113).

[00103] For examples of some of the above fusion partners (and more) used in the context of fusions with Cas9, Zinc Finger, and/or TALE proteins (for site specific target nucleic modification, modulation of transcription, and/or target protein modification, e.g., histone modification), see, e.g.: Nomura et al, J Am Chem Soc. 2007 Jul 18;129(28):8676-7; Rivenbark et al., Epigenetics. 2012 Apr;7(4):350-60; Nucleic Acids Res. 2016 Jul 8;44(12):5615-28; Gilbert et al., Cell. 2013 Jul 18;154(2):442-51; Kearns et al., Nat Methods. 2015 May;12(5):401-3; Mendenhall et al., Nat Biotechnol. 2013 Dec;31(12):1133-6; Hilton et al., Nat Biotechnol. 2015 May;33(5):510-7; Gordley et al., Proc Natl Acad Sci U S A. 2009 Mar 31;106(13):5053-8; Akopian et al., Proc Natl Acad Sci U S A. 2003 Jul 22;100(15):8688-91; Tan et., al., J Virol. 2006 Feb;80(4):1939-48; Tan et al., Proc Natl Acad Sci U S A. 2003 Oct 14;100(21):11997-2002; Papworth et al., Proc Natl Acad Sci U S A. 2003 Feb 18;100(4):1621-6; Sanjana et al., Nat Protoc. 2012 Jan 5;7(1):171-92; Beerli et al., Proc Natl Acad Sci U S A. 1998 Dec 8;95(25):14628-33; Snowden et al., Curr Biol. 2002 Dec 23;12(24):2159-66; Xu et.al., Xu et al., Cell Discov. 2016 May 3;2:16009; Komor et al., Nature. 2016 Apr 20;533(7603):420-4; Chaikind et al., Nucleic Acids Res. 2016 Aug 11; Choudhury at. al., Oncotarget. 2016 Jun 23; Du et al., Cold Spring Harb Protoc. 2016 Jan 4; Pham et al., Methods Mol Biol. 2016;1358:43-57; Balboa et al., Stem Cell Reports. 2015 Sep 8;5(3):448-59; Hara et al., Sci Rep. 2015 Jun 9;5:11221; Piatek et al., Plant Biotechnol J. 2015 May;13(4):578-89; Hu et al., Nucleic Acids Res. 2014 Apr;42(7):4375-90; Cheng et al., Cell Res. 2013 Oct;23(10):1163-71; and Maeder et al., Nat Methods. 2013 Oct;10(10):977-9.

[00104] Additional suitable heterologous polypeptides include, but are not limited to, a polypeptide that directly and/or indirectly provides for increased transcription and/or translation of a target nucleic acid (e.g., a transcription activator or a fragment thereof, a protein or fragment thereof that recruits a transcription activator, a small molecule/drug-responsive transcription and/or translation regulator, a translation-regulating protein, etc.). Non-limiting examples of heterologous polypeptides to accomplish increased or decreased transcription include transcription activator and transcription repressor domains. In some such cases, a chimeric CasZ polypeptide is targeted by the guide nucleic acid

(guide RNA) to a specific location (i.e., sequence) in the target nucleic acid and exerts locus-specific regulation such as blocking RNA polymerase binding to a promoter (which selectively inhibits transcription activator function), and/or modifying the local chromatin status (e.g., when a fusion sequence is used that modifies the target nucleic acid or modifies a polypeptide associated with the target nucleic acid). In some cases, the changes are transient (e.g., transcription repression or activation). In some cases, the changes are inheritable (e.g., when epigenetic modifications are made to the target nucleic acid or to proteins associated with the target nucleic acid, e.g., nucleosomal histones).

[00105] Non-limiting examples of heterologous polypeptides for use when targeting ssRNA target nucleic acids include (but are not limited to): splicing factors (e.g., RS domains); protein translation components (e.g., translation initiation, elongation, and/or release factors; e.g., eIF4G); RNA methylases; RNA editing enzymes (e.g., RNA deaminases, e.g., adenosine deaminase acting on RNA (ADAR), including A to I and/or C to U editing enzymes); helicases; RNA-binding proteins; and the like. It is understood that a heterologous polypeptide can include the entire protein or in some cases can include a fragment of the protein (e.g., a functional domain).

[00106] The heterologous polypeptide of a subject chimeric CasZ polypeptide can be any domain capable of interacting with ssRNA (which, for the purposes of this disclosure, includes intramolecular and/or intermolecular secondary structures, e.g., double-stranded RNA duplexes such as hairpins, stem-loops, etc.), whether transiently or irreversibly, directly or indirectly, including but not limited to an effector domain selected from the group comprising; Endonucleases (for example RNase III, the CRR22 DYW domain, Dicer, and PIN (PilT N-terminus) domains from proteins such as SMG5 and SMG6); proteins and protein domains responsible for stimulating RNA cleavage (for example CPSF, CstF, CFIm and CFIIm); Exonucleases (for example XRN-1 or Exonuclease T) ; Deadenylases (for example HNT3); proteins and protein domains responsible for nonsense mediated RNA decay (for example UPF1, UPF2, UPF3, UPF3b, RNP S1, Y14, DEK, REF2, and SRm160); proteins and protein domains responsible for stabilizing RNA (for example PABP) ; proteins and protein domains responsible for repressing translation (for example Ago2 and Ago4); proteins and protein domains responsible for stimulating translation (for example Staufen); proteins and protein domains responsible for (e.g., capable of) modulating translation (e.g., translation factors such as initiation factors, elongation factors, release factors, etc., e.g., eIF4G); proteins and protein domains responsible for polyadenylation of RNA (for example PAP1, GLD-2, and Star- PAP) ; proteins and protein domains responsible for polyuridinylation of RNA (for example CI D1 and terminal uridylate transferase) ; proteins and protein domains responsible for RNA localization (for example from IMP1, ZBP1, She2p, She3p, and Bicaudal-D); proteins and protein domains responsible for nuclear retention of RNA (for example Rrp6); proteins and protein domains responsible for nuclear export of RNA (for example TAP, NXF1, THO, TREX, REF, and Aly) ; proteins and protein domains responsible for repression of RNA splicing (for example PTB,

Sam68, and hnRNP A1) ; proteins and protein domains responsible for stimulation of RNA splicing (for example Serine/Arginine-rich (SR) domains) ; proteins and protein domains responsible for reducing the efficiency of transcription (for example FUS (TLS)); and proteins and protein domains responsible for stimulating transcription (for example CDK7 and HIV Tat). Alternatively, the effector domain may be selected from the group comprising Endonucleases; proteins and protein domains capable of stimulating RNA cleavage; Exonucleases; Deadenylases; proteins and protein domains having nonsense mediated RNA decay activity; proteins and protein domains capable of stabilizing RNA; proteins and protein domains capable of repressing translation; proteins and protein domains capable of stimulating translation; proteins and protein domains capable of modulating translation (e.g., translation factors such as initiation factors, elongation factors, release factors, etc., e.g., eIF4G); proteins and protein domains capable of polyadenylation of RNA; proteins and protein domains capable of polyuridinylation of RNA; proteins and protein domains having RNA localization activity; proteins and protein domains capable of nuclear retention of RNA; proteins and protein domains having RNA nuclear export activity; proteins and protein domains capable of repression of RNA splicing; proteins and protein domains capable of stimulation of RNA splicing; proteins and protein domains capable of reducing the efficiency of transcription ; and proteins and protein domains capable of stimulating transcription. Another suitable heterologous polypeptide is a PUF RNA-binding domain, which is described in more detail in WO2012068627, which is hereby incorporated by reference in its entirety.

[00107] Some RNA splicing factors that can be used (in whole or as fragments thereof) as heterologous polypeptides for a chimeric CasZ polypeptide have modular organization, with separate sequence-specific RNA binding modules and splicing effector domains. For example, members of the Serine/ Arginine-rich (SR) protein family contain N-terminal RNA recognition motifs (RRMs) that bind to exonic splicing enhancers (ESEs) in pre-mRNAs and C-terminal RS domains that promote exon inclusion. As another example, the hnRNP protein hnRNP Al binds to exonic splicing silencers (ESSs) through its RRM domains and inhibits exon inclusion through a C-terminal Glycine-rich domain. Some splicing factors can regulate alternative use of splice site (ss) by binding to regulatory sequences between the two alternative sites. For example, ASF/SF2 can recognize ESEs and promote the use of intron proximal sites, whereas hnRNP Al can bind to ESSs and shift splicing towards the use of intron distal sites. One application for such factors is to generate ESFs that modulate alternative splicing of endogenous genes, particularly disease associated genes. For example, Bcl-x pre-mRNA produces two splicing isoforms with two alternative 5' splice sites to encode proteins of opposite functions. The long splicing isoform Bcl-xL is a potent apoptosis inhibitor expressed in long-lived postmitotic cells and is up-regulated in many cancer cells, protecting cells against apoptotic signals. The short isoform Bcl-xS is a pro-apoptotic isoform and expressed at high levels in cells with a high turnover rate (e.g., developing lymphocytes). The ratio of the two Bcl-x splicing isoforms is regulated by multiple cώ-elements that are

located in either the core exon region or the exon extension region (i.e., between the two alternative 5' splice sites). For more examples, see WO2010075303, which is hereby incorporated by reference in its entirety.

[00108] Further suitable fusion partners include, but are not limited to, proteins (or fragments thereof) that are boundary elements (e.g., CTCF), proteins and fragments thereof that provide periphery recruitment (e.g., Lamin A, Lamin B, etc.), protein docking elements (e.g., FKBP/FRB, Pil1/Aby1, etc.).

[00109] Examples of various additional suitable heterologous polypeptide (or fragments thereof) for a subject chimeric CasZ polypeptide include, but are not limited to those described in the following applications (which publications are related to other CRISPR endonucleases such as Cas9, but the described fusion partners can also be used with CasZ instead): PCT patent applications: WO2010075303, WO2012068627, and WO2013155555, and can be found, for example, in U.S. patents and patent applications: 8,906,616; 8,895,308; 8,889,418; 8,889,356; 8,871,445; 8,865,406; 8,795,965; 8,771,945; 8,697,359; 20140068797; 20140170753; 20140179006; 20140179770; 20140186843; 20140186919; 20140186958; 20140189896; 20140227787; 20140234972; 20140242664; 20140242699; 20140242700; 20140242702; 20140248702; 20140256046; 20140273037; 20140273226; 20140273230; 20140273231; 20140273232; 20140273233; 20140273234; 20140273235; 20140287938; 20140295556; 20140295557; 20140298547; 20140304853; 20140309487; 20140310828; 20140310830; 20140315985; 20140335063; 20140335620; 20140342456; 20140342457; 20140342458; 20140349400; 20140349405; 20140356867; 20140356956; 20140356958; 20140356959; 20140357523; 20140357530; 20140364333; and 20140377868; all of which are hereby incorporated by reference in their entirety.

[00110] In some cases, a heterologous polypeptide (a fusion partner) provides for subcellular localization, i.e., the heterologous polypeptide contains a subcellular localization sequence (e.g., a nuclear localization signal (NLS) for targeting to the nucleus, a sequence to keep the fusion protein out of the nucleus, e.g., a nuclear export sequence (NES), a sequence to keep the fusion protein retained in the cytoplasm, a mitochondrial localization signal for targeting to the mitochondria, a chloroplast localization signal for targeting to a chloroplast, an ER retention signal, and the like). In some embodiments, a CasZ fusion polypeptide does not include a NLS so that the protein is not targeted to the nucleus (which can be advantageous, e.g., when the target nucleic acid is an RNA that is present in the cyosol). In some embodiments, the heterologous polypeptide can provide a tag (i.e., the heterologous polypeptide is a detectable label) for ease of tracking and/or purification (e.g., a fluorescent protein, e.g., green fluorescent protein (GFP), YFP, RFP, CFP, mCherry, tdTomato, and the like; a histidine tag, e.g., a 6XHis tag; a hemagglutinin (HA) tag; a FLAG tag; a Myc tag; and the like).

[00111] In some cases a CasZ protein (e.g., a wild type CasZ protein, a variant CasZ protein, a chimeric CasZ protein, a dCasZ protein, a chimeric CasZ protein where the CasZ portion has reduced nuclease activity - such as a dCasZ protein fused to a fusion partner, and the like) includes (is fused to) a

nuclear localization signal (NLS) (e.g, in some cases 2 or more, 3 or more, 4 or more, or 5 or more NLSs). Thus, in some cases, a CasZ polypeptide includes one or more NLSs (e.g., 2 or more, 3 or more, 4 or more, or 5 or more NLSs). In some cases, one or more NLSs (2 or more, 3 or more, 4 or more, or 5 or more NLSs) are positioned at or near (e.g., within 50 amino acids of) the N-terminus and/or the C-terminus. In some cases, one or more NLSs (2 or more, 3 or more, 4 or more, or 5 or more NLSs) are positioned at or near (e.g., within 50 amino acids of) the N-terminus. In some cases, one or more NLSs (2 or more, 3 or more, 4 or more, or 5 or more NLSs) are positioned at or near (e.g., within 50 amino acids of) the C-terminus. In some cases, one or more NLSs (3 or more, 4 or more, or 5 or more NLSs) are positioned at or near (e.g., within 50 amino acids of) both the N-terminus and the C-terminus. In some cases, an NLS is positioned at the N-terminus and an NLS is positioned at the C-terminus.

[00112] In some cases a CasZ protein (e.g., a wild type CasZ protein, a variant CasZ protein, a chimeric CasZ protein, a dCasZ protein, a chimeric CasZ protein where the CasZ portion has reduced nuclease activity - such as a dCasZ protein fused to a fusion partner, and the like) includes (is fused to) between 1 and 10 NLSs (e.g., 1-9, 1-8, 1-7, 1-6, 1-5, 2-10, 2-9, 2-8, 2-7, 2-6, or 2-5 NLSs). In some cases a CasZ protein (e.g., a wild type CasZ protein, a variant CasZ protein, a chimeric CasZ protein, a dCasZ protein, a chimeric CasZ protein where the CasZ portion has reduced nuclease activity - such as a dCasZ protein fused to a fusion partner, and the like) includes (is fused to) between 2 and 5 NLSs (e.g., 2-4, or 2-3 NLSs).

[00113] Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV (SEQ ID NO: 114); the NLS from nucleoplasmin (e.g., the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK (SEQ ID NO: 115)); the c-myc NLS having the amino acid sequence PAAKRVKLD (SEQ ID NO: 116) or RQRRNELKRSP (SEQ ID NO: 117); the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 118); the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO: 119) of the IBB domain from importin-alpha; the sequences VSRKRPRP (SEQ ID NO: 120) and PPKKARED (SEQ ID NO: 121) of the myoma T protein; the sequence PQPKKKPL (SEQ ID NO: 122) of human p53; the sequence SALIKKKKKMAP (SEQ ID NO: 123) of mouse c-abl IV; the sequences DRLRR (SEQ ID NO: 124) and PKQKKRK (SEQ ID NO: 125) of the influenza virus NS1; the sequence RKLKKKIKKL (SEQ ID NO: 126) of the Hepatitis virus delta antigen; the sequence REKKKFLKRR (SEQ ID NO: 127) of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 128) of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK (SEQ ID NO: 129) of the steroid hormone receptors (human) glucocorticoid. In general, NLS (or multiple NLSs) are of sufficient strength to drive accumulation of the CasZ protein in a detectable amount in the nucleus of a eukaryotic cell. Detection of accumulation in the

nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the CasZ protein such that location within a cell may be visualized. Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly.

[00114] In some cases, a CasZ fusion polypeptide includes a "Protein Transduction Domain" or PTD (also known as a CPP – cell penetrating peptide), which refers to a polypeptide, polynucleotide, carbohydrate, or organic or inorganic compound that facilitates traversing a lipid bilayer, micelle, cell membrane, organelle membrane, or vesicle membrane. A PTD attached to another molecule, which can range from a small polar molecule to a large macromolecule and/or a nanoparticle, facilitates the molecule traversing a membrane, for example going from extracellular space to intracellular space, or cytosol to within an organelle. In some embodiments, a PTD is covalently linked to the amino terminus a polypeptide (e.g., linked to a wild type CasZ to generate a fusino protein, or linked to a variant CasZ protein such as a dCasZ, nickase CasZ, or chimeric CasZ protein to generate a fusion protein). In some embodiments, a PTD is covalently linked to the carboxyl terminus of a polypeptide (e.g., linked to a wild type CasZ to generate a fusino protein, or linked to a variant CasZ protein such as a dCasZ, nickase CasZ, or chimeric CasZ protein to generate a fusion protein). In some cases, the PTD is inserted interally in the CasZ fusion polypeptide (i.e., is not at the N- or C-terminus of the CasZ fusion polypeptide) at a suitable insertion site. In some cases, a subject CasZ fusion polypeptide includes (is conjugated to, is fused to) one or more PTDs (e.g., two or more, three or more, four or more PTDs). In some cases, a PTD includes a nuclear localization signal (NLS) (e.g, in some cases 2 or more, 3 or more, 4 or more, or 5 or more NLSs). Thus, in some cases, a CasZ fusion polypeptide includes one or more NLSs (e.g., 2 or more, 3 or more, 4 or more, or 5 or more NLSs). In some cases, a PTD is covalently linked to a nucleic acid (e.g., a CasZ guide nucleic acid, a polynucleotide encoding a CasZ guide nucleic acid, a polynucleotide encoding a CasZ fusion polypeptide, a donor polynucleotide, etc.). Examples of PTDs include but are not limited to a minimal undecapeptide protein transduction domain (corresponding to residues 47-57 of HIV-1 TAT comprising YGRKKRRQRRR; SEQ ID NO: 130); a polyarginine sequence comprising a number of arginines sufficient to direct entry into a cell (e.g., 3, 4, 5, 6, 7, 8, 9, 10, or 10-50 arginines); a VP22 domain (Zender et al. (2002) *Cancer Gene Ther.* 9(6):489-96); an Drosophila Antennapedia protein transduction domain (Noguchi et al. (2003) *Diabetes* 52(7):1732-1737); a truncated human calcitonin peptide (Trehin et al. (2004) *Pharm. Research* 21:1248-1256); polylysine (Wender et al. (2000) *Proc. Natl. Acad. Sci. USA* 97:13003-13008); RRQRRTSKLMKR (SEQ ID NO: 131); Transportan GWTLNSAGYLLGKINLKALAALAKKIL (SEQ ID NO: 132); KALAWEAKLAKALAKALAKHLAKALAKALKCEA (SEQ ID NO: 133); and RQIKIWFQNRRMKWKK (SEQ ID NO: 134). Exemplary PTDs include but are not limited to,

YGRKKRRQRRR (SEQ ID NO: 130), RKKRRQRRR (SEQ ID NO: 135); an arginine homopolymer of from 3 arginine residues to 50 arginine residues; Exemplary PTD domain amino acid sequences include, but are not limited to, any of the following: YGRKKRRQRRR (SEQ ID NO: 130); RKKRRQRR (SEQ ID NO: 136); YARAAARQARA (SEQ ID NO: 137); THRLPRRRRRR (SEQ ID NO: 138); and GGRRARRRRRR (SEQ ID NO: 139). In some embodiments, the PTD is an activatable CPP (ACPP) (Aguilera et al. (2009) *Integr Biol (Camb)* June; 1(5-6): 371-381). ACPPs comprise a polycationic CPP (e.g., Arg9 or "R9") connected via a cleavable linker to a matching polyanion (e.g., Glu9 or "E9"), which reduces the net charge to nearly zero and thereby inhibits adhesion and uptake into cells. Upon cleavage of the linker, the polyanion is released, locally unmasking the polyarginine and its inherent adhesiveness, thus "activating" the ACPP to traverse the membrane.

Linkers (e.g., for fusion partners)

[00115] In some instances, a subject CasZ protein is fused to a fusion partner via a linker polypeptide (e.g., one or more linker polypeptides). The linker polypeptide may have any of a variety of amino acid sequences. Proteins can be joined by a spacer peptide, generally of a flexible nature, although other chemical linkages are not excluded. Suitable linkers include polypeptides of between 4 amino acids and 40 amino acids in length, or between 4 amino acids and 25 amino acids in length. These linkers can be produced by using synthetic, linker-encoding oligonucleotides to couple the proteins, or can be encoded by a nucleic acid sequence encoding the fusion protein. Peptide linkers with a degree of flexibility can be used. The linking peptides may have virtually any amino acid sequence, bearing in mind that the preferred linkers will have a sequence that results in a generally flexible peptide. The use of small amino acids, such as glycine and alanine, are of use in creating a flexible peptide. The creation of such sequences is routine to those of skill in the art. A variety of different linkers are commercially available and are considered suitable for use.

[00116] Examples of linker polypeptides include glycine polymers $(G)_n$, glycine-serine polymers (including, for example, $(GS)_n$, $GSGGS_n$ (SEQ ID NO: 140), $GGSGGS_n$ (SEQ ID NO: 141), and $GGGS_n$ (SEQ ID NO: 142), where n is an integer of at least one), glycine-alanine polymers, alanine-serine polymers. Exemplary linkers can comprise amino acid sequences including, but not limited to, GGSG (SEQ ID NO: 143), GGSGG (SEQ ID NO: 144), GSGSG (SEQ ID NO: 145), GSGGG (SEQ ID NO: 146), GGGSG (SEQ ID NO: 147), GSSSG (SEQ ID NO: 148), and the like. The ordinarily skilled artisan will recognize that design of a peptide conjugated to any desired element can include linkers that are all or partially flexible, such that the linker can include a flexible linker as well as one or more portions that confer less flexible structure.

Detectable labels

[00117] In some cases, a CasZ polypeptide of the present disclosure comprises (e.g., can be attached/fused to) a detectable label. Suitable detectable labels and/or moieties that can provide a

detectable signal can include, but are not limited to, an enzyme, a radioisotope, a member of a specific binding pair; a fluorophore; a fluorescent protein; a quantum dot; and the like.

[00118] Suitable fluorescent proteins include, but are not limited to, green fluorescent protein (GFP) or variants thereof, blue fluorescent variant of GFP (BFP), cyan fluorescent variant of GFP (CFP), yellow fluorescent variant of GFP (YFP), enhanced GFP (EGFP), enhanced CFP (ECFP), enhanced YFP (EYFP), GFPS65T, Emerald, Topaz (TYFP), Venus, Citrine, mCitrine, GFPuv, destabilised EGFP (dEGFP), destabilised ECFP (dECFP), destabilised EYFP (dEYFP), mCFPm, Cerulean, T-Sapphire, CyPet, YPet, mKO, HcRed, t-HcRed, DsRed, DsRed2, DsRed-monomer, J-Red, dimer2, t-dimer2(12), mRFP1, pocilloporin, Renilla GFP, Monster GFP, paGFP, Kaede protein and kindling protein, Phycobiliproteins and Phycobiliprotein conjugates including B-Phycoerythrin, R-Phycoerythrin and Allophycocyanin. Other examples of fluorescent proteins include mHoneydew, mBanana, mOrange, dTomato, tdTomato, mTangerine, mStrawberry, mCherry, mGrape1, mRaspberry, mGrape2, mPlum (Shaner et al. (2005) *Nat. Methods* 2:905-909), and the like. Any of a variety of fluorescent and colored proteins from Anthozoan species, as described in, e.g., Matz et al. (1999) *Nature Biotechnol.* 17:969-973, is suitable for use.

[00119] Suitable enzymes include, but are not limited to, horse radish peroxidase (HRP), alkaline phosphatase (AP), beta-galactosidase (GAL), glucose-6-phosphate dehydrogenase, beta-N-acetylglucosaminidase, β-glucuronidase, invertase, Xanthine Oxidase, firefly luciferase, glucose oxidase (GO), and the like.

Protospacer Adjacent Motif (PAM)

[00120] A natural CasZ protein binds to target DNA at a target sequence defined by the region of complementarity between the DNA-targeting RNA and the target DNA. As is the case for many CRISPR endonucleases, site-specific binding (and/or cleavage) of a double stranded target DNA occurs at locations determined by both (i) base-pairing complementarity between the guide RNA and the target DNA; and (ii) a short motif [referred to as the protospacer adjacent motif (PAM)] in the target DNA.

[00121] In some cases, the PAM for a CasZ protein is immediately 5' of the target sequence of the non-complementary strand of the target DNA (also referred to as the non-target strand; the complementary strand hybridizes to the guide sequence of the guide RNA while the non-complementary strand does not directly hybridize with the guide RNA and is the reverse complement of the non-complementary strand). In some cases (e.g., for CasZc), the PAM sequence of the non-complementary strand is 5'-TTA-3'. In some cases (e.g., for CasZb), the PAM sequence of the non-complementary strand is 5'-TTTN-3'. In some cases (e.g., for CasZb), the PAM sequence of the non-complementary strand is 5'-TTTA-3'.

[00122] In some cases, different CasZ proteins (i.e., CasZ proteins from various species) may be advantageous to use in the various provided methods in order to capitalize on various enzymatic characteristics of the different CasZ proteins (e.g., for different PAM sequence preferences; for increased or decreased enzymatic activity; for an increased or decreased level of cellular toxicity; to change the balance between NHEJ, homology-directed repair, single strand breaks, double strand breaks, etc.; to take advantage of a short total sequence; and the like). CasZ proteins from different species may require different PAM sequences in the target DNA. Thus, for a particular CasZ protein of choice, the PAM sequence preference may be different than the sequence(s) described above. Various methods (including in silico and/or wet lab methods) for identification of the appropriate PAM sequence are known in the art and are routine, and any convenient method can be used.

CasZ Guide RNA

[00123] A nucleic acid molecule that binds to a CasZ protein, forming a ribonucleoprotein complex (RNP), and targets the complex to a specific location within a target nucleic acid (e.g., a target DNA) is referred to herein as a "CasZ guide RNA" or simply as a "guide RNA." It is to be understood that in some cases, a hybrid DNA/RNA can be made such that a CasZ guide RNA includes DNA bases in addition to RNA bases, but the term "CasZ guide RNA" is still used to encompass such a molecule herein.

[00124] A CasZ guide RNA can be said to include two segments, a targeting segment and a protein-binding segment. The targeting segment of a CasZ guide RNA includes a nucleotide sequence (a guide sequence) that is complementary to (and therefore hybridizes with) a specific sequence (a target site) within a target nucleic acid (e.g., a target ssRNA, a target ssDNA, the complementary strand of a double stranded target DNA, etc.). The protein-binding segment (or "protein-binding sequence") interacts with (binds to) a CasZ polypeptide. The protein-binding segment of a subject CasZ guide RNA includes two complementary stretches of nucleotides that hybridize to one another to form a double stranded RNA duplex (dsRNA duplex). Site-specific binding and/or cleavage of a target nucleic acid (e.g., genomic DNA) can occur at locations (e.g., target sequence of a target locus) determined by base-pairing complementarity between the CasZ guide RNA (the guide sequence of the CasZ guide RNA) and the target nucleic acid.

[00125] A CasZ guide RNA and a CasZ protein, e.g., a fusion CasZ polypeptide, form a complex (e.g., bind via non-covalent interactions). The CasZ guide RNA provides target specificity to the complex by including a targeting segment, which includes a guide sequence (a nucleotide sequence that is complementary to a sequence of a target nucleic acid). The CasZ protein of the complex provides the site-specific activity (e.g., cleavage activity provided by the CasZ protein and/or an activity provided by the fusion partner in the case of a chimeric CasZ protein). In other words, the CasZ protein is guided to a

target nucleic acid sequence (e.g. a target sequence) by virtue of its association with the CasZ guide RNA.

[00126] The "guide sequence" also referred to as the "targeting sequence" of a CasZ guide RNA can be modified so that the CasZ guide RNA can target a CasZ protein (e.g., a naturally occurring CasZ protein, a fusion CasZ polypeptide (chimeric CasZ), and the like) to any desired sequence of any desired target nucleic acid, with the exception (e.g., as described herein) that the PAM sequence can be taken into account. Thus, for example, a CasZ guide RNA can have a guide sequence with complementarity to (e.g., can hybridize to) a sequence in a nucleic acid in a eukaryotic cell, e.g., a viral nucleic acid, a eukaryotic nucleic acid (e.g., a eukaryotic chromosome, chromosomal sequence, a eukaryotic RNA, etc.), and the like.

[00127] In some cases, a CasZ guide RNA has a length of 30 nucleotides (nt) or more (e.g., 35 nt or more, 40 nt or more, 45 nt or more, 50 nt or more, 55 nt or more, or 60 nt or more). In some embodiments, a CasZ guide RNA has a length of 40 nucleotides (nt) or more (e.g., 45 nt or more, 50 nt or more, 55 nt or more, or 60 nt or more). In some cases, a CasZ guide RNA has a length of from 30 nucleotides (nt) to 100 nt (e.g., 30-90, 30-80, 30-75, 30-70, 30-65, 40-100, 40-90, 40-80, 40-75, 40-70, or 40-65 nt). In some cases, a CasZ guide RNA has a length of from 40 nucleotides (nt) to 100 nt (e.g., 40-90, 40-80, 40-75, 40-70, or 40-65 nt).

Guide sequence of a CasZ guide RNA

[00128] A subject CasZ guide RNA includes a guide sequence (i.e., a targeting sequence), which is a nucleotide sequence that is complementary to a sequence (a target site) in a target nucleic acid. In other words, the guide sequence of a CasZ guide RNA can interact with a target nucleic acid (e.g., double stranded DNA (dsDNA), single stranded DNA (ssDNA), single stranded RNA (ssRNA), or double stranded RNA (dsRNA)) in a sequence-specific manner via hybridization (i.e., base pairing). The guide sequence of a CasZ guide RNA can be modified (e.g., by genetic engineering)/designed to hybridize to any desired target sequence (e.g., while taking the PAM into account, e.g., when targeting a dsDNA target) within a target nucleic acid (e.g., a eukaryotic target nucleic acid such as genomic DNA).

[00129] In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 60% or more (e.g., 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100%). In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 80% or more (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100%). In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 90% or more (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100%). In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 100%.

[00130] In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 100% over the seven contiguous 3'-most nucleotides of the target site of the target nucleic acid.

[00131] In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 60% or more (e.g., 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100%) over 17 or more (e.g., 18 or more, 19 or more, 20 or more, 21 or more, 22 or more) contiguous nucleotides. In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 80% or more (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100%) over 17 or more (e.g., 18 or more, 19 or more, 20 or more, 21 or more, 22 or more) contiguous nucleotides. In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 90% or more (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100%) over 17 or more (e.g., 18 or more, 19 or more, 20 or more, 21 or more, 22 or more) contiguous nucleotides. In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 100% over 17 or more (e.g., 18 or more, 19 or more, 20 or more, 21 or more, 22 or more) contiguous nucleotides.

[00132] In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 60% or more (e.g., 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100%) over 19 or more (e.g., 20 or more, 21 or more, 22 or more) contiguous nucleotides. In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 80% or more (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100%) over 19 or more (e.g., 20 or more, 21 or more, 22 or more) contiguous nucleotides. In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 90% or more (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100%) over 19 or more (e.g., 20 or more, 21 or more, 22 or more) contiguous nucleotides. In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 100% over 19 or more (e.g., 20 or more, 21 or more, 22 or more) contiguous nucleotides.

[00133] In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 60% or more (e.g., 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100%) over 17-25 contiguous nucleotides. In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 80% or more (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100%) over 17-25 contiguous nucleotides. In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 90%

or more (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100%) over 17-25 contiguous nucleotides. In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 100% over 17-25 contiguous nucleotides.

[00134] In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 60% or more (e.g., 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100%) over 19-25 contiguous nucleotides. In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 80% or more (e.g., 85% or more, 90% or more, 95% or more, 97% or more, 98% or more, 99% or more, or 100%) over 19-25 contiguous nucleotides. In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 90% or more (e.g., 95% or more, 97% or more, 98% or more, 99% or more, or 100%) over 19-25 contiguous nucleotides. In some cases, the percent complementarity between the guide sequence and the target site of the target nucleic acid is 100% over 19-25 contiguous nucleotides.

[00135] In some cases, the guide sequence has a length in a range of from 17-30 nucleotides (nt) (e.g., from 17-25, 17-22, 17-20, 19-30, 19-25, 19-22, 19-20, 20-30, 20-25, or 20-22 nt). In some cases, the guide sequence has a length in a range of from 17-25 nucleotides (nt) (e.g., from 17-22, 17-20, 19-25, 19-22, 19-20, 20-25, or 20-22 nt). In some cases, the guide sequence has a length of 17 or more nt (e.g., 18 or more, 19 or more, 20 or more, 21 or more, or 22 or more nt; 19 nt, 20 nt, 21 nt, 22 nt, 23 nt, 24 nt, 25 nt, etc.). In some cases, the guide sequence has a length of 19 or more nt (e.g., 20 or more, 21 or more, or 22 or more nt; 19 nt, 20 nt, 21 nt, 22 nt, 23 nt, 24 nt, 25 nt, etc.). In some cases, the guide sequence has a length of 17 nt. In some cases, the guide sequence has a length of 18 nt. In some cases, the guide sequence has a length of 19 nt. In some cases, the guide sequence has a length of 20 nt. In some cases, the guide sequence has a length of 21 nt. In some cases, the guide sequence has a length of 22 nt. In some cases, the guide sequence has a length of 23 nt.

Protein-binding segment of a CasZ guide RNA

[00136] The protein-binding segment of a subject CasZ guide RNA interacts with a CasZ protein. The CasZ guide RNA guides the bound CasZ protein to a specific nucleotide sequence within target nucleic acid via the above-mentioned guide sequence. The protein-binding segment of a CasZ guide RNA comprises two stretches of nucleotides that are complementary to one another and hybridize to form a double stranded RNA duplex (dsRNA duplex). Thus, the protein-binding segment includes a dsRNA duplex.

[00137] In some cases, the dsRNA duplex region includes a range of from 5-25 base pairs (bp) (e.g., from 5-22, 5-20, 5-18, 5-15, 5-12, 5-10, 5-8, 8-25, 8-22, 8-18, 8-15, 8-12, 12-25, 12-22, 12-18, 12-15, 13-25, 13-22, 13-18, 13-15, 14-25, 14-22, 14-18, 14-15, 15-25, 15-22, 15-18, 17-25, 17-22, or 17-18 bp, e.g., 5 bp, 6 bp, 7 bp, 8 bp, 9 bp, 10 bp, etc.). In some cases, the dsRNA duplex region includes a

range of from 6-15 base pairs (bp) (e.g., from 6-12, 6-10, or 6-8 bp, e.g., 6 bp, 7 bp, 8 bp, 9 bp, 10 bp, etc.). In some cases, the duplex region includes 5 or more bp (e.g., 6 or more, 7 or more, or 8 or more bp). In some cases, the duplex region includes 6 or more bp (e.g., 7 or more, or 8 or more bp). In some cases, not all nucleotides of the duplex region are paired, and therefore the duplex forming region can include a bulge. The term "bulge" herein is used to mean a stretch of nucleotides (which can be one nucleotide or multiple nucleotides) that do not contribute to a double stranded duplex, but which are surround 5' and 3' by nucleotides that do contribute, and as such a bulge is considered part of the duplex region. In some cases, the dsRNA includes 1 or more bulges (e.g., 2 or more, 3 or more, 4 or more bulges). In some cases, the dsRNA duplex includes 2 or more bulges (e.g., 3 or more, 4 or more bulges). In some cases, the dsRNA duplex includes 1-5 bulges (e.g., 1-4, 1-3, 2-5, 2-4, or 2-3 bulges).

[00138] Thus, in some cases, the stretches of nucleotides that hybridize to one another to form the dsRNA duplex have 70%-100% complementarity (e.g., 75%-100%, 80%-10%, 85%-100%, 90%-100%, 95%-100% complementarity) with one another. In some cases, the stretches of nucleotides that hybridize to one another to form the dsRNA duplex have 70%-100% complementarity (e.g., 75%-100%, 80%-10%, 85%-100%, 90%-100%, 95%-100% complementarity) with one another. In some cases, the stretches of nucleotides that hybridize to one another to form the dsRNA duplex have 85%-100% complementarity (e.g., 90%-100%, 95%-100% complementarity) with one another. In some cases, the stretches of nucleotides that hybridize to one another to form the dsRNA duplex have 70%-95% complementarity (e.g., 75%-95%, 80%-95%, 85%-95%, 90%-95% complementarity) with one another.

[00139] In other words, in some embodiments, the dsRNA duplex includes two stretches of nucleotides that have 70%-100% complementarity (e.g., 75%-100%, 80%-10%, 85%-100%, 90%-100%, 95%-100% complementarity) with one another. In some cases, the dsRNA duplex includes two stretches of nucleotides that have 85%-100% complementarity (e.g., 90%-100%, 95%-100% complementarity) with one another. In some cases, the dsRNA duplex includes two stretches of nucleotides that have 70%-95% complementarity (e.g., 75%-95%, 80%-95%, 85%-95%, 90%-95% complementarity) with one another.

[00140] The duplex region of a subject CasZ guide RNA can include one or more (1, 2, 3, 4, 5, etc) mutations relative to a naturally occurring duplex region. For example, in some cases a base pair can be maintained while the nucleotides contributing to the base pair from each segment can be different. In some cases, the duplex region of a subject CasZ guide RNA includes more paired bases, less paired bases, a smaller bulge, a larger bulge, fewer bulges, more bulges, or any convenient combination thereof, as compared to a naturally occurring duplex region (of a naturally occurring CasZ guide RNA).

[00141] Examples of various Cas9 guide RNAs and cpf1 guide RNAs can be found in the art, and in some cases variations similar to those introduced into Cas9 guide RNAs can also be introduced into CasZ guide RNAs of the present disclosure (e.g., mutations to the dsRNA duplex region, extension

of the 5' or 3' end for added stability for to provide for interaction with another protein, and the like). For example, see Jinek et al., Science. 2012 Aug 17;337(6096):816-21; Chylinski et al., RNA Biol. 2013 May;10(5):726-37; Ma et al., Biomed Res Int. 2013;2013:270805; Hou et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15644-9; Jinek et al., Elife. 2013;2:e00471; Pattanayak et al., Nat Biotechnol. 2013 Sep;31(9):839-43; Qi et al, Cell. 2013 Feb 28;152(5):1173-83; Wang et al., Cell. 2013 May 9;153(4):910-8; Auer et al., Genome Res. 2013 Oct 31; Chen et al., Nucleic Acids Res. 2013 Nov 1;41(20):e19; Cheng et al., Cell Res. 2013 Oct;23(10):1163-71; Cho et al., Genetics. 2013 Nov;195(3):1177-80; DiCarlo et al., Nucleic Acids Res. 2013 Apr;41(7):4336-43; Dickinson et al., Nat Methods. 2013 Oct;10(10):1028-34; Ebina et al., Sci Rep. 2013;3:2510; Fujii et. al, Nucleic Acids Res. 2013 Nov 1;41(20):e187; Hu et al., Cell Res. 2013 Nov;23(11):1322-5; Jiang et al., Nucleic Acids Res. 2013 Nov 1;41(20):e188; Larson et al., Nat Protoc. 2013 Nov;8(11):2180-96; Mali et. at., Nat Methods. 2013 Oct;10(10):957-63; Nakayama et al., Genesis. 2013 Dec;51(12):835-43; Ran et al., Nat Protoc. 2013 Nov;8(11):2281-308; Ran et al., Cell. 2013 Sep 12;154(6):1380-9; Upadhyay et al., G3 (Bethesda). 2013 Dec 9;3(12):2233-8; Walsh et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15514-5; Xie et al., Mol Plant. 2013 Oct 9; Yang et al., Cell. 2013 Sep 12;154(6):1370-9; Briner et al., Mol Cell. 2014 Oct 23;56(2):333-9; and U.S. patents and patent applications: 8,906,616; 8,895,308; 8,889,418; 8,889,356; 8,871,445; 8,865,406; 8,795,965; 8,771,945; 8,697,359; 20140068797; 20140170753; 20140179006; 20140179770; 20140186843; 20140186919; 20140186958; 20140189896; 20140227787; 20140234972; 20140242664; 20140242699; 20140242700; 20140242702; 20140248702; 20140256046; 20140273037; 20140273226; 20140273230; 20140273231; 20140273232; 20140273233; 20140273234; 20140273235; 20140287938; 20140295556; 20140295557; 20140298547; 20140304853; 20140309487; 20140310828; 20140310830; 20140315985; 20140335063; 20140335620; 20140342456; 20140342457; 20140342458; 20140349400; 20140349405; 20140356867; 20140356956; 20140356958; 20140356959; 20140357523; 20140357530; 20140364333; and 20140377868; all of which are hereby incorporated by reference in their entirety.

[00142] A CasZ guide RNA comprises both the guide sequence and two stretches ("duplex-forming segments") of nucleotides that hybridize to form the dsRNA duplex of the protein-binding segment. The particular sequence of a given CasZ guide RNA can be characteristic of the species in which a crRNA is found. Examples of suitable CasZ guide RNAs are provided herein.

Example Guide RNA sequences

[00143] Repeat sequences (non-guide sequence portion of a CasZ guide RNA) of crRNAs for naturally existing CasZ proteins (e.g., see Figure 1 and Figure 7) are shown in Table 1 and Table 3.

[00144] **Table 1.** crRNA repeat sequences for CasZ proteins

CasZ Protein	Repeat sequence	SEQ ID NO:
Za.1	GTTGCATTCCTTCATTCGTCTATTCGGGTTCTGCAAC	51
Za.2	GTTGCATTCCTTCATTCGTCTATCCGGGTTCTGCAAG	52
Za.3	GTTGCAGAACCCGAATAGACGAATGAAGGAATGCAAC	53
Za.4	CTATCATATTCAGAACAAAGGGATTAAGGAATGCAAC	54
Za.5	CTTTCATACTCAGAACAAAGGGATTAAGGAATGCAAC	55
Za.6	GTCTACAACTCATTGATAGAAATCAATGAGTTAGACA	56
Za.7	GTTATAAAGGCGGGGATCGCGACCGAGCGATTGAAAG	57
Zb.1	GTTGCATTCCTTAATTCATTTTCTCAATATCGGAAAC	58
Zb.2	GTTGCAGAAATAGAATAAAGGAATTAAGGAATGCAAC	59
Zb.3	CTTTCATACTCAGAACAAAGGGATTAAGGAATGCAAC	55
Zb.4	ATTTCATACTCAGAACAAAGGGATTAAGGAATGCAAC	61
Zb.5	GTTTCAGCGCACGAATTAACGAGATGAGAGATGCAACT	62
Zb.6	CTTGCAGAAGCTGAATAGACGAATCAAGGAATGCAAC	63
Zb.7	CACTTGCAGGCCTTGAATAGAGGAGTTAAGGAATGCAAC	64
Zb.8	GTCTCCATGACTGAAAAGTCGTGGCCGAATTGAAAC	65
Zb.9	GTTGCAGCGCCCGAACTGACGAGACGAGAGATGCAAC	66
Zb.10	GTTGCGCGAATAGAATAAAGGAATTAAGGAATGCAAC	67
Zb.11	AGTTGCATTCCTTAATCCCTCTGTTCAGTTTGTGCAAT	68
Zc.1	GTTGCATTCCTAGTTTCTCTAATTAGCACTGTGCAAC	69
Zc.2	GTTGCGGCGCGCGAATAAACGAGACTAGGAATGCAAC	70
Zc.3	ACTAGTTGCATTCCTTAATCCCTTTGTTCTGAATATGCTAG	71
Zc.4	CTTTCATATTCAGAACAAAGGGATTAAGGAATGCAAC	72
Zc.5	GTTGCAGTCCTTAACCCCTAGTTTCTGAATATGAAAGAT	73
Zc.6	GTTGCAGCCCCCGAACTAACGAGATGAGAGATGCAAC	74
Zc.7	CTTGCAGAACAATCATATATGACTAATCAGACTGCAAC	75
Zd.1	GTTGCACTCACCGGTGCTCACGACGTAGGGATGCAAC	76
Zd.2	GTCCCTACTCGCTAGGGAAACTAATTGAATGGAAAC	77
Ze.1	GTTGCATTCGGGTGCAAAACAGGGAGTAGAGTGTAAC	78
Ze.2	CTTCCAAACTCGAGCCAGTGGGGAGAGAAGTGGCA	79
Ze.3	CCTGTAGACCGGTCTCATTCTGAGAGGGGTATGCAACT	80
Ze.4	GTCTCGAGACCCTACAGATTTTGGAGAGGGGTGGGAC	81
Ze.4b	GTCCCACCCCTCTCCAAAATCTGTAGGGTCTCGAGAC	82
Zf.1	GTAGCAGGACTCTCCTCGAGAGAAACAGGGGTATGCT	83
Zf.2	GTACAATACCTCTCCTTTAAGAGAGGGAGGGGTACGCTAC	84
Zf.3	CCCCCTCGTTTCCTTCAGGGGATTCCTTTCC	85
Zg.1	GGTTCCCCCGGGCGCGGGTGGGGTGGCG	86
Zg.2	GGCTGCTCCGGGTGCGCGTGGAGCGAGG	87

CasZ Protein	Repeat sequence	SEQ ID NO:
Zh.1	GTTTTATACCCTTTAGAATTTAAACTGTCTAAAAG	88
Zi.1	ATTGCACCGGCCAACGCAAATCTGATTGATGGACAC	89
Zi.2	GCCGCAGCGGCCGACGCGGCCCTGATCGATGGACAC	90
Zj.1	GTCGAAATGCCCGCGCGGGGGCGTCGTACCCGCGAC	91
Zk.1	GGCTAGCCCGTGCGCGCAGGGACGAGTGG	92
Zk.2	GCCCGTGCGCGCAGGGACGAGTGG	93
Zk.3	GTTGCAGCGGCCGACGGAGCGCGAGCGTGGATGCCAC	94
Zk.4	CCATCGCCCCGCGCGCACGTGGATGAGCC	95
Zl.1	CTTTAGACTTCTCCGGAAGTCGAATTAATGGAAAC	96
Zl.2	GGGCGCCCCGCGCGAGCGGGGGTTGAAG	97
Za.8	CTTGCAGAACCCGGATAGACGAATGAAGGAATGCAAC	295
Zb.12	CTTGCAGGCCTTGAATAGAGGAGTTAAGGAATGCAAC	296
Zb.13	GTTGCACAGTGCTAATTAGAGAAACTAGGAATGCAAC	297
Zb.14	CTAGCATATTCAGAACAAAGGGATTAAGGAATGCAAC	298
Zb.15	CTTTCATATTCAGAAACTAGGGGTTAAGGACTGCAAC	299
Zc.8	GTTGCATCCCTACGTCGTGAGCACCGGTGAGTGCAAC	300
Ze.5	GGAAAGGAATCCCCTGAAGGAAACGAGGGGG	301
Zg.3	GTGTCCATCAATCAGATTTGCGTTGGCCGGTGCAAT	302
Zb.16	GTTTCAGCGCACGAATTAACGAGATGAGAGATGCAAC	303
Zj.2	CTTTTAGACAGTTTAAATTCTAAAGGGTATAAAAC	307

[00145] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a crRNA sequence of Table 1 or Table 3.

[00146] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, or CasZi crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity)

with a CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, or CasZi crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, or CasZi crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZa, CasZb, CasZc, CasZd, CasZe, CasZf, CasZg, CasZh, or CasZi crRNA sequence of Table 1 or Table 3.

[00147] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZa crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZa crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZa crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZa crRNA sequence of Table 1 or Table 3.

[00148] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZb crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZb crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZb crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZb crRNA sequence of Table 1 or Table 3.

[00149] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZc crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZc crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a

CasZc crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZc crRNA sequence of Table 1 or Table 3.

[00150] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZd crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZd crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZd crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZd crRNA sequence of Table 1 or Table 3.

[00151] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZe crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZe crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZe crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZe crRNA sequence of Table 1 or Table 3.

[00152] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZf crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZf crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZf crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZf crRNA sequence of Table 1 or Table 3.

[00153] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZg crRNA sequence of Table 1 or Table 3. In

some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZg crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZg crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZg crRNA sequence of Table 1 or Table 3.

[00154] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZh crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZh crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZh crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZh crRNA sequence of Table 1 or Table 3.

[00155] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZi crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZi crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZi crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZi crRNA sequence of Table 1 or Table 3.

[00156] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZj crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZj crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a

CasZj crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZj crRNA sequence of Table 1 or Table 3.

[00157] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZk crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZk crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZk crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZk crRNA sequence of Table 1 or Table 3.

[00158] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZl crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZl crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZl crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZl crRNA sequence of Table 1 or Table 3.

[00159] In some cases, a subject CasZ guide RNA comprises (e.g., in addition to a guide sequence, e.g., as part of the protein-binding region) a CasZj, CasZl, or CasZk crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZj, CasZl, or CasZk crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZj, CasZl, or CasZk crRNA sequence of Table 1 or Table 3. In some cases, a subject CasZ guide RNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZj, CasZl, or CasZk crRNA sequence of Table 1 or Table 3.

CasZ transactivating noncoding RNA (trancRNA)

[00160] Composisions and methods of the present disclosure include a CasZ transactivating noncoding RNA ("trancRNA"; also referred to herein as a "CasZ trancRNA"). In some cases, a trancRNA forms a complex with a CasZ polypeptide of the present disclosure and a CasZ guide RNA. A trancRNA can be identified as a highly transcribed RNA encoded by a nucleotide sequence present in a CasZ locus. The sequence encoding a trancRNA is usually located between the *cas* genes and the array of the CasZ locus (the repeats) (e.g., can be located adjacent to the repeat sequences). Examples below demonstrate detection of a CasZ trancRNA. In some cases, a CasZ trancRNA co-immunoprecipitates (forms a complex with) with a CasZ polypeptide. In some cases, the presence of a CasZ trancRNA is required for function of the system. Data related to trancRNAs (e.g., their expression and their location on naturally occurring arrays) is presented in the examples section below.

[00161] In some cases, a CasZ trancRNA has a length of from 60 nucleotides (nt) to 270 nt (e.g., 60-260, 70-270, 70-260, or 75-255 nt). In some cases, a CasZ trancRNA (e.g., a CasZa trancRNA) has a length of from 60-150 nt (e.g., 60-140, 60-130, 65-150, 65-140, 65-130, 70-150, 70-140, or 70-130 nt). In some cases, a CasZ trancRNA (e.g., a CasZa trancRNA) has a length of from 70-130 nt. In some cases, a CasZ trancRNA (e.g., a CasZa trancRNA) has a length of about 80 nt. In some cases, a CasZ trancRNA (e.g., a CasZa trancRNA) has a length of about 90 nt. In some cases, a CasZ trancRNA (e.g., a CasZa trancRNA) has a length of about 120 nt.

[00162] In some cases, a CasZ trancRNA (e.g., a CasZb trancRNA) has a length of from 85-240 nt (e.g., 85-230, 85-220, 85-150, 85-130, 95-240, 95-230, 95-220, 95-150, or 95-130 nt). In some cases, a CasZ trancRNA (e.g., a CasZb trancRNA) has a length of from 95-120 nt. In some cases, a CasZ trancRNA (e.g., a CasZb trancRNA) has a length of about 105 nt. In some cases, a CasZ trancRNA (e.g., a CasZb trancRNA) has a length of about 115 nt. In some cases, a CasZ trancRNA (e.g., a CasZb trancRNA) has a length of about 215 nt.

[00163] In some cases, a CasZ trancRNA (e.g., a CasZc trancRNA) has a length of from 80-275 nt (e.g., 85-260 nt). In some cases, a CasZ trancRNA (e.g., a CasZc trancRNA) has a length of from 80-110 nt (e.g., 85-105 nt). In some cases, a CasZ trancRNA (e.g., a CasZc trancRNA) has a length of from 235-270 nt (e.g., 240-260 nt). In some cases, a CasZ trancRNA (e.g., a CasZc trancRNA) has a length of about 95 nt. In some cases, a CasZ trancRNA (e.g., a CasZc trancRNA) has a length of about 250 nt.

Example trancRNA sequences

[00164] Examples of trancRNA sequences for naturally existing CasZ proteins are shown in Table 2.

[00165] **Table 2. CasZ trancRNA sequences**

CasZ Protein	trancRNA sequence	SEQ ID NO

CasZ Protein	trancRNA sequence	SEQ ID NO
Za.1	CGATTCCTCCCTACAGTAGTTAGGTATAGCCGAAAGGTAG AGACTAAATCTGTAGTTGGAGTGGGCCGCTTGCATCGGCC	151
Za.2	TCGTCTCGAGGGTTACCAAAATTGGCACTTCTCGACTTTAG GCCGATGCAAGCGGCCCACTCCACTACAGATTTAGTCTCTA CCTTGCGGCTATACCTAACTTACTGTAGGGAGGAATCGTG	152
Za.3	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGT CCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTG CATCAGCCTAA	153
Zb.2	CAGAATAATACTGACTTACTAAGATATCTTGAGGGTATACC CGAAAAGATTGGCGTTGTTGCAACGCAATAAGATGTAAAT CTGAAAAGGTTTGGAATCATATAAATAATTTTA	154
Zb.4	AAGCCAAGATATGGAATGCCATTGTAATATTATGGTGTTG ACTTAGTTTAGATTTAAACAATCTTCGATGGCTATATGCGG AAGGTTTGGCGTCGTTGTAACGC	155
Zb.6	CAGTGTGCATAGCTATAACACTACGCAAAGACTGCTAAAG AGCGATGTGCTCTATCGCAGTCTCACCTTTAATGGACTTAC GGATCTTTTGGAGCACTAAGCTCCGCTGCGGTGCAACACC GCCCTTTTCTTGCCTCTGCTTGCCCTTTCCGGTTATTATAGC CGGGAGAGTGCGGAAGATTACCGCTCTAGCTCGCAGCATG TTACTGAGTC	156
Zc.3	GCAAGTCATTCGGGGACACTTTTTGTTATTTAAAGTGTTTT AGATAAATCAGTGTCATGCTGAATAACGACCCGACCTATA AATAACATAATCC	157
Zc.5	GTCCTTAAGGTACTACACATTACATGTGAACGTGGAGCTA ATAATAGAAATATTATTAGACTACACCTTATTAATAACGGT AGGAGATCTATATGGTCTTGAATGGAATAGTAATTGTGAA ATTATAATTTCTGTTCTTAGCTACTTAAGATGGCTCGTTGC AAGCCACTCGGGGGCTCTCTTGAAGTCAAAGAGCTTTAGA CAAATCAGTGTCAAACTGAATAACGACCCGACCATGACTT CATAATCCCG	158

[00166] In some cases, a subject CasZ trancRNA comprises a CasZa trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZa trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZa trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZa trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZa trancRNA sequence above, and has a length of from 60-150 nt (e.g., 60-140, 60-130, 65-150, 65-140, 65-130, 70-150, 70-140, or 70-130 nt).

[00167] In some cases, a subject CasZ trancRNA comprises a CasZb trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZb trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZb trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZb trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZb trancRNA sequence above, and has a length of from 85-240 nt (e.g., 85-230, 85-220, 85-150, 85-130, 95-240, 95-230, 95-220, 95-150, or 95-130 nt).

[00168] In some cases, a subject CasZ trancRNA comprises a CasZc trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZc trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZc trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZc trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZc trancRNA sequence above, and has a length of from 80-110 nt (e.g., 85-105 nt) or from 235-270 nt (e.g., 240-260 nt).

[00169] In some cases, a subject CasZ trancRNA comprises a CasZa, CasZb, or CasZc trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 70% or more identity (e.g., 75% or more, 80% or more, 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZa, CasZb, or CasZc trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZa, CasZb, or CasZc trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide sequence having 90% or more identity (e.g., 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZa, CasZb, or CasZc trancRNA sequence above. In some cases, a subject CasZ trancRNA comprises a nucleotide

sequence having 80% or more identity (e.g., 85% or more, 90% or more, 93% or more, 95% or more, 97% or more, 98% or more, or 100% identity) with a CasZa, CasZb, or CasZc trancRNA sequence above, and has a length of from 60 nucleotides (nt) to 270 nt (e.g., 60-260, 70-270, 70-260, or 75-255 nt).

[00170] In some cases, a CasZ trancRNA comprises a modified nucleotide (e.g., methylated). In some cases, a CasZ trancRNA comprises one or more of: i) a base modification or substitution; ii) a backbone modification; iii) a modified internucleoside linkage; and iv) a modified sugar moiety. Possible nucleic acid modifications are described below.

CASZ SYSTEMS

[00171] The present disclosure provides a CasZ system. A CasZ system of the present disclosure can comprise one or more of: (1) a CasZ transactivating noncoding RNA (trancRNA) (referred to herein as a "CasZ trancRNA") or a nucleic acid encoding the CasZ trancRNA (e.g., an expression vector); (2) a CasZ protein (e.g., a wild type protein, a variant, a catalytically compromised variant, a CasZ fusion protein, and the like) or a nucleic acid encoding the CasZ protein (e.g., an RNA, an expression vector, and the like); and (3) a CasZ guide RNA (that binds to and provides sequence specificity to the CasZ protein, e.g., a guide RNA that can bind to a target sequence of a eukaryotic genome) or a nucleic acid encoding the CasZ guide RNA)(e.g., an expression vector). A CasZ system can include a host cell (e.g., a eukaryotic cell, a plant cell, a mammalian cell, a human cell) that comprises one or more of (1), (2), and (3) (in any combination), e.g., in some cases the host cell comprises a trancRNA and/or a nucleic acid encoding the trancRNA. In some cases, a CasZ system includes (e.g., in addition to the above) a donor template nucleic acid. In some cases, the CasZ system is a system of one or more nucleic acids (e.g., one or more expression vectors encoding any combination of the above).

NUCLEIC ACIDS

[00172] The present disclosure provides one or more nucleic acids comprising one or more of: a CasZ trancRNA sequence, a nucleotide sequence encoding a CasZ trancRNA, a nucleotide sequence encoding a CasZ polypeptide (e.g., a wild type CasZ protein, a nickase CasZ protein, a dCasZ protein, chimeric CasZ protein/CasZ fusion protein, and the like), a CasZ guide RNA sequence, a nucleotide sequence encoding a CasZ guide RNA, and a donor polynucleotide (donor template, donor DNA) sequence. In some cases, a subject nucleic acid (e.g., the one or more nucleic acids) is a recombinant expression vector (e.g., plasmid, viral vector, minicircle DNA, and the like). In some cases, the nucleotide sequence encoding the CasZ trancRNA, the nucleotide sequence encoding the CasZ protein, and/or the nucleotide sequence encoding the CasZ guide RNA is (are) operably linked to a promoter (e.g., an inducible promoter), e.g., one that is operable in a cell type of choice (e.g., a prokaryotic cell, a eukaryotic cell, a plant cell, an animal cell, a mammalian cell, a primate cell, a rodent cell, a human cell, etc.).

[00173] In some cases, a nucleotide sequence encoding a CasZ polypeptide of the present disclosure is codon optimized. This type of optimization can entail a mutation of a CasZ-encoding nucleotide sequence to mimic the codon preferences of the intended host organism or cell while encoding the same protein. Thus, the codons can be changed, but the encoded protein remains unchanged. For example, if the intended target cell was a human cell, a human codon-optimized CasZ-encoding nucleotide sequence could be used. As another non-limiting example, if the intended host cell were a mouse cell, then a mouse codon-optimized CasZ-encoding nucleotide sequence could be generated. As another non-limiting example, if the intended host cell were a plant cell, then a plant codon-optimized CasZ-encoding nucleotide sequence could be generated. As another non-limiting example, if the intended host cell were an insect cell, then an insect codon-optimized CasZ-encoding nucleotide sequence could be generated.

[00174] The present disclosure provides one or more recombinant expression vectors that include (in different recombinant expression vectors in some cases, and in the same recombinant expression vector in some cases): a CasZ trancRNA sequence, a nucleotide sequence encoding a CasZ trancRNA, a nucleotide sequence encoding a CasZ polypeptide (e.g., a wild type CasZ protein, a nickase CasZ protein, a dCasZ protein, chimeric CasZ protein/CasZ fusion protein, and the like), a CasZ guide RNA sequence, a nucleotide sequence encoding a CasZ guide RNA, and a donor polynucleotide (donor template, donor DNA) sequence. In some cases, a subject nucleic acid (e.g., the one or more nucleic acids) is a recombinant expression vector (e.g., plasmid, viral vector, minicircle DNA, and the like). In some cases, the nucleotide sequence encoding the CasZ trancRNA, the nucleotide sequence encoding the CasZ protein, and/or the nucleotide sequence encoding the CasZ guide RNA is (are) operably linked to a promoter (e.g., an inducible promoter), e.g., one that is operable in a cell type of choice (e.g., a prokarytoic cell, a eukaryotic cell, a plant cell, an animal cell, a mammalian cell, a primate cell, a rodent cell, a human cell, etc.).

[00175] Suitable expression vectors include viral expression vectors (e.g. viral vectors based on vaccinia virus; poliovirus; adenovirus (see, e.g., Li et al., Invest Opthalmol Vis Sci 35:2543 2549, 1994; Borras et al., Gene Ther 6:515 524, 1999; Li and Davidson, PNAS 92:7700 7704, 1995; Sakamoto et al., H Gene Ther 5:1088 1097, 1999; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655); adeno-associated virus (AAV) (see, e.g., Ali et al., Hum Gene Ther 9:81 86, 1998, Flannery et al., PNAS 94:6916 6921, 1997; Bennett et al., Invest Opthalmol Vis Sci 38:2857 2863, 1997; Jomary et al., Gene Ther 4:683 690, 1997, Rolling et al., Hum Gene Ther 10:641 648, 1999; Ali et al., Hum Mol Genet 5:591 594, 1996; Srivastava in WO 93/09239, Samulski et al., J. Vir. (1989) 63:3822-3828; Mendelson et al., Virol. (1988) 166:154-165; and Flotte et al., PNAS (1993) 90:10613-10617); SV40; herpes simplex virus; human immunodeficiency virus (see, e.g., Miyoshi et al., PNAS 94:10319 23, 1997; Takahashi et al., J Virol 73:7812 7816, 1999); a retroviral vector (e.g., Murine

Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus); and the like. In some cases, a recombinant expression vector of the present disclosure is a recombinant adeno-associated virus (AAV) vector. In some cases, a recombinant expression vector of the present disclosure is a recombinant lentivirus vector. In some cases, a recombinant expression vector of the present disclosure is a recombinant retroviral vector.

[00176] Depending on the host/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector.

[00177] In some embodiments, a nucleotide sequence encoding a CasZ guide RNA is operably linked to a control element, e.g., a transcriptional control element, such as a promoter. In some embodiments, a nucleotide sequence encoding a CasZ protein or a CasZ fusion polypeptide is operably linked to a control element, e.g., a transcriptional control element, such as a promoter.

[00178] The transcriptional control element can be a promoter. In some cases, the promoter is a constitutively active promoter. In some cases, the promoter is a regulatable promoter. In some cases, the promoter is an inducible promoter. In some cases, the promoter is a tissue-specific promoter. In some cases, the promoter is a cell type-specific promoter. In some cases, the transcriptional control element (e.g., the promoter) is functional in a targeted cell type or targeted cell population. For example, in some cases, the transcriptional control element can be functional in eukaryotic cells, e.g., hematopoietic stem cells (e.g., mobilized peripheral blood (mPB) CD34(+) cell, bone marrow (BM) CD34(+) cell, etc.).

[00179] Non-limiting examples of eukaryotic promoters (promoters functional in a eukaryotic cell) include EF1α, those from cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, early and late SV40, long terminal repeats (LTRs) from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. The expression vector may also contain a ribosome binding site for translation initiation and a transcription terminator. The expression vector may also include appropriate sequences for amplifying expression. The expression vector may also include nucleotide sequences encoding protein tags (e.g., 6xHis tag, hemagglutinin tag, fluorescent protein, etc.) that can be fused to the CasZ protein, thus resulting in a chimeric CasZ polypeptide.

[00180] In some cases, a nucleotide sequence encoding a CasZ guide RNA and/or a CasZ fusion polypeptide is operably linked to an inducible promoter. In some cases, a nucleotide sequence encoding a CasZ guide RNA and/or a CasZ fusion protein is operably linked to a constitutive promoter.

[00181] A promoter can be a constitutively active promoter (i.e., a promoter that is constitutively in an active/"ON" state), it may be an inducible promoter (i.e., a promoter whose state, active/"ON" or inactive/"OFF", is controlled by an external stimulus, e.g., the presence of a particular temperature, compound, or protein.), it may be a spatially restricted promoter (i.e., transcriptional control element, enhancer, etc.)(e.g., tissue specific promoter, cell type specific promoter, etc.), and it may be a temporally restricted promoter (i.e., the promoter is in the "ON" state or "OFF" state during specific stages of embryonic development or during specific stages of a biological process, e.g., hair follicle cycle in mice).

[00182] Suitable promoters can be derived from viruses and can therefore be referred to as viral promoters, or they can be derived from any organism, including prokaryotic or eukaryotic organisms. Suitable promoters can be used to drive expression by any RNA polymerase (e.g., pol I, pol II, pol III). Exemplary promoters include, but are not limited to the SV40 early promoter, mouse mammary tumor virus long terminal repeat (LTR) promoter; adenovirus major late promoter (Ad MLP); a herpes simplex virus (HSV) promoter, a cytomegalovirus (CMV) promoter such as the CMV immediate early promoter region (CMVIE), a rous sarcoma virus (RSV) promoter, a human U6 small nuclear promoter (U6) (Miyagishi et al., Nature Biotechnology 20, 497 - 500 (2002)), an enhanced U6 promoter (e.g., Xia et al., Nucleic Acids Res. 2003 Sep 1;31(17)), a human H1 promoter (H1), and the like.

[00183] In some cases, a nucleotide sequence encoding a CasZ guide RNA is operably linked to (under the control of) a promoter operable in a eukaryotic cell (e.g., a U6 promoter, an enhanced U6 promoter, an H1 promoter, and the like). As would be understood by one of ordinary skill in the art, when expressing an RNA (e.g., a guide RNA) from a nucleic acid (e.g., an expression vector) using a U6 promoter (e.g., in a eukaryotic cell), or another PolIII promoter, the RNA may need to be mutated if there are several Ts in a row (coding for Us in the RNA). This is because a string of Ts (e.g., 5 Ts) in DNA can act as a terminator for polymerase III (PolIII). Thus, in order to ensure transcription of a guide RNA in a eukaryotic cell it may sometimes be necessary to modify the sequence encoding the guide RNA to eliminate runs of Ts. In some cases, a nucleotide sequence encoding a CasZ protein (e.g., a wild type CasZ protein, a nickase CasZ protein, a dCasZ protein, a chimeric CasZ protein and the like) is operably linked to a promoter operable in a eukaryotic cell (e.g., a CMV promoter, an EF1α promoter, an estrogen receptor-regulated promoter, and the like).

[00184] Examples of inducible promoters include, but are not limited toT7 RNA polymerase promoter, T3 RNA polymerase promoter, Isopropyl-beta-D-thiogalactopyranoside (IPTG)-regulated promoter, lactose induced promoter, heat shock promoter, Tetracycline-regulated promoter, Steroid-regulated promoter, Metal-regulated promoter, estrogen receptor-regulated promoter, etc. Inducible promoters can therefore be regulated by molecules including, but not limited to, doxycycline; estrogen and/or an estrogen analog; IPTG; etc.

[00185] Inducible promoters suitable for use include any inducible promoter described herein or known to one of ordinary skill in the art. Examples of inducible promoters include, without limitation, chemically/biochemically-regulated and physically-regulated promoters such as alcohol-regulated promoters, tetracycline-regulated promoters (e.g., anhydrotetracycline (aTc)-responsive promoters and other tetracycline-responsive promoter systems, which include a tetracycline repressor protein (tetR), a tetracycline operator sequence (tetO) and a tetracycline transactivator fusion protein (tTA)), steroid-regulated promoters (e.g., promoters based on the rat glucocorticoid receptor, human estrogen receptor, moth ecdysone receptors, and promoters from the steroid/retinoid/thyroid receptor superfamily), metal-regulated promoters (e.g., promoters derived from metallothionein (proteins that bind and sequester metal ions) genes from yeast, mouse and human), pathogenesis-regulated promoters (e.g., induced by salicylic acid, ethylene or benzothiadiazole (BTH)), temperature/heat-inducible promoters (e.g., heat shock promoters), and light-regulated promoters (e.g., light responsive promoters from plant cells).

[00186] In some cases, the promoter is a spatially restricted promoter (i.e., cell type specific promoter, tissue specific promoter, etc.) such that in a multi-cellular organism, the promoter is active (i.e., "ON") in a subset of specific cells. Spatially restricted promoters may also be referred to as enhancers, transcriptional control elements, control sequences, etc. Any convenient spatially restricted promoter may be used as long as the promoter is functional in the targeted host cell (e.g., eukaryotic cell; prokaryotic cell).

[00187] In some cases, the promoter is a reversible promoter. Suitable reversible promoters, including reversible inducible promoters are known in the art. Such reversible promoters may be isolated and derived from many organisms, e.g., eukaryotes and prokaryotes. Modification of reversible promoters derived from a first organism for use in a second organism, e.g., a first prokaryote and a second a eukaryote, a first eukaryote and a second a prokaryote, etc., is well known in the art. Such reversible promoters, and systems based on such reversible promoters but also comprising additional control proteins, include, but are not limited to, alcohol regulated promoters (e.g., alcohol dehydrogenase I (alcA) gene promoter, promoters responsive to alcohol transactivator proteins (AlcR), etc.), tetracycline regulated promoters, (e.g., promoter systems including TetActivators, TetON, TetOFF, etc.), steroid regulated promoters (e.g., rat glucocorticoid receptor promoter systems, human estrogen receptor promoter systems, retinoid promoter systems, thyroid promoter systems, ecdysone promoter systems, mifepristone promoter systems, etc.), metal regulated promoters (e.g., metallothionein promoter systems, etc.), pathogenesis-related regulated promoters (e.g., salicylic acid regulated promoters, ethylene regulated promoters, benzothiadiazole regulated promoters, etc.), temperature regulated promoters (e.g., heat shock inducible promoters (e.g., HSP-70, HSP-90, soybean heat shock promoter, etc.), light regulated promoters, synthetic inducible promoters, and the like.

[00188] Methods of introducing a nucleic acid (e.g., DNA or RNA) (e.g., a nucleic acid comprising a donor polynucleotide sequence, one or more nucleic acids encoding a CasZ protein and/or a CasZ guide RNA and/or a CasZ trancRNA, and the like) into a host cell are known in the art, and any convenient method can be used to introduce a nucleic acid (e.g., an expression construct) into a cell. Suitable methods include e.g., viral infection, transfection, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct microinjection, nanoparticle-mediated nucleic acid delivery, and the like.

[00189] Introducing the recombinant expression vector into cells can occur in any culture media and under any culture conditions that promote the survival of the cells. Introducing the recombinant expression vector into a target cell can be carried out *in vivo* or *ex vivo*. Introducing the recombinant expression vector into a target cell can be carried out *in vitro*.

[00190] In some cases, a CasZ protein can be provided as RNA. The RNA can be provided by direct chemical synthesis or may be transcribed in vitro from a DNA (e.g., encoding the CasZ protein). Once synthesized, the RNA may be introduced into a cell by any of the well-known techniques for introducing nucleic acids into cells (e.g., microinjection, electroporation, transfection, etc.).

[00191] Nucleic acids may be provided to the cells using well-developed transfection techniques; see, e.g. Angel and Yanik (2010) PLoS ONE 5(7): e11756, and the commercially available TransMessenger® reagents from Qiagen, Stemfect™ RNA Transfection Kit from Stemgent, and TransIT®-mRNA Transfection Kit from Mirus Bio LLC. See also Beumer et al. (2008) PNAS 105(50):19821-19826.

[00192] Vectors may be provided directly to a target host cell. In other words, the cells are contacted with vectors comprising the subject nucleic acids (e.g., recombinant expression vectors having the donor template sequence and encoding the CasZ guide RNA; recombinant expression vectors encoding the CasZ protein; etc.) such that the vectors are taken up by the cells. Methods for contacting cells with nucleic acid vectors that are plasmids, include electroporation, calcium chloride transfection, microinjection, and lipofection are well known in the art. For viral vector delivery, cells can be contacted with viral particles comprising the subject viral expression vectors.

[00193] Retroviruses, for example, lentiviruses, are suitable for use in methods of the present disclosure. Commonly used retroviral vectors are "defective", i.e. unable to produce viral proteins required for productive infection. Rather, replication of the vector requires growth in a packaging cell line. To generate viral particles comprising nucleic acids of interest, the retroviral nucleic acids comprising the nucleic acid are packaged into viral capsids by a packaging cell line. Different packaging cell lines provide a different envelope protein (ecotropic, amphotropic or xenotropic) to be incorporated

into the capsid, this envelope protein determining the specificity of the viral particle for the cells (ecotropic for murine and rat; amphotropic for most mammalian cell types including human, dog and mouse; and xenotropic for most mammalian cell types except murine cells). The appropriate packaging cell line may be used to ensure that the cells are targeted by the packaged viral particles. Methods of introducing subject vector expression vectors into packaging cell lines and of collecting the viral particles that are generated by the packaging lines are well known in the art. Nucleic acids can also introduced by direct micro-injection (e.g., injection of RNA).

[00194] Vectors used for providing the nucleic acids encoding CasZ guide RNA and/or a CasZ polypeptide to a target host cell can include suitable promoters for driving the expression, that is, transcriptional activation, of the nucleic acid of interest. In other words, in some cases, the nucleic acid of interest will be operably linked to a promoter. This may include ubiquitously acting promoters, for example, the CMV-β-actin promoter, or inducible promoters, such as promoters that are active in particular cell populations or that respond to the presence of drugs such as tetracycline. By transcriptional activation, it is intended that transcription will be increased above basal levels in the target cell by 10 fold, by 100 fold, more usually by 1000 fold. In addition, vectors used for providing a nucleic acid encoding a CasZ guide RNA and/or a CasZ protein to a cell may include nucleic acid sequences that encode for selectable markers in the target cells, so as to identify cells that have taken up the CasZ guide RNA and/or CasZ protein.

[00195] A nucleic acid comprising a nucleotide sequence encoding a CasZ polypeptide, or a CasZ fusion polypeptide, is in some cases an RNA. Thus, a CasZ fusion protein can be introduced into cells as RNA. Methods of introducing RNA into cells are known in the art and may include, for example, direct injection, transfection, or any other method used for the introduction of DNA. A CasZ protein may instead be provided to cells as a polypeptide. Such a polypeptide may optionally be fused to a polypeptide domain that increases solubility of the product. The domain may be linked to the polypeptide through a defined protease cleavage site, e.g. a TEV sequence, which is cleaved by TEV protease. The linker may also include one or more flexible sequences, e.g. from 1 to 10 glycine residues. In some embodiments, the cleavage of the fusion protein is performed in a buffer that maintains solubility of the product, e.g. in the presence of from 0.5 to 2 M urea, in the presence of polypeptides and/or polynucleotides that increase solubility, and the like. Domains of interest include endosomolytic domains, e.g. influenza HA domain; and other polypeptides that aid in production, e.g. IF2 domain, GST domain, GRPE domain, and the like. The polypeptide may be formulated for improved stability. For example, the peptides may be PEGylated, where the polyethyleneoxy group provides for enhanced lifetime in the blood stream.

[00196] Additionally, or alternatively, a CasZ polypeptide of the present disclosure may be fused to a polypeptide permeant domain to promote uptake by the cell. A number of permeant domains are

known in the art and may be used in the non-integrating polypeptides of the present disclosure, including peptides, peptidomimetics, and non-peptide carriers. For example, a permeant peptide may be derived from the third alpha helix of *Drosophila melanogaster* transcription factor Antennapaedia, referred to as penetratin, which comprises the amino acid sequence RQIKIWFQNRRMKWKK (SEQ ID NO: 134). As another example, the permeant peptide comprises the HIV-1 tat basic region amino acid sequence, which may include, for example, amino acids 49-57 of naturally-occurring tat protein. Other permeant domains include poly-arginine motifs, for example, the region of amino acids 34-56 of HIV-1 rev protein, nona-arginine, octa-arginine, and the like. (See, for example, Futaki et al. (2003) Curr Protein Pept Sci. 2003 Apr; 4(2): 87-9 and 446; and Wender et al. (2000) Proc. Natl. Acad. Sci. U.S.A 2000 Nov. 21; 97(24):13003-8; published U.S. Patent applications 20030220334; 20030083256; 20030032593; and 20030022831, herein specifically incorporated by reference for the teachings of translocation peptides and peptoids). The nona-arginine (R9) sequence is one of the more efficient PTDs that have been characterized (Wender et al. 2000; Uemura et al. 2002). The site at which the fusion is made may be selected in order to optimize the biological activity, secretion or binding characteristics of the polypeptide. The optimal site will be determined by routine experimentation.

[00197] A CasZ polypeptide of the present disclosure may be produced *in vitro* or by eukaryotic cells or by prokaryotic cells, and it may be further processed by unfolding, e.g. heat denaturation, dithiothreitol reduction, etc. and may be further refolded, using methods known in the art.

[00198] Modifications of interest that do not alter primary sequence include chemical derivatization of polypeptides, e.g., acylation, acetylation, carboxylation, amidation, etc. Also included are modifications of glycosylation, e.g. those made by modifying the glycosylation patterns of a polypeptide during its synthesis and processing or in further processing steps; e.g. by exposing the polypeptide to enzymes which affect glycosylation, such as mammalian glycosylating or deglycosylating enzymes. Also embraced are sequences that have phosphorylated amino acid residues, e.g. phosphotyrosine, phosphoserine, or phosphothreonine.

[00199] Also suitable for inclusion in embodiments of the present disclosure are nucleic acids (e.g., encoding a CasZ guide RNA, encoding a CasZ fusion protein, etc.) and proteins (e.g., a CasZ fusion protein derived from a wild type protein or a variant protein) that have been modified using ordinary molecular biological techniques and synthetic chemistry so as to improve their resistance to proteolytic degradation, to change the target sequence specificity, to optimize solubility properties, to alter protein activity (e.g., transcription modulatory activity, enzymatic activity, etc.) or to render them more suitable. Analogs of such polypeptides include those containing residues other than naturally occurring L-amino acids, e.g. D-amino acids or non-naturally occurring synthetic amino acids. D-amino acids may be substituted for some or all of the amino acid residues.

[00200] A CasZ polypeptide of the present disclosure may be prepared by *in vitro* synthesis, using conventional methods as known in the art. Various commercial synthetic apparatuses are available, for example, automated synthesizers by Applied Biosystems, Inc., Beckman, etc. By using synthesizers, naturally occurring amino acids may be substituted with unnatural amino acids. The particular sequence and the manner of preparation will be determined by convenience, economics, purity required, and the like.

[00201] If desired, various groups may be introduced into the peptide during synthesis or during expression, which allow for linking to other molecules or to a surface. Thus cysteines can be used to make thioethers, histidines for linking to a metal ion complex, carboxyl groups for forming amides or esters, amino groups for forming amides, and the like.

[00202] A CasZ polypeptide of the present disclosure may also be isolated and purified in accordance with conventional methods of recombinant synthesis. A lysate may be prepared of the expression host and the lysate purified using high performance liquid chromatography (HPLC), exclusion chromatography, gel electrophoresis, affinity chromatography, or other purification technique. For the most part, the compositions which are used will comprise 20% or more by weight of the desired product, more usually 75% or more by weight, preferably 95% or more by weight, and for therapeutic purposes, usually 99.5% or more by weight, in relation to contaminants related to the method of preparation of the product and its purification. Usually, the percentages will be based upon total protein. Thus, in some cases, a CasZ polypeptide, or a CasZ fusion polypeptide, of the present disclosure is at least 80% pure, at least 85% pure, at least 90% pure, at least 95% pure, at least 98% pure, or at least 99% pure (e.g., free of contaminants, non-CasZ proteins or other macromolecules, etc.).

[00203] To induce cleavage or any desired modification to a target nucleic acid (e.g., genomic DNA), or any desired modification to a polypeptide associated with target nucleic acid, the CasZ guide RNA and/or the CasZ polypeptide and/or the CasZ trancRNA, and/or the donor template sequence, whether they be introduced as nucleic acids or polypeptides, can be provided to the cells for about 30 minutes to about 24 hours, e.g., 1 hour, 1.5 hours, 2 hours, 2.5 hours, 3 hours, 3.5 hours 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 12 hours, 16 hours, 18 hours, 20 hours, or any other period from about 30 minutes to about 24 hours, which may be repeated with a frequency of about every day to about every 4 days, e.g., every 1.5 days, every 2 days, every 3 days, or any other frequency from about every day to about every four days. The agent(s) may be provided to the subject cells one or more times, e.g. one time, twice, three times, or more than three times, and the cells allowed to incubate with the agent(s) for some amount of time following each contacting event e.g. 16-24 hours, after which time the media is replaced with fresh media and the cells are cultured further.

[00204] In cases in which two or more different targeting complexes are provided to the cell (e.g., two different CasZ guide RNAs that are complementary to different sequences within the same or different target nucleic acid), the complexes may be provided simultaneously (e.g. as two polypeptides and/or nucleic acids), or delivered simultaneously. Alternatively, they may be provided consecutively, e.g. the targeting complex being provided first, followed by the second targeting complex, etc. or vice versa.

[00205] To improve the delivery of a DNA vector into a target cell, the DNA can be protected from damage and its entry into the cell facilitated, for example, by using lipoplexes and polyplexes. Thus, in some cases, a nucleic acid of the present disclosure (e.g., a recombinant expression vector of the present disclosure) can be covered with lipids in an organized structure like a micelle or a liposome. When the organized structure is complexed with DNA it is called a lipoplex. There are three types of lipids, anionic (negatively-charged), neutral, or cationic (positively-charged). Lipoplexes that utilize cationic lipids have proven utility for gene transfer. Cationic lipids, due to their positive charge, naturally complex with the negatively charged DNA. Also as a result of their charge, they interact with the cell membrane. Endocytosis of the lipoplex then occurs, and the DNA is released into the cytoplasm. The cationic lipids also protect against degradation of the DNA by the cell.

[00206] Complexes of polymers with DNA are called polyplexes. Most polyplexes consist of cationic polymers and their production is regulated by ionic interactions. One large difference between the methods of action of polyplexes and lipoplexes is that polyplexes cannot release their DNA load into the cytoplasm, so to this end, co-transfection with endosome-lytic agents (to lyse the endosome that is made during endocytosis) such as inactivated adenovirus must occur. However, this is not always the case; polymers such as polyethylenimine have their own method of endosome disruption as does chitosan and trimethylchitosan.

[00207] Dendrimers, a highly branched macromolecule with a spherical shape, may be also be used to genetically modify stem cells. The surface of the dendrimer particle may be functionalized to alter its properties. In particular, it is possible to construct a cationic dendrimer (i.e., one with a positive surface charge). When in the presence of genetic material such as a DNA plasmid, charge complementarity leads to a temporary association of the nucleic acid with the cationic dendrimer. On reaching its destination, the dendrimer-nucleic acid complex can be taken up into a cell by endocytosis.

[00208] In some cases, a nucleic acid of the disclosure (e.g., an expression vector) includes an insertion site for a guide sequence of interest. For example, a nucleic acid can include an insertion site for a guide sequence of interest, where the insertion site is immediately adjacent to a nucleotide sequence encoding the portion of a CasZ guide RNA that does not change when the guide sequence is changed to hybrized to a desired target sequence (e.g., sequences that contribute to the CasZ binding aspect of the

guide RNA, e.g, the sequences that contribute to the dsRNA duplex(es) of the CasZ guide RNA – this portion of the guide RNA can also be referred to as the 'scaffold' or 'constant region' of the guide RNA). Thus, in some cases, a subject nucleic acid (e.g., an expression vector) includes a nucleotide sequence encoding a CasZ guide RNA, except that the portion encoding the guide sequence portion of the guide RNA is an insertion sequence (an insertion site). An insertion site is any nucleotide sequence used for the insertion of a desired sequence. "Insertion sites" for use with various technologies are known to those of ordinary skill in the art and any convenient insertion site can be used. An insertion site can be for any method for manipulating nucleic acid sequences. For example, in some cases the insertion site is a multiple cloning site (MCS) (e.g., a site including one or more restriction enzyme recognition sequences), a site for ligation independent cloning, a site for recombination-based cloning (e.g., recombination based on *att* sites), a nucleotide sequence recognized by a CRISPR/Cas (e.g. Cas9) based technology, and the like.

[00209] An insertion site can be any desirable length, and can depend on the type of insertion site (e.g., can depend on whether (and how many) the site includes one or more restriction enzyme recognition sequences, whether the site includes a target site for a CRISPR/Cas protein, etc.). In some cases, an insertion site of a subject nucleic acid is 3 or more nucleotides (nt) in length (e.g., 5 or more, 8 or more, 10 or more, 15 or more, 17 or more, 18 or more, 19 or more, 20 or more or 25 or more, or 30 or more nt in length). In some cases, the length of an insertion site of a subject nucleic acid has a length in a range of from 2 to 50 nucleotides (nt) (e.g., from 2 to 40 nt, from 2 to 30 nt, from 2 to 25 nt, from 2 to 20 nt, from 5 to 50 nt, from 5 to 40 nt, from 5 to 30 nt, from 5 to 25 nt, from 5 to 20 nt, from 10 to 50 nt, from 10 to 40 nt, from 10 to 30 nt, from 10 to 25 nt, from 10 to 20 nt, from 17 to 50 nt, from 17 to 40 nt, from 17 to 30 nt, from 17 to 25 nt). In some cases, the length of an insertion site of a subject nucleic acid has a length in a range of from 5 to 40 nt.

Nucleic acid modifications

[00210] In some embodiments, a subject nucleic acid (e.g., a CasZ guide RNA or trancRNA) has one or more modifications, e.g., a base modification, a backbone modification, etc., to provide the nucleic acid with a new or enhanced feature (e.g., improved stability). A nucleoside is a base-sugar combination. The base portion of the nucleoside is normally a heterocyclic base. The two most common classes of such heterocyclic bases are the purines and the pyrimidines. Nucleotides are nucleosides that further include a phosphate group covalently linked to the sugar portion of the nucleoside. For those nucleosides that include a pentofuranosyl sugar, the phosphate group can be linked to the 2', the 3', or the 5' hydroxyl moiety of the sugar. In forming oligonucleotides, the phosphate groups covalently link adjacent nucleosides to one another to form a linear polymeric compound. In turn, the respective ends of this linear polymeric compound can be further joined to form a circular compound, however, linear compounds are suitable. In addition, linear compounds may have internal nucleotide base

complementarity and may therefore fold in a manner as to produce a fully or partially double-stranded compound. Within oligonucleotides, the phosphate groups are commonly referred to as forming the internucleoside backbone of the oligonucleotide. The normal linkage or backbone of RNA and DNA is a 3' to 5' phosphodiester linkage.

[00211] Suitable nucleic acid modifications include, but are not limited to: 2'Omethyl modified nucleotides, 2' Fluoro modified nucleotides, locked nucleic acid (LNA) modified nucleotides, peptide nucleic acid (PNA) modified nucleotides, nucleotides with phosphorothioate linkages, and a 5' cap (e.g., a 7-methylguanylate cap (m7G)). Additional details and additional modifications are described below.

[00212] A 2'-O-Methyl modified nucleotide (also referred to as 2'-O-Methyl RNA) is a naturally occurring modification of RNA found in tRNA and other small RNAs that arises as a post-transcriptional modification. Oligonucleotides can be directly synthesized that contain 2'-O-Methyl RNA. This modification increases Tm of RNA:RNA duplexes but results in only small changes in RNA:DNA stability. It is stabile with respect to attack by single-stranded ribonucleases and is typically 5 to 10-fold less susceptible to DNases than DNA. It is commonly used in antisense oligos as a means to increase stability and binding affinity to the target message.

[00213] 2' Fluoro modified nucleotides (e.g., 2' Fluoro bases) have a fluorine modified ribose which increases binding affinity (Tm) and also confers some relative nuclease resistance when compared to native RNA. These modifications are commonly employed in ribozymes and siRNAs to improve stability in serum or other biological fluids.

[00214] LNA bases have a modification to the ribose backbone that locks the base in the C3'-endo position, which favors RNA A-type helix duplex geometry. This modification significantly increases Tm and is also very nuclease resistant. Multiple LNA insertions can be placed in an oligo at any position except the 3'-end. Applications have been described ranging from antisense oligos to hybridization probes to SNP detection and allele specific PCR. Due to the large increase in Tm conferred by LNAs, they also can cause an increase in primer dimer formation as well as self-hairpin formation. In some cases, the number of LNAs incorporated into a single oligo is 10 bases or less.

[00215] The phosphorothioate (PS) bond (i.e., a phosphorothioate linkage) substitutes a sulfur atom for a non-bridging oxygen in the phosphate backbone of a nucleic acid (e.g., an oligo). This modification renders the internucleotide linkage resistant to nuclease degradation. Phosphorothioate bonds can be introduced between the last 3-5 nucleotides at the 5'- or 3'-end of the oligo to inhibit exonuclease degradation. Including phosphorothioate bonds within the oligo (e.g., throughout the entire oligo) can help reduce attack by endonucleases as well.

[00216] In some embodiments, a subject nucleic acid has one or more nucleotides that are 2'-O-Methyl modified nucleotides. In some embodiments, a subject nucleic acid (e.g., a guide RNA, a tranc

RNA, etc.) has one or more 2' Fluoro modified nucleotides. In some embodiments, a subject nucleic acid (e.g., a dsRNA, a siNA, etc.) has one or more LNA bases. In some embodiments, a subject nucleic acid (e.g., a dsRNA, a siNA, etc.) has one or more nucleotides that are linked by a phosphorothioate bond (i.e., the subject nucleic acid has one or more phosphorothioate linkages). In some embodiments, a subject nucleic acid (e.g., a dsRNA, a siNA, etc.) has a 5' cap (e.g., a 7-methylguanylate cap (m7G)). In some embodiments, a subject nucleic acid (e.g., a guide RNA, a tranc RNA, etc.) has a combination of modified nucleotides. For example, a subject nucleic acid (e.g., a guide RNA, a tranc RNA, etc.) can have a 5' cap (e.g., a 7-methylguanylate cap (m7G)) in addition to having one or more nucleotides with other modifications (e.g., a 2'-O-Methyl nucleotide and/or a 2' Fluoro modified nucleotide and/or a LNA base and/or a phosphorothioate linkage).

Modified backbones and modified internucleoside linkages

[00217] Examples of suitable nucleic acids (e.g., a CasZ guide RNA and/or CasZ trancRNA) containing modifications include nucleic acids containing modified backbones or non-natural internucleoside linkages. Nucleic acids having modified backbones include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus atom in the backbone.

[00218] Suitable modified oligonucleotide backbones containing a phosphorus atom therein include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates, 5'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, phosphorodiamidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, selenophosphates and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein one or more internucleotide linkages is a 3' to 3', 5' to 5' or 2' to 2' linkage. Suitable oligonucleotides having inverted polarity comprise a single 3' to 3' linkage at the 3'-most internucleotide linkage i.e. a single inverted nucleoside residue which may be a basic (the nucleobase is missing or has a hydroxyl group in place thereof). Various salts (such as, for example, potassium or sodium), mixed salts and free acid forms are also included.

[00219] In some embodiments, a subject nucleic acid comprises one or more phosphorothioate and/or heteroatom internucleoside linkages, in particular $-CH_2-NH-O-CH_2-$, $-CH_2-N(CH_3)-O-CH_2-$ (known as a methylene (methylimino) or MMI backbone), $-CH_2-O-N(CH_3)-CH_2-$, $-CH_2-N(CH_3)-N(CH_3)-CH_2-$ and $-O-N(CH_3)-CH_2-CH_2-$ (wherein the native phosphodiester internucleotide linkage is represented as $-O-P(=O)(OH)-O-CH_2-$). MMI type internucleoside linkages are disclosed in the above referenced U.S. Pat. No. 5,489,677, the disclosure of which is incorporated herein by reference in its entirety. Suitable amide internucleoside linkages are disclosed in U.S. Pat. No. 5,602,240, the disclosure of which is incorporated herein by reference in its entirety.

[00220] Also suitable are nucleic acids having morpholino backbone structures as described in, e.g., U.S. Pat. No. 5,034,506. For example, in some embodiments, a subject nucleic acid comprises a 6-membered morpholino ring in place of a ribose ring. In some of these embodiments, a phosphorodiamidate or other non-phosphodiester internucleoside linkage replaces a phosphodiester linkage.

[00221] Suitable modified polynucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; riboacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts.

Mimetics

[00222] A subject nucleic acid can be a nucleic acid mimetic. The term "mimetic" as it is applied to polynucleotides is intended to include polynucleotides wherein only the furanose ring or both the furanose ring and the internucleotide linkage are replaced with non-furanose groups, replacement of only the furanose ring is also referred to in the art as being a sugar surrogate. The heterocyclic base moiety or a modified heterocyclic base moiety is maintained for hybridization with an appropriate target nucleic acid. One such nucleic acid, a polynucleotide mimetic that has been shown to have excellent hybridization properties, is referred to as a peptide nucleic acid (PNA). In PNA, the sugar-backbone of a polynucleotide is replaced with an amide containing backbone, in particular an aminoethylglycine backbone. The nucleotides are retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone.

[00223] One polynucleotide mimetic that has been reported to have excellent hybridization properties is a peptide nucleic acid (PNA). The backbone in PNA compounds is two or more linked aminoethylglycine units which gives PNA an amide containing backbone. The heterocyclic base moieties are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone. Representative U.S. patents that describe the preparation of PNA compounds include, but are not limited to: U.S. Pat. Nos. 5,539,082; 5,714,331; and 5,719,262, the disclosures of which are incorporated herein by reference in their entirety.

[00224] Another class of polynucleotide mimetic that has been studied is based on linked morpholino units (morpholino nucleic acid) having heterocyclic bases attached to the morpholino ring. A number of

linking groups have been reported that link the morpholino monomeric units in a morpholino nucleic acid. One class of linking groups has been selected to give a non-ionic oligomeric compound. The non-ionic morpholino-based oligomeric compounds are less likely to have undesired interactions with cellular proteins. Morpholino-based polynucleotides are non-ionic mimics of oligonucleotides which are less likely to form undesired interactions with cellular proteins (Dwaine A. Braasch and David R. Corey, *Biochemistry*, 2002, 41(14), 4503-4510). Morpholino-based polynucleotides are disclosed in U.S. Pat. No. 5,034,506, the disclosure of which is incorporated herein by reference in its entirety. A variety of compounds within the morpholino class of polynucleotides have been prepared, having a variety of different linking groups joining the monomeric subunits.

[00225] A further class of polynucleotide mimetic is referred to as cyclohexenyl nucleic acids (CeNA). The furanose ring normally present in a DNA/RNA molecule is replaced with a cyclohexenyl ring. CeNA DMT protected phosphoramidite monomers have been prepared and used for oligomeric compound synthesis following classical phosphoramidite chemistry. Fully modified CeNA oligomeric compounds and oligonucleotides having specific positions modified with CeNA have been prepared and studied (see Wang et al., *J. Am. Chem. Soc.*, 2000, 122, 8595-8602, the disclosure of which is incorporated herein by reference in its entirety). In general, the incorporation of CeNA monomers into a DNA chain increases its stability of a DNA/RNA hybrid. CeNA oligoadenylates formed complexes with RNA and DNA complements with similar stability to the native complexes. The study of incorporating CeNA structures into natural nucleic acid structures was shown by NMR and circular dichroism to proceed with easy conformational adaptation.

[00226] A further modification includes Locked Nucleic Acids (LNAs) in which the 2'-hydroxyl group is linked to the 4' carbon atom of the sugar ring thereby forming a 2'-C,4'-C-oxymethylene linkage thereby forming a bicyclic sugar moiety. The linkage can be a methylene (-CH_2-), group bridging the 2' oxygen atom and the 4' carbon atom wherein n is 1 or 2 (Singh et al., *Chem. Commun.*, 1998, 4, 455-456, the disclosure of which is incorporated herein by reference in its entirety). LNA and LNA analogs display very high duplex thermal stabilities with complementary DNA and RNA (Tm=+3 to +10° C), stability towards 3'-exonucleolytic degradation and good solubility properties. Potent and nontoxic antisense oligonucleotides containing LNAs have been described (e.g., Wahlestedt et al., *Proc. Natl. Acad. Sci.* U.S.A., 2000, 97, 5633-5638, the disclosure of which is incorporated herein by reference in its entirety).

[00227] The synthesis and preparation of the LNA monomers adenine, cytosine, guanine, 5-methyl-cytosine, thymine and uracil, along with their oligomerization, and nucleic acid recognition properties have been described (e.g., Koshkin et al., *Tetrahedron*, 1998, 54, 3607-3630, the disclosure of which is incorporated herein by reference in its entirety). LNAs and preparation thereof are also described in WO 98/39352 and WO 99/14226, as well as U.S. applications 20120165514, 20100216983, 20090041809,

20060117410, 20040014959, 20020094555, and 20020086998, the disclosures of which are incorporated herein by reference in their entirety.

Modified sugar moieties

[00228] A subject nucleic acid can also include one or more substituted sugar moieties. Suitable polynucleotides comprise a sugar substituent group selected from: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C.sub.1 to C_{10} alkyl or C_2 to C_{10} alkenyl and alkynyl. Particularly suitable are $O((CH_2)_nO)_mCH_3$, $O(CH_2)_nOCH_3$, $O(CH_2)_nNH_2$, $O(CH_2)_nCH_3$, $O(CH_2)_nONH_2$, and $O(CH_2)_nON((CH_2)_nCH_3)_2$, where n and m are from 1 to about 10. Other suitable polynucleotides comprise a sugar substituent group selected from: C_1 to C_{10} lower alkyl, substituted lower alkyl, alkenyl, alkynyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH_3, OCN, Cl, Br, CN, CF_3, OCF_3, $SOCH_3$, SO_2CH_3, ONO_2, NO_2, N_3, NH_2, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and other substituents having similar properties. A suitable modification includes 2'-methoxyethoxy (2'-O-CH_2 CH_2OCH_3, also known as 2'-O-(2-methoxyethyl) or 2'-MOE) (Martin et al., *Helv. Chim. Acta*, 1995, 78, 486-504, the disclosure of which is incorporated herein by reference in its entirety) i.e., an alkoxyalkoxy group. A further suitable modification includes 2'-dimethylaminooxyethoxy, i.e., a $O(CH_2)_2ON(CH_3)_2$ group, also known as 2'-DMAOE, as described in examples hereinbelow, and 2'-dimethylaminoethoxyethoxy (also known in the art as 2'-O-dimethyl-amino-ethoxy-ethyl or 2'-DMAEOE), i.e., 2'-O-CH_2-O-CH_2-$N(CH_3)_2$.

[00229] Other suitable sugar substituent groups include methoxy (-O-CH_3), aminopropoxy (--O CH_2 CH_2 CH_2NH_2), allyl (-CH_2-CH=CH_2), -O-allyl (--O-- CH_2—CH=CH_2) and fluoro (F). 2'-sugar substituent groups may be in the arabino (up) position or ribo (down) position. A suitable 2'-arabino modification is 2'-F. Similar modifications may also be made at other positions on the oligomeric compound, particularly the 3' position of the sugar on the 3' terminal nucleoside or in 2'-5' linked oligonucleotides and the 5' position of 5' terminal nucleotide. Oligomeric compounds may also have sugar mimetics such as cyclobutyl moieties in place of the pentofuranosyl sugar.

Base modifications and substitutions

[00230] A subject nucleic acid may also include nucleobase (often referred to in the art simply as "base") modifications or substitutions. As used herein, "unmodified" or "natural" nucleobases include the purine bases adenine (A) and guanine (G), and the pyrimidine bases thymine (T), cytosine (C) and uracil (U). Modified nucleobases include other synthetic and natural nucleobases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-

thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl (-C=C-CH_3) uracil and cytosine and other alkynyl derivatives of pyrimidine bases, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 2-F-adenine, 2-amino-adenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Further modified nucleobases include tricyclic pyrimidines such as phenoxazine cytidine(1H-pyrimido(5,4-b)(1,4)benzoxazin-2(3H)-one), phenothiazine cytidine (1H-pyrimido(5,4-b)(1,4)benzothiazin-2(3H)-one), G-clamps such as a substituted phenoxazine cytidine (e.g. 9-(2-aminoethoxy)-H-pyrimido(5,4-(b) (1,4)benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido(4,5-b)indol-2-one), pyridoindole cytidine (H-pyrido(3',2':4,5)pyrrolo(2,3-d)pyrimidin-2-one).

[00231] Heterocyclic base moieties may also include those in which the purine or pyrimidine base is replaced with other heterocycles, for example 7-deaza-adenine, 7-deazaguanosine, 2-aminopyridine and 2-pyridone. Further nucleobases include those disclosed in U.S. Pat. No. 3,687,808, those disclosed in *The Concise Encyclopedia Of Polymer Science And Engineering*, pages 858-859, Kroschwitz, J. I., ed. John Wiley & Sons, 1990, those disclosed by Englisch et al., *Angewandte Chemie, International Edition*, 1991, 30, 613, and those disclosed by Sanghvi, Y. S., *Chapter 15, Antisense Research and Applications*, pages 289-302, Crooke, S. T. and Lebleu, B., ed., CRC Press, 1993; the disclosures of which are incorporated herein by reference in their entirety. Certain of these nucleobases are useful for increasing the binding affinity of an oligomeric compound. These include 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-propynyluracil and 5-propynylcytosine. 5-methylcytosine substitutions have been shown to increase nucleic acid duplex stability by 0.6-1.2° C. (Sanghvi et al., eds., *Antisense Research and Applications*, CRC Press, Boca Raton, 1993, pp. 276-278; the disclosure of which is incorporated herein by reference in its entirety) and are suitable base substitutions, e.g., when combined with 2'-O-methoxyethyl sugar modifications.

Conjugates

[00232] Another possible modification of a subject nucleic acid involves chemically linking to the polynucleotide one or more moieties or conjugates which enhance the activity, cellular distribution or cellular uptake of the oligonucleotide. These moieties or conjugates can include conjugate groups covalently bound to functional groups such as primary or secondary hydroxyl groups. Conjugate groups include, but are not limited to, intercalators, reporter molecules, polyamines, polyamides, polyethylene glycols, polyethers, groups that enhance the pharmacodynamic properties of oligomers, and groups that enhance the pharmacokinetic properties of oligomers. Suitable conjugate groups include, but are not limited to, cholesterols, lipids, phospholipids, biotin, phenazine, folate, phenanthridine, anthraquinone,

acridine, fluoresceins, rhodamines, coumarins, and dyes. Groups that enhance the pharmacodynamic properties include groups that improve uptake, enhance resistance to degradation, and/or strengthen sequence-specific hybridization with the target nucleic acid. Groups that enhance the pharmacokinetic properties include groups that improve uptake, distribution, metabolism or excretion of a subject nucleic acid.

[00233] Conjugate moieties include but are not limited to lipid moieties such as a cholesterol moiety (Letsinger et al., *Proc. Natl. Acad. Sci.* USA, 1989, 86, 6553-6556), cholic acid (Manoharan et al., *Bioorg. Med. Chem. Let.*, 1994, 4, 1053-1060), a thioether, e.g., hexyl-S-tritylthiol (Manoharan et al., *Ann. N.Y. Acad. Sci.*, 1992, 660, 306-309; Manoharan et al., Bioorg. *Med. Chem. Let.*, 1993, 3, 2765-2770), a thiocholesterol (Oberhauser et al., *Nucl. Acids Res.*, 1992, 20, 533-538), an aliphatic chain, e.g., dodecandiol or undecyl residues (Saison-Behmoaras et al., *EMBO J.*, 1991, 10, 1111-1118; Kabanov et al., *FEBS Lett.*, 1990, 259, 327-330; Svinarchuk et al., *Biochimie*, 1993, 75, 49-54), a phospholipid, e.g., di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate (Manoharan et al., *Tetrahedron Lett.*, 1995, 36, 3651-3654; Shea et al., *Nucl. Acids Res.*, 1990, 18, 3777-3783), a polyamine or a polyethylene glycol chain (Manoharan et al., *Nucleosides & Nucleotides*, 1995, 14, 969-973), or adamantane acetic acid (Manoharan et al., *Tetrahedron Lett.*, 1995, 36, 3651-3654), a palmityl moiety (Mishra et al., *Biochim. Biophys. Acta*, 1995, 1264, 229-237), or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety (Crooke et al., *J. Pharmacol. Exp. Ther.*, 1996, 277, 923-937).

[00234] A conjugate may include a "Protein Transduction Domain" or PTD (also known as a CPP – cell penetrating peptide), which may refer to a polypeptide, polynucleotide, carbohydrate, or organic or inorganic compound that facilitates traversing a lipid bilayer, micelle, cell membrane, organelle membrane, or vesicle membrane. A PTD attached to another molecule, which can range from a small polar molecule to a large macromolecule and/or a nanoparticle, facilitates the molecule traversing a membrane, for example going from extracellular space to intracellular space, or cytosol to within an organelle (e.g., the nucleus). In some cases, a PTD is covalently linked to the 3' end of an exogenous polynucleotide. In some cases, a PTD is covalently linked to the 5' end of an exogenous polynucleotide. Exemplary PTDs include but are not limited to a minimal undecapeptide protein transduction domain (corresponding to residues 47-57 of HIV-1 TAT comprising YGRKKRRQRRR; SEQ ID NO: 130); a polyarginine sequence comprising a number of arginines sufficient to direct entry into a cell (e.g., 3, 4, 5, 6, 7, 8, 9, 10, or 10-50 arginines); a VP22 domain (Zender et al. (2002) *Cancer Gene Ther.* 9(6):489-96); an Drosophila Antennapedia protein transduction domain (Noguchi et al. (2003) *Diabetes* 52(7):1732-1737); a truncated human calcitonin peptide (Trehin et al. (2004) *Pharm. Research* 21:1248-1256); polylysine (Wender et al. (2000) *Proc. Natl. Acad. Sci. USA* 97:13003-13008); RRQRRTSKLMKR SEQ ID NO: 131); Transportan GWTLNSAGYLLGKINLKALAALAKKIL SEQ ID NO: 132);

KALAWEAKLAKALAKALAKHLAKALAKALKCEA SEQ ID NO: 133); and RQIKIWFQNRRMKWKK SEQ ID NO: 134). Exemplary PTDs include but are not limited to, YGRKKRRQRRR SEQ ID NO: 130), RKKRRQRRR SEQ ID NO: 135); an arginine homopolymer of from 3 arginine residues to 50 arginine residues; Exemplary PTD domain amino acid sequences include, but are not limited to, any of the following: YGRKKRRQRRR SEQ ID NO: 130); RKKRRQRR SEQ ID NO: 136); YARAAARQARA SEQ ID NO: 137); THRLPRRRRRR SEQ ID NO: 138); and GGRRARRRRRR SEQ ID NO: 139). In some cases, the PTD is an activatable CPP (ACPP) (Aguilera et al. (2009) *Integr Biol (Camb)* June; 1(5-6): 371-381). ACPPs comprise a polycationic CPP (e.g., Arg9 or "R9") connected via a cleavable linker to a matching polyanion (e.g., Glu9 or "E9"), which reduces the net charge to nearly zero and thereby inhibits adhesion and uptake into cells. Upon cleavage of the linker, the polyanion is released, locally unmasking the polyarginine and its inherent adhesiveness, thus "activating" the ACPP to traverse the membrane.

Introducing components into a target cell

[00235] A CasZ guide RNA (or a nucleic acid comprising a nucleotide sequence encoding same) and/or a CasZ polypeptide (or a nucleic acid comprising a nucleotide sequence encoding same) and/or a CasZ trancRNA (or a nucleic acid that includes a nucleotide sequence encoding same) and/or a donor polynucleotide (donor template) can be introduced into a host cell by any of a variety of well-known methods.

[00236] Any of a variety of compounds and methods can be used to deliver to a target cell a CasZ system of the present disclosure. As a non-limiting example, a CasZ system of the present disclosure can be combined with a lipid. As another non-limiting example, a CasZ system of the present disclosure can be combined with a particle, or formulated into a particle.

[00237] Methods of introducing a nucleic acid into a host cell are known in the art, and any convenient method can be used to introduce a subject nucleic acid (e.g., an expression construct/vector) into a target cell (e.g., prokaryotic cell, eukaryotic cell, plant cell, animal cell, mammalian cell, human cell, and the like). Suitable methods include, e.g., viral infection, transfection, conjugation, protoplast fusion, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, particle gun technology, calcium phosphate precipitation, direct micro injection, nanoparticle-mediated nucleic acid delivery (see, e.g., Panyam et., al Adv Drug Deliv Rev. 2012 Sep 13. pii: S0169-409X(12)00283-9. doi: 10.1016/j.addr.2012.09.023), and the like.

[00238] In some cases, a CasZ polypeptide of the present disclosure (e.g., wild type protein, variant protein, chimeric/fusion protein, dCasZ, etc.) is provided as a nucleic acid (e.g., an mRNA, a DNA, a plasmid, an expression vector, a viral vector, etc.) that encodes the CasZ polypeptide. In some cases, the CasZ polypeptide of the present disclosure is provided directly as a protein (e.g., without an

associated guide RNA or with an associate guide RNA, i.e., as a ribonucleoprotein complex). A CasZ polypeptide of the present disclosure can be introduced into a cell (provided to the cell) by any convenient method; such methods are known to those of ordinary skill in the art. As an illustrative example, a CasZ polypeptide of the present disclosure can be injected directly into a cell (e.g., with or without a CasZ guide RNA or nucleic acid encoding a CasZ guide RNA, and with or without a donor polynucleotide and with or without a CasZ trancRNA). As another example, a preformed complex of a CasZ polypeptide of the present disclosure and a CasZ guide RNA (an RNP) can be introduced into a cell (e.g, eukaryotic cell) (e.g., via injection, via nucleofection; via a protein transduction domain (PTD) conjugated to one or more components, e.g., conjugated to the CasZ protein, conjugated to a guide RNA, conjugated to a CasZ trancRNA, conjugated to a CasZ polypeptide of the present disclosure and a guide RNA; etc.).

[00239] In some cases, a nucleic acid (e.g., a CasZ guide RNA and/or a nucleic acid encoding it, a nucleic acid encoding a CasZ protein, a CasZ trancRNA and/or a nucleic acid encoding it, and the like) and/or a polypeptide (e.g., a CasZ polypeptide; a CasZ fusion polypeptide) is delivered to a cell (e.g., a target host cell) in a particle, or associated with a particle. In some cases, a CasZ system of the present disclosure is delivered to a cell in a particle, or associated with a particle. The terms "particle" and nanoparticle" can be used interchangeable, as appropriate. For example, a recombinant expression vector comprising a nucleotide sequence encoding a CasZ polypeptide of the present disclosure and/or a CasZ guide RNA, an mRNA comprising a nucleotide sequence encoding a CasZ polypeptide of the present disclosure, and guide RNA may be delivered simultaneously using particles or lipid envelopes; for instance, a CasZ polypeptide and/or a CasZ guide RNA and/or a trancRNA, e.g., as a complex (e.g., a ribonucleoprotein (RNP) complex), can be delivered via a particle, e.g., a delivery particle comprising lipid or lipidoid and hydrophilic polymer, e.g., a cationic lipid and a hydrophilic polymer, for instance wherein the cationic lipid comprises 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP) or 1,2-ditetradecanoyl-sn-glycero-3-phosphocholine (DMPC) and/or wherein the hydrophilic polymer comprises ethylene glycol or polyethylene glycol (PEG); and/or wherein the particle further comprises cholesterol (e.g., particle from formulation 1=DOTAP 100, DMPC 0, PEG 0, Cholesterol 0; formulation number 2=DOTAP 90, DMPC 0, PEG 10, Cholesterol 0; formulation number 3=DOTAP 90, DMPC 0, PEG 5, Cholesterol 5). For example, a particle can be formed using a multistep process in which a CasZ polypepide and a CasZ guideRNA are mixed together, e.g., at a 1:1 molar ratio, e.g., at room temperature, e.g., for 30 minutes, e.g., in sterile, nuclease free 1 x phosphate-buffered saline (PBS); and separately, DOTAP, DMPC, PEG, and cholesterol as applicable for the formulation are dissolved in alcohol, e.g., 100% ethanol; and, the two solutions are mixed together to form particles containing the complexes).

[00240] A CasZ polypeptide of the present disclosure (or an mRNA comprising a nucleotide sequence encoding a CasZ polypeptide of the present disclosure; or a recombinant expression vector comprising a nucleotide sequence encoding a CasZ polypeptide of the present disclosure) and/or CasZ guide RNA (or a nucleic acid such as one or more expression vectors encoding the CasZ guide RNA) may be delivered simultaneously using particles or lipid envelopes. For example, a biodegradable core-shell structured nanoparticle with a poly (β-amino ester) (PBAE) core enveloped by a phospholipid bilayer shell can be used. In some cases, particles/nanoparticles based on self assembling bioadhesive polymers are used; such particles/nanoparticles may be applied to oral delivery of peptides, intravenous delivery of peptides and nasal delivery of peptides, e.g., to the brain. Other embodiments, such as oral absorption and ocular delivery of hydrophobic drugs are also contemplated. A molecular envelope technology, which involves an engineered polymer envelope which is protected and delivered to the site of the disease, can be used. Doses of about 5 mg/kg can be used, with single or multiple doses, depending on various factors, e.g., the target tissue.

[00241] Lipidoid compounds (e.g., as described in US patent application 20110293703) are also useful in the administration of polynucleotides, and can be used to deliver a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure, or a CasZ system of the present disclosure. In one aspect, the aminoalcohol lipidoid compounds are combined with an agent to be delivered to a cell or a subject to form microparticles, nanoparticles, liposomes, or micelles. The aminoalcohol lipidoid compounds may be combined with other aminoalcohol lipidoid compounds, polymers (synthetic or natural), surfactants, cholesterol, carbohydrates, proteins, lipids, etc. to form the particles. These particles may then optionally be combined with a pharmaceutical excipient to form a pharmaceutical composition.

[00242] A poly(beta-amino alcohol) (PBAA) can be used to deliver a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA), or a CasZ system of the present disclosure, to a target cell. US Patent Publication No. 20130302401 relates to a class of poly(beta-amino alcohols) (PBAAs) that has been prepared using combinatorial polymerization.

[00243] Sugar-based particles may be used, for example GalNAc, as described with reference to WO2014118272 (incorporated herein by reference) and Nair, J K et al., 2014, Journal of the American Chemical Society 136 (49), 16958-16961) can be used to deliver a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure, or a CasZ system of the present disclosure, to a target cell.

[00244] In some cases, lipid nanoparticles (LNPs) are used to deliver a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure,

a nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA), or a CasZ system of the present disclosure, to a target cell. Negatively charged polymers such as RNA may be loaded into LNPs at low pH values (e.g., pH 4) where the ionizable lipids display a positive charge. However, at physiological pH values, the LNPs exhibit a low surface charge compatible with longer circulation times. Four species of ionizable cationic lipids have been focused upon, namely 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxy-keto-N,N-dimethyl-3-aminopropane (DLinKDMA), and 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA). Preparation of LNPs and is described in, e.g., Rosin et al. (2011) Molecular Therapy 19:1286-2200). The cationic lipids 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA), (3-o-[2"-(methoxypolyethyleneglycol 2000) succinoyl]-1,2-dimyristoyl-sn-glycol (PEG-S-DMG), and R-3-[(.omega.-methoxy-poly(ethylene glycol)2000) carbamoyl]-1,2-dimyristyloxlpropyl-3-amine (PEG-C-DOMG) may be used. A nucleic acid (e.g., a CasZ guide RNA; a nucleic acid of the present disclosure; etc.) may be encapsulated in LNPs containing DLinDAP, DLinDMA, DLinK-DMA, and DLinKC2-DMA (cationic lipid:DSPC:CHOL: PEGS-DMG or PEG-C-DOMG at 40:10:40:10 molar ratios). In some cases, 0.2% SP-DiOC18 is incorporated.

[00245] Spherical Nucleic Acid (SNA™) constructs and other nanoparticles (particularly gold nanoparticles) can be used to deliver a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA), or a CasZ system of the present disclosure, to a target cell.. See, e.g., Cutler et al., J. Am. Chem. Soc. 2011 133:9254-9257, Hao et al., Small. 2011 7:3158-3162, Zhang et al., ACS Nano. 2011 5:6962-6970, Cutler et al., J. Am. Chem. Soc. 2012 134:1376-1391, Young et al., Nano Lett. 2012 12:3867-71, Zheng et al., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80, Mirkin, Nanomedicine 2012 7:635-638 Zhang et al., J. Am. Chem. Soc. 2012 134:16488-1691, Weintraub, Nature 2013 495:S14-S16, Choi et al., Proc. Natl. Acad. Sci. USA. 2013 110(19): 7625-7630, Jensen et al., Sci. Transl. Med. 5, 209ra152 (2013) and Mirkin, et al., Small, 10:186-192.

[00246] Self-assembling nanoparticles with RNA may be constructed with polyethyleneimine (PEI) that is PEGylated with an Arg-Gly-Asp (RGD) peptide ligand attached at the distal end of the polyethylene glycol (PEG).

[00247] In general, a "nanoparticle" refers to any particle having a diameter of less than 1000 nm. In some cases, nanoparticles suitable for use in delivering a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a

nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA), or a CasZ system of the present disclosure, to a target cell have a diameter of 500 nm or less, e.g., from 25 nm to 35 nm, from 35 nm to 50 nm, from 50 nm to 75 nm, from 75 nm to 100 nm, from 100 nm to 150 nm, from 150 nm to 200 nm, from 200 nm to 300 nm, from 300 nm to 400 nm, or from 400 nm to 500 nm. In some cases, nanoparticles suitable for use in delivering a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure, or a CasZ system of the present disclosure, to a target cell have a diameter of from 25 nm to 200 nm. In some cases, nanoparticles suitable for use in delivering a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure, or a CasZ system of the present disclosure, to a target cell have a diameter of 100 nm or less In some cases, nanoparticles suitable for use in delivering a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure, or a CasZ system of the present disclosure, to a target cell have a diameter of from 35 nm to 60 nm.

[00248] Nanoparticles suitable for use in delivering a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA), or a CasZ system of the present disclosure, to a target cell may be provided in different forms, e.g., as solid nanoparticles (e.g., metal such as silver, gold, iron, titanium), non-metal, lipid-based solids, polymers), suspensions of nanoparticles, or combinations thereof. Metal, dielectric, and semiconductor nanoparticles may be prepared, as well as hybrid structures (e.g., core-shell nanoparticles). Nanoparticles made of semiconducting material may also be labeled quantum dots if they are small enough (typically below 10 nm) that quantization of electronic energy levels occurs. Such nanoscale particles are used in biomedical applications as drug carriers or imaging agents and may be adapted for similar purposes in the present disclosure.

[00249] Semi-solid and soft nanoparticles are also suitable for use in delivering a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA), or a CasZ system of the present disclosure, to a target cell. A prototype nanoparticle of semi-solid nature is the liposome.

[00250] In some cases, an exosome is used to deliver a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA), or a CasZ system of the present disclosure, to a target cell. Exosomes are endogenous nano-vesicles that transport RNAs and proteins, and which can deliver RNA to the brain and other target organs.

[00251] In some cases, a liposome is used to deliver a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA), or a CasZ system of the present disclosure, to a target cell. Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. Liposomes can be made from several different types of lipids; however, phospholipids are most commonly used to generate liposomes. Although liposome formation is spontaneous when a lipid film is mixed with an aqueous solution, it can also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus. Several other additives may be added to liposomes in order to modify their structure and properties. For instance, either cholesterol or sphingomyelin may be added to the liposomal mixture in order to help stabilize the liposomal structure and to prevent the leakage of the liposomal inner cargo. A liposome formulation may be mainly comprised of natural phospholipids and lipids such as 1,2-distearoryl-sn-glycero-3-phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines and monosialoganglioside.

[00252] A stable nucleic-acid-lipid particle (SNALP) can be used to deliver a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA), or a CasZ system of the present disclosure, to a target cell. The SNALP formulation may contain the lipids 3-N-[(methoxypoly(ethylene glycol) 2000) carbamoyl]-1,2-dimyristyloxy-propylamine (PEG-C-DMA), 1,2-dilinoleyloxy-N,N-dimethyl-3-aminopropane (DLinDMA), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC) and cholesterol, in a 2:40:10:48 molar percent ratio. The SNALP liposomes may be prepared by formulating D-Lin-DMA and PEG-C-DMA with distearoylphosphatidylcholine (DSPC), Cholesterol and siRNA using a 25:1 lipid/siRNA ratio and a 48/40/10/2 molar ratio of Cholesterol/D-Lin-DMA/DSPC/PEG-C-DMA. The resulting SNALP liposomes can be about 80-100 nm in size. A SNALP may comprise synthetic cholesterol (Sigma-Aldrich, St Louis, Mo., USA), dipalmitoylphosphatidylcholine (Avanti Polar Lipids, Alabaster, Ala., USA), 3-N-[(w-methoxy poly(ethylene glycol)2000)carbamoyl]-1,2-dimyrestyloxypropylamine, and cationic 1,2-dilinoleyloxy-3-N,Ndimethylaminopropane. A SNALP may comprise synthetic cholesterol (Sigma-Aldrich), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC; Avanti Polar Lipids Inc.), PEG-cDMA, and 1,2-dilinoleyloxy-3-(N;N-dimethyl)aminopropane (DLinDMA).

[00253] Other cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane (DLin-KC2-DMA) can be used to deliver a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA), or a CasZ system of the present

disclosure, to a target cell. A preformed vesicle with the following lipid composition may be contemplated: amino lipid, distearoylphosphatidylcholine (DSPC), cholesterol and (R)-2,3-bis(octadecyloxy) propyl-1-(methoxy poly(ethylene glycol)2000)propylcarbamate (PEG-lipid) in the molar ratio 40/10/40/10, respectively, and a FVII siRNA/total lipid ratio of approximately 0.05 (w/w). To ensure a narrow particle size distribution in the range of 70-90 nm and a low polydispersity index of 0.11.+-.0.04 (n=56), the particles may be extruded up to three times through 80 nm membranes prior to adding the guide RNA. Particles containing the highly potent amino lipid 16 may be used, in which the molar ratio of the four lipid components 16, DSPC, cholesterol and PEG-lipid (50/10/38.5/1.5) which may be further optimized to enhance in vivo activity.

[00254] Lipids may be formulated with a CasZ system of the present disclosure or component(s) thereof or nucleic acids encoding the same to form lipid nanoparticles (LNPs). Suitable lipids include, but are not limited to, DLin-KC2-DMA4, C12-200 and colipids disteroylphosphatidyl choline, cholesterol, and PEG-DMG may be formulated with a CasZ system, or component thereof, of the present disclosure, using a spontaneous vesicle formation procedure. The component molar ratio may be about 50/10/38.5/1.5 (DLin-KC2-DMA or C12-200/disteroylphosphatidyl choline/cholesterol/PEG-DMG).

[00255] A CasZ system of the present disclosure, or a component thereof, may be delivered encapsulated in PLGA microspheres such as that further described in US published applications 20130252281 and 20130245107 and 20130244279.

[00256] Supercharged proteins can be used to deliver a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA), or a CasZ system of the present disclosure, to a target cell. Supercharged proteins are a class of engineered or naturally occurring proteins with unusually high positive or negative net theoretical charge. Both supernegatively and superpositively charged proteins exhibit the ability to withstand thermally or chemically induced aggregation. Superpositively charged proteins are also able to penetrate mammalian cells. Associating cargo with these proteins, such as plasmid DNA, RNA, or other proteins, can enable the functional delivery of these macromolecules into mammalian cells both in vitro and in vivo.

[00257] Cell Penetrating Peptides (CPPs) can be used to deliver a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA), or a CasZ system of the present disclosure, to a target cell. CPPs typically have an amino acid composition that either contains a high relative abundance of positively charged amino acids such as lysine or arginine or has sequences that contain an alternating pattern of polar/charged amino acids and non-polar, hydrophobic amino acids.

[00258] An implantable device can be used to deliver a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA) (e.g., a CasZ guide RNA, a nucleic acid encoding a CasZ guide RNA, a nucleic acid encoding CasZ polypeptide, a donor template, and the like), or a CasZ system of the present disclosure, to a target cell (e.g., a target cell *in vivo*, where the target cell is a target cell in circulation, a target cell in a tissue, a target cell in an organ, etc.). An implantable device suitable for use in delivering a CasZ polypeptide of the present disclosure, a CasZ fusion polypeptide of the present disclosure, an RNP of the present disclosure, a nucleic acid of the present disclosure (e.g., a CasZ guide RNA and/or a CasZ trancRNA), or a CasZ system of the present disclosure, to a target cell (e.g., a target cell *in vivo*, where the target cell is a target cell in circulation, a target cell in a tissue, a target cell in an organ, etc.) can include a container (e.g., a reservoir, a matrix, etc.) that comprises the CasZ polypeptide, the CasZ fusion polypeptide, the RNP, or the CasZ system (or component thereof, e.g., a nucleic acid of the present disclosure).

[00259] A suitable implantable device can comprise a polymeric substrate, such as a matrix for example, that is used as the device body, and in some cases additional scaffolding materials, such as metals or additional polymers, and materials to enhance visibility and imaging. An implantable delivery device can be advantageous in providing release locally and over a prolonged period, where the polypeptide and/or nucleic acid to be delivered is released directly to a target site, e.g., the extracellular matrix (ECM), the vasculature surrounding a tumor, a diseased tissue, etc. Suitable implantable delivery devices include devices suitable for use in delivering to a cavity such as the abdominal cavity and/or any other type of administration in which the drug delivery system is not anchored or attached, comprising a biostable and/or degradable and/or bioabsorbable polymeric substrate, which may for example optionally be a matrix. In some cases, a suitable implantable drug delivery device comprises degradable polymers, wherein the main release mechanism is bulk erosion. In some cases, a suitable implantable drug delivery device comprises non degradable, or slowly degraded polymers, wherein the main release mechanism is diffusion rather than bulk erosion, so that the outer part functions as membrane, and its internal part functions as a drug reservoir, which practically is not affected by the surroundings for an extended period (for example from about a week to about a few months). Combinations of different polymers with different release mechanisms may also optionally be used. The concentration gradient at the can be maintained effectively constant during a significant period of the total releasing period, and therefore the diffusion rate is effectively constant (termed "zero mode" diffusion). By the term "constant" it is meant a diffusion rate that is maintained above the lower threshold of therapeutic effectiveness, but which may still optionally feature an initial burst and/or may fluctuate, for example increasing and decreasing to a certain degree. The diffusion rate can be so maintained for a prolonged period, and it can be considered

constant to a certain level to optimize the therapeutically effective period, for example the effective silencing period.

[00260] In some cases, the implantable delivery system is designed to shield the nucleotide based therapeutic agent from degradation, whether chemical in nature or due to attack from enzymes and other factors in the body of the subject.

[00261] The site for implantation of the device, or target site, can be selected for maximum therapeutic efficacy. For example, a delivery device can be implanted within or in the proximity of a tumor environment, or the blood supply associated with a tumor. The target location can be, e.g.: 1) the brain at degenerative sites like in Parkinson or Alzheimer disease at the basal ganglia, white and gray matter; 2) the spine, as in the case of amyotrophic lateral sclerosis (ALS); 3) uterine cervix; 4) active and chronic inflammatory joints; 5) dermis as in the case of psoriasis; 7) sympathetic and sensoric nervous sites for analgesic effect; 7) a bone; 8) a site of acute or chronic infection; 9) Intra vaginal; 10) Inner ear--auditory system, labyrinth of the inner ear, vestibular system; 11) Intra tracheal; 12) Intra-cardiac; coronary, epicardiac; 13) urinary tract or bladder; 14) biliary system; 15) parenchymal tissue including and not limited to the kidney, liver, spleen; 16) lymph nodes; 17) salivary glands; 18) dental gums; 19) Intra-articular (into joints); 20) Intra-ocular; 21) Brain tissue; 22) Brain ventricles; 23) Cavities, including abdominal cavity (for example but without limitation, for ovary cancer); 24) Intra esophageal; and 25) Intra rectal; and 26) into the vasculature.

[00262] The method of insertion, such as implantation, may optionally already be used for other types of tissue implantation and/or for insertions and/or for sampling tissues, optionally without modifications, or alternatively optionally only with non-major modifications in such methods. Such methods optionally include but are not limited to brachytherapy methods, biopsy, endoscopy with and/or without ultrasound, such as stereotactic methods into the brain tissue, laparoscopy, including implantation with a laparoscope into joints, abdominal organs, the bladder wall and body cavities.

MODIFIED HOST CELLS

[00263] The present disclosure provides a modified cell comprising a CasZ polypeptide of the present disclosure and/or a nucleic acid comprising a nucleotide sequence encoding a CasZ polypeptide of the present disclosure. The present disclosure provides a modified cell comprising a CasZ polypeptide of the present disclosure, where the modified cell is a cell that does not normally comprise a CasZ polypeptide of the present disclosure. The present disclosure provides a modified cell (e.g., a genetically modified cell) comprising nucleic acid comprising a nucleotide sequence encoding a CasZ polypeptide of the present disclosure. The present disclosure provides a genetically modified cell that is genetically modified with an mRNA comprising a nucleotide sequence encoding a CasZ polypeptide of the present disclosure. The present disclosure provides a genetically modified cell that is genetically modified with a recombinant expression vector comprising a nucleotide sequence encoding a CasZ polypeptide of the

present disclosure. The present disclosure provides a genetically modified cell that is genetically modified with a recombinant expression vector comprising: a) a nucleotide sequence encoding a CasZ polypeptide of the present disclosure; and b) a nucleotide sequence encoding a CasZ guide RNA of the present disclosure. The present disclosure provides a genetically modified cell that is genetically modified with a recombinant expression vector comprising: a) a nucleotide sequence encoding a CasZ polypeptide of the present disclosure; b) a nucleotide sequence encoding a CasZ guide RNA of the present disclosure; and c) a nucleotide sequence encoding a donor template.

[00264] A cell that serves as a recipient for a CasZ polypeptide of the present disclosure and/or a nucleic acid comprising a nucleotide sequence encoding a CasZ polypeptide of the present disclosure and/or a CasZ guide RNA of the present disclosure (or a nucleic acid encoding it) and/or a CasZ trancRNA (or a nucleic acid encoding it), can be any of a variety of cells, including, e.g., *in vitro* cells; *in vivo* cells; *ex vivo* cells; primary cells; cancer cells; animal cells; plant cells; algal cells; fungal cells; etc. A cell that serves as a recipient for a CasZ polypeptide of the present disclosure and/or a nucleic acid comprising a nucleotide sequence encoding a CasZ polypeptide of the present disclosure and/or a CasZ guide RNA of the present disclosure is referred to as a "host cell" or a "target cell." A host cell or a target cell can be a recipient of a CasZ system of the present disclosure. A host cell or a target cell can be a recipient of a CasZ RNP of the present disclosure. A host cell or a target cell can be a recipient of a single component of a CasZ system of the present disclosure.

[00265] Non-limiting examples of cells (target cells) include: a prokaryotic cell, eukaryotic cell, a bacterial cell, an archaeal cell, a cell of a single-cell eukaryotic organism, a protozoa cell, a cell from a plant (e.g., cells from plant crops, fruits, vegetables, grains, soy bean, corn, maize, wheat, seeds, tomatos, rice, cassava, sugarcane, pumpkin, hay, potatos, cotton, cannabis, tobacco, flowering plants, conifers, gymnosperms, angiosperms, ferns, clubmosses, hornworts, liverworts, mosses, dicotyledons, monocotyledons, etc.), an algal cell, (e.g., *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens*, *C. agardh*, and the like), seaweeds (e.g. kelp) a fungal cell (e.g., a yeast cell, a cell from a mushroom), an animal cell, a cell from an invertebrate animal (e.g., fruit fly, cnidarian, echinoderm, nematode, etc.), a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal), a cell from a mammal (e.g., an ungulate (e.g., a pig, a cow, a goat, a sheep); a rodent (e.g., a rat, a mouse); a non-human primate; a human; a feline (e.g., a cat); a canine (e.g., a dog); etc.), and the like. In some cases, the cell is a cell that does not originate from a natural organism (e.g., the cell can be a synthetically made cell; also referred to as an artificial cell).

[00266] A cell can be an *in vitro* cell (e.g., a cell in culture, e.g., an established cultured cell line). A cell can be an *ex vivo* cell (cultured cell from an individual). A cell can be and *in vivo* cell (e.g., a cell in an individual). A cell can be an isolated cell. A cell can be a cell inside of an organism. A cell can be an organism. A cell can be a cell in a cell culture (e.g., *in vitro* cell culture). A cell can be one of a

collection of cells. A cell can be a prokaryotic cell or derived from a prokaryotic cell. A cell can be a bacterial cell or can be derived from a bacterial cell. A cell can be an archaeal cell or derived from an archaeal cell. A cell can be a eukaryotic cell or derived from a eukaryotic cell. A cell can be a plant cell or derived from a plant cell. A cell can be an animal cell or derived from an animal cell. A cell can be an invertebrate cell or derived from an invertebrate cell. A cell can be a vertebrate cell or derived from a vertebrate cell. A cell can be a mammalian cell or derived from a mammalian cell. A cell can be a rodent cell or derived from a rodent cell. A cell can be a human cell or derived from a human cell. A cell can be a microbe cell or derived from a microbe cell. A cell can be a fungi cell or derived from a fungi cell. A cell can be an insect cell. A cell can be an arthropod cell. A cell can be a protozoan cell. A cell can be a helminth cell.

[00267] Suitable cells include a stem cell (e.g. an embryonic stem (ES) cell, an induced pluripotent stem (iPS) cell; a germ cell (e.g., an oocyte, a sperm, an oogonia, a spermatogonia, etc.); a somatic cell, e.g. a fibroblast, an oligodendrocyte, a glial cell, a hematopoietic cell, a neuron, a muscle cell, a bone cell, a hepatocyte, a pancreatic cell, etc.

[00268] Suitable cells include human embryonic stem cells, fetal cardiomyocytes, myofibroblasts, mesenchymal stem cells, autotransplated expanded cardiomyocytes, adipocytes, totipotent cells, pluripotent cells, blood stem cells, myoblasts, adult stem cells, bone marrow cells, mesenchymal cells, embryonic stem cells, parenchymal cells, epithelial cells, endothelial cells, mesothelial cells, fibroblasts, osteoblasts, chondrocytes, exogenous cells, endogenous cells, stem cells, hematopoietic stem cells, bone-marrow derived progenitor cells, myocardial cells, skeletal cells, fetal cells, undifferentiated cells, multi-potent progenitor cells, unipotent progenitor cells, monocytes, cardiac myoblasts, skeletal myoblasts, macrophages, capillary endothelial cells, xenogenic cells, allogenic cells, and post-natal stem cells.

[00269] In some cases, the cell is an immune cell, a neuron, an epithelial cell, and endothelial cell, or a stem cell. In some cases, the immune cell is a T cell, a B cell, a monocyte, a natural killer cell, a dendritic cell, or a macrophage. In some cases, the immune cell is a cytotoxic T cell. In some cases, the immune cell is a helper T cell. In some cases, the immune cell is a regulatory T cell (Treg).

[00270] In some cases, the cell is a stem cell. Stem cells include adult stem cells. Adult stem cells are also referred to as somatic stem cells.

[00271] Adult stem cells are resident in differentiated tissue, but retain the properties of self-renewal and ability to give rise to multiple cell types, usually cell types typical of the tissue in which the stem cells are found. Numerous examples of somatic stem cells are known to those of skill in the art, including muscle stem cells; hematopoietic stem cells; epithelial stem cells; neural stem cells;

mesenchymal stem cells; mammary stem cells; intestinal stem cells; mesodermal stem cells; endothelial stem cells; olfactory stem cells; neural crest stem cells; and the like.

[00272] Stem cells of interest include mammalian stem cells, where the term "mammalian" refers to any animal classified as a mammal, including humans; non-human primates; domestic and farm animals; and zoo, laboratory, sports, or pet animals, such as dogs, horses, cats, cows, mice, rats, rabbits, etc. In some cases, the stem cell is a human stem cell. In some cases, the stem cell is a rodent (e.g., a mouse; a rat) stem cell. In some cases, the stem cell is a non-human primate stem cell.

[00273] Stem cells can express one or more stem cell markers, e.g., SOX9, KRT19, KRT7, LGR5, CA9, FXYD2, CDH6, CLDN18, TSPAN8, BPIFB1, OLFM4, CDH17, and PPARGC1A.

[00274] In some cases, the stem cell is a hematopoietic stem cell (HSC). HSCs are mesoderm-derived cells that can be isolated from bone marrow, blood, cord blood, fetal liver and yolk sac. HSCs are characterized as $CD34^+$ and $CD3^-$. HSCs can repopulate the erythroid, neutrophil-macrophage, megakaryocyte and lymphoid hematopoietic cell lineages *in vivo*. *In vitro*, HSCs can be induced to undergo at least some self-renewing cell divisions and can be induced to differentiate to the same lineages as is seen *in vivo*. As such, HSCs can be induced to differentiate into one or more of erythroid cells, megakaryocytes, neutrophils, macrophages, and lymphoid cells.

[00275] In other cases, the stem cell is a neural stem cell (NSC). Neural stem cells (NSCs) are capable of differentiating into neurons, and glia (including oligodendrocytes, and astrocytes). A neural stem cell is a multipotent stem cell which is capable of multiple divisions, and under specific conditions can produce daughter cells which are neural stem cells, or neural progenitor cells that can be neuroblasts or glioblasts, e.g., cells committed to become one or more types of neurons and glial cells respectively. Methods of obtaining NSCs are known in the art.

[00276] In other cases, the stem cell is a mesenchymal stem cell (MSC). MSCs originally derived from the embryonal mesoderm and isolated from adult bone marrow, can differentiate to form muscle, bone, cartilage, fat, marrow stroma, and tendon. Methods of isolating MSC are known in the art; and any known method can be used to obtain MSC. See, e.g., U.S. Pat. No. 5,736,396, which describes isolation of human MSC.

[00277] A cell is in some cases a plant cell. A plant cell can be a cell of a monocotyledon. A cell can be a cell of a dicotyledon.

[00278] In some cases, the cell is a plant cell. For example, the cell can be a cell of a major agricultural plant, e.g., Barley, Beans (Dry Edible), Canola, Corn, Cotton (Pima), Cotton (Upland), Flaxseed, Hay (Alfalfa), Hay (Non-Alfalfa), Oats, Peanuts, Rice, Sorghum, Soybeans, Sugarbeets, Sugarcane, Sunflowers (Oil), Sunflowers (Non-Oil), Sweet Potatoes , Tobacco (Burley), Tobacco (Flue-cured), Tomatoes, Wheat (Durum), Wheat (Spring), Wheat (Winter), and the like. As another example,

the cell is a cell of a vegetable crops which include but are not limited to, e.g., alfalfa sprouts, aloe leaves, arrow root, arrowhead, artichokes, asparagus, bamboo shoots, banana flowers, bean sprouts, beans, beet tops, beets, bittermelon, bok choy, broccoli, broccoli rabe (rappini), brussels sprouts, cabbage, cabbage sprouts, cactus leaf (nopales), calabaza, cardoon, carrots, cauliflower, celery, chayote, chinese artichoke (crosnes), chinese cabbage, chinese celery, chinese chives, choy sum, chrysanthemum leaves (tung ho), collard greens, corn stalks, corn-sweet, cucumbers, daikon, dandelion greens, dasheen, dau mue (pea tips), donqua (winter melon), eggplant, endive, escarole, fiddle head ferns, field cress, frisee, gai choy (chinese mustard), gailon, galanga (siam, thai ginger), garlic, ginger root, gobo, greens, hanover salad greens, huauzontle, jerusalem artichokes, jicama, kale greens, kohlrabi, lamb's quarters (quilete), lettuce (bibb), lettuce (boston), lettuce (boston red), lettuce (green leaf), lettuce (iceberg), lettuce (lolla rossa), lettuce (oak leaf - green), lettuce (oak leaf - red), lettuce (processed), lettuce (red leaf), lettuce (romaine), lettuce (ruby romaine), lettuce (russian red mustard), linkok, lo bok, long beans, lotus root, mache, maguey (agave) leaves, malanga, mesculin mix, mizuna, moap (smooth luffa), moo, moqua (fuzzy squash), mushrooms, mustard, nagaimo, okra, ong choy, onions green, opo (long squash), ornamental corn, ornamental gourds, parsley, parsnips, peas, peppers (bell type), peppers, pumpkins, radicchio, radish sprouts, radishes, rape greens, rape greens, rhubarb, romaine (baby red), rutabagas, salicornia (sea bean), sinqua (angled/ridged luffa), spinach, squash, straw bales, sugarcane, sweet potatoes, swiss chard, tamarindo, taro, taro leaf, taro shoots, tatsoi, tepeguaje (guaje), tindora, tomatillos, tomatoes, tomatoes (cherry), tomatoes (grape type), tomatoes (plum type), tumeric, turnip tops greens, turnips, water chestnuts, yampi, yams (names), yu choy, yuca (cassava), and the like.

[00279] A cell is in some cases an arthropod cell. For example, the cell can be a cell of a sub-order, a family, a sub-family, a group, a sub-group, or a species of, e.g., *Chelicerata*, *Myriapodia*, *Hexipodia*, *Arachnida*, *Insecta*, *Archaeognatha*, *Thysanura*, *Palaeoptera*, *Ephemeroptera*, *Odonata*, *Anisoptera*, *Zygoptera*, *Neoptera*, *Exopterygota*, *Plecoptera* , *Embioptera* , *Orthoptera*, *Zoraptera* , *Dermaptera*, *Dictyoptera*, *Notoptera*, *Grylloblattidae*, *Mantophasmatidae*, *Phasmatodea* , *Blattaria*, *Isoptera*, *Mantodea*, *Parapneuroptera*, *Psocoptera*, *Thysanoptera*, *Phthiraptera*, *Hemiptera*, *Endopterygota or Holometabola* , *Hymenoptera* , *Coleoptera*, *Strepsiptera*, *Raphidioptera*, *Megaloptera*, *Neuroptera* , *Mecoptera* , *Siphonaptera*, *Diptera*, *Trichoptera*, or *Lepidoptera.*

[00280] A cell is in some cases an insect cell. For example, in some cases, the cell is a cell of a mosquito, a grasshopper, a true bug, a fly, a flea, a bee, a wasp, an ant, a louse, a moth, or a beetle.

KITS

[00281] The present disclosure provides a kit comprising a CasZ system of the present disclosure, or a component of a CasZ system of the present disclosure.

[00282] A kit of the present disclosure can comprise any combination as listed for a CasZ system (e.g., see above). A kit of the present disclosure can comprise: a) a component, as described above, of a

CasZ system of the present disclosure, or can comprise a CasZ system of the present disclosure; and b) one or more additional reagents, e.g., i) a buffer; ii) a protease inhibitor; iii) a nuclease inhibitor; iv) a reagent required to develop or visualize a detectable label; v) a positive and/or negative control target DNA; vi) a positive and/or negative control CasZ guide RNA; vii) a CasZ trancRNA; and the like. A kit of the present disclosure can comprise: a) a component, as described above, of a CasZ system of the present disclosure, or can comprise a CasZ system of the present disclosure; and b) a therapeutic agent.

[00283] A kit of the present disclosure can comprise a recombinant expression vector comprising: a) an insertion site for inserting a nucleic acid comprising a nucleotide sequence encoding a portion of a CasZ guide RNA that hybridizes to a target nucleotide sequence in a target nucleic acid; and b) a nucleotide sequence encoding the CasZ-binding portion of a CasZ guide RNA. A kit of the present disclosure can comprise a recombinant expression vector comprising: a) an insertion site for inserting a nucleic acid comprising a nucleotide sequence encoding a portion of a CasZ guide RNA that hybridizes to a target nucleotide sequence in a target nucleic acid; b) a nucleotide sequence encoding the CasZ-binding portion of a CasZ guide RNA; and c) a nucleotide sequence encoding a CasZ polypeptide of the present disclosure. A kit of the present disclosure can comprise a recombinant expression vector comprising a nucleotide sequence encoding a CasZ trancRNA.

DETECTION OF SSDNA

[00284] A CasZ (Cas14) polypeptide of the present disclosure, once activated by detection of a target DNA (double or single stranded), can promiscuously cleave non-targeted single stranded DNA (ssDNA). Once a CasZ (Cas14) is activated by a guide RNA, which occurs when the guide RNA hybridizes to a target sequence of a target DNA (i.e., the sample includes the target DNA, e.g., target ssDNA), the protein becomes a nuclease that promiscuously cleaves ssDNAs (i.e., the nuclease cleaves non-target ssDNAs, i.e., ssDNAs to which the guide sequence of the guide RNA does not hybridize). Thus, when the target DNA is present in the sample (e.g., in some cases above a threshold amount), the result is cleavage of ssDNAs in the sample, which can be detected using any convenient detection method (e.g., using a labeled single stranded detector DNA). In some cases, a CasZ polypeptide requires, in addition to a CasZ guide RNA, a tranc RNA for activation.

[00285] Provided are compositions and methods for detecting a target DNA (double stranded or single stranded) in a sample. In some cases, a detector DNA is used that is single stranded (ssDNA) and does not hybridize with the guide sequence of the guide RNA (i.e., the detector ssDNA is a non-target ssDNA). Such methods can include (a) contacting the sample with: (i) a CasZ polypeptide; (ii) a guide RNA comprising: a region that binds to the CasZ polypeptide, and a guide sequence that hybridizes with the target DNA; and (iii) a detector DNA that is single stranded and does not hybridize with the guide sequence of the guide RNA; and (b) measuring a detectable signal produced by cleavage of the single stranded detector DNA by the CasZ polypeptide, thereby detecting the target DNA. . In some cases, the

methods include can include (a) contacting the sample with: (i) a CasZ polypeptide; (ii) a guide RNA comprising: a region that binds to the CasZ polypeptide, and a guide sequence that hybridizes with the target DNA; (iii) a CasZ tranc RNA; and (iv) a detector DNA that is single stranded and does not hybridize with the guide sequence of the guide RNA; and (b) measuring a detectable signal produced by cleavage of the single stranded detector DNA by the CasZ polypeptide, thereby detecting the target DNA. As noted above, once a subject CasZ polypeptide protein is activated by a guide RNA, which occurs when the sample includes a target DNA to which the guide RNA hybridizes (i.e., the sample includes the targeted target DNA), the CasZ polypeptide is activated and functions as an endoribonuclease that non-specifically cleaves ssDNAs (including non-target ssDNAs) present in the sample. Thus, when the targeted target DNA is present in the sample (e.g., in some cases above a threshold amount), the result is cleavage of ssDNA (including non-target ssDNA) in the sample, which can be detected using any convenient detection method (e.g., using a labeled detector ssDNA).

[00286] Also provided are compositions and methods for cleaving single stranded DNAs (ssDNAs) (e.g., non-target ssDNAs). Such methods can include contacting a population of nucleic acids, wherein said population comprises a target DNA and a plurality of non-target ssDNAs, with: (i) a CasZ polypeptide; and (ii) a guide RNA comprising: a region that binds to the CasZ polypeptide, and a guide sequence that hybridizes with the target DNA, wherein the CasZ polypeptide cleaves non-target ssDNAs of said plurality. Such methods can include contacting a population of nucleic acids, wherein said population comprises a target DNA and a plurality of non-target ssDNAs, with: (i) a CasZ polypeptide; (ii) a guide RNA comprising: a region that binds to the CasZ polypeptide, and a guide sequence that hybridizes with the target DNA, and (iii) a CasZ tranc RNA, wherein the CasZ polypeptide cleaves non-target ssDNAs of said plurality. Such methods can be used, e.g., to cleave foreign ssDNAs (e.g., viral DNAs) in a cell.

[00287] The contacting step of a subject method can be carried out in a composition comprising divalent metal ions. The contacting step can be carried out in an acellular environment, e.g., outside of a cell. The contacting step can be carried out inside a cell. The contacting step can be carried out in a cell *in vitro*. The contacting step can be carried out in a cell *ex vivo*. The contacting step can be carried out in a cell *in vivo*.

[00288] The guide RNA can be provided as RNA or as a nucleic acid encoding the guide RNA (e.g., a DNA such as a recombinant expression vector). The tranc RNA can be provided as RNA or as a nucleic acid encoding the guide RNA (e.g., a DNA such as a recombinant expression vector). The CasZ polypeptide can be provided as a protein *per se* or as a nucleic acid encoding the protein (e.g., an mRNA, a DNA such as a recombinant expression vector). In some cases, two or more (e.g., 3 or more, 4 or more, 5 or more, or 6 or more) guide RNAs can be provided. In some cases, a single-molecule RNA

comprising: i) a CasZ guide RNA; and ii) a tranc RNA (or a nucleic acid comprising a nucleotide sequence encoding the single-molecule RNA) is used.

[00289] In some cases (e.g., when contacting a sample with a guide RNA and a CasZ polypeptide; or when contacting a sample with a guide RNA, a CasZ polypeptide, and a tranc RNA), the sample is contacted for 2 hours or less (e.g., 1.5 hours or less, 1 hour or less, 40 minutes or less, 30 minutes or less, 20 minutes or less, 10 minutes or less, or 5 minutes or less, or 1 minute or less) prior to the measuring step. For example, in some cases the sample is contacted for 40 minutes or less prior to the measuring step. In some cases, the sample is contacted for 20 minutes or less prior to the measuring step. In some cases, the sample is contacted for 10 minutes or less prior to the measuring step. In some cases, the sample is contacted for 5 minutes or less prior to the measuring step. In some cases, the sample is contacted for 1 minute or less prior to the measuring step. In some cases, the sample is contacted for from 50 seconds to 60 seconds prior to the measuring step. In some cases, the sample is contacted for from 40 seconds to 50 seconds prior to the measuring step. In some cases, the sample is contacted for from 30 seconds to 40 seconds prior to the measuring step. In some cases, the sample is contacted for from 20 seconds to 30 seconds prior to the measuring step. In some cases, the sample is contacted for from 10 seconds to 20 seconds prior to the measuring step.

[00290] In some cases, a method of the present disclosure for detecting a target DNA comprises: a) contacting a sample with a guide RNA, a CasZ polypeptide, and a detector DNA), where the sample is contacted for 2 hours or less (e.g., 1.5 hours or less, 1 hour or less, 40 minutes or less, 30 minutes or less, 20 minutes or less, 10 minutes or less, or 5 minutes or less, or 1 minute or less), under conditions that provide for *trans* cleavage of the detector DNA; b) maintaining the sample from step (a) for a period of time under conditions that do not provide for *trans* cleavage of the detector RNA; and c) after the time period of step (b), measuring a detectable signal produced by cleavage of the single stranded detector DNA by the CasZ polypeptide, thereby detecting the target DNA. Conditions that provide for *trans* cleavage of the detector DNA include temperature conditions such as from 17°C to about 39°C (e.g., about 37°C). Conditions that do not provide for *trans* cleavage of the detector DNA include temperatures of 10°C or less, 5°C or less, 4°C or less, or 0°C.

[00291] In some cases, a method of the present disclosure for detecting a target DNA comprises: a) contacting a sample with a guide RNA, a tranc RNA, a CasZ polypeptide, and a detector DNA (or contacting a sample with: i) a single-molecule RNA comprising a guide RNA and a tranc RNA; i) a CasZ polypeptide; and iii) a detector DNA), where the sample is contacted for 2 hours or less (e.g., 1.5 hours or less, 1 hour or less, 40 minutes or less, 30 minutes or less, 20 minutes or less, 10 minutes or less, or 5 minutes or less, or 1 minute or less), under conditions that provide for *trans* cleavage of the detector DNA; b) maintaining the sample from step (a) for a period of time under conditions that do not provide for *trans* cleavage of the detector RNA; and c) after the time period of step (b), measuring a

detectable signal produced by cleavage of the single stranded detector DNA by the CasZ polypeptide, thereby detecting the target DNA. Conditions that provide for *trans* cleavage of the detector DNA include temperature conditions such as from 17°C to about 39°C (e.g., about 37°C). Conditions that do not provide for *trans* cleavage of the detector DNA include temperatures of 10°C or less, 5°C or less, 4°C or less, or 0°C.

[00292] In some cases, a detectable signal produced by cleavage of a single-stranded detector DNA is produced for no more than 60 minutes. For example, in some cases, a detectable signal produced by cleavage of a single-stranded detector DNA is produced for no more than 60 minutes, no more than 45 minutes, no more than 30 minutes, no more than 15 minutes, no more than 10 minutes, or no more than 5 minutes. For example, in some cases, a detectable signal produced by cleavage of a single-stranded detector DNA is produced for a period of time of from 1 minute to 60 minutes, e.g., from 1 minute to 5 minutes, from 5 minutes to 10 minutes, from 10 minutes to 15 minutes, from 15 minutes to 30 minutes, from 30 minutes to 45 minutes, or from 45 minutes to 60 minutes. In some cases, after the detectable signal is produced (e.g., produced for no more than 60 minutes), production of the detectable signal can be stopped, e.g., by lowering the temperature of the sample (e.g., lowering the temperature to 10°C or less, 5°C or less, 4°C or less, or 0°C), by adding an inhibitor to the sample, by lyophilizing the sample, by heating the sample to over 40°C, and the like. The measuring step can occur at any time after production of the detectable signal has been stopped. For example, the measuring step can occur from 5 minutes to 48 hours after production of the detectable signal has been stopped. For example, the measuring step can occur from 5 minutes to 15 minutes, from 15 minutes to 30 minutes, from 30 minutes to 60 minutes, from 1 hour to 4 hours, from 4 hours to 8 hours, from 8 hours to 12 hours, from 12 hours to 24 hours, from 24 hours to 36 hours, or from 36 hours to 48 hours, after production of the detectable signal has been stopped. The measuring step can occur more than 48 hours after production of the detectable signal has been stopped.

[00293] A method of the present disclosure for detecting a target DNA (single-stranded or double-stranded) in a sample can detect a target DNA with a high degree of sensitivity. In some cases, a method of the present disclosure can be used to detect a target DNA present in a sample comprising a plurality of DNAs (including the target DNA and a plurality of non-target DNAs), where the target DNA is present at one or more copies per 10^7 non-target DNAs (e.g., one or more copies per 10^6 non-target DNAs, one or more copies per 10^5 non-target DNAs, one or more copies per 10^4 non-target DNAs, one or more copies per 10^3 non-target DNAs, one or more copies per 10^2 non-target DNAs, one or more copies per 50 non-target DNAs, one or more copies per 20 non-target DNAs, one or more copies per 10 non-target DNAs, or one or more copies per 5 non-target DNAs). In some cases, a method of the present disclosure can be used to detect a target DNA present in a sample comprising a plurality of DNAs (including the target DNA and a plurality of non-target DNAs), where the target DNA is present at one

or more copies per 10^{18} non-target DNAs (e.g., one or more copies per 10^{15} non-target DNAs, one or more copies per 10^{12} non-target DNAs, one or more copies per 10^{9} non-target DNAs, one or more copies per 10^{6} non-target DNAs, one or more copies per 10^{5} non-target DNAs, one or more copies per 10^{4} non-target DNAs, one or more copies per 10^{3} non-target DNAs, one or more copies per 10^{2} non-target DNAs, one or more copies per 50 non-target DNAs, one or more copies per 20 non-target DNAs, one or more copies per 10 non-target DNAs, or one or more copies per 5 non-target DNAs).

[00294] In some cases, a method of the present disclosure can detect a target DNA present in a sample, where the target DNA is present at from one copy per 10^{7} non-target DNAs to one copy per 10 non-target DNAs (e.g., from 1 copy per 10^{7} non-target DNAs to 1 copy per 10^{2} non-target DNAs, from 1 copy per 10^{7} non-target DNAs to 1 copy per 10^{3} non-target DNAs, from 1 copy per 10^{7} non-target DNAs to 1 copy per 10^{4} non-target DNAs, from 1 copy per 10^{7} non-target DNAs to 1 copy per 10^{5} non-target DNAs, from 1 copy per 10^{7} non-target DNAs to 1 copy per 10^{6} non-target DNAs, from 1 copy per 10^{6} non-target DNAs to 1 copy per 10 non-target DNAs, from 1 copy per 10^{6} non-target DNAs to 1 copy per 10^{2} non-target DNAs, from 1 copy per 10^{6} non-target DNAs to 1 copy per 10^{3} non-target DNAs, from 1 copy per 10^{6} non-target DNAs to 1 copy per 10^{4} non-target DNAs, from 1 copy per 10^{6} non-target DNAs to 1 copy per 10^{5} non-target DNAs, from 1 copy per 10^{5} non-target DNAs to 1 copy per 10 non-target DNAs, from 1 copy per 10^{5} non-target DNAs to 1 copy per 10^{2} non-target DNAs, from 1 copy per 10^{5} non-target DNAs to 1 copy per 10^{3} non-target DNAs, or from 1 copy per 10^{5} non-target DNAs to 1 copy per 10^{4} non-target DNAs).

[00295] In some cases, a method of the present disclosure can detect a target DNA present in a sample, where the target DNA is present at from one copy per 10^{18} non-target DNAs to one copy per 10 non-target DNAs (e.g., from 1 copy per 10^{18} non-target DNAs to 1 copy per 10^{2} non-target DNAs, from 1 copy per 10^{15} non-target DNAs to 1 copy per 10^{2} non-target DNAs, from 1 copy per 10^{12} non-target DNAs to 1 copy per 10^{2} non-target DNAs, from 1 copy per 10^{9} non-target DNAs to 1 copy per 10^{2} non-target DNAs, from 1 copy per 10^{7} non-target DNAs to 1 copy per 10^{2} non-target DNAs, from 1 copy per 10^{7} non-target DNAs to 1 copy per 10^{3} non-target DNAs, from 1 copy per 10^{7} non-target DNAs to 1 copy per 10^{4} non-target DNAs, from 1 copy per 10^{7} non-target DNAs to 1 copy per 10^{5} non-target DNAs, from 1 copy per 10^{7} non-target DNAs to 1 copy per 10^{6} non-target DNAs, from 1 copy per 10^{6} non-target DNAs to 1 copy per 10 non-target DNAs, from 1 copy per 10^{6} non-target DNAs to 1 copy per 10^{2} non-target DNAs, from 1 copy per 10^{6} non-target DNAs to 1 copy per 10^{3} non-target DNAs, from 1 copy per 10^{6} non-target DNAs to 1 copy per 10^{4} non-target DNAs, from 1 copy per 10^{6} non-target DNAs to 1 copy per 10^{5} non-target DNAs, from 1 copy per 10^{5} non-target DNAs to 1 copy per 10 non-target DNAs, from 1 copy per 10^{5} non-target DNAs to 1 copy per 10^{2} non-target DNAs, from 1 copy per 10^{5} non-target DNAs to 1 copy per 10^{3} non-target DNAs, or from 1 copy per 10^{5} non-target DNAs to 1 copy per 10^{4} non-target DNAs).

[00296] In some cases, a method of the present disclosure can detect a target DNA present in a sample, where the target DNA is present at from one copy per 10^7 non-target DNAs to one copy per 100 non-target DNAs (e.g., from 1 copy per 10^7 non-target DNAs to 1 copy per 10^2 non-target DNAs, from 1 copy per 10^7 non-target DNAs to 1 copy per 10^3 non-target DNAs, from 1 copy per 10^7 non-target DNAs to 1 copy per 10^4 non-target DNAs, from 1 copy per 10^7 non-target DNAs to 1 copy per 10^5 non-target DNAs, from 1 copy per 10^7 non-target DNAs to 1 copy per 10^6 non-target DNAs, from 1 copy per 10^6 non-target DNAs to 1 copy per 100 non-target DNAs, from 1 copy per 10^6 non-target DNAs to 1 copy per 10^2 non-target DNAs, from 1 copy per 10^6 non-target DNAs to 1 copy per 10^3 non-target DNAs, from 1 copy per 10^6 non-target DNAs to 1 copy per 10^4 non-target DNAs, from 1 copy per 10^6 non-target DNAs to 1 copy per 10^5 non-target DNAs, from 1 copy per 10^5 non-target DNAs to 1 copy per 100 non-target DNAs, from 1 copy per 10^5 non-target DNAs to 1 copy per 10^2 non-target DNAs, from 1 copy per 10^5 non-target DNAs to 1 copy per 10^3 non-target DNAs, or from 1 copy per 10^5 non-target DNAs to 1 copy per 10^4 non-target DNAs).

[00297] In some cases, the threshold of detection, for a subject method of detecting a target DNA in a sample, is 10 nM or less. Thus, e.g., the target DNA can be present in the sample in a concentration of 10 nM or less. The term "threshold of detection" is used herein to describe the minimal amount of target DNA that must be present in a sample in order for detection to occur. Thus, as an illustrative example, when a threshold of detection is 10 nM, then a signal can be detected when a target DNA is present in the sample at a concentration of 10 nM or more. In some cases, a method of the present disclosure has a threshold of detection of 5 nM or less. In some cases, a method of the present disclosure has a threshold of detection of 1 nM or less. In some cases, a method of the present disclosure has a threshold of detection of 0.5 nM or less. In some cases, a method of the present disclosure has a threshold of detection of 0.1 nM or less. In some cases, a method of the present disclosure has a threshold of detection of 0.05 nM or less. In some cases, a method of the present disclosure has a threshold of detection of 0.01 nM or less. In some cases, a method of the present disclosure has a threshold of detection of 0.005 nM or less. In some cases, a method of the present disclosure has a threshold of detection of 0.001 nM or less. In some cases, a method of the present disclosure has a threshold of detection of 0.0005 nM or less. In some cases, a method of the present disclosure has a threshold of detection of 0.0001 nM or less. In some cases, a method of the present disclosure has a threshold of detection of 0.00005 nM or less. In some cases, a method of the present disclosure has a threshold of detection of 0.00001 nM or less. In some cases, a method of the present disclosure has a threshold of detection of 10 pM or less. In some cases, a method of the present disclosure has a threshold of detection of 1 pM or less. In some cases, a method of the present disclosure has a threshold of detection of 500 fM or less. In some cases, a method of the present disclosure has a threshold of detection of 250 fM or less. In some cases, a method of the present disclosure has a threshold of

detection of 100 fM or less. In some cases, a method of the present disclosure has a threshold of detection of 50 fM or less. In some cases, a method of the present disclosure has a threshold of detection of 500 aM (attomolar) or less. In some cases, a method of the present disclosure has a threshold of detection of 250 aM or less. In some cases, a method of the present disclosure has a threshold of detection of 100 aM or less. In some cases, a method of the present disclosure has a threshold of detection of 50 aM or less. In some cases, a method of the present disclosure has a threshold of detection of 10 aM or less. In some cases, a method of the present disclosure has a threshold of detection of 1 aM or less.

[00298] In some cases, the threshold of detection (for detecting the target DNA in a subject method), is in a range of from 500 fM to 1 nM (e.g., from 500 fM to 500 pM, from 500 fM to 200 pM, from 500 fM to 100 pM, from 500 fM to 10 pM, from 500 fM to 1 pM, from 800 fM to 1 nM, from 800 fM to 500 pM, from 800 fM to 200 pM, from 800 fM to 100 pM, from 800 fM to 10 pM, from 800 fM to 1 pM, from 1 pM to 1 nM, from 1 pM to 500 pM, from 1 pM to 200 pM, from 1 pM to 100 pM, or from 1 pM to 10 pM) (where the concentration refers to the threshold concentration of target DNA at which the target DNA can be detected). In some cases, a method of the present disclosure has a threshold of detection in a range of from 800 fM to 100 pM. In some cases, a method of the present disclosure has a threshold of detection in a range of from 1 pM to 10 pM. In some cases, a method of the present disclosure has a threshold of detection in a range of from 10 fM to 500 fM, e.g., from 10 fM to 50 fM, from 50 fM to 100 fM, from 100 fM to 250 fM, or from 250 fM to 500 fM.

[00299] In some cases, the minimum concentration at which a target DNA can be detected in a sample is in a range of from 500 fM to 1 nM (e.g., from 500 fM to 500 pM, from 500 fM to 200 pM, from 500 fM to 100 pM, from 500 fM to 10 pM, from 500 fM to 1 pM, from 800 fM to 1 nM, from 800 fM to 500 pM, from 800 fM to 200 pM, from 800 fM to 100 pM, from 800 fM to 10 pM, from 800 fM to 1 pM, from 1 pM to 1 nM, from 1 pM to 500 pM, from 1 pM to 200 pM, from 1 pM to 100 pM, or from 1 pM to 10 pM). In some cases, the minimum concentration at which a target DNA can be detected in a sample is in a range of from 800 fM to 100 pM. In some cases, the minimum concentration at which a target DNA can be detected in a sample is in a range of from 1 pM to 10 pM.

[00300] In some cases, the threshold of detection (for detecting the target DNA in a subject method), is in a range of from 1 aM to 1 nM (e.g., from 1 aM to 500 pM, from 1 aM to 200 pM, from 1 aM to 100 pM, from 1 aM to 10 pM, from 1 aM to 1 pM, from 100 aM to 1 nM, from 100 aM to 500 pM, from 100 aM to 200 pM, from 100 aM to 100 pM, from 100 aM to 10 pM, from 100 aM to 1 pM, from 250 aM to 1 nM, from 250 aM to 500 pM, from 250 aM to 200 pM, from 250 aM to 100 pM, from 250 aM to 10 pM, from 250 aM to 1 pM, from 500 aM to 1 nM, from 500 aM to 500 pM, from 500 aM to 200 pM, from 500 aM to 100 pM, from 500 aM to 10 pM, from 500 aM to 1 pM, from 750 aM to 1 nM, from 750 aM to 500 pM, from 750 aM to 200 pM, from 750 aM to 100 pM, from 750 aM to 10 pM,

from 750 aM to 1 pM, from 1 fM to 1 nM, from 1 fM to 500 pM, from 1 fM to 200 pM, from 1 fM to 100 pM, from 1 fM to 10 pM, from 1 fM to 1 pM, from 500 fM to 500 pM, from 500 fM to 200 pM, from 500 fM to 100 pM, from 500 fM to 10 pM, from 500 fM to 1 pM, from 800 fM to 1 nM, from 800 fM to 500 pM, from 800 fM to 200 pM, from 800 fM to 100 pM, from 800 fM to 10 pM, from 800 fM to 1 pM, from 1 pM to 1 nM, from 1 pM to 500 pM, from 1 pM to 200 pM, from 1 pM to 100 pM, or from 1 pM to 10 pM) (where the concentration refers to the threshold concentration of target DNA at which the target DNA can be detected). In some cases, a method of the present disclosure has a threshold of detection in a range of from 1 aM to 800 aM. In some cases, a method of the present disclosure has a threshold of detection in a range of from 50 aM to 1 pM. In some cases, a method of the present disclosure has a threshold of detection in a range of from 50 aM to 500 fM.

[00301] In some cases, a target DNA is present in a sample in a range of from 1 aM to 1 nM (e.g., from 1 aM to 500 pM, from 1 aM to 200 pM, from 1 aM to 100 pM, from 1 aM to 10 pM, from 1 aM to 1 pM, from 100 aM to 1 nM, from 100 aM to 500 pM, from 100 aM to 200 pM, from 100 aM to 100 pM, from 100 aM to 10 pM, from 100 aM to 1 pM, from 250 aM to 1 nM, from 250 aM to 500 pM, from 250 aM to 200 pM, from 250 aM to 100 pM, from 250 aM to 10 pM, from 250 aM to 1 pM, from 500 aM to 1 nM, from 500 aM to 500 pM, from 500 aM to 200 pM, from 500 aM to 100 pM, from 500 aM to 10 pM, from 500 aM to 1 pM, from 750 aM to 1 nM, from 750 aM to 500 pM, from 750 aM to 200 pM, from 750 aM to 100 pM, from 750 aM to 10 pM, from 750 aM to 1 pM, from 1 fM to 1 nM, from 1 fM to 500 pM, from 1 fM to 200 pM, from 1 fM to 100 pM, from 1 fM to 10 pM, from 1 fM to 1 pM, from 500 fM to 500 pM, from 500 fM to 200 pM, from 500 fM to 100 pM, from 500 fM to 10 pM, from 500 fM to 1 pM, from 800 fM to 1 nM, from 800 fM to 500 pM, from 800 fM to 200 pM, from 800 fM to 100 pM, from 800 fM to 10 pM, from 800 fM to 1 pM, from 1 pM to 1 nM, from 1 pM to 500 pM, from 1 pM to 200 pM, from 1 pM to 100 pM, or from 1 pM to 10 pM). In some cases, a target DNA is present in a sample in a range of from 1 aM to 800 aM. In some cases, a target DNA is present in a sample in a range of from 50 aM to 1 pM. In some cases, a target DNA is present in a sample in a range of from 50 aM to 500 fM.

[00302] In some cases, the minimum concentration at which a target DNA can be detected in a sample is in a range of from 1 aM to 1 nM (e.g., from 1 aM to 500 pM, from 1 aM to 200 pM, from 1 aM to 100 pM, from 1 aM to 10 pM, from 1 aM to 1 pM, from 100 aM to 1 nM, from 100 aM to 500 pM, from 100 aM to 200 pM, from 100 aM to 100 pM, from 100 aM to 10 pM, from 100 aM to 1 pM, from 250 aM to 1 nM, from 250 aM to 500 pM, from 250 aM to 200 pM, from 250 aM to 100 pM, from 250 aM to 10 pM, from 250 aM to 1 pM, from 500 aM to 1 nM, from 500 aM to 500 pM, from 500 aM to 200 pM, from 500 aM to 100 pM, from 500 aM to 10 pM, from 500 aM to 1 pM, from 750 aM to 1 nM, from 750 aM to 500 pM, from 750 aM to 200 pM, from 750 aM to 100 pM, from 750 aM to 10 pM, from 750 aM to 1 pM, from 1 fM to 1 nM, from 1 fM to 500 pM, from 1 fM to 200 pM, from 1 fM to

100 pM, from 1 fM to 10 pM, from 1 fM to 1 pM, from 500 fM to 500 pM, from 500 fM to 200 pM, from 500 fM to 100 pM, from 500 fM to 10 pM, from 500 fM to 1 pM, from 800 fM to 1 nM, from 800 fM to 500 pM, from 800 fM to 200 pM, from 800 fM to 100 pM, from 800 fM to 10 pM, from 800 fM to 1 pM, from 1 pM to 1 nM, from 1 pM to 500 pM, from 1 pM to 200 pM, from 1 pM to 100 pM, or from 1 pM to 10 pM). In some cases, the minimum concentration at which a target DNA can be detected in a sample is in a range of from 1 aM to 500 pM. In some cases, the minimum concentration at which a target DNA can be detected in a sample is in a range of from 100 aM to 500 pM.

[00303] In some cases, a target DNA is present in a sample in a range of from 1 aM to 1 nM (e.g., from 1 aM to 500 pM, from 1 aM to 200 pM, from 1 aM to 100 pM, from 1 aM to 10 pM, from 1 aM to 1 pM, from 100 aM to 1 nM, from 100 aM to 500 pM, from 100 aM to 200 pM, from 100 aM to 100 pM, from 100 aM to 10 pM, from 100 aM to 1 pM, from 250 aM to 1 nM, from 250 aM to 500 pM, from 250 aM to 200 pM, from 250 aM to 100 pM, from 250 aM to 10 pM, from 250 aM to 1 pM, from 500 aM to 1 nM, from 500 aM to 500 pM, from 500 aM to 200 pM, from 500 aM to 100 pM, from 500 aM to 10 pM, from 500 aM to 1 pM, from 750 aM to 1 nM, from 750 aM to 500 pM, from 750 aM to 200 pM, from 750 aM to 100 pM, from 750 aM to 10 pM, from 750 aM to 1 pM, from 1 fM to 1 nM, from 1 fM to 500 pM, from 1 fM to 200 pM, from 1 fM to 100 pM, from 1 fM to 10 pM, from 1 fM to 1 pM, from 500 fM to 500 pM, from 500 fM to 200 pM, from 500 fM to 100 pM, from 500 fM to 10 pM, from 500 fM to 1 pM, from 800 fM to 1 nM, from 800 fM to 500 pM, from 800 fM to 200 pM, from 800 fM to 100 pM, from 800 fM to 10 pM, from 800 fM to 1 pM, from 1 pM to 1 nM, from 1 pM to 500 pM, from 1 pM to 200 pM, from 1 pM to 100 pM, or from 1 pM to 10 pM). In some cases, a target DNA is present in a sample in a range of from 1 aM to 500 pM. In some cases, a target DNA is present in a sample in a range of from 100 aM to 500 pM.

[00304] In some cases, a subject composition or method exhibits an attomolar (aM) sensitivity of detection. In some cases, a subject composition or method exhibits a femtomolar (fM) sensitivity of detection. In some cases, a subject composition or method exhibits a picomolar (pM) sensitivity of detection. In some cases, a subject composition or method exhibits a nanomolar (nM) sensitivity of detection.

Target DNA

[00305] A target DNA can be single stranded (ssDNA) or double stranded (dsDNA). When the target DNA is single stranded, there is no preference or requirement for a PAM sequence in the target DNA. However, when the target DNA is dsDNA, a PAM is usually present adjacent to the target sequence of the target DNA (e.g., see discussion of the PAM elsewhere herein). The source of the target DNA can be the same as the source of the sample, e.g., as described below.

[00306] The source of the target DNA can be any source. In some cases, the target DNA is a viral DNA (e.g., a genomic DNA of a DNA virus). As such, subject method can be for detecting the

presence of a viral DNA amongst a population of nucleic acids (e.g., in a sample). A subject method can also be used for the cleavage of non-target ssDNAs in the present of a target DNA. For example, if a method takes place in a cell, a subject method can be used to promiscuously cleave non-target ssDNAs in the cell (ssDNAs that do not hybridize with the guide sequence of the guide RNA) when a particular target DNA is present in the cell (e.g., when the cell is infected with a virus and viral target DNA is detected).

[00307] Examples of possible target DNAs include, but are not limited to, viral DNAs such as: a papovavirus (e.g., human papillomavirus (HPV), polyomavirus); a hepadnavirus (e.g., Hepatitis B Virus (HBV)); a herpesvirus (e.g., herpes simplex virus (HSV), varicella zoster virus (VZV), Epstein-Barr virus (EBV), cytomegalovirus (CMV), herpes lymphotropic virus, Pityriasis Rosea, kaposi's sarcoma-associated herpesvirus); an adenovirus (e.g., atadenovirus, aviadenovirus, ichtadenovirus, mastadenovirus, siadenovirus); a poxvirus (e.g., smallpox, vaccinia virus, cowpox virus, monkeypox virus, orf virus, pseudocowpox, bovine papular stomatitis virus; tanapox virus, yaba monkey tumor virus; molluscum contagiosum virus (MCV)); a parvovirus (e.g., adeno-associated virus (AAV), Parvovirus B19, human bocavirus, bufavirus, human parv4 G1); Geminiviridae; Nanoviridae; Phycodnaviridae; and the like. In some cases, the target DNA is parasite DNA. In some cases, the target DNA is bacterial DNA, e.g., DNA of a pathogenic bacterium.

Samples

[00308] A subject sample includes nucleic acid (e.g., a plurality of nucleic acids). The term "plurality" is used herein to mean two or more. Thus, in some cases a sample includes two or more (e.g., 3 or more, 5 or more, 10 or more, 20 or more, 50 or more, 100 or more, 500 or more, 1,000 or more, or 5,000 or more) nucleic acids (e.g., DNAs). A subject method can be used as a very sensitive way to detect a target DNA present in a sample (e.g., in a complex mixture of nucleic acids such as DNAs). In some cases, the sample includes 5 or more DNAs (e.g., 10 or more, 20 or more, 50 or more, 100 or more, 500 or more, 1,000 or more, or 5,000 or more DNAs) that differ from one another in sequence. In some cases, the sample includes 10 or more, 20 or more, 50 or more, 100 or more, 500 or more, 10^3 or more, 5 x 10^3 or more, 10^4 or more, 5 x 10^4 or more, 10^5 or more, 5 x 10^5 or more, 10^6 or more 5 x 10^6 or more, or 10^7 or more, DNAs. In some cases, the sample comprises from 10 to 20, from 20 to 50, from 50 to 100, from 100 to 500, from 500 to 10^3, from 10^3 to 5 x 10^3, from 5 x 10^3 to 10^4, from 10^4 to 5 x 10^4, from 5 x 10^4 to 10^5, from 10^5 to 5 x 10^5, from 5 x 10^5 to 10^6, from 10^6 to 5 x 10^6, or from 5 x 10^6 to 10^7, or more than 10^7, DNAs. In some cases, the sample comprises from 5 to 10^7 DNAs (e.g., that differ from one another in sequence)(e.g., from 5 to 10^6, from 5 to 10^5, from 5 to 50,000, from 5 to 30,000, from 10 to 10^6, from 10 to 10^5, from 10 to 50,000, from 10 to 30,000, from 20 to 10^6, from 20 to 10^5, from 20 to 50,000, or from 20 to 30,000 DNAs). In some cases, the sample includes 20 or more DNAs that differ from one another in sequence. In some cases, the sample includes DNAs from a cell lysate (e.g., a

eukaryotic cell lysate, a mammalian cell lysate, a human cell lysate, a prokaryotic cell lysate, a plant cell lysate, and the like). For example, in some cases the sample includes DNA from a cell such as a eukaryotic cell, e.g., a mammalian cell such as a human cell.

[00309] The term "sample" is used herein to mean any sample that includes DNA (e.g., in order to determine whether a target DNA is present among a population of DNAs). The sample can be derived from any source, e.g., the sample can be a synthetic combination of purified DNAs; the sample can be a cell lysate, an DNA-enriched cell lysate, or DNAs isolated and/or purified from a cell lysate. The sample can be from a patient (e.g., for the purpose of diagnosis). The sample can be from permeabilized cells. The sample can be from crosslinked cells. The sample can be in tissue sections. The sample can be from tissues prepared by crosslinking followed by delipidation and adjustment to make a uniform refractive index. Examples of tissue preparation by crosslinking followed by delipidation and adjustment to make a uniform refractive index have been described in, for example, Shah et al., Development (2016) 143, 2862-2867 doi:10.1242/dev.138560.

[00310] A "sample" can include a target DNA and a plurality of non-target DNAs. In some cases, the target DNA is present in the sample at one copy per 10 non-target DNAs, one copy per 20 non-target DNAs, one copy per 25 non-target DNAs, one copy per 50 non-target DNAs, one copy per 100 non-target DNAs, one copy per 500 non-target DNAs, one copy per 10^3 non-target DNAs, one copy per 5×10^3 non-target DNAs, one copy per 10^4 non-target DNAs, one copy per 5×10^4 non-target DNAs, one copy per 10^5 non-target DNAs, one copy per 5×10^5 non-target DNAs, one copy per 10^6 non-target DNAs, or less than one copy per 10^6 non-target DNAs. In some cases, the target DNA is present in the sample at from one copy per 10 non-target DNAs to 1 copy per 20 non-target DNAs, from 1 copy per 20 non-target DNAs to 1 copy per 50 non-target DNAs, from 1 copy per 50 non-target DNAs to 1 copy per 100 non-target DNAs, from 1 copy per 100 non-target DNAs to 1 copy per 500 non-target DNAs, from 1 copy per 500 non-target DNAs to 1 copy per 10^3 non-target DNAs, from 1 copy per 10^3 non-target DNAs to 1 copy per 5×10^3 non-target DNAs, from 1 copy per 5×10^3 non-target DNAs to 1 copy per 10^4 non-target DNAs, from 1 copy per 10^4 non-target DNAs to 1 copy per 10^5 non-target DNAs, from 1 copy per 10^5 non-target DNAs to 1 copy per 10^6 non-target DNAs, or from 1 copy per 10^6 non-target DNAs to 1 copy per 10^7 non-target DNAs.

[00311] Suitable samples include but are not limited to saliva, blood, serum, plasma, urine, aspirate, and biopsy samples. Thus, the term "sample" with respect to a patient encompasses blood and other liquid samples of biological origin, solid tissue samples such as a biopsy specimen or tissue cultures or cells derived therefrom and the progeny thereof. The definition also includes samples that have been manipulated in any way after their procurement, such as by treatment with reagents; washed; or enrichment for certain cell populations, such as cancer cells. The definition also includes sample that have been enriched for particular types of molecules, *e.g.*, DNAs. The term "sample" encompasses

biological samples such as a clinical sample such as blood, plasma, serum, aspirate, cerebral spinal fluid (CSF), and also includes tissue obtained by surgical resection, tissue obtained by biopsy, cells in culture, cell supernatants, cell lysates, tissue samples, organs, bone marrow, and the like. A "biological sample" includes biological fluids derived therefrom (e.g., cancerous cell, infected cell, etc.), *e.g.*, a sample comprising DNAs that is obtained from such cells (*e.g.*, a cell lysate or other cell extract comprising DNAs).

[00312] A sample can comprise, or can be obtained from, any of a variety of cells, tissues, organs, or acellular fluids. Suitable sample sources include eukaryotic cells, bacterial cells, and archaeal cells. Suitable sample sources include single-celled organisms and multi-cellular organisms. Suitable sample sources include single-cell eukaryotic organisms; a plant or a plant cell; an algal cell, e.g., *Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Sargassum patens, C. agardh*, and the like; a fungal cell (e.g., a yeast cell); an animal cell, tissue, or organ; a cell, tissue, or organ from an invertebrate animal (e.g. fruit fly, cnidarian, echinoderm, nematode, an insect, an arachnid, etc.); a cell, tissue, fluid, or organ from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal); a cell, tissue, fluid, or organ from a mammal (e.g., a human; a non-human primate; an ungulate; a feline; a bovine; an ovine; a caprine; etc.). Suitable sample sources include nematodes, protozoans, and the like. Suitable sample sources include parasites such as helminths, malarial parasites, etc.

[00313] Suitable sample sources include a cell, tissue, or organism of any of the six kingdoms, e.g., Bacteria (e.g., Eubacteria); Archaebacteria; Protista; Fungi; Plantae; and Animalia. Suitable sample sources include plant-like members of the kingdom Protista, including, but not limited to, algae (e.g., green algae, red algae, glaucophytes, cyanobacteria); fungus-like members of Protista, e.g., slime molds, water molds, etc.; animal-like members of Protista, e.g., flagellates (e.g., Euglena), amoeboids (e.g., amoeba), sporozoans (e.g., Apicomplexa, Myxozoa, Microsporidia), and ciliates (e.g., Paramecium). Suitable sample sources include include members of the kingdom Fungi, including, but not limited to, members of any of the phyla: Basidiomycota (club fungi; e.g., members of Agaricus, Amanita, Boletus, Cantherellus, etc.); Ascomycota (sac fungi, including, e.g., Saccharomyces); Mycophycophyta (lichens); Zygomycota (conjugation fungi); and Deuteromycota. Suitable sample sources include include members of the kingdom Plantae, including, but not limited to, members of any of the following divisions: Bryophyta (e.g., mosses), Anthocerotophyta (e.g., hornworts), Hepaticophyta (e.g., liverworts), Lycophyta (e.g., club mosses), Sphenophyta (e.g., horsetails), Psilophyta (e.g., whisk ferns), Ophioglossophyta, Pterophyta (e.g., ferns), Cycadophyta, Gingkophyta, Pinophyta, Gnetophyta, and Magnoliophyta (e.g., flowering plants). Suitable sample sources include include members of the kingdom Animalia, including, but not limited to, members of any of the following phyla: Porifera (sponges); Placozoa; Orthonectida (parasites of marine invertebrates); Rhombozoa; Cnidaria (corals,

anemones, jellyfish, sea pens, sea pansies, sea wasps); Ctenophora (comb jellies); Platyhelminthes (flatworms); Nemertina (ribbon worms); Ngathostomulida (jawed worms)p Gastrotricha; Rotifera; Priapulida; Kinorhyncha; Loricifera; Acanthocephala; Entoprocta; Nemotoda; Nematomorpha; Cycliophora; Mollusca (mollusks); Sipuncula (peanut worms); Annelida (segmented worms); Tardigrada (water bears); Onychophora (velvet worms); Arthropoda (including the subphyla: Chelicerata, Myriapoda, Hexapoda, and Crustacea, where the Chelicerata include, e.g., arachnids, Merostomata, and Pycnogonida, where the Myriapoda include, e.g., Chilopoda (centipedes), Diplopoda (millipedes), Paropoda, and Symphyla, where the Hexapoda include insects, and where the Crustacea include shrimp, krill, barnacles, etc.; Phoronida; Ectoprocta (moss animals); Brachiopoda; Echinodermata (e.g. starfish, sea daisies, feather stars, sea urchins, sea cucumbers, brittle stars, brittle baskets, etc.); Chaetognatha (arrow worms); Hemichordata (acorn worms); and Chordata. Suitable members of Chordata include any member of the following subphyla: Urochordata (sea squirts; including Ascidiacea, Thaliacea, and Larvacea); Cephalochordata (lancelets); Myxini (hagfish); and Vertebrata, where members of Vertebrata include, e.g., members of Petromyzontida (lampreys), Chondrichthyces (cartilaginous fish), Actinopterygii (ray-finned fish), Actinista (coelocanths), Dipnoi (lungfish), Reptilia (reptiles, e.g., snakes, alligators, crocodiles, lizards, etc.), Aves (birds); and Mammalian (mammals). Suitable plants include any monocotyledon and any dicotyledon.

[00314] Suitable sources of a sample include cells, fluid, tissue, or organ taken from an organism; from a particular cell or group of cells isolated from an organism; etc. For example, where the organism is a plant, suitable sources include xylem, the phloem, the cambium layer, leaves, roots, etc. Where the organism is an animal, suitable sources include particular tissues (e.g., lung, liver, heart, kidney, brain, spleen, skin, fetal tissue, etc.), or a particular cell type (e.g., neuronal cells, epithelial cells, endothelial cells, astrocytes, macrophages, glial cells, islet cells, T lymphocytes, B lymphocytes, etc.).

[00315] In some cases, the source of the sample is a (or is suspected of being a diseased cell, fluid, tissue, or organ. In some cases, the source of the sample is a normal (non-diseased) cell, fluid, tissue, or organ. In some cases, the source of the sample is a (or is suspected of being a pathogen-infected cell, tissue, or organ. For example, the source of a sample can be an individual who may or may not be infected – and the sample could be any biological sample (e.g., blood, saliva, biopsy, plasma, serum, bronchoalveolar lavage, sputum, a fecal sample, cerebrospinal fluid, a fine needle aspirate, a swab sample (e.g., a buccal swab, a cervical swab, a nasal swab), interstitial fluid, synovial fluid, nasal discharge, tears, buffy coat, a mucous membrane sample, an epithelial cell sample (e.g., epithelial cell scraping), etc.) collected from the individual. In some cases, the sample is a cell-free liquid sample. In some cases, the sample is a liquid sample that can comprise cells. Pathogens include viruses, fungi, helminths, protozoa, malarial parasites, Plasmodium parasites, Toxoplasma parasites, Schistosoma parasites, and the like. "Helminths" include roundworms, heartworms, and phytophagous nematodes

(Nematoda), flukes (Tematoda), Acanthocephala, and tapeworms (Cestoda). Protozoan infections include infections from Giardia spp., Trichomonas spp., African trypanosomiasis, amoebic dysentery, babesiosis, balantidial dysentery, Chaga's disease, coccidiosis, malaria and toxoplasmosis. Examples of pathogens such as parasitic/protozoan pathogens include, but are not limited to: *Plasmodium falciparum*, *Plasmodium vivax*, *Trypanosoma cruzi* and *Toxoplasma gondii*. Fungal pathogens include, but are not limited to: *Cryptococcus neoformans, Histoplasma capsulatum, Coccidioides immitis, Blastomyces dermatitidis, Chlamydia trachomatis,* and *Candida albicans*. Pathogenic viruses include, e.g., immunodeficiency virus (e.g., HIV); influenza virus; dengue; West Nile virus; herpes virus; yellow fever virus; Hepatitis Virus C; Hepatitis Virus A; Hepatitis Virus B; papillomavirus; and the like. Pathogenic viruses can include DNA viruses such as: a papovavirus (e.g., human papillomavirus (HPV), polyomavirus); a hepadnavirus (e.g., Hepatitis B Virus (HBV)); a herpesvirus (e.g., herpes simplex virus (HSV), varicella zoster virus (VZV), epstein-barr virus (EBV), cytomegalovirus (CMV), herpes lymphotropic virus, Pityriasis Rosea, kaposi's sarcoma-associated herpesvirus); an adenovirus (e.g., atadenovirus, aviadenovirus, ichtadenovirus, mastadenovirus, siadenovirus); a poxvirus (e.g., smallpox, vaccinia virus, cowpox virus, monkeypox virus, orf virus, pseudocowpox, bovine papular stomatitis virus; tanapox virus, yaba monkey tumor virus; molluscum contagiosum virus (MCV)); a parvovirus (e.g., adeno-associated virus (AAV), Parvovirus B19, human bocavirus, bufavirus, human parv4 G1); Geminiviridae; Nanoviridae; Phycodnaviridae; and the like. Pathogens can include, e.g., DNAviruses [e.g.: a papovavirus (e.g., human papillomavirus (HPV), polyomavirus); a hepadnavirus (e.g., Hepatitis B Virus (HBV)); a herpesvirus (e.g., herpes simplex virus (HSV), varicella zoster virus (VZV), Epstein-barr virus (EBV), cytomegalovirus (CMV), herpes lymphotropic virus, Pityriasis Rosea, kaposi's sarcoma-associated herpesvirus); an adenovirus (e.g., atadenovirus, aviadenovirus, ichtadenovirus, mastadenovirus, siadenovirus); a poxvirus (e.g., smallpox, vaccinia virus, cowpox virus, monkeypox virus, orf virus, pseudocowpox, bovine papular stomatitis virus; tanapox virus, yaba monkey tumor virus; molluscum contagiosum virus (MCV)); a parvovirus (e.g., adeno-associated virus (AAV), Parvovirus B19, human bocavirus, bufavirus, human parv4 G1); Geminiviridae; Nanoviridae; Phycodnaviridae; and the like], *Mycobacterium tuberculosis*, *Streptococcus agalactiae*, methicillin-resistant *Staphylococcus aureus*, *Legionella pneumophila*, *Streptococcus pyogenes*, *Escherichia coli*, *Neisseria gonorrhoeae*, *Neisseria meningitidis*, Pneumococcus, *Cryptococcus neoformans*, *Histoplasma capsulatum*, Hemophilus influenzae B, *Treponema pallidum*, Lyme disease spirochetes, *Pseudomonas aeruginosa*, *Mycobacterium leprae*, *Brucella abortus*, rabies virus, influenza virus, cytomegalovirus, herpes simplex virus I, herpes simplex virus II, human serum parvo-like virus, respiratory syncytial virus, varicella-zoster virus, hepatitis B virus, hepatitis C virus, measles virus, adenovirus, human T-cell leukemia viruses, Epstein-Barr virus, murine leukemia virus, mumps virus, vesicular stomatitis virus, Sindbis virus, lymphocytic choriomeningitis virus, wart virus, blue tongue virus, Sendai virus, feline

leukemia virus, Reovirus, polio virus, simian virus 40, mouse mammary tumor virus, dengue virus, rubella virus, West Nile virus, *Plasmodium falciparum, Plasmodium vivax, Toxoplasma gondii, Trypanosoma rangeli, Trypanosoma cruzi, Trypanosoma rhodesiense, Trypanosoma brucei, Schistosoma mansoni, Schistosoma japonicum, Babesia bovis, Eimeria tenella, Onchocerca volvulus, Leishmania tropica, Mycobacterium tuberculosis, Trichinella spiralis, Theileria parva, Taenia hydatigena, Taenia ovis, Taenia saginata, Echinococcus granulosus, Mesocestoides corti, Mycoplasma arthritidis, M. hyorhinis, M. orale, M. arginini, Acholeplasma laidlawii, M. salivarium* and *M. pneumoniae*.

Measuring a detectable signal

[00316] In some cases, a subject method includes a step of measuring (e.g., measuring a detectable signal produced by CasZ-mediated ssDNA cleavage). Because a CasZ polypeptide cleaves non-targeted ssDNA once activated, which occurs when a guide RNA hybridizes with a target DNA in the presence of a CasZ polypeptide (and, in some cases, also including a tranc RNA), a detectable signal can be any signal that is produced when ssDNA is cleaved. For example, in some cases the step of measuring can include one or more of: gold nanoparticle based detection (e.g., see Xu et al., Angew Chem Int Ed Engl. 2007;46(19):3468-70; and Xia et al., Proc Natl Acad Sci U S A. 2010 Jun 15;107(24):10837-41), fluorescence polarization, colloid phase transition/dispersion (e.g., Baksh et al., Nature. 2004 Jan 8;427(6970):139-41), electrochemical detection, semiconductor-based sensing (e.g., Rothberg et al., Nature. 2011 Jul 20;475(7356):348-52; e.g., one could use a phosphatase to generate a pH change after ssDNA cleavage reactions, by opening 2'-3' cyclic phosphates, and by releasing inorganic phosphate into solution), and detection of a labeled detector ssDNA (see elsewhere herein for more details). 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.

[00317] The measuring can in some cases be quantitative, e.g., in the sense that the amount of signal detected can be used to determine the amount of target DNA present in the sample. The measuring can in some cases be qualitative, e.g., in the sense that the presence or absence of detectable signal can indicate the presence or absence of targeted DNA (e.g., virus, SNP, etc.). In some cases, a detectable signal will not be present (e.g., above a given threshold level) unless the targeted DNA(s) (e.g., virus, SNP, etc.) is present above a particular threshold concentration. In some cases, the threshold of detection can be titrated by modifying the amount of CasZ polypeptide, guide RNA, sample volume, and/or detector ssDNA (if one is used). As such, for example, as would be understood by one of ordinary skill in the art, a number of controls can be used if desired in order to set up one or more reactions, each set up

to detect a different threshold level of target DNA, and thus such a series of reactions could be used to determine the amount of target DNA present in a sample (e.g., one could use such a series of reactions to determine that a target DNA is present in the sample 'at a concentration of at least X'). Non-limiting examples of applications of/uses for the compositions and methods of the disclosure include single-nucleotide polymorphism (SNP) detection, cancer screening, detection of bacterial infection, detection of antibiotic resistance, detection of viral infection, and the like. The compositions and methods of this disclosure can be used to detect any DNA target. For example, any virus that integrates nucleic acid material into the genome can be detected because a subject sample can include cellular genomic DNA – and the guide RNA can be designed to detect integrated nucleotide sequence. A method of the present disclosure in some cases does not include an amplification step. A method of the present disclosure in some cases includes an amplification step.

[00318] In some cases, a method of the present disclosure can be used to determine the amount of a target DNA in a sample (e.g., a sample comprising the target DNA and a plurality of non-target DNAs). Determining the amount of a target DNA in a sample can comprise comparing the amount of detectable signal generated from a test sample to the amount of detectable signal generated from a reference sample. Determining the amount of a target DNA in a sample can comprise: measuring the detectable signal to generate a test measurement; measuring a detectable signal produced by a reference sample to generate a reference measurement; and comparing the test measurement to the reference measurement to determine an amount of target DNA present in the sample.

[00319] For example, in some cases, a method of the present disclosure for determining the amount of a target DNA in a sample comprises: a) contacting the sample (e.g., a sample comprising the target DNA and a plurality of non-target DNAs) with: (i) a guide RNA that hybridizes with the target DNA, (ii) a CasZ polypeptide that cleaves DNAs present in the sample, and (iii) a detector ssDNA; b) measuring a detectable signal produced by CasZ polypeptide-mediated ssDNA cleavage (e.g., cleavage of the detector ssDNA), generating a test measurement; c) measuring a detectable signal produced by a reference sample to generate a reference measurement; and d) comparing the test measurement to the reference measurement to determine an amount of target DNA present in the sample.

[00320] As another example, in some cases, a method of the present disclosure for determining the amount of a target DNA in a sample comprises: a) contacting the sample (e.g., a sample comprising the target DNA and a plurality of non-target DNAs) with: (i) a guide RNA that hybridizes with the target DNA, (ii) a CasZ polypeptide that cleaves DNAs present in the sample, (iii) a tranc RNA; (iv) a detector ssDNA; b) measuring a detectable signal produced by CasZ polypeptide-mediated ssDNA cleavage (e.g., cleavage of the detector ssDNA), generating a test measurement; c) measuring a detectable signal produced by a reference sample to generate a reference measurement; and d) comparing the test measurement to the reference measurement to determine an amount of target DNA present in the sample.

Amplification of nucleic acids in the sample

[00321] In some embodiments, sensitivity of a subject composition and/or method (e.g., for detecting the presence of a target DNA, such as viral DNA or a SNP, in cellular genomic DNA) can be increased by coupling detection with nucleic acid amplification. In some cases, the nucleic acids in a sample are amplified prior to contact with a CasZ polypeptide that cleaves ssDNA (e.g., amplification of nucleic acids in the sample can begin prior to contact with a CasZ polypeptide). In some cases, the nucleic acids in a sample are amplified simultaneous with contact with a CasZ polypeptide. For example, in some cases a subject method includes amplifying nucleic acids of a sample (e.g., by contacting the sample with amplification components) prior to contacting the amplified sample with a CasZ polypeptide. In some cases, a subject method includes contacting a sample with amplification components at the same time (simultaneous with) that the sample is contacted with a CasZ polypeptide. If all components are added simultaneously (amplification components and detection components such as a CasZ polypeptide, a guide RNA, and a detector DNA), it is possible that the trans-cleavage activity of the CasZ polypeptide, will begin to degrade the nucleic acids of the sample at the same time the nucleic acids are undergoing amplification. However, even if this is the case, amplifying and detecting simultaneously can still increase sensitivity compared to performing the method without amplification.

[00322] In some cases, specific sequences (e.g., sequences of a virus, sequences that include a SNP of interest) are amplified from the sample, e.g., using primers. As such, a sequence to which the guide RNA will hybridize can be amplified in order to increase sensitivity of a subject detection method – this could achieve biased amplification of a desired sequence in order to increase the number of copies of the sequence of interest present in the sample relative to other sequences present in the sample. As one illustrative example, if a subject method is being used to determine whether a given sample includes a particular virus (or a particular SNP), a desired region of viral sequence (or non-viral genomic sequence) can be amplified, and the region amplified will include the sequence that would hybridize to the guide RNA if the viral sequence (or SNP) were in fact present in the sample.

[00323] As noted, in some cases the nucleic acids are amplified (e.g., by contact with amplification components) prior to contacting the amplified nucleic acids with a CasZ polypeptide. In some cases, amplification occurs for 10 seconds or more, (e.g., 30 seconds or more, 45 seconds or more, 1 minute or more, 2 minutes or more, 3 minutes or more, 4 minutes or more, 5 minutes or more, 7.5 minutes or more, 10 minutes or more, etc.) prior to contact with an enzymatically active CasZ polypeptide. In some cases, amplification occurs for 2 minutes or more (e.g., 3 minutes or more, 4 minutes or more, 5 minutes or more, 7.5 minutes or more, 10 minutes or more, etc.) prior to contact with an active CasZ polypeptide. In some cases, amplification occurs for a period of time in a range of from 10 seconds to 60 minutes (e.g., 10 seconds to 40 minutes, 10 seconds to 30 minutes, 10 seconds to 20 minutes, 10 seconds to 15 minutes, 10 seconds to 10 minutes, 10 seconds to 5 minutes, 30 seconds to 40

minutes, 30 seconds to 30 minutes, 30 seconds to 20 minutes, 30 seconds to 15 minutes, 30 seconds to 10 minutes, 30 seconds to 5 minutes, 1 minute to 40 minutes, 1 minute to 30 minutes, 1 minute to 20 minutes, 1 minute to 15 minutes, 1 minute to 10 minutes, 1 minute to 5 minutes, 2 minutes to 40 minutes, 2 minutes to 30 minutes, 2 minutes to 20 minutes, 2 minutes to 15 minutes, 2 minutes to 10 minutes, 2 minutes to 5 minutes, 5 minutes to 40 minutes, 5 minutes to 30 minutes, 5 minutes to 20 minutes, 5 minutes to 15 minutes, or 5 minutes to 10 minutes). In some cases, amplification occurs for a period of time in a range of from 5 minutes to 15 minutes. In some cases, amplification occurs for a period of time in a range of from 7 minutes to 12 minutes.

[00324] In some cases, a sample is contacted with amplification components at the same time as contact with a CasZ polypeptide. In some such cases, the CasZ polypeptide is inactive at the time of contact and is activated once nucleic acids in the sample have been amplified.

[00325] Various amplification methods and components will be known to one of ordinary skill in the art and any convenient method can be used (see, e.g., Zanoli and Spoto, Biosensors (Basel). 2013 Mar; 3(1): 18–43; Gill and Ghaemi, Nucleosides, Nucleotides, and Nucleic Acids, 2008, 27: 224-243; Craw and Balachandrana, Lab Chip, 2012, 12, 2469-2486; which are herein incorporated by reference in their entirety). Nucleic acid amplification can comprise polymerase chain reaction (PCR), reverse transcription PCR (RT-PCR), quantitative PCR (qPCR), reverse transcription qPCR (RT-qPCR), nested PCR, multiplex PCR, asymmetric PCR, touchdown PCR, random primer PCR, hemi-nested PCR, polymerase cycling assembly (PCA), colony PCR, ligase chain reaction (LCR), digital PCR, methylation specific-PCR (MSP),co-amplification at lower denaturation temperature-PCR (COLD-PCR), allele-specific PCR, intersequence-specific PCR (ISS-PCR), whole genome amplification (WGA), inverse PCR, and thermal asymmetric interlaced PCR (TAIL-PCR).

[00326] In some cases, the amplification is isothermal amplification. The term "isothermal amplification" indicates a method of nucleic acid (e.g., DNA) amplification (e.g., using enzymatic chain reaction) that can use a single temperature incubation thereby obviating the need for a thermal cycler. Isothermal amplification is a form of nucleic acid amplification which does not rely on the thermal denaturation of the target nucleic acid during the amplification reaction and hence may not require multiple rapid changes in temperature. Isothermal nucleic acid amplification methods can therefore be carried out inside or outside of a laboratory environment. By combining with a reverse transcription step, these amplification methods can be used to isothermally amplify RNA.

[00327] Examples of isothermal amplification methods include but are not limited to: loop-mediated isothermal Amplification (LAMP), helicase-dependent Amplification (HDA), recombinase polymerase amplification (RPA), strand displacement amplification (SDA), nucleic acid sequence-based amplification (NASBA), transcription mediated amplification (TMA), nicking enzyme amplification reaction (NEAR), rolling circle amplification (RCA), multiple

displacement amplification (MDA), Ramification (RAM), circular helicase-dependent amplification (cHDA), single primer isothermal amplification (SPIA), signal mediated amplification of RNA technology (SMART), self-sustained sequence replication (3SR), genome exponential amplification reaction (GEAR) and isothermal multiple displacement amplification (IMDA).

[00328] In some cases, the amplification is recombinase polymerase amplification (RPA) (see, e.g., U.S. patent Nos. 8,030,000; 8,426,134; 8,945,845; 9,309,502; and 9,663,820, which are hereby incorporated by reference in their entirety). Recombinase polymerase amplification (RPA) uses two opposing primers (much like PCR) and employs three enzymes – a recombinase, a single-stranded DNA-binding protein (SSB) and a strand-displacing polymerase. The recombinase pairs oligonucleotide primers with homologous sequence in duplex DNA, SSB binds to displaced strands of DNA to prevent the primers from being displaced, and the strand displacing polymerase begins DNA synthesis where the primer has bound to the target DNA. Adding a reverse transcriptase enzyme to an RPA reaction can facilitate detection RNA as well as DNA, without the need for a separate step to produce cDNA. One example of components for an RPA reaction is as follows (see, e.g., U.S. patent Nos. 8,030,000; 8,426,134; 8,945,845; 9,309,502; 9,663,820): 50mM Tris pH 8.4, 80mM Potassium actetate, 10mM Magnesium acetate, 2 mM DTT, 5% PEG compound (Carbowax-20M), 3mM ATP, 30 mM Phosphocreatine, 100 ng/μl creatine kinase, 420 ng/μl gp32, 140 ng/μl UvsX, 35 ng/μl UvsY, 2000M dNTPs, 300 nM each oligonucleotide, 35 ng/μl Bsu polymerase, and a nucleic acid-containing sample).

[00329] In a transcription mediated amplification (TMA), an RNA polymerase is used to make RNA from a promoter engineered in the primer region, and then a reverse transcriptase synthesizes cDNA from the primer. A third enzyme, e.g., Rnase H can then be used to degrade the RNA target from cDNA without the heat-denatured step. This amplification technique is similar to Self-Sustained Sequence Replication (3SR) and Nucleic Acid Sequence Based Amplification (NASBA), but varies in the enzymes employed. For another example, helicase-dependent amplification (HDA) utilizes a thermostable helicase (Tte-UvrD) rather than heat to unwind dsDNA to create single-strands that are then available for hybridization and extension of primers by polymerase. For yet another example, a loop mediated amplification (LAMP) employs a thermostable polymerase with strand displacement capabilities and a set of four or more specific designed primers. Each primer is designed to have hairpin ends that, once displaced, snap into a hairpin to facilitate self-priming and further polymerase extension. In a LAMP reaction, though the reaction proceeds under isothermal conditions, an initial heat denaturation step is required for double-stranded targets. In addition, amplification yields a ladder pattern of various length products. For yet another example, a strand displacement amplification (SDA) combines the ability of a restriction endonuclease to nick the unmodified strand of its target DNA and an exonuclease-deficient DNA polymerase to extend the 3' end at the nick and displace the downstream DNA strand.

Detector DNA

[00330] In some cases, a subject method includes contacting a sample (e.g., a sample comprising a target DNA and a plurality of non-target ssDNAs) with: i) a CasZ polypeptide; ii) a guide RNA; and iii) a detector DNA that is single stranded and does not hybridize with the guide sequence of the guide RNA.

[00331] A suitable single-stranded detector DNA has a length of from 7 nucleotides to 25 nucleotides. For example, a suitable single-stranded detector DNA has a length of from 7 nucleotides to 10 nucleotides, from 11 nucleotides to 15 nucleotides, from 15 nucleotides to 20 nucleotides, or from 20 nucleotides to 25 nucleotides. In some cases, a suitable single-stranded detector DNA has a length of from 10 nucleotides to 15 nucleotides. In some cases, a suitable single-stranded detector DNA has a length of 10 nucleotides. In some cases, a suitable single-stranded detector DNA has a length of 11 nucleotides. In some cases, a suitable single-stranded detector DNA has a length of 12 nucleotides. In some cases, a suitable single-stranded detector DNA has a length of 13 nucleotides. In some cases, a suitable single-stranded detector DNA has a length of 14 nucleotides. In some cases, a suitable single-stranded detector DNA has a length of 15 nucleotides.

[00332] In some cases, a subject method includes: a) contacting a sample with a labeled single stranded detector DNA (detector ssDNA) that includes a fluorescence-emitting dye pair; a CasZ polypeptide that cleaves the labeled detector ssDNA after it is activated (by binding to the guide RNA in the context of the guide RNA hybridizing to a target DNA); and b) measuring the detectable signal that is produced by the fluorescence-emitting dye pair. For example, in some cases, a subject method includes contacting a sample with a labeled detector ssDNA comprising a fluorescence resonance energy transfer (FRET) pair or a quencher/fluor pair, or both. In some cases, a subject method includes contacting a sample with a labeled detector ssDNA comprising a FRET pair. In some cases, a subject method includes contacting a sample with a labeled detector ssDNA comprising a fluor/quencher pair.

[00333] Fluorescence-emitting dye pairs comprise a FRET pair or a quencher/fluor pair. In both cases of a FRET pair and a quencher/fluor pair, the emission spectrum of one of the dyes overlaps a region of the absorption spectrum of the other dye in the pair. As used herein, the term "fluorescence-emitting dye pair" is a generic term used to encompass both a "fluorescence resonance energy transfer (FRET) pair" and a "quencher/fluor pair," both of which terms are discussed in more detail below. The term "fluorescence-emitting dye pair" is used interchangeably with the phrase "a FRET pair and/or a quencher/fluor pair."

[00334] In some cases (e.g., when the detector ssDNA includes a FRET pair) the labeled detector ssDNA produces an amount of detectable signal prior to being cleaved, and the amount of detectable signal that is measured is reduced when the labeled detector ssDNA is cleaved. In some cases, the

labeled detector ssDNA produces a first detectable signal prior to being cleaved (e.g., from a FRET pair) and a second detectable signal when the labeled detector ssDNA is cleaved (e.g., from a quencher/fluor pair). As such, in some cases, the labeled detector ssDNA comprises a FRET pair and a quencher/fluor pair.

[00335] In some cases, the labeled detector ssDNA comprises a FRET pair. FRET is a process by which radiationless transfer of energy occurs from an excited state fluorophore to a second chromophore in close proximity. The range over which the energy transfer can take place is limited to approximately 10 nanometers (100 angstroms), and the efficiency of transfer is extremely sensitive to the separation distance between fluorophores. Thus, as used herein, the term "FRET" ("fluorescence resonance energy transfer"; also known as "Förster resonance energy transfer") refers to a physical phenomenon involving a donor fluorophore and a matching acceptor fluorophore selected so that the emission spectrum of the donor overlaps the excitation spectrum of the acceptor, and further selected so that when donor and acceptor are in close proximity (usually 10 nm or less) to one another, excitation of the donor will cause excitation of and emission from the acceptor, as some of the energy passes from donor to acceptor via a quantum coupling effect. Thus, a FRET signal serves as a proximity gauge of the donor and acceptor; only when they are in close proximity to one another is a signal generated. The FRET donor moiety (e.g., donor fluorophore) and FRET acceptor moiety (e.g., acceptor fluorophore) are collectively referred to herein as a "FRET pair".

[00336] The donor-acceptor pair (a FRET donor moiety and a FRET acceptor moiety) is referred to herein as a “FRET pair” or a “signal FRET pair.” Thus, in some cases, a subject labeled detector ssDNA includes two signal partners (a signal pair), when one signal partner is a FRET donor moiety and the other signal partner is a FRET acceptor moiety. A subject labeled detector ssDNA that includes such a FRET pair (a FRET donor moiety and a FRET acceptor moiety) will thus exhibit a detectable signal (a FRET signal) when the signal partners are in close proximity (e.g., while on the same RNA molecule), but the signal will be reduced (or absent) when the partners are separated (e.g., after cleavage of the RNA molecule by a CasZ polypeptide).

[00337] FRET donor and acceptor moieties (FRET pairs) will be known to one of ordinary skill in the art and any convenient FRET pair (e.g., any convenient donor and acceptor moiety pair) can be used. Examples of suitable FRET pairs include but are not limited to those presented in Table 1. See also: Bajar et al. Sensors (Basel). 2016 Sep 14;16(9); and Abraham et al. PLoS One. 2015 Aug 3;10(8):e0134436.

[00338] **Table 6.** Examples of FRET pairs (donor and acceptor FRET moieties)

Donor	**Acceptor**
Tryptophan	Dansyl
IAEDANS (1)	DDPM (2)

Donor	Acceptor
BFP	DsRFP
Dansyl	Fluorescein isothiocyanate (FITC)
Dansyl	Octadecylrhodamine
Cyan fluorescent protein (CFP)	Green fluorescent protein (GFP)
CF (3)	Texas Red
Fluorescein	Tetramethylrhodamine
Cy3	Cy5
GFP	Yellow fluorescent protein (YFP)
BODIPY FL (4)	BODIPY FL (4)
Rhodamine 110	Cy3
Rhodamine 6G	Malachite Green
FITC	Eosin Thiosemicarbazide
B-Phycoerythrin	Cy5
Cy5	Cy5.5

(1) 5-(2-iodoacetylaminoethyl)aminonaphthalene-1-sulfonic acid
(2) N-(4-dimethylamino-3,5-dinitrophenyl)maleimide
(3) carboxyfluorescein succinimidyl ester
(4) 4,4-difluoro-4-bora-3a,4a-diaza-s-indacene

[00339] In some cases, a detectable signal is produced when the labeled detector ssDNA is cleaved (e.g., in some cases, the labeled detector ssDNA comprises a quencher/fluor pair). One signal partner of a signal quenching pair produces a detectable signal and the other signal partner is a quencher moiety that quenches the detectable signal of the first signal partner (i.e., the quencher moiety quenches the signal of the signal moiety such that the signal from the signal moiety is reduced (quenched) when the signal partners are in proximity to one another, e.g., when the signal partners of the signal pair are in close proximity).

[00340] For example, in some cases, an amount of detectable signal increases when the labeled detector ssDNA is cleaved. For example, in some cases, the signal exhibited by one signal partner (a signal moiety) is quenched by the other signal partner (a quencher signal moiety), e.g., when both are present on the same ssDNA molecule prior to cleavage by a CasZ polypeptide. Such a signal pair is referred to herein as a "quencher/fluor pair", "quenching pair", or "signal quenching pair." For example, in some cases, one signal partner (e.g., the first signal partner) is a signal moiety that produces a detectable signal that is quenched by the second signal partner (e.g., a quencher moiety). The signal partners of such a quencher/fluor pair will thus produce a detectable signal when the partners are separated (e.g., after cleavage of the detector ssDNA by a CasZ polypeptide), but the signal will be quenched when the partners are in close proximity (e.g., prior to cleavage of the detector ssDNA by a CasZ polypeptide).

[00341] A quencher moiety can quench a signal from the signal moiety (e.g., prior to cleavage of the detector ssDNA by a CasZ polypeptide) to various degrees. In some cases, a quencher moiety quenches the signal from the signal moiety where the signal detected in the presence of the quencher moiety (when the signal partners are in proximity to one another) is 95% or less of the signal detected in the absence of the quencher moiety (when the signal partners are separated). For example, in some cases, the signal detected in the presence of the quencher moiety can be 90% or less, 80% or less, 70% or less, 60% or less, 50% or less, 40% or less, 30% or less, 20% or less, 15% or less, 10% or less, or 5% or less of the signal detected in the absence of the quencher moiety. In some cases, no signal (e.g., above background) is detected in the presence of the quencher moiety.

[00342] In some cases, the signal detected in the absence of the quencher moiety (when the signal partners are separated) is at least 1.2 fold greater (e.g., at least 1.3fold, at least 1.5 fold, at least 1.7 fold, at least 2 fold, at least 2.5 fold, at least 3 fold, at least 3.5 fold, at least 4 fold, at least 5 fold, at least 7 fold, at least 10 fold, at least 20 fold, or at least 50 fold greater) than the signal detected in the presence of the quencher moiety (when the signal partners are in proximity to one another).

[00343] In some cases, the signal moiety is a fluorescent label. In some such cases, the quencher moiety quenches the signal (the light signal) from the fluorescent label (e.g., by absorbing energy in the emission spectra of the label). Thus, when the quencher moiety is not in proximity with the signal moiety, the emission (the signal) from the fluorescent label is detectable because the signal is not absorbed by the quencher moiety. Any convenient donor acceptor pair (signal moiety /quencher moiety pair) can be used and many suitable pairs are known in the art.

[00344] In some cases, the quencher moiety absorbs energy from the signal moiety (also referred to herein as a "detectable label") and then emits a signal (e.g., light at a different wavelength). Thus, in some cases, the quencher moiety is itself a signal moiety (e.g., a signal moiety can be 6-carboxyfluorescein while the quencher moiety can be 6-carboxy-tetramethylrhodamine), and in some such cases, the pair could also be a FRET pair. In some cases, a quencher moiety is a dark quencher. A dark quencher can absorb excitation energy and dissipate the energy in a different way (e.g., as heat). Thus, a dark quencher has minimal to no fluorescence of its own (does not emit fluorescence). Examples of dark quenchers are further described in U.S. patent numbers 8,822,673 and 8,586,718; U.S. patent publications 20140378330, 20140349295, and 20140194611; and international patent applications: WO200142505 and WO200186001, all if which are hereby incorporated by reference in their entirety.

[00345] Examples of fluorescent labels include, but are not limited to: an Alexa Fluor® dye, an ATTO dye (e.g., ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 514, ATTO 520, ATTO 532, ATTO Rho6G, ATTO 542, ATTO 550, ATTO 565, ATTO Rho3B, ATTO Rho11, ATTO Rho12, ATTO Thio12, ATTO Rho101, ATTO 590, ATTO 594, ATTO Rho13, ATTO 610, ATTO 620,

ATTO Rho14, ATTO 633, ATTO 647, ATTO 647N, ATTO 655, ATTO Oxa12, ATTO 665, ATTO 680, ATTO 700, ATTO 725, ATTO 740), a DyLight dye, a cyanine dye (e.g., Cy2, Cy3, Cy3.5, Cy3b, Cy5, Cy5.5, Cy7, Cy7.5), a FluoProbes dye, a Sulfo Cy dye, a Seta dye, an IRIS Dye, a SeTau dye, an SRfluor dye, a Square dye, fluorescein isothiocyanate (FITC), tetramethylrhodamine (TRITC), Texas Red, Oregon Green, Pacific Blue, Pacific Green, Pacific Orange, quantum dots, and a tethered fluorescent protein.

[00346] In some cases, a detectable label is a fluorescent label selected from: an Alexa Fluor® dye, an ATTO dye (e.g., ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 514, ATTO 520, ATTO 532, ATTO Rho6G, ATTO 542, ATTO 550, ATTO 565, ATTO Rho3B, ATTO Rho11, ATTO Rho12, ATTO Thio12, ATTO Rho101, ATTO 590, ATTO 594, ATTO Rho13, ATTO 610, ATTO 620, ATTO Rho14, ATTO 633, ATTO 647, ATTO 647N, ATTO 655, ATTO Oxa12, ATTO 665, ATTO 680, ATTO 700, ATTO 725, ATTO 740), a DyLight dye, a cyanine dye (e.g., Cy2, Cy3, Cy3.5, Cy3b, Cy5, Cy5.5, Cy7, Cy7.5), a FluoProbes dye, a Sulfo Cy dye, a Seta dye, an IRIS Dye, a SeTau dye, an SRfluor dye, a Square dye, fluorescein (FITC), tetramethylrhodamine (TRITC), Texas Red, Oregon Green, Pacific Blue, Pacific Green, and Pacific Orange.

[00347] In some cases, a detectable label is a fluorescent label selected from: an Alexa Fluor® dye, an ATTO dye (e.g., ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 514, ATTO 520, ATTO 532, ATTO Rho6G, ATTO 542, ATTO 550, ATTO 565, ATTO Rho3B, ATTO Rho11, ATTO Rho12, ATTO Thio12, ATTO Rho101, ATTO 590, ATTO 594, ATTO Rho13, ATTO 610, ATTO 620, ATTO Rho14, ATTO 633, ATTO 647, ATTO 647N, ATTO 655, ATTO Oxa12, ATTO 665, ATTO 680, ATTO 700, ATTO 725, ATTO 740), a DyLight dye, a cyanine dye (e.g., Cy2, Cy3, Cy3.5, Cy3b, Cy5, Cy5.5, Cy7, Cy7.5), a FluoProbes dye, a Sulfo Cy dye, a Seta dye, an IRIS Dye, a SeTau dye, an SRfluor dye, a Square dye, fluorescein (FITC), tetramethylrhodamine (TRITC), Texas Red, Oregon Green, Pacific Blue, Pacific Green, Pacific Orange, a quantum dot, and a tethered fluorescent protein.

[00348] Examples of ATTO dyes include, but are not limited to: ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 514, ATTO 520, ATTO 532, ATTO Rho6G, ATTO 542, ATTO 550, ATTO 565, ATTO Rho3B, ATTO Rho11, ATTO Rho12, ATTO Thio12, ATTO Rho101, ATTO 590, ATTO 594, ATTO Rho13, ATTO 610, ATTO 620, ATTO Rho14, ATTO 633, ATTO 647, ATTO 647N, ATTO 655, ATTO Oxa12, ATTO 665, ATTO 680, ATTO 700, ATTO 725, and ATTO 740.

[00349] Examples of AlexaFluor dyes include, but are not limited to: Alexa Fluor® 350, Alexa Fluor® 405, Alexa Fluor® 430, Alexa Fluor® 488, Alexa Fluor® 500, Alexa Fluor® 514, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 555, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 610, Alexa Fluor® 633, Alexa Fluor® 635, Alexa Fluor® 647, Alexa Fluor® 660, Alexa Fluor® 680, Alexa Fluor® 700, Alexa Fluor® 750, Alexa Fluor® 790, and the like.

[00350] Examples of quencher moieties include, but are not limited to: a dark quencher, a Black Hole Quencher® (BHQ®) (e.g., BHQ-0, BHQ-1, BHQ-2, BHQ-3), a Qxl quencher, an ATTO quencher (e.g., ATTO 540Q, ATTO 580Q, and ATTO 612Q), dimethylaminoazobenzenesulfonic acid (Dabsyl), Iowa Black RQ, Iowa Black FQ, IRDye QC-1, a QSY dye (e.g., QSY 7, QSY 9, QSY 21), AbsoluteQuencher, Eclipse, and metal clusters such as gold nanoparticles, and the like.

[00351] In some cases, a quencher moiety is selected from: a dark quencher, a Black Hole Quencher® (BHQ®) (e.g., BHQ-0, BHQ-1, BHQ-2, BHQ-3), a Qxl quencher, an ATTO quencher (e.g., ATTO 540Q, ATTO 580Q, and ATTO 612Q), dimethylaminoazobenzenesulfonic acid (Dabsyl), Iowa Black RQ, Iowa Black FQ, IRDye QC-1, a QSY dye (e.g., QSY 7, QSY 9, QSY 21), AbsoluteQuencher, Eclipse, and a metal cluster.

[00352] Examples of an ATTO quencher include, but are not limited to: ATTO 540Q, ATTO 580Q, and ATTO 612Q. Examples of a Black Hole Quencher® (BHQ®) include, but are not limited to: BHQ-0 (493 nm), BHQ-1 (534 nm), BHQ-2 (579 nm) and BHQ-3 (672 nm).

[00353] For examples of some detectable labels (e.g., fluorescent dyes) and/or quencher moieties, see, e.g., Bao et al., Annu Rev Biomed Eng. 2009;11:25-47; as well as U.S. patent numbers 8,822,673 and 8,586,718; U.S. patent publications 20140378330, 20140349295, 20140194611, 20130323851, 20130224871, 20110223677, 20110190486, 20110172420, 20060179585 and 20030003486; and international patent applications: WO200142505 and WO200186001, all of which are hereby incorporated by reference in their entirety.

[00354] In some cases, cleavage of a labeled detector ssDNA can be detected by measuring a colorimetric read-out. For example, the liberation of a fluorophore (e.g., liberation from a FRET pair, liberation from a quencher/fluor pair, and the like) can result in a wavelength shift (and thus color shift) of a detectable signal. Thus, in some cases, cleavage of a subject labeled detector ssDNA can be detected by a color-shift. Such a shift can be expressed as a loss of an amount of signal of one color (wavelength), a gain in the amount of another color, a change in the ration of one color to another, and the like.

KITS FOR DETECTING TARGET DNA

[00355] The present disclosure provides a kit for detecting a target DNA, e.g., in a sample comprising a plurality of DNAs. In some cases, the kit comprises: (a) a labeled detector ssDNA (e.g., a labeled detector ssDNA comprising a fluorescence-emitting dye pair, e.g., a FRET pair and/or a quencher/fluor pair); and (b) one or more of: (i) a guide RNA, and/or a nucleic acid encoding said guide RNA; and ii) a CasZ polypeptide, and/or a nucleic acid encoding said CasZ polypeptide. In some cases, a nucleic acid encoding a guide RNA includes sequence insertion sites for the insertion of guide sequences by a user.

[00356] In some cases, the kit comprises: (a) a labeled detector ssDNA (e.g., a labeled detector ssDNA comprising a fluorescence-emitting dye pair, e.g., a FRET pair and/or a quencher/fluor pair); and (b) one or more of: (i) a guide RNA, and/or a nucleic acid encoding said guide RNA; ii) a tranc RNA and/or a nucleic acid encoding said guide RNA; and iii) a CasZ polypeptide, and/or a nucleic acid encoding said CasZ polypeptide. In some cases, a nucleic acid encoding a guide RNA includes sequence insertion sites for the insertion of guide sequences by a user.

[00357] In some cases, the kit comprises: (a) a labeled detector ssDNA (e.g., a labeled detector ssDNA comprising a fluorescence-emitting dye pair, e.g., a FRET pair and/or a quencher/fluor pair); and (b) one or more of: (i) a single-molecule RNA comprising a guide RNA and a tranc RNA, and/or a nucleic acid encoding single-molecule RNA; and iii) a CasZ polypeptide, and/or a nucleic acid encoding said CasZ polypeptide. In some cases, a nucleic acid encoding a single-molecule RNA includes sequence insertion sites for the insertion of guide sequences by a user.

[00358] In some cases, a subject kit comprises: (a) a labeled detector ssDNA comprising a fluorescence-emitting dye pair, e.g., a FRET pair and/or a quencher/fluor pair; and (b) one or more of: (i) a guide RNA, and/or a nucleic acid encoding said guide RNA; and/or i) a CasZ polypeptide.

Positive controls

[00359] A kit of the present disclosure (e.g., one that comprises a labeled detector ssDNA and a CasZ polypeptide) can also include a positive control target DNA. In some cases, the kit also includes a positive control guide RNA that comprises a nucleotide sequence that hybridizes to the control target DNA. In some cases, the positive control target DNA is provided in various amounts, in separate containers. In some cases, the positive control target DNA is provided in various known concentrations, in separate containers, along with control non-target DNAs.

Nucleic acids

[00360] While the RNAs of the disclosure (e.g., guide RNAs, tranc RNAs, single-molecule RNAs comprising a guide RNA and a tranc RNA) can be synthesized using any convenient method (e.g., chemical synthesis, *in vitro* using an RNA polymerase enzyme, e.g., T7 polymerase, T3 polymerase, SP6 polymerase, etc.), nucleic acids encoding such RNAs are also envisioned. Additionally, while a CasZ polypeptide of the disclosure can be provided (e.g., as part of a kit) in protein form, nucleic acids (such as mRNA and/or DNA) encoding the CasZ polypeptide can also be provided.

[00361] In some cases, a kit of the present disclosure comprises a nucleic acid (e.g., a DNA, e.g., a recombinant expression vector) that comprises a nucleotide sequence encoding a single-molecule RNA comprising: i) a guide RNA; and ii) a tranc RNA. In some cases, the nucleotide sequence encodes the guide RNA portion of the single-molecule RNA without a guide sequence. For example, in some cases, the nucleic acid comprises a nucleotide sequence encoding: i) a constant region of a guide RNA (a guide

RNA without a guide sequence), and comprises an insertion site for a nucleic acid encoding a guide sequence; and ii) a tranc RNA.

[00362] For example, in some cases, a kit of the present disclosure comprises a nucleic acid (e.g., a DNA, e.g., a recombinant expression vector) that comprises a nucleotide sequence encoding a guide RNA. In some cases, the nucleotide sequence encodes a guide RNA without a guide sequence. For example, in some cases, the nucleic acid comprises a nucleotide sequence encoding a constant region of a guide RNA (a guide RNA without a guide sequence), and comprises an insertion site for a nucleic acid encoding a guide sequence. In some cases, a kit of the present disclosure comprises a nucleic acid (e.g., an mRNA, a DNA, e.g., a recombinant expression vector) that comprises a nucleotide sequence encoding a CasZ polypeptide.

[00363] In some cases, the guide RNA-encoding nucleotide sequence is operably linked to a promoter, e.g., a promoter that is functional in a prokaryotic cell, a promoter that is functional in a eukaryotic cell, a promoter that is functional in a mammalian cell, a promoter that is functional in a human cell, and the like. In some cases, a nucleotide sequence encoding a CasZ polypeptide is operably linked to a promoter, e.g., a promoter that is functional in a prokaryotic cell, a promoter that is functional in a eukaryotic cell, a promoter that is functional in a mammalian cell, a promoter that is functional in a human cell, a cell type-specific promoter, a regulatable promoter, a tissue-specific promoter, and the like.

UTILITY

[00364] CasZ compositions (e.g., expression vectors, kits, compositions, nucleic acids, and the like) find use in a variety of methods. For example, a CasZ compositions of the present disclosure can be used to (i) modify (e.g., cleave, e.g., nick; methylate; etc.) target nucleic acid (DNA or RNA; single stranded or double stranded); (ii) modulate transcription of a target nucleic acid; (iii) label a target nucleic acid; (iv) bind a target nucleic acid (e.g., for purposes of isolation, labeling, imaging, tracking, etc.); (v) modify a polypeptide (e.g., a histone) associated with a target nucleic acid; and the like. Thus, the present disclosure provides a method of modifying a target nucleic acid. In some cases, a method of the present disclosure for modifying a target nucleic acid comprises contacting the target nucleic acid with: a) a CasZ polypeptide of the present disclosure; and b) one or more (e.g., two) CasZ guide RNAs. In some cases, a method of the present disclosure for modifying a target nucleic acid comprises contacting the target nucleic acid with: a) a CasZ polypeptide, and b) one or more (e.g., two) CasZ guide RNAs, and c) a CasZ trancRNA. In some cases, a method of the present disclosure for modifying a target nucleic acid comprises contacting the target nucleic acid with: a) a CasZ polypeptide of the present disclosure; b) a CasZ guide RNA; and c) a donor nucleic acid (e.g, a donor template). In some cases, a method of the present disclosure for modifying a target nucleic acid comprises contacting the target nucleic acid with: a) a CasZ polypeptide; b) a CasZ guide RNA; c) a CasZ trancRNA, and d) a donor nucleic acid (e.g, a donor template).In some cases, the contacting step is carried out in a cell *in vitro*. In

some cases, the contacting step is carried out in a cell *in vivo*. In some cases, the contacting step is carried out in a cell *ex vivo*.

[00365] Because a method that uses a CasZ polypeptide includes binding of the CasZ polypeptide to a particular region in a target nucleic acid (by virtue of being targeted there by an associated CasZ guide RNA), the methods are generally referred to herein as methods of binding (e.g., a method of binding a target nucleic acid). However, it is to be understood that in some cases, while a method of binding may result in nothing more than binding of the target nucleic acid, in other cases, the method can have different final results (e.g., the method can result in modification of the target nucleic acid, e.g., cleavage/methylation/etc., modulation of transcription from the target nucleic acid; modulation of translation of the target nucleic acid; genome editing; modulation of a protein associated with the target nucleic acid; isolation of the target nucleic acid; etc.).

[00366] For examples of suitable methods (e.g., that are used with CRISPR/Cas9 systems), see, for example, Jinek et al., Science. 2012 Aug 17;337(6096):816-21; Chylinski et al., RNA Biol. 2013 May;10(5):726-37; Ma et al., Biomed Res Int. 2013;2013:270805; Hou et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15644-9; Jinek et al., Elife. 2013;2:e00471; Pattanayak et al., Nat Biotechnol. 2013 Sep;31(9):839-43; Qi et al, Cell. 2013 Feb 28;152(5):1173-83; Wang et al., Cell. 2013 May 9;153(4):910-8; Auer et al., Genome Res. 2013 Oct 31; Chen et al., Nucleic Acids Res. 2013 Nov 1;41(20):e19; Cheng et al., Cell Res. 2013 Oct;23(10):1163-71; Cho et al., Genetics. 2013 Nov;195(3):1177-80; DiCarlo et al., Nucleic Acids Res. 2013 Apr;41(7):4336-43; Dickinson et al., Nat Methods. 2013 Oct;10(10):1028-34; Ebina et al., Sci Rep. 2013;3:2510; Fujii et al, Nucleic Acids Res. 2013 Nov 1;41(20):e187; Hu et al., Cell Res. 2013 Nov;23(11):1322-5; Jiang et al., Nucleic Acids Res. 2013 Nov 1;41(20):e188; Larson et al., Nat Protoc. 2013 Nov;8(11):2180-96; Mali et. at., Nat Methods. 2013 Oct;10(10):957-63; Nakayama et al., Genesis. 2013 Dec;51(12):835-43; Ran et al., Nat Protoc. 2013 Nov;8(11):2281-308; Ran et al., Cell. 2013 Sep 12;154(6):1380-9; Upadhyay et al., G3 (Bethesda). 2013 Dec 9;3(12):2233-8; Walsh et al., Proc Natl Acad Sci U S A. 2013 Sep 24;110(39):15514-5; Xie et al., Mol Plant. 2013 Oct 9; Yang et al., Cell. 2013 Sep 12;154(6):1370-9; and U.S. patents and patent applications: 8,906,616; 8,895,308; 8,889,418; 8,889,356; 8,871,445; 8,865,406; 8,795,965; 8,771,945; 8,697,359; 20140068797; 20140170753; 20140179006; 20140179770; 20140186843; 20140186919; 20140186958; 20140189896; 20140227787; 20140234972; 20140242664; 20140242699; 20140242700; 20140242702; 20140248702; 20140256046; 20140273037; 20140273226; 20140273230; 20140273231; 20140273232; 20140273233; 20140273234; 20140273235; 20140287938; 20140295556; 20140295557; 20140298547; 20140304853; 20140309487; 20140310828; 20140310830; 20140315985; 20140335063; 20140335620; 20140342456; 20140342457; 20140342458; 20140349400; 20140349405; 20140356867; 20140356956; 20140356958; 20140356959; 20140357523; 20140357530; 20140364333; and 20140377868; each of which is hereby incorporated by reference in its entirety.

[00367] For example, the present disclosure provides (but is not limited to) methods of cleaving a target nucleic acid; methods of editing a target nucleic acid; methods of modulating transcription from a target nucleic acid; methods of isolating a target nucleic acid, methods of binding a target nucleic acid, methods of imaging a target nucleic acid, methods of modifying a target nucleic acid, and the like.

[00368] As used herein, the terms/phrases "contact a target nucleic acid" and "contacting a target nucleic acid", for example, with a CasZ polypeptide or with a CasZ fusion polypeptide, etc., encompass all methods for contacting the target nucleic acid. For example, a CasZ polypeptide can be provided to a cell as protein, RNA (encoding the CasZ polypeptide), or DNA (encoding the CasZ polypeptide); while a CasZ guide RNA can be provided as a guide RNA or as a nucleic acid encoding the guide RNA and a CasZ trancRNA can be provided as a trancRNA or as a nucleic acid encoding the trancRNA. As such, when, for example, performing a method in a cell (e.g., inside of a cell *in vitro*, inside of a cell *in vivo*, inside of a cell *ex vivo*), a method that includes contacting the target nucleic acid encompasses the introduction into the cell of any or all of the components in their active/final state (e.g., in the form of a protein(s) for CasZ polypeptide; in the form of a protein for a CasZ fusion polypeptide; in the form of an RNA in some cases for the guide RNA), and also encompasses the introduction into the cell of one or more nucleic acids encoding one or more of the components (e.g., nucleic acid(s) comprising nucleotide sequence(s) encoding a CasZ polypeptide or a CasZ fusion polypeptide, nucleic acid(s) comprising nucleotide sequence(s) encoding guide RNA(s), nucleic acid comprising a nucleotide sequence encoding a donor template, and the like). Because the methods can also be performed *in vitro* outside of a cell, a method that includes contacting a target nucleic acid, (unless otherwise specified) encompasses contacting outside of a cell *in vitro*, inside of a cell *in vitro*, inside of a cell *in vivo*, inside of a cell *ex vivo*, etc.

[00369] In some cases, a method of the present disclosure for modifying a target nucleic acid comprises introducing into a target cell a CasZ locus, e.g., a nucleic acid comprising a nucleotide sequence encoding a CasZ polypeptide as well as nucleotide sequences of about 1 kilobase (kb) to 5 kb in length surrounding the CasZ-encoding nucleotide sequence from a cell (e.g., in some cases a cell that in its natural state (the state in which it occurs in nature) comprises a CasZ locus) comprising a CasZ locus, where the target cell does not normally (in its natural state) comprise a CasZ locus (e.g., in some cases the locus includes a CasZ trancRNA. However, one or more spacer sequences, encoding guide sequences for the encoded crRNA(s), can be modified such that one or more target sequences of interest are targeted. Thus, for example, in some cases, a method of the present disclosure for modifying a target nucleic acid comprises introducing into a target cell a CasZ locus, e.g., a nucleic acid obtained from a source cell (e.g., in some cases a cell that in its natural state (the state in which it occurs in nature) comprises a CasZ locus), where the nucleic acid has a length of from 100 nucleotides (nt) to 5 kb in length (e.g., from 100 nt to 500 nt, from 500 nt to 1 kb, from 1 kb to 1.5 kb, from 1.5 kb to 2 kb, from 2

kb to 2.5 kb, from 2.5 kb to 3 kb, from 3 kb to 3.5 kb, from 3.5 kb to 4 kb, or from 4 kb to 5 kb in length) and comprises a nucleotide sequence encoding a CasZ polypeptide. As noted above, in some such cases, one or more spacer sequences, encoding guide sequences for the encoded crRNA(s), can be modified such that one or more target sequences of interest are targeted. In some cases, the method comprises introducing into a target cell: i) a CasZ locus; and ii) a donor DNA template. In some cases, the target nucleic acid is in a cell-free composition *in vitro*. In some cases, the target nucleic acid is present in a target cell. In some cases, the target nucleic acid is present in a target cell, where the target cell is a prokaryotic cell. In some cases, the target nucleic acid is present in a target cell, where the target cell is a eukaryotic cell. In some cases, the target nucleic acid is present in a target cell, where the target cell is a mammalian cell. In some cases, the target nucleic acid is present in a target cell, where the target cell is a plant cell.

[00370] In some cases, a method of the present disclosure for modifying a target nucleic acid comprises contacting a target nucleic acid with a CasZ polypeptide of the present disclosure, or with a CasZ fusion polypeptide of the present disclosure. In some cases, abmethod of the present disclosure for modifying a target nucleic acid comprises contacting a target nucleic acid with a CasZ polypeptide and a CasZ guide RNA. In some cases, abmethod of the present disclosure for modifying a target nucleic acid comprises contacting a target nucleic acid with a CasZ polypeptide, a CasZ guide RNA, and a CasZ trancRNA. In some cases, abmethod of the present disclosure for modifying a target nucleic acid comprises contacting a target nucleic acid with a CasZ polypeptide, a first CasZ guide RNA, and a second CasZ guide RNA (and in some cases a CasZ trancRNA). In some cases, a method of the present disclosure for modifying a target nucleic acid comprises contacting a target nucleic acid with a CasZ polypeptide of the present disclosure and a CasZ guide RNA and a donor DNA template. In some cases, a method of the present disclosure for modifying a target nucleic acid comprises contacting a target nucleic acid with a CasZ polypeptide of the present disclosure and a CasZ guide RNA and a CasZ trancRNA and a donor DNA template.

[00371] In some cases, the target nucleic acid is in a cell-free composition *in vitro*. In some cases, the target nucleic acid is present in a target cell. In some cases, the target nucleic acid is present in a target cell, where the target cell is a prokaryotic cell. In some cases, the target nucleic acid is present in a target cell, where the target cell is a eukaryotic cell. In some cases, the target nucleic acid is present in a target cell, where the target cell is a mammalian cell. In some cases, the target nucleic acid is present in a target cell, where the target cell is a plant cell.

Target nucleic acids and target cells of interest

[00372] A target nucleic acid can be any nucleic acid (e.g., DNA, RNA), can be double stranded or single stranded, can be any type of nucleic acid (e.g., a chromosome (genomic DNA), derived from a chromosome, chromosomal DNA, plasmid, viral, extracellular, intracellular, mitochondrial, chloroplast,

linear, circular, etc.) and can be from any organism (e.g., as long as the CasZ guide RNA comprises a nucleotide sequence that hybridizes to a target sequence in a target nucleic acid, such that the target nucleic acid can be targeted).

[00373] A target nucleic acid can be DNA or RNA. A target nucleic acid can be double stranded (e.g., dsDNA, dsRNA) or single stranded (e.g., ssRNA, ssDNA). In some cases, a target nucleic acid is single stranded. In some cases, a target nucleic acid is a single stranded RNA (ssRNA). In some cases, a target ssRNA (e.g., a target cell ssRNA, a viral ssRNA, etc.) is selected from: mRNA, rRNA, tRNA, non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and microRNA (miRNA). In some cases, a target nucleic acid is a single stranded DNA (ssDNA) (e.g., a viral DNA). As noted above, in some cases, a target nucleic acid is single stranded.

[00374] A target nucleic acid can be located anywhere, for example, outside of a cell *in vitro*, inside of a cell *in vitro*, inside of a cell *in vivo*, inside of a cell *ex vivo*. Suitable target cells (which can comprise target nucleic acids such as genomic DNA) include, but are not limited to: a bacterial cell; an archaeal cell; a cell of a single-cell eukaryotic organism; a plant cell; an algal cell, e.g., *Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Chlorella pyrenoidosa*, *Sargassum patens, C. agardh*, and the like; a fungal cell (e.g., a yeast cell); an animal cell; a cell from an invertebrate animal (e.g. fruit fly, a cnidarian, an echinoderm, a nematode, etc.); a cell of an insect (e.g., a mosquito; a bee; an agricultural pest; etc.); a cell of an arachnid (e.g., a spider; a tick; etc.); a cell from a vertebrate animal (e.g., a fish, an amphibian, a reptile, a bird, a mammal); a cell from a mammal (e.g., a cell from a rodent; a cell from a human; a cell of a non-human mammal; a cell of a rodent (e.g., a mouse, a rat); a cell of a lagomorph (e.g., a rabbit); a cell of an ungulate (e.g., a cow, a horse, a camel, a llama, a vicuña, a sheep, a goat, etc.); a cell of a marine mammal (e.g., a whale, a seal, an elephant seal, a dolphin, a sea lion; etc.) and the like. Any type of cell may be of interest (e.g. a stem cell, e.g. an embryonic stem (ES) cell, an induced pluripotent stem (iPS) cell, a germ cell (e.g., an oocyte, a sperm, an oogonia, a spermatogonia, etc.), an adult stem cell, a somatic cell, e.g. a fibroblast, a hematopoietic cell, a neuron, a muscle cell, a bone cell, a hepatocyte, a pancreatic cell; an in vitro or in vivo embryonic cell of an embryo at any stage, e.g., a 1-cell, 2-cell, 4-cell, 8-cell, etc. stage zebrafish embryo; etc.).

[00375] Cells may be from established cell lines or they may be primary cells, where "primary cells", "primary cell lines", and "primary cultures" are used interchangeably herein to refer to cells and cells cultures that have been derived from a subject and allowed to grow in vitro for a limited number of passages, i.e. splittings, of the culture. For example, primary cultures are cultures that may have been passaged 0 times, 1 time, 2 times, 4 times, 5 times, 10 times, or 15 times, but not enough times go through the crisis stage. Typically, the primary cell lines are maintained for fewer than 10 passages *in vitro*. Target cells can be unicellular organisms and/or can be grown in culture. If the cells are primary cells, they may be harvest from an individual by any convenient method. For example, leukocytes may

be conveniently harvested by apheresis, leukocytapheresis, density gradient separation, etc., while cells from tissues such as skin, muscle, bone marrow, spleen, liver, pancreas, lung, intestine, stomach, etc. can be conveniently harvested by biopsy.

[00376] In some of the above applications, the subject methods may be employed to induce target nucleic acid cleavage, target nucleic acid modification, and/or to bind target nucleic acids (e.g., for visualization, for collecting and/or analyzing, etc.) in mitotic or post-mitotic cells *in vivo* and/or *ex vivo* and/or *in vitro* (e.g., to disrupt production of a protein encoded by a targeted mRNA, to cleave or otherwise modify target DNA, to geneically modify a target cell, and the like). Because the guide RNA provides specificity by hybridizing to target nucleic acid, a mitotic and/or post-mitotic cell of interest in the disclosed methods may include a cell from any organism (e.g. a bacterial cell, an archaeal cell, a cell of a single-cell eukaryotic organism, a plant cell, an algal cell, e.g., *Botryococcus braunii, Chlamydomonas reinhardtii, Nannochloropsis gaditana, Chlorella pyrenoidosa*, *Sargassum patens, C. agardh*, and the like, a fungal cell (e.g., a yeast cell), an animal cell, a cell from an invertebrate animal (e.g. fruit fly, cnidarian, echinoderm, nematode, etc.), a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal), a cell from a mammal, a cell from a rodent, a cell from a human, etc.). In some cases, a subject CasZ protein (and/or nucleic acid encoding the protein such as DNA and/or RNA), and/or CasZ guide RNA (and/or a DNA encoding the guide RNA), and/or donor template, and/or RNP can be intrduced into an individual (i.e., the target cell can be in vivo) (e.g., a mammal, a rat, a mouse, a pig, a primate, a non-human primate, a human, etc.). In some cases, such an administration can be for the purpose of treating and/or preventing a disease, e.g., by editing the genome of targeted cells.

[00377] Plant cells include cells of a monocotyledon, and cells of a dicotyledon. The cells can be root cells, leaf cells, cells of the xylem, cells of the phloem, cells of the cambium, apical meristem cells, parenchyma cells, collenchyma cells, sclerenchyma cells, and the like. Plant cells include cells of agricultural crops such as wheat, corn, rice, sorghum, millet, soybean, etc. Plant cells include cells of agricultural fruit and nut plants, e.g., plant that produce apricots, oranges, lemons, apples, plums, pears, almonds, etc.

[00378] Additional examples of target cells are listed above in the section titled "Modified cells." Non-limiting examples of cells (target cells) include: a prokaryotic cell, eukaryotic cell, a bacterial cell, an archaeal cell, a cell of a single-cell eukaryotic organism, a protozoa cell, a cell from a plant (e.g., cells from plant crops, fruits, vegetables, grains, soy bean, corn, maize, wheat, seeds, tomatos, rice, cassava, sugarcane, pumpkin, hay, potatos, cotton, cannabis, tobacco, flowering plants, conifers, gymnosperms, angiosperms, ferns, clubmosses, hornworts, liverworts, mosses, dicotyledons, monocotyledons, etc.), an algal cell, (e.g., *Botryococcus braunii*, *Chlamydomonas reinhardtii*, *Nannochloropsis gaditana*, *Chlorella pyrenoidosa*, *Sargassum patens*, *C. agardh*, and the like), seaweeds (e.g. kelp) a fungal cell

(e.g., a yeast cell, a cell from a mushroom), an animal cell, a cell from an invertebrate animal (e.g., fruit fly, cnidarian, echinoderm, nematode, etc.), a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal), a cell from a mammal (e.g., an ungulate (e.g., a pig, a cow, a goat, a sheep); a rodent (e.g., a rat, a mouse); a non-human primate; a human; a feline (e.g., a cat); a canine (e.g., a dog); etc.), and the like. In some cases, the cell is a cell that does not originate from a natural organism (e.g., the cell can be a synthetically made cell; also referred to as an artificial cell).

[00379] A cell can be an *in vitro* cell (e.g., established cultured cell line). A cell can be an *ex vivo* cell (cultured cell from an individual). A cell can be and *in vivo* cell (e.g., a cell in an individual). A cell can be an isolated cell. A cell can be a cell inside of an organism. A cell can be an organism. A cell can be a cell in a cell culture (e.g., *in vitro* cell culture). A cell can be one of a collection of cells. A cell can be a prokaryotic cell or derived from a prokaryotic cell. A cell can be a bacterial cell or can be derived from a bacterial cell. A cell can be an archaeal cell or derived from an archaeal cell. A cell can be a eukaryotic cell or derived from a eukaryotic cell. A cell can be a plant cell or derived from a plant cell. A cell can be an animal cell or derived from an animal cell. A cell can be an invertebrate cell or derived from an invertebrate cell. A cell can be a vertebrate cell or derived from a vertebrate cell. A cell can be a mammalian cell or derived from a mammalian cell. A cell can be a rodent cell or derived from a rodent cell. A cell can be a human cell or derived from a human cell. A cell can be a microbe cell or derived from a microbe cell. A cell can be a fungi cell or derived from a fungi cell. A cell can be an insect cell. A cell can be an arthropod cell. A cell can be a protozoan cell. A cell can be a helminth cell.

[00380] Suitable cells include a stem cell (e.g. an embryonic stem (ES) cell, an induced pluripotent stem (iPS) cell; a germ cell (e.g., an oocyte, a sperm, an oogonia, a spermatogonia, etc.); a somatic cell, e.g. a fibroblast, an oligodendrocyte, a glial cell, a hematopoietic cell, a neuron, a muscle cell, a bone cell, a hepatocyte, a pancreatic cell, etc.

[00381] Suitable cells include human embryonic stem cells, fetal cardiomyocytes, myofibroblasts, mesenchymal stem cells, autotransplated expanded cardiomyocytes, adipocytes, totipotent cells, pluripotent cells, blood stem cells, myoblasts, adult stem cells, bone marrow cells, mesenchymal cells, embryonic stem cells, parenchymal cells, epithelial cells, endothelial cells, mesothelial cells, fibroblasts, osteoblasts, chondrocytes, exogenous cells, endogenous cells, stem cells, hematopoietic stem cells, bone-marrow derived progenitor cells, myocardial cells, skeletal cells, fetal cells, undifferentiated cells, multi-potent progenitor cells, unipotent progenitor cells, monocytes, cardiac myoblasts, skeletal myoblasts, macrophages, capillary endothelial cells, xenogenic cells, allogenic cells, and post-natal stem cells.

[00382] In some cases, the cell is an immune cell, a neuron, an epithelial cell, and endothelial cell, or a stem cell. In some cases, the immune cell is a T cell, a B cell, a monocyte, a natural killer cell, a

dendritic cell, or a macrophage. In some cases, the immune cell is a cytotoxic T cell. In some cases, the immune cell is a helper T cell. In some cases, the immune cell is a regulatory T cell (Treg).

[00383] In some cases, the cell is a stem cell. Stem cells include adult stem cells. Adult stem cells are also referred to as somatic stem cells.

[00384] Adult stem cells are resident in differentiated tissue, but retain the properties of self-renewal and ability to give rise to multiple cell types, usually cell types typical of the tissue in which the stem cells are found. Numerous examples of somatic stem cells are known to those of skill in the art, including muscle stem cells; hematopoietic stem cells; epithelial stem cells; neural stem cells; mesenchymal stem cells; mammary stem cells; intestinal stem cells; mesodermal stem cells; endothelial stem cells; olfactory stem cells; neural crest stem cells; and the like.

[00385] Stem cells of interest include mammalian stem cells, where the term "mammalian" refers to any animal classified as a mammal, including humans; non-human primates; domestic and farm animals; and zoo, laboratory, sports, or pet animals, such as dogs, horses, cats, cows, mice, rats, rabbits, etc. In some cases, the stem cell is a human stem cell. In some cases, the stem cell is a rodent (e.g., a mouse; a rat) stem cell. In some cases, the stem cell is a non-human primate stem cell.

[00386] Stem cells can express one or more stem cell markers, e.g., SOX9, KRT19, KRT7, LGR5, CA9, FXYD2, CDH6, CLDN18, TSPAN8, BPIFB1, OLFM4, CDH17, and PPARGC1A.

[00387] In some cases, the stem cell is a hematopoietic stem cell (HSC). HSCs are mesoderm-derived cells that can be isolated from bone marrow, blood, cord blood, fetal liver and yolk sac. HSCs are characterized as $CD34^+$ and $CD3^-$. HSCs can repopulate the erythroid, neutrophil-macrophage, megakaryocyte and lymphoid hematopoietic cell lineages *in vivo*. *In vitro*, HSCs can be induced to undergo at least some self-renewing cell divisions and can be induced to differentiate to the same lineages as is seen *in vivo*. As such, HSCs can be induced to differentiate into one or more of erythroid cells, megakaryocytes, neutrophils, macrophages, and lymphoid cells.

[00388] In other cases, the stem cell is a neural stem cell (NSC). Neural stem cells (NSCs) are capable of differentiating into neurons, and glia (including oligodendrocytes, and astrocytes). A neural stem cell is a multipotent stem cell which is capable of multiple divisions, and under specific conditions can produce daughter cells which are neural stem cells, or neural progenitor cells that can be neuroblasts or glioblasts, e.g., cells committed to become one or more types of neurons and glial cells respectively. Methods of obtaining NSCs are known in the art.

[00389] In other cases, the stem cell is a mesenchymal stem cell (MSC). MSCs originally derived from the embryonal mesoderm and isolated from adult bone marrow, can differentiate to form muscle, bone, cartilage, fat, marrow stroma, and tendon. Methods of isolating MSC are known in the art;

and any known method can be used to obtain MSC. See, e.g., U.S. Pat. No. 5,736,396, which describes isolation of human MSC.

[00390] A cell is in some cases a plant cell. A plant cell can be a cell of a monocotyledon. A cell can be a cell of a dicotyledon.

[00391] In some cases, the cell is a plant cell. For example, the cell can be a cell of a major agricultural plant, e.g., Barley, Beans (Dry Edible), Canola, Corn, Cotton (Pima), Cotton (Upland), Flaxseed, Hay (Alfalfa), Hay (Non-Alfalfa), Oats, Peanuts, Rice, Sorghum, Soybeans, Sugarbeets, Sugarcane, Sunflowers (Oil), Sunflowers (Non-Oil), Sweet Potatoes , Tobacco (Burley), Tobacco (Flue-cured), Tomatoes, Wheat (Durum), Wheat (Spring), Wheat (Winter), and the like. As another example, the cell is a cell of a vegetable crops which include but are not limited to, e.g., alfalfa sprouts, aloe leaves, arrow root, arrowhead, artichokes, asparagus, bamboo shoots, banana flowers, bean sprouts, beans, beet tops, beets, bittermelon, bok choy, broccoli, broccoli rabe (rappini), brussels sprouts, cabbage, cabbage sprouts, cactus leaf (nopales), calabaza, cardoon, carrots, cauliflower, celery, chayote, chinese artichoke (crosnes), chinese cabbage, chinese celery, chinese chives, choy sum, chrysanthemum leaves (tung ho), collard greens, corn stalks, corn-sweet, cucumbers, daikon, dandelion greens, dasheen, dau mue (pea tips), donqua (winter melon), eggplant, endive, escarole, fiddle head ferns, field cress, frisee, gai choy (chinese mustard), gailon, galanga (siam, thai ginger), garlic, ginger root, gobo, greens, hanover salad greens, huauzontle, jerusalem artichokes, jicama, kale greens, kohlrabi, lamb's quarters (quilete), lettuce (bibb), lettuce (boston), lettuce (boston red), lettuce (green leaf), lettuce (iceberg), lettuce (lolla rossa), lettuce (oak leaf - green), lettuce (oak leaf - red), lettuce (processed), lettuce (red leaf), lettuce (romaine), lettuce (ruby romaine), lettuce (russian red mustard), linkok, lo bok, long beans, lotus root, mache, maguey (agave) leaves, malanga, mesculin mix, mizuna, moap (smooth luffa), moo, moqua (fuzzy squash), mushrooms, mustard, nagaimo, okra, ong choy, onions green, opo (long squash), ornamental corn, ornamental gourds, parsley, parsnips, peas, peppers (bell type), peppers, pumpkins, radicchio, radish sprouts, radishes, rape greens, rape greens, rhubarb, romaine (baby red), rutabagas, salicornia (sea bean), sinqua (angled/ridged luffa), spinach, squash, straw bales, sugarcane, sweet potatoes, swiss chard, tamarindo, taro, taro leaf, taro shoots, tatsoi, tepeguaje (guaje), tindora, tomatillos, tomatoes, tomatoes (cherry), tomatoes (grape type), tomatoes (plum type), tumeric, turnip tops greens, turnips, water chestnuts, yampi, yams (names), yu choy, yuca (cassava), and the like.

[00392] A cell is in some cases an arthropod cell. For example, the cell can be a cell of a sub-order, a family, a sub-family, a group, a sub-group, or a species of, e.g., *Chelicerata*, *Myriapodia*, *Hexipodia*, *Arachnida*, *Insecta*, *Archaeognatha*, *Thysanura*, *Palaeoptera*, *Ephemeroptera*, *Odonata*, *Anisoptera*, *Zygoptera*, *Neoptera*, *Exopterygota*, *Plecoptera* , *Embioptera* , *Orthoptera*, *Zoraptera* , *Dermaptera*, *Dictyoptera*, *Notoptera*, *Grylloblattidae*, *Mantophasmatidae*, *Phasmatodea* , *Blattaria*, *Isoptera*, *Mantodea*, *Parapneuroptera*, *Psocoptera*, *Thysanoptera*, *Phthiraptera*, *Hemiptera*,

Endopterygota or Holometabola , *Hymenoptera* , *Coleoptera*, *Strepsiptera*, *Raphidioptera*, *Megaloptera*, *Neuroptera* , *Mecoptera* , *Siphonaptera*, *Diptera*, *Trichoptera*, or *Lepidoptera.*

[00393] A cell is in some cases an insect cell. For example, in some cases, the cell is a cell of a mosquito, a grasshopper, a true bug, a fly, a flea, a bee, a wasp, an ant, a louse, a moth, or a beetle.

Donor Polynucleotide (donor template)

[00394] Guided by a CasZ guide RNA, a CasZ protein in some cases generates site-specific double strand breaks (DSBs) or single strand breaks (SSBs) (e.g., when the CasZ protein is a nickase variant) within double-stranded DNA (dsDNA) target nucleic acids, which are repaired either by non-homologous end joining (NHEJ) or homology-directed recombination (HDR).

[00395] In some cases, contacting a target DNA (with a CasZ protein and a CasZ guide RNA) occurs under conditions that are permissive for nonhomologous end joining or homology-directed repair. Thus, in some cases, a subject method includes contacting the target DNA with a donor polynucleotide (e.g., by introducing the donor polynucleotide into a cell), wherein the donor polynucleotide, a portion of the donor polynucleotide, a copy of the donor polynucleotide, or a portion of a copy of the donor polynucleotide integrates into the target DNA. In some cases, the method does not comprise contacting a cell with a donor polynucleotide, and the target DNA is modified such that nucleotides within the target DNA are deleted.

[00396] In some cases, a CasZ trancRNA (or nucleic acid encoding same), a CasZ guide RNA (or nucleic acid encoding same), and/or a CasZ protein (or a nucleic acid encoding same, such as an RNA or a DNA, e.g, one or more expression vectors) are coadministered (e.g., contacted with a target nucleic acid, administered to cells, etc.) with a donor polynucleotide sequence that includes at least a segment with homology to the target DNA sequence, the subject methods may be used to add, i.e. insert or replace, nucleic acid material to a target DNA sequence (e.g. to "knock in" a nucleic acid, e.g., one that encodes for a protein, an siRNA, an miRNA, etc.), to add a tag (e.g., 6xHis, a fluorescent protein (e.g., a green fluorescent protein; a yellow fluorescent protein, etc.), hemagglutinin (HA), FLAG, etc.), to add a regulatory sequence to a gene (e.g. promoter, polyadenylation signal, internal ribosome entry sequence (IRES), 2A peptide, start codon, stop codon, splice signal, localization signal, etc.), to modify a nucleic acid sequence (e.g., introduce a mutation, remove a disease causnig mutation by introducing a correct sequence), and the like. As such, a complex comprising a CasZ guide RNA and CasZ protein (or CasZ guide RNA and CasZ trancRNA and CasZ protein) is useful in any *in vitro* or *in vivo* application in which it is desirable to modify DNA in a site-specific, i.e. "targeted", way, for example gene knock-out, gene knock-in, gene editing, gene tagging, etc., as used in, for example, gene therapy, e.g. to treat a disease or as an antiviral, antipathogenic, or anticancer therapeutic, the production of genetically modified organisms in agriculture, the large scale production of proteins by cells for therapeutic,

diagnostic, or research purposes, the induction of iPS cells, biological research, the targeting of genes of pathogens for deletion or replacement, etc.

[00397] In applications in which it is desirable to insert a polynucleotide sequence into he genome where a target sequence is cleaved, a donor polynucleotide (a nucleic acid comprising a donor sequence) can also be provided to the cell. By a "donor sequence" or "donor polynucleotide" or "donor template" it is meant a nucleic acid sequence to be inserted at the site cleaved by the CasZ protein (e.g., after dsDNA cleavage, after nicking a target DNA, after dual nicking a target DNA, and the like). The donor polynucleotide can contain sufficient homology to a genomic sequence at the target site, e.g. 70%, 80%, 85%, 90%, 95%, or 100% homology with the nucleotide sequences flanking the target site, e.g. within about 50 bases or less of the target site, e.g. within about 30 bases, within about 15 bases, within about 10 bases, within about 5 bases, or immediately flanking the target site, to support homology-directed repair between it and the genomic sequence to which it bears homology. Approximately 25, 50, 100, or 200 nucleotides, or more than 200 nucleotides, of sequence homology between a donor and a genomic sequence (or any integral value between 10 and 200 nucleotides, or more) can support homology-directed repair. Donor polynucleotides can be of any length, e.g. 10 nucleotides or more, 50 nucleotides or more, 100 nucleotides or more, 250 nucleotides or more, 500 nucleotides or more, 1000 nucleotides or more, 5000 nucleotides or more, etc.

[00398] The donor sequence is typically not identical to the genomic sequence that it replaces. Rather, the donor sequence may contain at least one or more single base changes, insertions, deletions, inversions or rearrangements with respect to the genomic sequence, so long as sufficient homology is present to support homology-directed repair (e.g., for gene correction, e.g., to convert a disease-causing base pair ot a non disease-causing base pair). In some embodiments, the donor sequence comprises a non-homologous sequence flanked by two regions of homology, such that homology-directed repair between the target DNA region and the two flanking sequences results in insertion of the non-homologous sequence at the target region. Donor sequences may also comprise a vector backbone containing sequences that are not homologous to the DNA region of interest and that are not intended for insertion into the DNA region of interest. Generally, the homologous region(s) of a donor sequence will have at least 50% sequence identity to a genomic sequence with which recombination is desired. In certain embodiments, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or 99.9% sequence identity is present. Any value between 1% and 100% sequence identity can be present, depending upon the length of the donor polynucleotide.

[00399] The donor sequence may comprise certain sequence differences as compared to the genomic sequence, e.g. restriction sites, nucleotide polymorphisms, selectable markers (e.g., drug resistance genes, fluorescent proteins, enzymes etc.), etc., which may be used to assess for successful insertion of the donor sequence at the cleavage site or in some cases may be used for other purposes

(e.g., to signify expression at the targeted genomic locus). In some cases, if located in a coding region, such nucleotide sequence differences will not change the amino acid sequence, or will make silent amino acid changes (i.e., changes which do not affect the structure or function of the protein). Alternatively, these sequences differences may include flanking recombination sequences such as FLPs, loxP sequences, or the like, that can be activated at a later time for removal of the marker sequence.

[00400] In some cases, the donor sequence is provided to the cell as single-stranded DNA. In some cases, the donor sequence is provided to the cell as double-stranded DNA. It may be introduced into a cell in linear or circular form. If introduced in linear form, the ends of the donor sequence may be protected (e.g., from exonucleolytic degradation) by any convenient method and such methods are known to those of skill in the art. For example, one or more dideoxynucleotide residues can be added to the 3' terminus of a linear molecule and/or self-complementary oligonucleotides can be ligated to one or both ends. See, for example, Chang et al. (1987) Proc. Natl. Acad Sci USA 84:4959-4963; Nehls et al. (1996) Science 272:886-889. Additional methods for protecting exogenous polynucleotides from degradation include, but are not limited to, addition of terminal amino group(s) and the use of modified internucleotide linkages such as, for example, phosphorothioates, phosphoramidates, and O-methyl ribose or deoxyribose residues. As an alternative to protecting the termini of a linear donor sequence, additional lengths of sequence may be included outside of the regions of homology that can be degraded without impacting recombination. A donor sequence can be introduced into a cell as part of a vector molecule having additional sequences such as, for example, replication origins, promoters and genes encoding antibiotic resistance. Moreover, donor sequences can be introduced as naked nucleic acid, as nucleic acid complexed with an agent such as a liposome or poloxamer, or can be delivered by viruses (e.g., adenovirus, AAV), as described elsewhere herein for nucleic acids encoding a CasZ guide RNA and/or a CasZ fusion polypeptide and/or donor polynucleotide.

TRANSGENIC, NON-HUMAN ORGANISMS

[00401] As described above, in some cases, a nucleic acid (e.g., a recombinant expression vector) of the present disclosure (e.g., a nucleic acid comprising a nucleotide sequence encoding a CasZ polypeptide; a nucleic acid comprising a nucleotide sequence encoding a CasZ fusion polypeptide; etc.), is used as a transgene to generate a transgenic non-human organism that produces a CasZ polypeptide, or a CasZ fusion polypeptide, of the present disclosure. The present disclosure provides a transgenic-non-human organism comprising a nucleotide sequence encoding a CasZ polypeptide, or a CasZ fusion polypeptide, of the present disclosure.

Transgenic, non-human animals

[00402] The present disclosure provides a transgenic non-human animal, which animal comprises a transgene comprising a nucleic acid comprising a nucleotide sequence encoding a CasZ polypeptide or a CasZ fusion polypeptide. In some embodiments, the genome of the transgenic non-

human animal comprises a nucleotide sequence encoding a CasZ polypeptide, or a CasZ fusion polypeptide, of the present disclosure. In some cases, the transgenic non-human animal is homozygous for the genetic modification. In some cases, the transgenic non-human animal is heterozygous for the genetic modification. In some embodiments, the transgenic non-human animal is a vertebrate, for example, a fish (e.g., salmon, trout, zebra fish, gold fish, puffer fish, cave fish, etc.), an amphibian (frog, newt, salamander, etc.), a bird (e.g., chicken, turkey, etc.), a reptile (e.g., snake, lizard, etc.), a non-human mammal (e.g., an ungulate, e.g., a pig, a cow, a goat, a sheep, etc.; a lagomorph (e.g., a rabbit); a rodent (e.g., a rat, a mouse); a non-human primate; etc.), etc. In some cases, the transgenic non-human animal is an invertebrate. In some cases, the transgenic non-human animal is an insect (e.g., a mosquito; an agricultural pest; etc.). In some cases, the transgenic non-human animal is an arachnid.

[00403] Nucleotide sequences encoding a a CasZ polypeptide,e or a CasZ fusion polypeptide, of the present disclosure can be under the control of (i.e., operably linked to) an unknown promoter (e.g., when the nucleic acid randomly integrates into a host cell genome) or can be under the control of (i.e., operably linked to) a known promoter. Suitable known promoters can be any known promoter and include constitutively active promoters (e.g., CMV promoter), inducible promoters (e.g., heat shock promoter, tetracycline-regulated promoter, steroid-regulated promoter, metal-regulated promoter, estrogen receptor-regulated promoter, etc.), spatially restricted and/or temporally restricted promoters (e.g., a tissue specific promoter, a cell type specific promoter, etc.), etc.

Transgenic plants

[00404] As described above, in some cases, a nucleic acid (e.g., a recombinant expression vector) of the present disclosure (e.g., a nucleic acid comprising a nucleotide sequence encoding a CasZ polypeptide of the present disclosure; a nucleic acid comprising a nucleotide sequence encoding a CasZ fusion polypeptide of the present disclosure; etc.), is used as a transgene to generate a transgenic plant that produces a CasZ polypeptide, or a CasZ fusion polypeptide, of the present disclosure. The present disclosure provides a transgenic plant comprising a nucleotide sequence encoding a CasZ polypeptide, or a CasZ fusion polypeptide, of the present disclosure.In some embodiments, the genome of the transgenic plant comprises a subject nucleic acid. In some embodiments, the transgenic plant is homozygous for the genetic modification. In some embodiments, the transgenic plant is heterozygous for the genetic modification.

[00405] Methods of introducing exogenous nucleic acids into plant cells are well known in the art. Such plant cells are considered "transformed," as defined above. Suitable methods include viral infection (such as double stranded DNA viruses), transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate precipitation, direct microinjection, silicon carbide whiskers technology, Agrobacterium-mediated transformation and the like. The choice of

method is generally dependent on the type of cell being transformed and the circumstances under which the transformation is taking place (i.e. in vitro, ex vivo, or in vivo).

[00406] Transformation methods based upon the soil bacterium *Agrobacterium tumefaciens* are particularly useful for introducing an exogenous nucleic acid molecule into a vascular plant. The wild type form of *Agrobacterium* contains a Ti (tumor-inducing) plasmid that directs production of tumorigenic crown gall growth on host plants. Transfer of the tumor-inducing T-DNA region of the Ti plasmid to a plant genome requires the Ti plasmid-encoded virulence genes as well as T-DNA borders, which are a set of direct DNA repeats that delineate the region to be transferred. An *Agrobacterium*-based vector is a modified form of a Ti plasmid, in which the tumor inducing functions are replaced by the nucleic acid sequence of interest to be introduced into the plant host.

[00407] *Agrobacterium*-mediated transformation generally employs cointegrate vectors or binary vector systems, in which the components of the Ti plasmid are divided between a helper vector, which resides permanently in the *Agrobacterium* host and carries the virulence genes, and a shuttle vector, which contains the gene of interest bounded by T-DNA sequences. A variety of binary vectors is well known in the art and are commercially available, for example, from Clontech (Palo Alto, Calif.). Methods of coculturing *Agrobacterium* with cultured plant cells or wounded tissue such as leaf tissue, root explants, hypocotyledons, stem pieces or tubers, for example, also are well known in the art. See, e.g., Glick and Thompson, (eds.), Methods in Plant Molecular Biology and Biotechnology, Boca Raton, Fla.: CRC Press (1993).

[00408] Microprojectile-mediated transformation also can be used to produce a subject transgenic plant. This method, first described by Klein et al. (Nature 327:70-73 (1987)), relies on microprojectiles such as gold or tungsten that are coated with the desired nucleic acid molecule by precipitation with calcium chloride, spermidine or polyethylene glycol. The microprojectile particles are accelerated at high speed into an angiosperm tissue using a device such as the BIOLISTIC PD-1000 (Biorad; Hercules Calif.).

[00409] A nucleic acid of the present disclosure (e.g., a nucleic acid (e.g., a recombinant expression vector) comprising a nucleotide sequence encoding a CasZ polypeptide, or a CasZ fusion polypeptide, of the present disclosure) may be introduced into a plant in a manner such that the nucleic acid is able to enter a plant cell(s), e.g., via an in vivo or ex vivo protocol. By "in vivo," it is meant in the nucleic acid is administered to a living body of a plant e.g. infiltration. By “ex vivo” it is meant that cells or explants are modified outside of the plant, and then such cells or organs are regenerated to a plant. A number of vectors suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described, including those described in Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology Academic Press, and Gelvin et al., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers. Specific examples include those derived from a Ti plasmid of

Agrobacterium tumefaciens, as well as those disclosed by Herrera-Estrella et al. (1983) Nature 303: 209, Bevan (1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technolo 3: 637-642. Alternatively, non-Ti vectors can be used to transfer the DNA into plants and cells by using free DNA delivery techniques. By using these methods transgenic plants such as wheat, rice (Christou (1991) Bio/Technology 9:957-9 and 4462) and corn (Gordon-Kamm (1990) Plant Cell 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks et al. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technolo 10: 667-674; Wan and Lemeaux (1994) Plant Physiol 104: 37-48 and for Agrobacterium-mediated DNA transfer (Ishida et al. (1996) Nature Biotech 14: 745-750). Exemplary methods for introduction of DNA into chloroplasts are biolistic bombardment, polyethylene glycol transformation of protoplasts, and microinjection (Danieli et al Nat. Biotechnol 16:345-348, 1998; Staub et al Nat. Biotechnol 18: 333-338, 2000; O'Neill et al Plant J. 3:729-738, 1993; Knoblauch et al Nat. Biotechnol 17: 906-909; U.S. Pat. Nos. 5,451,513, 5,545,817, 5,545,818, and 5,576,198; in Intl. Application No. WO 95/16783; and in Boynton et al., Methods in Enzymology 217: 510-536 (1993), Svab et al., Proc. Natl. Acad. Sci. USA 90: 913-917 (1993), and McBride et al., Proc. Natl. Acad. Sci. USA 91: 7301-7305 (1994)). Any vector suitable for the methods of biolistic bombardment, polyethylene glycol transformation of protoplasts and microinjection will be suitable as a targeting vector for chloroplast transformation. Any double stranded DNA vector may be used as a transformation vector, especially when the method of introduction does not utilize Agrobacterium.

[00410] Plants which can be genetically modified include grains, forage crops, fruits, vegetables, oil seed crops, palms, forestry, and vines. Specific examples of plants which can be modified follow: maize, banana, peanut, field peas, sunflower, tomato, canola, tobacco, wheat, barley, oats, potato, soybeans, cotton, carnations, sorghum, lupin and rice.

[00411] The present disclosure provides transformed plant cells, tissues, plants and products that contain the transformed plant cells. A feature of the subject transformed cells, and tissues and products that include the same is the presence of a subject nucleic acid integrated into the genome, and production by plant cells of a CasZ polypeptide, or a CasZ fusion polypeptide, of the present disclosure. Recombinant plant cells of the present invention are useful as populations of recombinant cells, or as a tissue, seed, whole plant, stem, fruit, leaf, root, flower, stem, tuber, grain, animal feed, a field of plants, and the like.

[00412] Nucleotide sequences encoding a CasZ polypeptide, or a CasZ fusion polypeptide, of the present disclosure can be under the control of (i.e., operably linked to) an unknown promoter (e.g., when the nucleic acid randomly integrates into a host cell genome) or can be under the control of (i.e., operably linked to) a known promoter. Suitable known promoters can be any known promoter and

include constitutively active promoters, inducible promoters, spatially restricted and/or temporally restricted promoters, etc.

Examples of Non-Limiting Aspects of the Disclosure

[00413] Aspects, including embodiments, of the present subject matter described above may be beneficial alone or in combination, with one or more other aspects or embodiments. Without limiting the foregoing description, certain non-limiting aspects of the disclosure, numbered 1-36 are provided below. As will be apparent to those of skill in the art upon reading this disclosure, each of the individually numbered aspects may be used or combined with any of the preceding or following individually numbered aspects. This is intended to provide support for all such combinations of aspects and is not limited to combinations of aspects explicitly provided below:

Aspects

[00414] Aspect 1. A method of guiding a CasZ polypeptide to a target sequence of a target nucleic acid, the method comprising contacting the target nucleic acid with an engineered and/or non-naturally occurring complex comprising:(a) a CasZ polypeptide; and (b) a CasZ guide RNA that comprises a guide sequence that hybridizes to a target sequence of the target nucleic acid, and comprises a region that binds to the CasZ polypeptide.

[00415] Aspect 2. The method of aspect 1, wherein the method results in modification of the target nucleic acid, modulation of transcription from the target nucleic acid, or modification of a polypeptide associated with a target nucleic acid.

[00416] Aspect 3. The method of aspect 2, wherein the target nucleic acid is modified by being cleaved.

[00417] Aspect 4. The method of any one of aspects 1-3, wherein the target nucleic acid is selected from: double stranded DNA, single stranded DNA, RNA, genomic DNA, and extrachromosomal DNA.

[00418] Aspect 5. The method of any one of aspects 1-4, wherein the guide sequence and the region that binds to the CasZ polypeptide are heterologous to one another.

[00419] Aspect 6. The method of any one of aspects 1-5, wherein said contacting results in genome editing.

[00420] Aspect 7. The method of any one of aspects 1-5, wherein said contacting takes place outside of a bacterial cell and outside of an archaeal cell.

[00421] Aspect 8. The method of any one of aspects 1-5, wherein said contacting takes place *in vitro* outside of a cell.

[00422] Aspect 9. The method of any one of aspects 1-7, wherein said contacting takes place inside of a target cell.

[00423] Aspect 10. The method of aspect 9, wherein said contacting comprises: introducing into the target cell at least one of: (a) the CasZ polypeptide, or a nucleic acid encoding the CasZ polypeptide; and (b) the CasZ guide RNA, or a nucleic acid encoding the CasZ guide RNA.

[00424] Aspect 11. The method of aspect 10, wherein the nucleic acid encoding the CasZ polypeptide is a non-naturally sequence that is codon optimized for expression in the target cell.

[00425] Aspect 12. The method of any one of aspects 9-11, wherein the target cell is a eukaryotic cell.

[00426] Aspect 13. The method of any one of aspects 9-12, wherein the target cell is in culture *in vitro*.

[00427] Aspect 14. The method of any one of aspects 9-12, wherein the target cell is *in vivo*.

[00428] Aspect 15. The method of any one of aspects 9-12, wherein the target cell is *ex vivo*.

[00429] Aspect 16. The method of aspect 12, wherein the eukaryotic cell is selected from the group consisting of: a plant cell, a fungal cell, a single cell eukaryotic organism, a mammalian cell, a reptile cell, an insect cell, an avian cell, a fish cell, a parasite cell, an arthropod cell, an arachnid cell, a cell of an invertebrate, a cell of a vertebrate, a rodent cell, a mouse cell, a rat cell, a primate cell, a non-human primate cell, and a human cell.

[00430] Aspect 17. The method of any one of aspects 9-16, wherein said contacting further comprises: introducing a DNA donor template into the target cell.

[00431] Aspect 18. The method of any one of aspects 1-17, wherein the method comprises contacting the target nucleic acid with a CasZ transactivating noncoding RNA (trancRNA).

[00432] Aspect 19. The method of any one of aspects 9-17, wherein said contacting comprises: introducing a CasZ transactivating noncoding RNA (trancRNA) and/or a nucleic acid encoding the CasZ trancRNA into the target cell.

[00433] Aspect 20. The method of aspect 18 or aspect 19, wherein the trancRNA comprises a nucleotide sequence having 70% or more (at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) nucleotide sequence identity with a trancRNA sequence of Table 2.

[00434] Aspect 21. A composition comprising an engineered and/or non-naturally occurring complex comprising: (a) a CasZ polypeptide, or a nucleic acid encoding said CasZ polypeptide; and (b) a CasZ guide RNA, or a nucleic acid encoding said CasZ guide RNA, wherein said CasZ guide RNA comprises a guide sequence that is complementary to a target sequence of a target nucleic acid, and comprises a region that can bind to the CasZ polypeptide.

[00435] Aspect 22. The composition of aspect 21, further comprising a CasZ transactivating noncoding RNA (trancRNA), or a nucleic acid encoding said CasZ trancRNA.

[00436] Aspect 23. A kit comprising an engineered and/or non-naturally occurring complex comprising: (a) a CasZ polypeptide, or a nucleic acid encoding said CasZ polypeptide; (b) a CasZ guide RNA, or a nucleic acid encoding said CasZ guide RNA, wherein said CasZ guide RNA comprises a guide sequence that is complementary to a target sequence of a target nucleic acid, and comprises a region that can bind to the CasZ polypeptide.

[00437] Aspect 24. The kit of aspect 23, further comprising a CasZ transactivating noncoding RNA (trancRNA), or a nucleic acid encoding said CasZ trancRNA.

[00438] Aspect 25. A genetically modified eukaryotic cell, comprising at least one of: (a) a CasZ polypeptide, or a nucleic acid encoding said CasZ polypeptide; (b) a CasZ guide RNA, or a nucleic acid encoding said CasZ guide RNA, wherein said CasZ guide RNA comprises a guide sequence that is complementary to a target sequence of a target nucleic acid, and comprises a region that can bind to the CasZ polypeptide; and (c) a CasZ transactivating noncoding RNA (trancRNA), or a nucleic acid encoding said CasZ trancRNA.

[00439] Aspect 26. The composition, kit, or eukaryotic cell of any one of the preceding aspects, characterized by at least one of: (a) the nucleic acid encoding said CasZ polypeptide comprises a nucleotide sequence that: (i) encodes the CasZ polypeptide and, (ii) is operably linked to a heterologous promoter; (b) the nucleic acid encoding said CasZ guide RNA comprises a nucleotide sequence that: (i) encodes the CasZ guide RNA and, (ii) is operably linked to a heterologous promoter; and (c) the nucleic acid encoding said CasZ trancRNA comprises a nucleotide sequence that: (i) encodes the CasZ trancRNA and, (ii) is operably linked to a heterologous promoter.

[00440] Aspect 27. The composition, kit, or eukaryotic cell of any one of the preceding aspects, for use in a method of therapeutic treatment of a patient.

[00441] Aspect 28. The method, composition, kit, or eukaryotic cell of any one of the preceding aspects, wherein at least one of: the nucleic acid encoding said CasZ polypeptide, the nucleic acid encoding said CasZ guide RNA, and the nucleic acid encoding said CasZ trancRNA, is a recombinant expression vector.

[00442] Aspect 29. The method, composition, kit, or eukaryotic cell of any one of the preceding aspects, wherein the CasZ guide RNA and/or the CasZ trancRNA comprises one or more of: a modified nucleobase, a modified backbone or non-natural internucleoside linkage, a modified sugar moiety, a Locked Nucleic Acid, a Peptide Nucleic Acid, and a deoxyribonucleotide.

[00443] Aspect 30. The method, composition, kit, or eukaryotic cell of any one of the preceding aspects, wherein the CasZ polypeptide is a variant CasZ polypeptide with reduced nuclease activity compared to a corresponding wild type CasZ protein.

[00444] Aspect 31. The method, composition, kit, or eukaryotic cell of any one of the preceding aspects, wherein at least one of: the CasZ polypeptide, the nucleic acid encoding the CasZ polypeptide, the CasZ guide RNA, the nucleic acid encoding the CasZ guide RNA, the CasZ trancRNA, and the nucleic acid encoding the CasZ trancRNA; is conjugated to a heterologous moiety.

[00445] Aspect 32. The method, composition, kit, or eukaryotic cell of aspect 31, wherein the heterologous moiety is a heterologous polypeptide.

[00446] Aspect 33. The method, composition, kit, or eukaryotic cell of any one of the preceding aspects, wherein the CasZ polypeptide has reduced nuclease activity compared to a corresponding wild type CasZ protein, and is fused to a heterologous polypeptide.

[00447] Aspect 34. The method, composition, kit, or eukaryotic cell of aspect 33, wherein the heterologous polypeptide: (i) has DNA modifying activity, (ii) exhibits the ability to increase or decrease transcription, and/or (iii) has enzymatic activity that modifies a polypeptide associated with DNA.

[00448] Aspect 35. The method, composition, kit, or eukaryotic cell of any one of the preceding aspects, wherein the CasZ polypeptide comprises an amino acid sequence having 70% or more (at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) amino acid sequence identity with a CasZ protein of Figure 1 or Figure 7.

[00449] Aspect 36. The method, composition, kit, or eukaryotic cell of any one of the preceding aspects, wherein the guide sequence and the region that binds to the CasZ polypeptide are heterologous to one another.

[00450] Aspect 37. A method of detecting a target DNA in a sample, the method comprising: (a) contacting the sample with: (i) a CasZ polypeptide; (ii) a guide RNA comprising: a region that binds to the CasZ polypeptide, and a guide sequence that hybridizes with the target DNA; and (iii) a detector DNA that is single stranded and does not hybridize with the guide sequence of the guide RNA; and (b) measuring a detectable signal produced by cleavage of the single stranded detector DNA by the CasZ, thereby detecting the target DNA.

[00451] Aspect 38. The method of aspect 37, wherein the target DNA is single stranded.

[00452] Aspect 39. The method of aspect 37 or 38, wherein the target DNA is double stranded.

[00453] Aspect 40. The method of any one of aspects 37-39, wherein the target DNA is viral DNA.

[00454] Aspect 41. The method of any one of aspects 37-40, wherein the target DNA is papovavirus, hepdnavirus, herpesvirus, adenovirus, poxvirus, or parvovirus DNA.

[00455] Aspect 42. The method of any one of aspects 37-41, wherein the CasZ polypeptide comprises an amino acid sequence having at least 85% amino acid sequence identity to the CasZ amino acid sequence set forth in any one of FIGs. 1 and 7.

[00456] Aspect 43. The method of any one of aspects 37-41, wherein the CasZ polypeptide is a Cas14a polypeptide.

[00457] Aspect 44. The method according to any one of aspects 37-43, wherein the sample comprises DNA molecules from a cell lysate.

[00458] Aspect 45. The method according to any one of aspects 37-44, wherein the sample comprises cells.

[00459] Aspect 46. The method according to any one of aspects 37-45, wherein said contacting is carried out inside of a cell *in vitro*, *ex vivo*, or *in vivo*.

[00460] Aspect 47. The method according to aspect 46, wherein the cell is a eukaryotic cell.

[00461] Aspect 48. The method according to any one of aspects 37-47, wherein the target DNA can be detected at a concentration as low as 10 aM.

[00462] Aspect 49. The method according to any one of aspects 37-48, comprising determining an amount of the target DNA present in the sample.

[00463] Aspect 50. The method according to aspect 49, wherein said determining comprises: measuring the detectable signal to generate a test measurement; measuring a detectable signal produced by a reference sample or cell to generate a reference measurement; and comparing the test measurement to the reference measurement to determine an amount of target DNA present in the sample.

[00464] Aspect 51. The method according to any one of aspects 37-50, wherein measuring a detectable signal comprises one or more of: gold nanoparticle based detection, fluorescence polarization, colloid phase transition/dispersion, electrochemical detection, and semiconductor-based sensing.

[00465] Aspect 52. The method according to any one of aspects 37-51, wherein the single stranded detector DNA comprises a fluorescence-emitting dye pair.

[00466] Aspect 53. The method according to aspect 52, wherein the fluorescence-emitting dye pair produces an amount of detectable signal prior to cleavage of the single stranded detector DNA, and the amount of detectable signal is reduced after cleavage of the single stranded detector DNA.

[00467] Aspect 54. The method according to aspect 52, wherein the single stranded detector DNA produces a first detectable signal prior to being cleaved and a second detectable signal after cleavage of the single stranded detector DNA.

[00468] Aspect 55. The method according to any one of aspects 52-54, wherein the fluorescence-emitting dye pair is a fluorescence resonance energy transfer (FRET) pair.

[00469] Aspect 56. The method according to aspect 18, wherein an amount of detectable signal increases after cleavage of the single stranded detector DNA.

[00470] Aspect 57. The method according to aspect 52 or aspect 56, wherein the fluorescence-emitting dye pair is a quencher/fluor pair.

[00471] Aspect 58. The method according to any one of aspects 52-57, wherein the single stranded detector DNA comprises two or more fluorescence-emitting dye pairs.

[00472] Aspect 59. The method according to aspect 58, wherein said two or more fluorescence-emitting dye pairs include a fluorescence resonance energy transfer (FRET) pair and a quencher/fluor pair.

[00473] Aspect 60. The method according to any one of aspects 37-59, wherein the single stranded detector DNA comprises a modified nucleobase, a modified sugar moiety, and/or a modified nucleic acid linkage.

[00474] Aspect 61. The method according to any one of aspects 37-60, wherein the method comprises amplifying nucleic acids in the sample.

[00475] Aspect 62. The method according to aspect 61, wherein said amplifying comprises isothermal amplification.

[00476] Aspect 63. The method according to aspect 62, wherein the isothermal amplification comprises recombinase polymerase amplification (RPA).

[00477] Aspect 64. The method according to any one of aspects 61-63, wherein said amplifying begins prior to the contacting of step (a).

[00478] Aspect 65. The method according to any one of aspects 61-63, wherein said amplifying begins together with the contacting of step (a).

[00479] Aspect 66. A kit for detecting a target DNA in a sample, the kit comprising: (a) a guide RNA, or a nucleic acid encoding the guide RNA; wherein the guide RNA comprises: a region that binds to a CasZ polypeptide, and a guide sequence that is complementary to a target DNA; and (b) a labeled detector DNA that is single stranded and does not hybridize with the guide sequence of the guide RNA.

[00480] Aspect 67. The kit of aspect 66, further comprising a CasZ polypeptide.

[00481] Aspect 68. The kit of aspect 67, wherein the CasZ polypeptide comprises an amino acid sequence having at least 85% amino acid sequence identity to the CasZ amino acid sequence set forth in any one of FIGs. 1 and 7.

[00482] Aspect 69. The kit of aspect 67, wherein the CasZ polypeptide is a Cas14a polypeptide.

[00483] Aspect 70. The kit of any one of aspects 66-69, wherein the single stranded detector DNA comprises a fluorescence-emitting dye pair.

[00484] Aspect 71. The kit of aspect 70, wherein the fluorescence-emitting dye pair is a FRET pair.

[00485] Aspect 72. The kit of aspect 70, wherein the fluorescence-emitting dye pair is a quencher/fluor pair.

[00486] Aspect 73. The kit of any one of aspects 70-72, wherein the single stranded detector DNA comprises two or more fluorescence-emitting dye pairs.

[00487] Aspect 74. The kit of aspect 73, wherein said two or more fluorescence-emitting dye pairs include a first fluorescence-emitting dye pair that produces a first detectable signal and a second fluorescence-emitting dye pair that produces a second detectable signal.

[00488] Aspect 75. The kit of any one of aspects 66-74, further comprising nucleic acid amplification components.

[00489] Aspect 76. The kit of aspect 75, wherein the nucleic acid amplification components are components for recombinase polymerase amplification (RPA).

[00490] Aspect 77. A method of cleaving single stranded DNAs (ssDNAs), the method comprising: contacting a population of nucleic acids, wherein said population comprises a target DNA and a plurality of non-target ssDNAs, with: (i) a CasZ polypeptide; and (ii) a guide RNA comprising: a region that binds to the CasZ polypeptide, and a guide sequence that hybridizes with the target DNA, wherein the CasZ polypeptide cleaves non-target ssDNAs of said plurality.

[00491] Aspect 78. The method of aspect 77, wherein said contacting is inside of a cell *in vitro*, *ex vivo*, or *in vivo*.

[00492] Aspect 79. The method of aspect 78, wherein the cell is a eukaryotic cell.

[00493] Aspect 80. The method of aspect 79, wherein the eukaryotic cell is a plant cell.

[00494] Aspect 81. The method of any one of aspects 78-80, wherein the non-target ssDNAs are foreign to the cell.

[00495] Aspect 82. The method of aspect 81, wherein the non-target ssDNAs are viral DNAs.

[00496] Aspect 83. The method of any one of aspects 77-82, wherein the target DNA is single stranded.

[00497] Aspect 84. The method of any one of aspects 77-82, wherein the target DNA is double stranded.

[00498] Aspect 85. The method of any one of aspects 77-84, wherein the target DNA is viral DNA.

[00499] Aspect 86. The method of any one of aspects 77-84, wherein the target DNA is papovavirus, hepdnavirus, herpesvirus, adenovirus, poxvirus, or parvovirus DNA.

EXAMPLES

[00500] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not

intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Celsius, and pressure is at or near atmospheric. Standard abbreviations may be used, e.g., bp, base pair(s); kb, kilobase(s); pl, picoliter(s); s or sec, second(s); min, minute(s); h or hr, hour(s); aa, amino acid(s); kb, kilobase(s); bp, base pair(s); nt, nucleotide(s); i.m., intramuscular(ly); i.p., intraperitoneal(ly); s.c., subcutaneous(ly); and the like.

Materials and Methods

[00501] The following materials and methods generally apply to the results presented in the Examples described herein except where noted otherwise.

Metagenomics and metatranscriptomics

[00502] The initial analysis was performed on previously assembled and binned metagenomes from two sites: the Rifle Integrated Field Research (IFRC) site, adjacent to the Colorado River near Rifle, Colorado and Crystal Geyser, a cold, CO_2-driven geyser on the Colorado Plateau in Utah. Metatranscriptomic data from IFRC site was used to detect transcription of non-coding elements in nature. Further mining of CRISPR-Cas14 systems was then performed on public metagenomes from IMG/M.

CRISPR-Cas computation analysis

[00503] The assembled contigs from the various samples were scanned with the HMMer suite for known Cas proteins using Hidden Markov Model (HMMs) profiles. Additional HMMs were constructed for Cas14 proteins based on the MAFFT alignments of putative type V effectors that contained less than 800 aa, and were adjacent to acquisition cas genes and CRISPR arrays. These HMMs were iteratively refined by augmenting them with manually selected novel putative Cas14 sequences that were found using the existing Cas14 HMM models. The sequence of Cas14 repeat sequences are provided in Table 3. CRISPR arrays were identified using a local version of the CrisprFinder software and CRISPRDetect. Phylogenetic trees of Cas1 and type V effector proteins were constructed using RAxML with PROTGAMMALG as the substitution model and 100 bootstrap samplings. Trees were visualized using FigTree 1.4.1 (http://tree.bio.ed.ac.uk/software/figtree/). Metatranscriptomic reads were mapped to assembled contigs using Bowtie2. RNase presence analysis was based on HMMs that were built from alignment of KEGG orthologous groups (KOs) downloaded from KEGG database.

[00504] Table 3. Repeat sequences (non-guide sequence portion of a Cas14 guide RNA) of all Cas14 proteins used herein (e.g., see Figure 7) are shown in Table 3.

Cas14 Protein	Repeat sequence	SEQ ID NO:	Scaffold Accession No:
Cas14a.1 (CasZa.3)	GTTGCAGAACCCGAATAGACGAATGAAGGAATGCAAC	53	**NCBI:** MK005734
Cas14a.2 (Za.8)	CTTGCAGAACCCGGATAGACGAATGAAGGAATGCAAC	295	**NCBI:** MK005733
Cas14a.3 (CasZa.3)	GTTGCAGAACCCGAATAGACGAATGAAGGAATGCAAC	53	**NCBI:** MK005732
Cas14a.4 (CasZa.4)	CTATCATATTCAGAACAAAGGGATTAAGGAATGCAAC	54	**NCBI:** MK005735
Cas14a.5 (CasZa.5, CasZb.3)	CTTTCATACTCAGAACAAAGGGATTAAGGAATGCAAC	55	**NCBI:** MK005736
Cas14a.6 (CasZa.6)	GTCTACAACTCATTGATAGAAATCAATGAGTTAGACA	56	**IMG/M:** Ga0137385 _10000156
Cas14b.1 (Cas Zb.2)	GTTGCAGAAATAGAATAAAGGAATTAAGGAATGCAAC	59	**NCBI:** MK005737
Cas14b.2 (CasZa.5, CasZb.3)	CTTTCATACTCAGAACAAAGGGATTAAGGAATGCAAC	55	**NCBI:** MK005738
Cas14b.3 (CasZb.4)	ATTTCATACTCAGAACAAAGGGATTAAGGAATGCAAC	61	**NCBI:** MK005739
Cas14b.4 (CasZb.16)	GTTTCAGCGCACGAATTAACGAGATGAGAGATGCAAC	303	**NCBI:** MK005740
Cas14b.5 (CasZb.6)	CTTGCAGAAGCTGAATAGACGAATCAAGGAATGCAAC	63	**NCBI:** MK005741
Cas14b.6 (Za.12)	CTTGCAGGCCTTGAATAGAGGAGTTAAGGAATGCAAC	296	**NCBI:** MK005742
Cas14b.7 (CasZb.9)	GTTGCAGCGCCCGAACTGACGAGACGAGAGATGCAAC	66	**IMG/M:** Ga0172369 _10000737
Cas14b.8 (Cas Zb.10)	GTTGCGCGAATAGAATAAAGGAATTAAGGAATGCAAC	67	**IMG/M:** Ga0172369 _10010464
Cas14b.9 (CasZb.11)	AGTTGCATTCCTTAATCCCTCTGTTCAGTTTGTGCAAT	68	**IMG/M:** Ga0172365 _10004421
Cas14b.10 (Zb.13)	GTTGCACAGTGCTAATTAGAGAAACTAGGAATGCAAC	297	**NCBI:** MK005743
Cas14b.11 (CasZc.2)	GTTGCGGCGCGCGAATAAACGAGACTAGGAATGCAAC	70	**NCBI:** MK005744
Cas14b.12 (Zb.14)	CTAGCATATTCAGAACAAAGGGATTAAGGAATGCAAC	298	**NCBI:** MK005745
Cas14b.13	CTTTCATATTCAGAACAAAGGGATTAAGGAATGCAAC	72	**NCBI:**

Cas14 Protein	Repeat sequence	SEQ ID NO:	Scaffold Accession No:
(CasZc.4)			MK005746
Cas14b.14 (Zb.15)	CTTTCATATTCAGAAACTAGGGGTTAAGGACTGCAAC	299	**NCBI:** MK005747
Cas14b.15 (CasZc.6)	GTTGCAGCCCCCGAACTAACGAGATGAGAGATGCAAC	74	**IMG/M:** Ga0116204 _1008574
Cas14b.16 (casZc.7)	CTTGCAGAACAATCATATATGACTAATCAGACTGCAAC	75	**IMG/M:** Ga0078972 _1001015a
Cas14c.1 (Zb.8)	GTTGCATCCCTACGTCGTGAGCACCGGTGAGTGCAAC	300	**NCBI:** MK005748
Cas14c.2 (CasZd.2)	GTCCCTACTCGCTAGGGAAACTAATTGAATGGAAAC	77	**IMG/M:** JGI12048 J13642_ 10201286
Cas14d.1 (CasZe.2)	CTTCCAAACTCGAGCCAGTGGGGAGAGAAGTGGCA	79	**NCBI:** MK005750
Cas14d.2 (CasZe.3)	CCTGTAGACCGGTCTCATTCTGAGAGGGGTATGCAACT	80	**NCBI:** MK005751
Cas14d.3 (CasZe.4)	GTCTCGAGACCCTACAGATTTTGGAGAGGGGTGGGAC	81	**NCBI:** MK005752
Cas14e.1 (CasZf.1)	GTAGCAGGACTCTCCTCGAGAGAAACAGGGGTATGCT	83	**NCBI:** MK005753
Cas14e.2 (CasZf.2)	GTACAATACCTCTCCTTTAAGAGAGGGAGGGGTACGCTAC	84	**NCBI:** MK005754
Cas14e.3 (Zc.5)	GGAAAGGAATCCCCTGAAGGAAACGAGGGGG	301	**NCBI:** MK005755
Cas14f.1 (CasZg.1)	GGTTCCCCCGGGCGCGGGTGGGGTGGCG	86	**NCBI:** MK005756
Cas14f.2 (CasZg.2)	GGCTGCTCCGGGTGCGCGTGGAGCGAGG	87	**IMG/M:** Ga0105042 _100140
Cas14g.1 (Ze.3)	GTGTCCATCAATCAGATTTGCGTTGGCCGGTGCAAT	302	**NCBI:** MK005758
Cas14g.2 (CasZi.2)	GCCGCAGCGGCCGACGCGGCCCTGATCGATGGACAC	90	**IMG/M:** Ga0123330 _1010394
Cas14h.1 (CasZk.1)	GGCTAGCCCGTGCGCGCAGGGACGAGTGG	92	**IMG/M:** Ga0070762 _10001740
Cas14h.2 (CasZk.2)	GCCCGTGCGCGCAGGGACGAGTGG	93	**IMG/M:** Ga0070766 _10011912
Cas14h.3 (CasZk.4)	CCATCGCCCCGCGCGCACGTGGATGAGCC	95	**IMG/M:** Ga0116216 _10000905
Cas14u.1 (CasZa.7)	GTTATAAAGGCGGGGATCGCGACCGAGCGATTGAAAG	57	**IMG/M:** Ga0066793

Cas14 Protein	Repeat sequence	SEQ ID NO:	Scaffold Accession No:
			_10010091
Cas14u.2 (CasZb.8)	GTCTCCATGACTGAAAAGTCGTGGCCGAATTGAAAC	65	**IMG/M:** JGI24730 J26740 _1002785
Cas14u.3 (CasZe.1)	GTTGCATTCGGGTGCAAAACAGGGAGTAGAGTGTAAC	78	**NCBI:** MK005749
Cas14u.4 (CasZj.2)	CTTTTAGACAGTTTAAATTCTAAAGGGTATAAAAC	307	**NCBI:** MK005757
Cas14u.5 (CasZj.1)	GTCGAAATGCCCGCGCGGGGGCGTCGTACCCGCGAC	308	**IMG/M:** Ga0137373 _10000316
Cas14u.6 (CasZk.3)	GTTGCAGCGGCCGACGGAGCGCGAGCGTGGATGCCAC	309	**IMG/M:** Ga0070717 _10000077
Cas14u.7 (CasZl.1)	CTTTAGACTTCTCCGGAAGTCGAATTAATGGAAAC	310	**IMG/M:** JGI12210 **IMG/M:** J13797 _10004690
Cas14u.8 (CasZl.2)	GGGCGCCCCGCGCGAGCGGGGGTTGAAG	311	**IMG/M:** Ga0073904 _10021651

Generation of expression plasmids, RNA and DNA substrates

[00505] Minimal CRISPR loci for putative systems were designed by removing acquisition proteins and generating minimal arrays with a single spacer. These minimal loci were ordered as gBlocks (IDT) and assembled into a plasmid with a tetracycline inducible promoter driving expression of the locus. Plasmid maps were available on Addgene and in the figures. All RNA was in vitro transcribed using T7 polymerase and PCR products as dsDNA template. Resulting IVTs were gel extracted and ethanol precipitated. DNA substrates were obtained from IDT and their sequences are available in Table 4. For radiolabeled cleavage assays DNA oligos were gel extracted from a PAGE gel before radiolabeling. For FQ assays, DNA substrates were used without further purification.

E. coli RNAseq

[00506] Small RNA sequencing was conducted as described previously with modification in Harrington et al. (2017). *E. coli* NEB Stable3 was transformed with a plasmid expressing Cas14a1 system with a tetracycline inducible promoter upstream of the Cas14a1 ORF or the same plasmid with an N-terminal 10x-histidine tag fused to Cas14. Starters were grown up overnight in SOB, diluted 1:100 in

5mL fresh SOB containing 214nM anhydrotetracycline and grown up overnight at 25°C. For sequencing of RNA pulled down with Cas14a, the plasmid containing an N-terminal His-tag fused to Cas14a1 was grown up at 18°C before lysis and purification as described in "Protein purification", stopping after the Ni-NTA elution. Cells were pelleted and RNA was extracted using hot phenol as previously described. Total nucleic acids were treated with TURBO DNase and phenol extracted. The resulting RNA was treated with rSAP which was heat inactivated before addition of T4 PNK. Adapters were ligated onto the small RNA using the NEBnext small RNA kit and gel-extracted on an 8% native PAGE gel. RNA was sequenced on a MiSeq with single end 300bp reads. For analysis, the resulting reads were trimmed using Cutadapt, discarding sequences <8nt and mapped to the plasmid reference using Bowtie2.

PAM depletion assays

[00507] PAM depletion assays were conducted as previously described in Burstein et al. (2017). Randomized plasmid libraries were generated using a primer containing a randomized PAM region adjacent to the target sequence. The randomized primers were hybridized with a primer that was complementary to the 3' end of the primer and the duplex was extended using Klenow Fragment (NEB). The dsDNA containing the target and were digested with EcoRI and NcoI, ligated into pUC19 backbone and transformed into *E. coli* DH5α and >107 cells were harvested. Next *E. coli* NEBstable was transformed with either a CRISPR plasmid or an empty vector control and these transformed *E. coli* were made electrocompetent by repeated washing with 10% glycerol. These electrocompetent cells were transformed with 200ng of the target library and plated on bioassay dishes containing selection for the target (carbenicillin, 100mg l-1) and CRISPR plasmid (chloramphenicol, 30mg l-1). Cells were harvested and prepared for amplicon sequencing on an Illumina MiSeq. The PAM region was extracted using Cutadapt and depletion values were calculated in python. PAMs were visualized using WebLogo.

Transcriptomic RNA mapping

[00508] RNA was extracted from 0.2 mm filters using the Invitrogen TRIzol reagent, followed by genomic DNA removal and cleaning using the Qiagen RNase-Free DNase Set kit and the Qiagen Mini RNeasy kit. An Agilent 2100 Bioanalyzer (Agilent Technologies) was used to assess the integrity of the RNA samples. The Applied Biosystems SOLiD Total RNA-Seq kit was used to generate the cDNA template library. The SOLiD EZ Bead system (Life Technologies) was used to perform emulsion clonal bead amplification to generate bead templates for SOLiD platform sequencing. Samples were sequenced at Pacific Northwest National Laboratory on the 5500XL SOLiD platform. The 50 bp single reads were trimmed using Sickle as in Brown et al. (2015).

Protein purification

[00509] Cas14a1 was purified as described previously with modification. *E. coli* BL21(DE3) RIL were transformed with 10xHis-MBP-Cas14a1 expression plasmid and grown up to OD600 =0.5 in Terrific Broth (TB) and induced with 0.5mM IPTG. Cells were grown overnight at 18°C, collected by centrifugation, resuspended in Lysis Buffer (50 mM Tris-HCl, pH 7.5, 20 mM imidazole, 0.5 mM TCEP, 500 mM NaCl) and broken by sonication. Lysate was batch loaded on to Ni-NTA resin, washed with the above buffer before elution with Elution Buffer (50 mM Tris-HCl, pH 7.5, 300 mM imidazole, 0.5 mM TCEP, 500 mM NaCl). The MBP and His-tag were removed by overnight incubation with TEV at 4°C. The resulting protein exchanged into Buffer A (20 mM HEPES, pH 7.5, 0.5 mM TCEP, 150 mM NaCl) and loaded over tandem MBP, heparin columns (GE, Hi-Trap) and eluted with a linear gradient from Buffer A to Buffer B (20 mM HEPES, pH 7.5, 0.5 mM TCEP, 1250 mM NaCl). The resulting fractions containing Cas14a1 were loaded onto an S200 gel filtration column, flash frozen and stored at -80°C until use.

***In vitro* cleavage assays**

Radiolabeled

[00510] Radiolabeled cleavage assays were conducted in 1× Cleavage Buffer (25 mM NaCl, 20 mM HEPES, pH 7.5, 1 mM DTT, 5% glycerol). 100nM Cas14a1 was complexed with 125nM crRNA and 125nM tracrRNA for 10min at RT. ~1nM radiolabeled DNA or RNA substrate was added and allowed to react for 30min at 37°C. The reaction was stopped by adding 2x Quench Buffer (90% formamide, 25mM EDTA and trace bromophenol blue), heated to 95°C for 2min and run on a 10% polyacrylamide gel containing 7M Urea and 0.5×TBE. Products were visualized by phosphorimaging.

M13 DNA cleavage

[00511] M13 DNA cleavage assays were conducted in 100 mM NaCl, 20 mM HEPES, pH 7.5, 1 mM DTT, 5% glycerol. 250nM Cas14a1 was complexed with 250nM crRNA and 250nM tracrRNA and 250nM ssDNA activator. The reaction was initiated by addition of 5 nM M13 ssDNA plasmid and was quenched by addition of loading buffer supplemented with 10mM EDTA. Products were separated on a 1.5% agarose TAE gel prestained with SYBR gold (Thermofisher).

FQ detection of trans-cleavage

[00512] FQ detection was conducted as previously described in Chen et al. (2018) with modification. 100nM Cas14a1 was complexed with 125nM crRNA, 125nM tracrRNA, 50nM FQ probe and 2nM ssDNA activator in 1× Cleavage Buffer at 37°C for 10 min. The reaction was then initiated by addition of activator DNA when for all reactions except for the RNA optimization experiments where the

variable RNA component was used to initiate. The reaction was monitored in a fluorescence plate reader for up to 120 minutes at 37°C with fluorescence measurements taken every 1 min (λex: 485 nm; λem: 535 nm). The resulting data were background subtracted using the readings taken in the absence of activator and fit using a single exponential decay curve.

Data Availability

[00513] Plasmids used herein are available on Addgene (plasmid numbers 112500, 112501, 112502 , 112503, 112504, 112505, 112506). Oligonucleotides used herein are provided in Table 4 and Table 5. The plasmids used herein are provided in Figure 24. The Cas14 protein sequences used herein are provided in Figure 7.

[00514] Table 4. Oligonucleotides and plasmids used herein.

DNA Name	Sequence	SEQ ID NO:
Radiolabeld DNA activator 1 T strand	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	220
Radiolabeld DNA activator 1 NT strand	GTGCCCGcaTTTTTtaccttacaccgcttgcgaaagttgtcagaagataatCGGCGTAg	221
Radiolabeld DNA activator 2 T strand	tttatatgtttctcctggagataacgcaatcgtgacaactttcgcaagcggtgtaaggtaGCAGGCTTCcgaattccgcgtttttacggc	222
Radiolabeld DNA activator 2 NT strand	gccgtaaaaacgcggaattcgGAAGCCTGCtaccttacaccgcttgcgaaagttgtcacgattgcgttatctccaggagaaacatataaa	223
Radiolabeld Activator 3	gatcttcagcTATACATTATTGCACCAACACTAAGGCAGAGTATGttacctggac	224
Radiolabeld Activator 4	gatcttcagcTTTGTATTACTGGAAGGATGCTTGCTTGAGGTGTAaaaacctggac	225
F-Q 5nt	/56-FAM/TTTTT/3IABkFQ/	226
F-Q 6nt	/56-FAM/TTTTTT/3IABkFQ/	227
F-Q 7nt	/56-FAM/TTTTTTT/3IABkFQ/	228
F-Q 8nt	/56-FAM/TTTTTTTT/3IABkFQ/	229
F-Q 9nt	/56-FAM/TTTTTTTTT/3IABkFQ/	230
F-Q 10nt	/56-FAM/TTTTTTTTTT/3IABkFQ/	231
F-Q 11nt	/56-FAM/TTTTTTTTTTT/3IABkFQ/	232
F-Q 12nt	/56-FAM/TTTTTTTTTTTT/3IABkFQ/	233
F-Q 10nt A/T	/56-FAM/TATATATATA/3IABkFQ/	371

F-Q 6nt A/T	/56-FAM/TATATA/3IABkFQ/	372
F-Q 5nt A/T	/56-FAM/TATAT/3IABkFQ/	373
Target 2, Perfect AAAT 3'	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGT AAACGGCCACAAGC	234
Target 2, Perfect TCGT 3'	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGT GCTCGGCCACAAGC	235
Target 2, 1-2MM	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGAGCT AAACGGCCACAAGC	236
Target 2, 3-4MM	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCCTCGT AAACGGCCACAAGC	237
Target 2, 5-6MM	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGCGGACGT AAACGGCCACAAGC	238
Target 2, 7-8MM	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGGAGCGCGACGT AAACGGCCACAAGC	239
Target 2, 9-10MM	GCCGGGGTGGTGCCCATCCTGGTCGAGCTGCTCGGCGACGT AAACGGCCACAAGC	240
Target 2, 11-12MM	GCCGGGGTGGTGCCCATCCTGGTCGAGCACGACGGCGACGT AAACGGCCACAAGC	241
Target 2, 13-14MM	GCCGGGGTGGTGCCCATCCTGGTCGACGTGGACGGCGACGT AAACGGCCACAAGC	242
Target 2, 15-16MM	GCCGGGGTGGTGCCCATCCTGGTCCTGCTGGACGGCGACGT AAACGGCCACAAGC	243
Target 2, 17-18MM	GCCGGGGTGGTGCCCATCCTGGAGGAGCTGGACGGCGACGT AAACGGCCACAAGC	244
Target 2, 19-20MM	GCCGGGGTGGTGCCCATCCTCCTCGAGCTGGACGGCGACGT AAACGGCCACAAGC	245
HERC2 Amp Fwd	G*T*G*T*TAATACAAAGGTACAGGAACAAAGAATTTG	246
HERC2 Amp Rev	CAAAGAGAAGCCTCGGCC	247
Target 1, Perfect AAAT 3'	TTTATTCAAGGCAATCACTATCAGCTGTGGAACACCCAGGT AAACTAACACAACT	248
Target 1, Perfect TCGT 3'	TTTATTCAAGGCAATCACTATCAGCTGTGGAACACCCAGGT GCTCTAACACAACT	249
Target 1, 1-2MM	TTTATTCAAGGCAATCACTATCAGCTGTGGAACACCCACCTA AACTAACACAACT	250
Target 1, 3-4MM	TTTATTCAAGGCAATCACTATCAGCTGTGGAACACCGTGGT AAACTAACACAACT	251
Target 1, 5-6MM	TTTATTCAAGGCAATCACTATCAGCTGTGGAACAGGCAGGT AAACTAACACAACT	252
Target 1, 7-8MM	TTTATTCAAGGCAATCACTATCAGCTGTGGAAGTCCCAGGT AAACTAACACAACT	253
Target 1, 9-10MM	TTTATTCAAGGCAATCACTATCAGCTGTGGTTCACCCAGGTA AACTAACACAACT	254
Target 1, 11-12MM	TTTATTCAAGGCAATCACTATCAGCTGTCCAACACCCAGGTA AACTAACACAACT	255

Target 1, 13-14MM	TTTATTCAAGGCAATCACTATCAGCTCAGGAACACCCAGGTAAACTAACACAACT	256
Target 1, 15-16MM	TTTATTCAAGGCAATCACTATCAGGAGTGGAACACCCAGGTAAACTAACACAACT	257
Target 1, 17-18MM	TTTATTCAAGGCAATCACTATCTCCTGTGGAACACCCAGGTAAACTAACACAACT	258
Target 1, 19-20MM	TTTATTCAAGGCAATCACTAAGAGCTGTGGAACACCCAGGTAAACTAACACAACT	259
Target 3, Perfect	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	220
Target 3, 1-2MM	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggCGAAAAAtgCGGGCAC	260
Target 3, 3-4MM	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaAAtaAAAAAtgCGGGCAC	261
Target 3, 5-6MM	cTACGCCGattatcttctgacaactttcgcaagcggtgtGGggtaAAAAAtgCGGGCAC	262
Target 3, 7-8MM	cTACGCCGattatcttctgacaactttcgcaagcggtACaaggtaAAAAAtgCGGGCAC	263
Target 3, 9-10MM	cTACGCCGattatcttctgacaactttcgcaagcgACgtaaggtaAAAAAtgCGGGCAC	264
Target 3, 11-12MM	cTACGCCGattatcttctgacaactttcgcaagTAgtgtaaggtaAAAAAtgCGGGCAC	265
Target 3, 13-14MM	cTACGCCGattatcttctgacaactttcgcaGAcggtgtaaggtaAAAAAtgCGGGCAC	266
Target 3, 15-16MM	cTACGCCGattatcttctgacaactttcgTGagcggtgtaaggtaAAAAAtgCGGGCAC	267
Target 3, 17-18MM	cTACGCCGattatcttctgacaactttTAcaagcggtgtaaggtaAAAAAtgCGGGCAC	268
Target 3, 19-20MM	cTACGCCGattatcttctgacaactCCcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	269
Target 3, 21-22MM	cTACGCCGattatcttctgacaaTCttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	270
Target 3, 23-24MM	cTACGCCGattatcttctgacGGctttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	271
Target 3, 25-26MM	cTACGCCGattatcttctgGTaactttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	272
Full Length Activator	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	220
-20 5′ activator target	tatcttctgacaactttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	273
-25 5′ activator target	tctgacaactttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	274
-30 5′ activator target	caactttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	275
-35 5′ activator target	ttcgcaagcggtgtaaggtaAAAAAtgCGGGCAC	276

-5 3′ activator target	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggtaAAAAAtgCG	277
-9 3′ activator target	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggtaAAAA	278
-14 3′ activator target	cTACGCCGattatcttctgacaactttcgcaagcggtgta	279
-19 3′ activator target	cTACGCCGattatcttctgacaactttcgcaagcg	280
-24 3′ activator target	cTACGCCGattatcttctgacaactttcgc	281
-29 3′ activator target	cTACGCCGattatcttctgacaact	282
-34 3′ activator target	cTACGCCGattatcttctga	283
No loop	caactttcgcaagcggtgtaaggtaAAAAAtgCG	284
0nt SL	atggaatgtggcgaacgctttcaacGAAAcaactttcgcaagcggtgtaaggtaAAAAAtgCG	285
5nt SL	atggaatgtggcgaacgcttagttgGAAAcaactttcgcaagcggtgtaaggtaAAAAAtgCG	286
10nt SL	atggaatgtggcgaagcgaaagttgGAAAcaactttcgcaagcggtgtaaggtaAAAAAtgCG	287
15nt SL	atggaatgtgcgcttgcgaaagttgGAAAcaactttcgcaagcggtgtaaggtaAAAAAtgCG	288
20nt SL	atggatacaccgcttgcgaaagttgGAAAcaactttcgcaagcggtgtaaggtaAAAAAtgCG	289
25nt SL	taccttacaccgcttgcgaaagttgGAAAcaactttcgcaagcggtgtaaggtaAAAAAtgCG	290
M13_1 Oligo	GTTTTATCTTCTGCTGGTGGTTCGTTCGGTATTTTTAATG	291
M13_2 Oligo	CATTAAAAATACCGAACGAACCACCAGCAGAAGATAAAAC	292
M13_3 Oligo	GACCATTTGCGAAATGTATCTAATGGTCAAACTAAATCTACTC	293
M13_4 Oligo	GAGTAGATTTAGTTTGACCATTAGATACATTTCGCAAATGGTC	294

[00515] Table 5. RNA Oligonucleotides.

RNA Name	**Sequence**	**SEQ ID NO:**
ssRNA target	cTACGCCGattatcttctgacaactttcgcaagcggtgtaaggtaAAAAAtgCGGGCACcc	332
crRNA	GGAATGCAACtaccttacaccgcttgcgaa	333
tracrRNA	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCTCGAAACAAATTCATTT	334

crRNA 1	GACGAATGAAGGAATGCAACtaccttacaccgcttgcgaa	335
crRNA 2	GACGAATGAAGGAATGCAACccttacaccgcttgcgaaag	336
crRNA 3	GACGAATGAAGGAATGCAACttacaccgcttgcgaaagtt	337
crRNA 4	GACGAATGAAGGAATGCAACacaccgcttgcgaaagttgt	338
crRNA MM target 2	GACGAATGAAGGAATGCAACCGTCGCCGTCCAGCTCGACCA	339
crRNA MM target 3	GACGAATGAAGGAATGCAACGATCGTTACGCTAACTATGA	340
HERC2 A sgRNA	TTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCC CTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCAT CAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAATTCATTTgaaaGAATGAAGGAATGCAACacttgacac ttaatgctcaa	341
LbCas12a HERC2 crRNA	GGGTAATTTCTACTAAGTGTAGATacttgacacttaatgctcaa	342
25nt spacer crRNA target 4	GACGAATGAAGGAATGCAACtaccttacaccgcttgcgaaagttg	343
20nt spacer crRNA target 4	GACGAATGAAGGAATGCAACtaccttacaccgcttgcgaa	335
18nt spacer crRNA target 4	GACGAATGAAGGAATGCAACtaccttacaccgcttgcg	344
16nt spacer crRNA target 4	GACGAATGAAGGAATGCAACtaccttacaccgcttg	345
14nt spacer crRNA target 4	GACGAATGAAGGAATGCAACtaccttacaccgct	346
12nt spacer crRNA target 4	GACGAATGAAGGAATGCAACtaccttacaccg	347
10nt spacer crRNA target 4	GACGAATGAAGGAATGCAACtaccttacac	348
Full repeat crRNA target 4	GTTGCAGAACCCGAATAGACGAATGAAGGAATGCAACtaccttaa caccgcttgcgaa	349
20nt repeat crRNA target 4	GACGAATGAAGGAATGCAACtaccttacaccgcttgcgaa	335
17nt repeat crRNA target 4	GAATGAAGGAATGCAACtaccttacaccgcttgcgaa	350
15nt repeat crRNA target 4	ATGAAGGAATGCAACtaccttacaccgcttgcgaa	351

10nt repeat crRNA target 4	GGAATGCAACtaccttacaccgcttgcgaa	333
tracrRNA +41nt	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTC CCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCA TCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAATTCATTTTTCCTCTCCAATTCTGCACAAAAAAA GGTGAGTCCTTAT	352
tracrRNA +3nt	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTC CCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCA TCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAATTCATTT	334
tracrRNA -26	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTC CCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCA TCAGCCTAATGTCGAGAAGTGCTTTCTT	353
tracrRNA -65	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTC CCTTAGGGGATTAGAACTTGAGTGAAGGTGG	354
tracrRNA +0	TTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTCC CTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCAT CAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAATTCA	355
tracrRNA +90nt	ttcacacTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCT GTCCCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTT GCATCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAA CCCTCGAAACAAATTCAttttttcctctccaattctgcacaaaaaaaggtgagtccttat aaaccggcgtgcagaacgccggctcacctttttttcttcattcgatttta	356
sgRNA 1	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTC CCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCA TCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAATTCATTTgaaaGAATGAAGGAATGCAACtaccttaca ccgcttgcgaa	357
sgRNA 2	CTTCACTGATAAAGTGGAGAACCGCTTCACCAAAAGCTGTC CCTTAGGGGATTAGAACTTGAGTGAAGGTGGGCTGCTTGCA TCAGCCTAATGTCGAGAAGTGCTTTCTTCGGAAAGTAACCCT CGAAACAAATTCATTTTTCCTCTCCAATTCTGCACAAgaaaGT TGCAGAACCCGAATAGACGAATGAAGGAATGCAACtaccttacac cgcttgcgaa	358
M13 target 1 crRNA	GACGAATGAAGGAATGCAACTACCGAACGAACCACCAGCA GAAGA	359
M13 target 2 crRNA	GACGAATGAAGGAATGCAACTCTTCTGCTGGTGGTTCGTTC GGTA	360
M13 target 3 crRNA	GACGAATGAAGGAATGCAACGTTTGACCATTAGATACATTT CG	361
M13 target 4 crRNA	GACGAATGAAGGAATGCAACCGAAATGTATCTAATGGTCAA AC	362

Example 1

[00516] **Figure 1** depicts examples of naturally occurring CasZ protein sequences.

[00517] **Figure 2** depicts schematic representations of CasZ loci, which include a Cas1 protein in addition to the CasZ protein.

[00518] **Figure 3** depicts a phylogenetic tree of CasZ sequences in relation to other Class 2 CRISPR/Cas effector protein sequences.

[00519] **Figure 4** depicts a phylogenetic tree of Cas1 sequences from CasZ loci in relation to Cas1 sequences from other Class 2 CRISPR/Cas loci.

[00520] **Figure 5** depicts transcriptomic RNA mapping data demonstrating expression of trancRNA from CasZ loci. The trancRNAs are adjacent to the CasZ repeat array, but do not include the repeat sequence and are not complementary to the repeat sequence. Shown are RNA mapping data for the following loci: CasZa3, CasZb4, CasZc5, CasZd1, and CasZe3. Small repeating aligned arrows represent the repeats of the CRISPR array (indicating the presence of guide RNA-encoding sequence); the peaks outside and adjacent to the repeat arrays represent highly transcribed trancRNAs.

[00521] This metatranscriptomic data was not 16S depleted, and hence large portions of the data were mapped to 16S, and mRNA, for example, was almost not represented at all in these reads. Nonetheless, RNA mapping to the predicted trancRNA regions was observed.

Example 2

[00522] A set of CRISPR-Cas systems from uncultivated archaea that contained Cas14, a family of exceptionally compact RNA-guided nucleases of just 400-700 amino acids were disclosed herein, including Cas1 and Cas2 proteins that are responsible for integrating DNA into CRISPR genomic loci and showed evidence of actively adapting their CRISPR arrays to new infections. Despite their small size, Cas14 proteins were capable of RNA-guided single-stranded DNA (ssDNA) cleavage without restrictive sequence requirements. Moreover, target recognition by Cas14 triggered non-specific cutting of ssDNA molecules. Metagenomic data showed that multiple CRISPR-Cas14 systems evolved independently and suggested a potential evolutionary origin of single-effector CRISPR-based adaptive immunity.

[00523] Competition between microbes and viruses stimulated the evolution of CRISPR-based adaptive immunity to provide protection against infectious agents. In class 2 CRISPR-Cas systems, a single 100-200 kilodalton (kD) CRISPR-associated (Cas) protein with multiple functional domains carried out RNA-guided binding and cutting of DNA or RNA substrates. To determine whether simpler, smaller RNA-guided proteins occurred in nature, terabase-scale metagenomic datasets were queried for uncharacterized genes proximal to both a CRISPR array and cas1, the gene that encoded the universal CRISPR integrase. This analysis identified a diverse family of CRISPR-Cas systems that contain cas1, cas2, cas4, and cas14, described herein, encoding a 40-70 kD polypeptide (Figure 8, Panel A). Twenty-

four (24) different cas14 gene variants have been identified that cluster into three subgroups (Cas14a–c) based on comparative sequence analysis (Figure 8, Panels A–B, Figure 9, Figure 10). Cas14 proteins were ~400–700 amino acids (aa), about half the size of previously known class 2 CRISPR RNA-guided enzymes (Figure 8, Panels C–D). While the identified Cas14 proteins exhibited considerable sequence diversity, all were united by the presence of a predicted RuvC nuclease domain, whose organization was characteristic of Type V CRISPR-Cas DNA-targeting enzymes (Figure 8, Panel D).

[00524] The identified Cas14 proteins occurred almost exclusively within DPANN, a super-phylum of symbiotic archaea characterized by small cell and genome sizes. Phylogenetic comparisons showed that Cas14 proteins were widely diverse with similarities to C2c10 and C2c9, families of bacterial RuvC-domain-containing proteins that were sometimes found near a CRISPR array but never together with other *cas* genes (Figure 8, Panel B and Figure 9). This observation and the small size of *c2c10* and *cas14* genes made it improbable that these systems could function as standalone CRISPR effectors.

[00525] Figure 8, Panels A-D depict architecture and phylogeny of CRISPR-Cas14 genomic loci. Figure 8, Panel A depicts a phylogenetic tree of Type V CRISPR systems. Newly identified miniature CRISPR systems are highlighted in orange. Figure 8, Panel B depicts representative loci architectures for C2c10 and CRISPR-Cas14 systems. Figure 8, Panel C depicts the length distribution of Cas14a-c systems compared to Cas12a-e and Cas9. Figure 8, Panel D depicts the domain organization of Cas14a compared to Cas9 and Cas12a. Protein lengths are drawn to scale.

[00526] Figure 9 depicts the maximum likelihood tree for known Type V CRISPR effectors and class 2 candidates containing a RuvC domain. Inset shows individual orthologs for each newly identified subtype. Figure 10 depicts a maximum likelihood tree for Cas1 from known CRISPR systems.

Example 3

[00527] Based on their proximity to conserved genes responsible for creating genetic memory of infection (*cas1*, *cas2*, *cas4*) (Figure 11, Panel A), it was explored whether CRISPR-Cas14 systems actively acquired DNA sequences into their CRISPR arrays. Assembled metagenomic contiguous DNA sequences (contigs) for multiple CRISPR-Cas14 loci revealed that otherwise identical CRISPR systems showed diversity in their CRISPR arrays, suggesting active adaptation to new infections (Figure 11, Panel B and Figure 12, Panel A). Without intending to be bound by any particular theory, it is proposed that the active acquisition of new DNA sequences indicated that these CRISPR-Cas14 loci encoded functional enzymes with nucleic acid targeting activity despite their small size. To test this possibility, it was investigated whether RNA components were produced from CRISPR-Cas14 loci. Environmental metatranscriptomic sequencing data were analyzed for the presence of RNA from the native archaeal host that contains CRISPR-Cas14a (Figure 12, Panel B and Figure 13, Panel A). In addition to CRISPR

RNAs (crRNAs), a highly abundant non-coding RNA was mapped to about a 130-base pair sequence located between *cas14a* and the adjacent CRISPR array. The 20 nucleotides (nts) at the 3′ end of this transcript were mostly complementary to the repeat segment of the crRNA (Figure 12, Panel C and Figure 13, Panel B), as observed for trans-activating CRISPR RNAs (tracrRNAs) found in association with Cas9, Cas12b and Cas12e CRISPR systems. In these previously studied systems, the double-stranded-RNA-cutting enzyme Ribonuclease III (RNase III) generated mature tracrRNAs and crRNAs, but no genes encoding RNase III were present in *cas14*-containing reconstructed genomes (Figure 14, Panel A). This observation implied that an alternative mechanism for CRISPR-associated RNA processing existed in these hosts.

[00528] To test whether the Cas14a proteins and associated RNA components could assemble in a heterologous organism, a plasmid was introduced into *E. coli* containing a minimal CRISPR-Cas14a locus that included the Cas14 gene, the CRISPR array and intergenic regions containing the putative tracrRNA. Affinity purification of the Cas14a protein from cell lysate and sequencing of co-purifying RNA revealed a highly abundant mature crRNA as well as the putative tracrRNA, suggesting that Cas14 associated with both crRNA and tracrRNA (Figure 14, Panel B). The calculated mass of the assembled Cas14a protein-tracrRNA-crRNA particle was 48% RNA by weight compared to just 17% for *S. pyogenes* Cas9 (SpCas9) (Figure 12, Panel D) and 8% for *F. novicida* Cas12a (FnCas12a), hinting at a central role of the RNA in the architecture of the Cas14a complex. Known class 2 CRISPR systems required a short sequence called a protospacer adjacent motif (PAM) to target double-stranded DNA (dsDNA). To test whether Cas14a required a PAM and could conduct dsDNA interference, *E. coli* was transformed expressing a minimal Cas14a locus with a dsDNA plasmid containing a randomized PAM region next to a sequence matching the target-encoding sequence (spacer) in the Cas14 array. No depletion of a PAM sequence was detected among *E. coli* transformants, suggesting that the CRISPR-Cas14a system was either unable to target dsDNA, could do so without requiring a PAM, or was inactive in this heterologous host (Figure 15, Panels A-D).

[00529] Figure 11, Panels A-B depict acquisition of new spacers by CRISPR-Cas14 systems. Figure 11, Panel A depicts alignment of Cas14 Cas1 orthologs. Expansion shows conservation of previously implicated active site residues highlighted in red boxes. Figure 11, Panel B depicts multiple CRISPR arrays assembled for various CRISPR-Cas14 systems revealing spacer diversity for these CRISPR systems. Orange arrows indicate repeats while variously colored boxes indicate unique spacers.

[00530] Figure 12, Panels A-D depict that CRISPR-Cas14a actively adapts and encodes a tracrRNA. Figure 12, Panel A depicts pacer diversity for Cas14a and Cas14b with CRISPR repeats diagramed in orange and unique spacers shown in different colors. Figure 12, Panel B depicts metatranscriptomics reads mapped to Cas14a1 and Cas14a3. Inset shows expansion of most abundant repeat and spacer sequence. Figure 12, Panel C depicts *in silico* predicted structure of Cas14a1 crRNA

and tracrRNA. RNase III orthologs were not identified in host genomes (Figure 14, Panel A). Figure 12, Panel D depicts fraction of various CRISPR complexes mass made up of by RNA and protein.

[00531] Figure 13, Panels A-B depict metatranscriptomics for CRISPR-Cas14 loci. Figure 13, Panel A depicts environmental RNA sequencing reads for Cas14a orthologs. Location of Cas14 and the CRISPR array indicated below. RNA structures to the right show the *in silico* predicted structure of the tracrRNA identified from metatranscriptomics. Figure 13, Panel B depicts predicted hybridization for Cas14a1 crRNA:tracrRNA duplex.

[00532] Figure 14, Panels A-B depict RNA processing and heterologous expression by CRISPR-Cas14. Figure 14, Panel A depicts the presence of common RNase orthologs in Cas14 containing genomes. Light purple represents hits that were significantly shorter than the expected length for the given RNase. Note that RNase III is absent in all investigated genomes. Figure 14, Panel B depicts small RNAseq reads from heterologous expression of Cas14a1 locus in *E. coli* (Figure 14, Panel B, bottom two graphs) compared to metatranscriptomic reads (Figure 14, Panel B, top graph). Pull down refers to RNA that copurified with Ni-NTA affinity purified Cas14a1.

[00533] Figure 15, Panels A-D depict plasmid depletion by Cas14a1 and SpCas9. Figure 15, Panel A depicts a diagram outlining a PAM discovery experiment. *E. coli* expressing the CRISPR system of interest was challenged with a plasmid containing a randomized PAM sequence flanking the target. The surviving (transformed) cells were harvested and sequenced along with a control harboring an empty vector. The depleted sequences were then sequenced and PAMs depleted more than the PAM Depletion Value Threshold (PDVT) were used to generate a Weblogo. Figure 15, Panels B-D depict PAM sequences depleted by heterologously expressed Cas14a1 transformed with a target plasmid containing a randomized PAM sequence 5' (Figure 15, Panel B) or 3' (Figure 15, Panel C) of the target. "No sequences" indicated that no sequences were found to be depleted at or above the given PDVT.

Example 4

[00534] It was tested whether purified Cas14a-tracrRNA-crRNA complexes were capable of RNA-guided nucleic acid cleavage *in vitro*. All currently reconstituted DNA-targeting class 2 interference complexes were able to recognize both dsDNA and ssDNA substrates. Purified Cas14a-tracrRNA-crRNA complexes were incubated with radiolabeled target oligonucleotides (ssDNA, dsDNA, and ssRNA) bearing 20-nucleotide sequence complementary to the crRNA guide sequence, or a non-complementary ssDNA, and these substrates were analyzed for Cas14a-mediated cleavage. Only in the presence of a complementary ssDNA substrate was any cleavage product detected (Figure 16, Panel A and Figure 17, Panels A-C), and cleavage was dependent on the presence of both tracrRNA and crRNA, which could also be combined into a single-guide RNA (sgRNA) (Figure 16, Panel B and Figure 18). The lack of detectable dsDNA cleavage suggested that Cas14a targeted ssDNA selectively, although it

was possible that some other factor or sequence requirement could enable dsDNA recognition in the native host. Mutation of the conserved active site residues in the Cas14a RuvC domain eliminated cleavage activity (Figure 17, Panel A), implicating RuvC as the domain responsible for DNA cutting. Moreover, Cas14a DNA cleavage was sensitive to truncation of the RNA components to lengths shorter than the naturally produced sequences (Figure 16, Panel B and Figure 19, Panels A-D). These results established Cas14a as the smallest class 2 CRISPR effector demonstrated to conduct programmable RNA-guided DNA cleavage.

[00535] It was tested whether Cas14a required a PAM for ssDNA cleavage *in vitro* by tiling Cas14a guides across a ssDNA substrate (Figure 16, Panel C). Despite sequence variation adjacent to the targets of these different guides, cleavage was observed for all four sequences. The cleavage sites occurred beyond the guide-complementary region of the ssDNA and shifted in response to guide binding position (Figure 16, Panel C). These data demonstrated Cas14a was an ssDNA-targeting CRISPR endonuclease that did not require a PAM for activation.

[00536] Based on the observation that Cas14a cut outside of the crRNA/DNA targeting heteroduplex, it was proposed that Cas14a may possess target-activated non-specific ssDNA cleavage activity, similar to the RuvC-containing enzyme Cas12a. To test this possibility, Cas14a-tracrRNA-crRNA was incubated with a complementary activator DNA and an aliquot of M13 bacteriophage ssDNA bearing no sequence complementarity to the Cas14a crRNA or activator (Figure 16, Panel D). The M13 ssDNA was rapidly degraded to small fragments, an activity that was eliminated by mutation of the conserved Cas14a RuvC active site, suggesting that activation of Cas14a resulted in non-specific ssDNA degradation.

[00537] To investigate the specificity of target-dependent non-specific DNA cutting activity by Cas14a, a fluorophore-quencher (FQ) assay was adapted in which cleavage of dye-labeled ssDNA generates a fluorescent signal (Figure 20, Panel A). When Cas14a was incubated with various guide RNA-target ssDNA pairs, a fluorescent signal was observed only in the presence of the cognate target and showed strong preference for longer FQ-containing substrates (Figure 19, Panel F and Figure 20, Panel A). Cas14a mismatch tolerance was tested by tiling 2-nt mismatches across the targeted region in various ssDNA substrates. Mismatches near the middle of the ssDNA target strongly inhibited Cas14a activity, revealing an internal seed sequence that was distinct from the PAM-proximal seed region observed for dsDNA-targeting CRISPR-Cas systems (Figure 20, Panel B and Figure 21, Panels A-D). Moreover, DNA substrates containing strong secondary structure resulted in reduced activation of Cas14a (Figure 21, Panel E). Truncation of ssDNA substrates also resulted in reduced or undetectable *trans* cleavage (Figure 21, Panel F). These results suggested a mechanism of fidelity distinct from dsDNA-targeting class 2 CRISPR systems, possibly utilizing a preordered region of the crRNA to gate cleavage activity similarly to the RNA-targeting Cas13a enzymes.

[00538] Further investigation of compact Type V systems in metagenomic data revealed a large diversity of systems that, like Cas14a–c, include a gene encoding a short RuvC-containing protein adjacent to acquisition-associated *cas* genes and a CRISPR array. Twenty (20) additional such systems were found that cluster into five main families (Cas14d–h). These families seemed to have evolved from independent domestication events of TnpB, the transposase-associated protein implicated as the evolutionary parent of type V CRISPR effectors. Excluding *cas*14g, which was related to *cas*12b, the *cas14*-like genes formed separate clades on the type V effector phylogeny (Figure 22, Panels A-B), and their *cas1* genes had different origins (Figure 10, Panel A). Altogether 38 CRISPR-Cas14 systems belonging to eight families (Cas14a–h) were identified and eight additional systems that could not be clustered with the analysis (termed Cas14u, Table 3).

[00539] The small size of the Cas14 proteins described herein and their resemblance to type V effector proteins suggested that RNA-guided ssDNA cleavage may have existed as an ancestral class 2 CRISPR system. In this scenario, a small, domesticated TnpB-like ssDNA interference complex may have gained additional domains over time, gradually improving dsDNA recognition and cleavage. Smaller Cas9 orthologs exhibited weaker dsDNA-targeting activity than their larger counterparts but retained the ability to robustly cleave ssDNA. Aside from the evolutionary implications, the ability of Cas14 to specifically target ssDNA suggested a role in defense against ssDNA viruses or mobile genetic elements (MGEs) that propagated through ssDNA intermediates. Without intending to be bound by any particular theory, an ssDNA-targeting CRISPR system may be particularly advantageous in certain marine environments where ssDNA viruses comprised the vast majority of viral abundance.

[00540] **Figure 16**, Panels A-D depict CRISPR-Cas14a as an RNA-guided DNA-endonuclease. Figure 16, Panel A depicts cleavage kinetics of Cas14a1 targeting ssDNA, dsDNA, ssRNA and off-target ssDNA. Figure 16, Panel B depicts a diagram of Cas14a RNP bound to target ssDNA and Cas14a1 cleavage kinetics of radiolabeled ssDNA in the presence of various RNA components. Figure 16, Panel C depicts tiling of a ssDNA substrate by Cas14a1 guide sequences. Figure 16, Panel D depicts cleavage of the ssDNA viral M13 genome with activated Cas14a1.

[00541] **Figure 17**, Panels A-E depict degradation of ssDNA by Cas14a1. Figure 17, Panel A depicts SDS-PAGE of purified Cas14a1 and Cas14a1 point mutants. Figure 17, Panel B depicts optimization of salt, cation and temperature for Cas14a1 cleavage of ssDNA targets. Figure 17, Panel C depicts radiolabled cleavage of ssDNA by Cas14a1 with spacer sequences of various lengths. Figure 17, Panel D depicts alignment of Cas14 with previously studied Cas12 proteins to identify RuvC active site residues and Figure 17, Panel E depicts cleavage of ssDNA by purified Cas14a1 RuvC point mutants.

[00542] **Figure 18** depicts the kinetics of Cas14a1 cleavage of ssDNA with various guide RNA components.

[00543] **Figure 19**, Panels A-F depict optimization of Cas14a1 guide RNA components. Figure 19, Panel A depicts a diagram of Cas14a1 targeting ssDNA. Impact on Cas14a1 cleavage of an FQ ssDNA substrate by varying the spacer length (Figure 19, Panel B), repeat length (Figure 19, Panel C), tracrRNA (Figure 19, Panel D), and fusing the crRNA and tracrRNA together (Figure 19, Panel E). Figure 19, Panel F depicts a heat map showing the background subtracted fluorescence resulting from cleavage of an ssDNA FQ reporter in the presence of various guide and target combinations.

[00544] **Figure 20**, Panels A-E depict high fidelity ssDNA SNP detection by CRISPR-Cas14a. Figure 20, Panel A depicts a fluorescent-quencher (FQ) assay for detection of ssDNA by Cas14a1 and the cleavage kinetics for various length FQ substrates. Figure 20, Panel B depicts cleavage kinetics for Cas14a1 with mismatches tiled across the substrate (individual points represent replicate measurements). Figure 20, panel C depicts a diagram of Cas14-DETECTR strategy and HERC2 eye color SNP. Figure 20, panel D depicts titration of T7 exonuclease and impact on Cas14a-DETECTR. Figure 20, panel E, depicts SNP detection using Cas14a-DETECTR with a blue-eye targeting guide for a blue-eyed and brown-eyed saliva sample compared to ssDNA detection using Cas12a.

[00545] **Figure 21**, Panels A-F depict the impact of various activators on Cas14a1 cleavage rate Figure 21, Panel A depicts a diagram of Cas14a1 targeting of ssDNA with position of mismatches used in panels A-D and raw rates for representative replicates of mismatch (MM) position for Target 1. Cleavage rates for Cas14a targeting substrates with mutations tiled across three different substrates (Figure 21, Panels B-D). Figure 21, Panel E depicts trans cleavage rates for substrates with increasing amounts of secondary structure. Figure 21, Panel F depicts trans leavage rates with truncated substrates. Points represent individual measurements.

[00546] **Figure 22**, Panels A-B depict diversity of CRISPR-Cas14 systems. Figure 22, Panel A depicts representative locus architecture for indicated Cas14 systems. Protein lengths are drawn to scale. Figure 22, Panel B depicts a maximum likelihood tree for Type V effectors including all eight identified subtypes of Cas14.

[00547] **Figure 23,** Panels A-C depict a test of Cas14a1 mediated interference in a heterologous host. Diagram of Cas14a1 and LbCas12a constructs to test interference in *E. coli*. (B) Plaques of ΦX174 spotted on *E. coli* revealing Cas12a- but not Cas14a1-mediated interference. Each spot represents a 10-fold dilution of the ΦX174 stock. (C) Growth curves of *E. coli* expressing Cas14a1 or LbCas12a infected with ΦX174 (T, targeting; NT, non-targeting). Figure 19, Panel F shows a heat map showing the background-subtracted fluorescence resulting from cleavage of a ssDNA FQ reporter in the presence of various guide and target combinations after a 30-minute incubation.

[00548] While the present invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications may be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present invention. All such modifications are intended to be within the scope of the claims appended hereto.

REFERENCES

1. R. Barrangou *et al.*, CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* **315**, 1709–12 (2007).
2. S. A. Jackson *et al.*, CRISPR-Cas: Adapting to change. *Science.* **356** (6333), pp. 1–9 (2017).
3. S. Shmakov *et al.*, Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
4. J. S. Chen, J. A. Doudna, The chemistry of Cas9 and its CRISPR colleagues. *Nat. Rev. Chem.* **1**, 0078 (2017).
5. C. T. Brown *et al.*, Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature.* **523**, 208–211 (2015).
6. K. Anantharaman *et al.*, Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
7. V. M. Markowitz *et al.*, IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* **42**, 568–573 (2014).
8. V. M. Markowitz *et al.*, IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, 115–122 (2012).
9. A. J. Probst *et al.*, Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO2concentrations. *Environ. Microbiol.* **19**, 459–474 (2017).
10. I. Yosef, M. G. Goren, U. Qimron, Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res.* **40**, 5569–5576 (2012).
11. J. K. Nuñez, A. S. Y. Lee, A. Engelman, J. a. Doudna, Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. *Nature.* **519**, 193–198 (2015).
12. S. Shmakov *et al.*, Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol. Cell.* **60**, 385–397 (2015).
13. D. Burstein *et al.*, New CRISPR-Cas systems from uncultivated microbes. *Nature.* **542**, 237–241 (2017).
14. C. Rinke *et al.*, Insights into the phylogeny and coding potential of microbial dark matter. *Nature.* **499**, 431–437 (2013).
15. C. J. Castelle *et al.*, Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690–701 (2015).
16. E. Deltcheva *et al.*, CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature.* **471**, 602–607 (2011).
17. K. E. Savell, J. J. Day, Applications of CRISPR/CAS9 in the mammalian central nervous system. *Yale J. Biol. Med.* **90** (2017), pp. 567–581.
18. F. J. M. Mojica, C. Díez-Villaseñor, J. García-Martínez, C. Almendros, Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology.* **155**, 733–740 (2009).
19. Y. Zhang, R. Rajan, H. S. Seifert, A. Mondragón, E. J. Sontheimer, DNase H Activity of Neisseria meningitidis Cas9. *Mol. Cell.* **60**, 242–255 (2015).
20. E. Ma, L. B. Harrington, M. R. O'Connell, K. Zhou, J. A. Doudna, Single-Stranded DNA

Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol. Cell.* **60**, 398–407 (2015).
21. J. S. Chen *et al.*, CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*. **360**, 436–439 (2018).
22. B. Zetsche *et al.*, Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*. **163**, 759–771 (2015).
23. A. East-Seletsky *et al.*, Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature*. **538**, 270–273 (2016).
24. L. Liu *et al.*, The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell*. **170**, 714–726.e10 (2017).
25. O. O. Abudayyeh *et al.*, C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science (80-.).* **353**, 1–9 (2016).
26. G. J. Knott *et al.*, Guide-bound structures of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme. *Nat. Struct. Mol. Biol.* **24**, 825–833 (2017).
27. S. Y. Li *et al.*, CRISPR-Cas12a has both cis- and trans-cleavage activities on single-stranded DNA. *Cell Res.* **28**, 491–493 (2018).
28. H. Eiberg *et al.*, Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum. Genet.* **123**, 177–187 (2008).
29. S. Shmakov *et al.*, Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).
30. E. V Koonin, K. S. Makarova, F. Zhang, Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.* **37**, 67–78 (2017).
31. K. S. Makarova *et al.*, An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol*, 1–15 (2015).
32. O. Barabas *et al.*, Mechanism of IS 200 / IS 605 Family DNA Transposases : Activation and Transposon- Directed Target Site Selection, 208–220 (2008).
33. M. Yoshida *et al.*, Quantitative viral community DNA analysis reveals the dominance of single-stranded DNA viruses in offshore upper bathyal sediment from Tohoku, Japan. *Front. Microbiol.* **9**, 1–10 (2018).
34. C. T. Brown *et al.*, Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature*. **523**, 208–211 (2015).
35. K. Anantharaman *et al.*, Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **7**, 13219 (2016).
36. A. J. Probst *et al.*, Genomic resolution of a cold subsurface aquifer community provides metabolic insights for novel microbes adapted to high CO2concentrations. *Environ. Microbiol.* **19**, 459–474 (2017).
37. V. M. Markowitz *et al.*, IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* **42**, 568–573 (2014).
38. V. M. Markowitz *et al.*, IMG: The integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, 115–122 (2012).
39. R. D. Finn, J. Clements, S. R. Eddy, HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
40. D. Burstein *et al.*, New CRISPR-Cas systems from uncultivated microbes. *Nature*. **542**, 237–241 (2017).
41. I. Grissa, G. Vergnaud, C. Pourcel, CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* **35**, W52–W57 (2007).
42. A. Biswas, R. H. J. Staals, S. E. Morales, P. C. Fineran, C. M. Brown, CRISPRDetect: A flexible algorithm to define CRISPR arrays. *BMC Genomics*. **17**, 1–14 (2016).
43. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. **30**, 1312–1313 (2014).
44. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods*. **9**, 357–359 (2012).

45. H. Ogata *et al.*, KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**, 29–34 (1999).
46. L. B. Harrington *et al.*, A thermostable Cas9 with increased lifetime in human plasma. *Nat. Commun.* **8**, 1–7 (2017).
47. G. Crooks, G. Hon, J. Chandonia, S. Brenner, NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
48. L. B. Harrington *et al.*, A Broad-Spectrum Inhibitor of CRISPR-Cas9. *Cell.* **170**, 1224–1233.e15 (2017).
49. J. S. Chen *et al.*, CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science.* **360**, 436–439 (2018).
50. Brown *et al.*, Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature.* **523**, 208–211 (2015), doi:10.1038/nature14486.

CLAIMS

1. A method of guiding a class 2 CRISPR/Cas endonuclease to a target sequence of a target nucleic acid, the method comprising contacting the target nucleic acid with a non-naturally occurring complex comprising:

(a) the class 2 CRISPR/Cas endonuclease, wherein the class 2 CRISPR/Cas endonuclease has a length of 350-900 amino acids and includes 3 partial RuvC subdomains that are not contiguous with respect to the primary amino acid sequence of the protein, but form a RuvC domain when the protein is produced and folds; and

(b) a guide RNA that comprises a guide sequence that hybridizes to a target sequence of the target nucleic acid and comprises a region that binds to the class 2 CRISPR/Cas endonuclease.

2. The method of claim 1, wherein the method results in modification of the target nucleic acid, modulation of transcription from the target nucleic acid, or modification of a polypeptide associated with a target nucleic acid.

3. The method of claim 2, wherein the target nucleic acid is modified by being cleaved.

4. The method of any one of claims 1-3, wherein the target nucleic acid is selected from: double stranded DNA, single stranded DNA, RNA, genomic DNA, and extrachromosomal DNA.

5. The method of any one of claims 1-4, wherein the guide sequence and the region that binds to the class 2 CRISPR/Cas endonuclease are heterologous to one another.

6. The method of any one of claims 1-5, wherein said contacting results in genome editing.

7. The method of any one of claims 1-5, wherein said contacting takes place:

a) outside of a bacterial cell and outside of an archaeal cell; or

b) *in vitro* outside of a cell; or

c) inside of a target cell.

8. The method of claim 7, wherein said contacting comprises: introducing into the target cell at least one of:

23 02 23

(a) the class 2 CRISPR/Cas endonuclease, or a nucleic acid encoding the class 2 CRISPR/Cas endonuclease; and

(b) the guide RNA, or a nucleic acid encoding the guide RNA.

9. The method of claim 8, wherein the nucleic acid encoding the class 2 CRISPR/Cas endonuclease is a non-naturally occurring sequence that is codon optimized for expression in the target cell.

10. The method of any one of claims 7-9, wherein the target cell is a eukaryotic cell.

11. The method of any one of claims 7-10, wherein the target cell is:
a) in culture *in vitro*; or
b) *in vivo*; or
c) *ex vivo*.

23 02 23

12. The method of claim 10, wherein the eukaryotic cell is selected from the group consisting of: a plant cell, a fungal cell, a single cell eukaryotic organism, a mammalian cell, a reptile cell, an insect cell, an avian cell, a fish cell, a parasite cell, an arthropod cell, a cell of an invertebrate, a cell of a vertebrate, a rodent cell, a mouse cell, a rat cell, a primate cell, a non-human primate cell, and a human cell.

13. The method of any one of claims 7-12, wherein said contacting further comprises: introducing a DNA donor template into the target cell.

14. The method of any one of claims 1-13, wherein the method comprises contacting the target nucleic acid with a transactivating noncoding RNA (trancRNA).

15. A composition comprising a non-naturally occurring complex comprising:

(a) a class 2 CRISPR/Cas endonuclease, or a nucleic acid encoding said class 2 CRISPR/Cas endonuclease; and

(b) a guide RNA, or a nucleic acid encoding said guide RNA, wherein said guide RNA comprises a guide sequence that is complementary to a target sequence of a target nucleic acid, and comprises a region that can bind to the class 2 CRISPR/Cas endonuclease;

wherein the class 2 CRISPR/Cas endonuclease:

i) has a length of 350-900 amino acids;

ii) includes 3 partial RuvC subdomains that are not contiguous with respect to the primary amino acid sequence of the protein, but form a RuvC domain when the protein is produced and folds; and

iii) interacts with the guide RNA to form a ribonucleoprotein complex that is targeted to the target sequence via base pairing between the guide RNA and the target sequence.

16. The composition of claim 15, further comprising a transactivating noncoding RNA (trancRNA), or a nucleic acid encoding said trancRNA.

17. A kit comprising a non-naturally occurring complex comprising:

(a) a class 2 CRISPR/Cas endonuclease, or a nucleic acid encoding said class 2 CRISPR/Cas endonuclease; and

(b) a guide RNA, or a nucleic acid encoding said guide RNA, wherein said guide RNA comprises a guide sequence that is complementary to a target sequence of a target nucleic acid, and comprises a region that can bind to the class 2 CRISPR/Cas endonuclease;

wherein the class 2 CRISPR/Cas endonuclease:

i) has a length of 350-900 amino acids;

ii) includes 3 partial RuvC subdomains that are not contiguous with respect to the primary amino acid sequence of the protein, but form a RuvC domain when the protein is produced and folds; and

iii) interacts with the guide RNA to form a ribonucleoprotein complex that is targeted to the target sequence via base pairing between the guide RNA and the target sequence.

18. The kit of claim 17, further comprising a transactivating noncoding RNA (trancRNA), or a nucleic acid encoding said trancRNA.

19. A genetically modified eukaryotic cell, comprising:

(a) a class 2 CRISPR/Cas endonuclease, or a nucleic acid encoding said class 2 CRISPR/Cas endonuclease; and

(b) a guide RNA, or a nucleic acid encoding said guide RNA, wherein said guide RNA comprises a guide sequence that is complementary to a target sequence of a target nucleic acid, and comprises a region that can bind to the class 2 CRISPR/Cas endonuclease;

wherein the class 2 CRISPR/Cas endonuclease:

23 02 23

i) has a length of 350-900 amino acids;

ii) includes 3 partial RuvC subdomains that are not contiguous with respect to the primary amino acid sequence of the protein, but form a RuvC domain when the protein is produced and folds; and

iii) interacts with the guide RNA to form a ribonucleoprotein complex that is targeted to the target sequence via base pairing between the guide RNA and the target sequence.

20. The eukaryotic cell of claim 19, further comprising a transactivating noncoding RNA (trancRNA), or a nucleic acid encoding said trancRNA.

21. The composition, kit, or eukaryotic cell of any one of claims 15-20, for use in a method of therapeutic treatment of a patient.

22. The method, composition, kit, or eukaryotic cell of any one of the preceding claims, wherein the guide RNA and/or the trancRNA comprises one or more of: a modified nucleobase, a modified backbone or non-natural internucleoside linkage, a modified sugar moiety, a Locked Nucleic Acid, a Peptide Nucleic Acid, and a deoxyribonucleotide.

23. The method, composition, kit, or eukaryotic cell of any one of the preceding claims, wherein the class 2 CRISPR/Cas endonuclease is a variant class 2 CRISPR/Cas endonuclease with reduced nuclease activity compared to a corresponding wild type class 2 CRISPR/Cas endonuclease.

24. The method, composition, kit, or eukaryotic cell of any one of the preceding claims, wherein at least one of: the class 2 CRISPR/Cas endonuclease, the nucleic acid encoding the class 2 CRISPR/Cas endonuclease, the guide RNA, the nucleic acid encoding the guide RNA, the trancRNA, and the nucleic acid encoding the trancRNA; is conjugated to a heterologous moiety.

25. A method of detecting a target DNA in a sample, the method comprising:

(a) contacting the sample with:

(i) a class 2 CRISPR/Cas endonuclease, wherein the class 2 CRISPR/Cas endonuclease has a length of 350-900 amino acids and includes 3 partial RuvC subdomains that are not contiguous with respect to the primary amino acid sequence of the protein, but form a RuvC domain when the protein is produced and folds;

23 02 23

(ii) a guide RNA comprising: a region that binds to the class 2 CRISPR/Cas endonuclease, and a guide sequence that hybridizes with the target DNA; and

(iii) a detector DNA that is single stranded and does not hybridize with the guide sequence of the guide RNA; and

(b) measuring a detectable signal produced by cleavage of the single stranded detector DNA by the class 2 CRISPR/Cas endonuclease, thereby detecting the target DNA.

26. The method of claim 25, wherein the target DNA is single stranded or double stranded.

27. The method according to claim 25 or 26, wherein measuring a detectable signal comprises one or more of: gold nanoparticle based detection, fluorescence polarization, colloid phase transition/dispersion, electrochemical detection, and semiconductor-based sensing.

28. The method according to any one of claims 25-27, wherein the single stranded detector DNA comprises a fluorescence-emitting dye pair.

29. The method according to claim 28, wherein the fluorescence-emitting dye pair is a quencher/fluor pair.

30. The method according to any one of claims 25-29, wherein the method comprises amplifying nucleic acids in the sample.

31. A kit for detecting a target DNA in a sample, the kit comprising:

(a) a guide RNA, or a nucleic acid encoding the guide RNA; wherein the guide RNA comprises: a region that binds to a class 2 CRISPR/Cas endonuclease, and a guide sequence that is complementary to a target DNA;

(b) a labeled detector DNA that is single stranded and does not hybridize with the guide sequence of the guide RNA; And

(c) the class 2 CRISPR/Cas endonuclease, wherein the class 2 CRISPR/Cas endonuclease:

i) has a length of 350-900 amino acids;

ii) includes 3 partial RuvC subdomains that are not contiguous with respect to the primary amino acid sequence of the protein, but form a RuvC domain when the protein is produced and folds; and

23 02 23

iii) interacts with the guide RNA to form a ribonucleoprotein complex that is targeted to the target sequence via base pairing between the guide RNA and the target sequence.

32. The kit of claim 31, wherein the class 2 CRISPR/Cas endonuclease comprises an amino acid sequence having at least 85% amino acid sequence identity to the class 2 CRISPR/Cas endonuclease amino acid sequence set forth in any one of FIGs. 1 and 7.

33. The kit of claim 31 or 32, wherein the single stranded detector DNA comprises a fluorescence-emitting dye pair.

34. The kit of claim 33, wherein the fluorescence-emitting dye pair is a quencher/fluor pair.

35. The kit of any one of claims 31-34, further comprising nucleic acid amplification components.

23 02 23

36. A method of cleaving single stranded DNAs (ssDNAs), the method comprising:

contacting a population of nucleic acids, wherein said population comprises a target DNA and a plurality of non-target ssDNAs, with:

(i) a class 2 CRISPR/Cas endonuclease, wherein the class 2 CRISPR/Cas endonuclease has a length of 350-900 amino acids and includes 3 partial RuvC subdomains that are not contiguous with respect to the primary amino acid sequence of the protein, but form a RuvC domain when the protein is produced and folds; and

(ii) a guide RNA comprising: a region that binds to the class 2 CRISPR/Cas endonuclease, and a guide sequence that hybridizes with the target DNA,

wherein the class 2 CRISPR/Cas endonuclease cleaves non-target ssDNAs of said plurality.

37. The method of claim 36, wherein said contacting is inside of a cell *in vitro*, *ex vivo*, or *in vivo*.

38. The method of claim 36 or claim 37, wherein the target DNA is single stranded or double stranded.