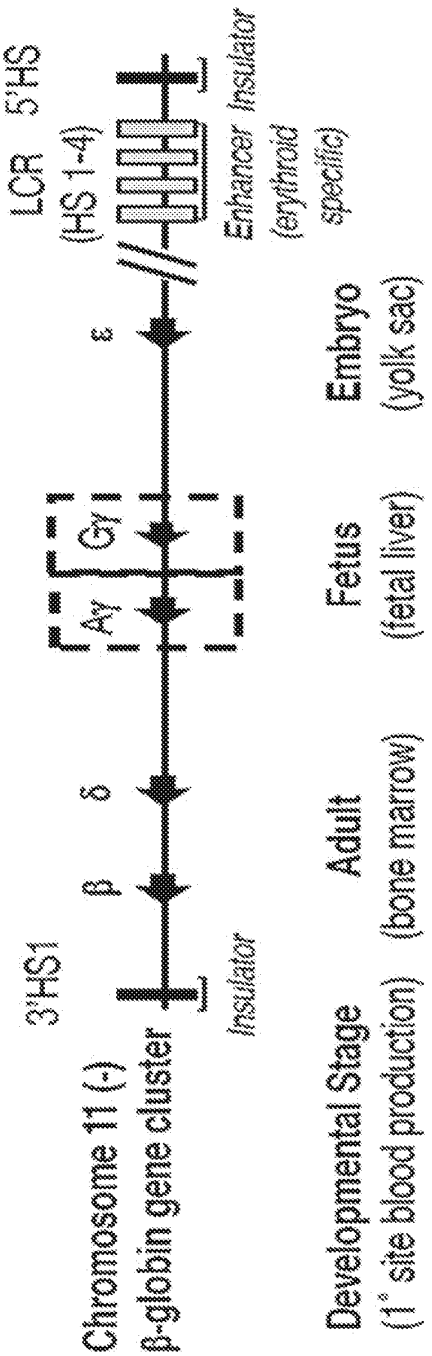

LCR 5'HS
(HS 1-4)
Enhancer Insulator
(erythroid
specific)
ε
Gγ
Aγ
δ
β
3'HS1
Chromosome 11 (-)
β-globin gene cluster
Insulator
Developmental Stage
(1° site blood production)
Adult
(bone marrow)
Fetus
(fetal liver)
Embryo
(yolk sac)

FIG 1

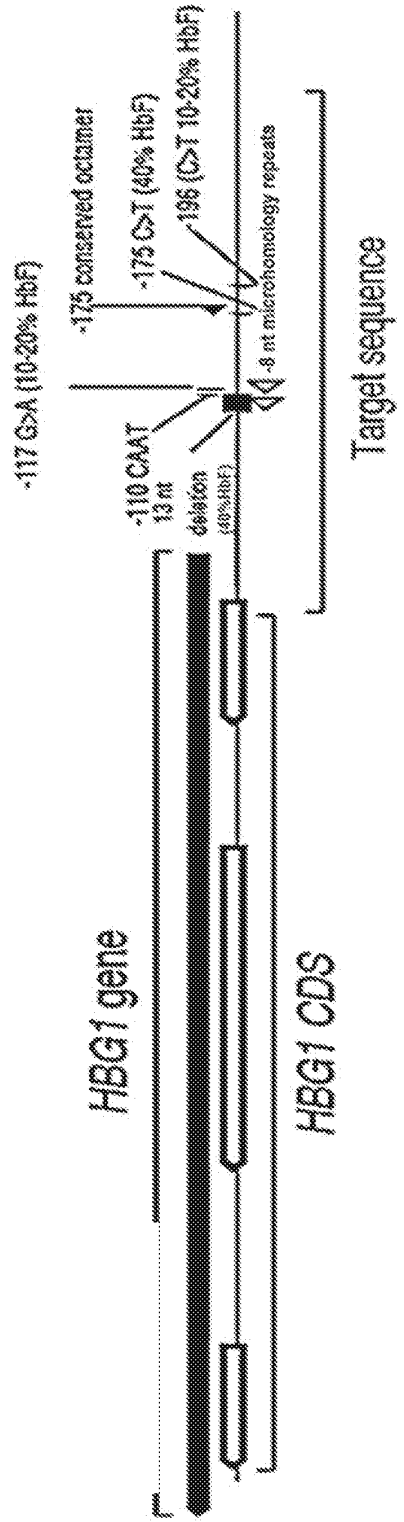

-117 G>A (10-20% HbF)
-175 conserved octamer
-175 C>T (40% HbF)
-196 (C>T 10-20% HbF)
-110 CAAT
13 nt
deletion
(40% HbF)
-8 nt microhomology repeats
Target sequence
HBG1 gene
HBG1 CDS

FIG 2A


-114 C>T(11-14% HbF)
-158 C>T 40% HbF)
-175 T>C(2-30% HbF)
-202 (C>G 15-25% HbF)
-110 CAAT
13 nt
deletion
(40% HbF)
-8 nt microhomology repeats
Target sequence
HBG2 gene
HBG2 CDS

FIG 2B

FIG 3A

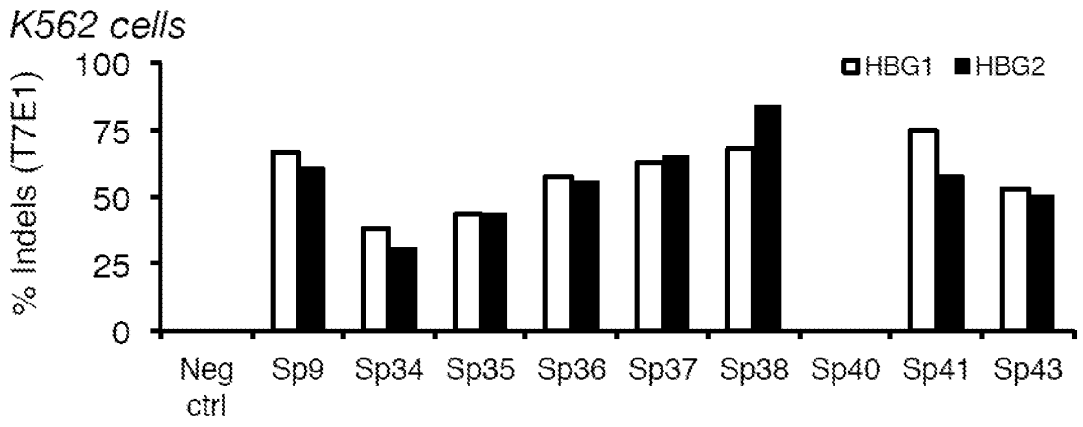

K562 cells
% Indels (T7E1)
100
75
50
25
0
HBG1
HBG2
Neg ctrl
Sp9
Sp34
Sp35
Sp36
Sp37
Sp38
Sp40
Sp41
Sp43

FIG 3B

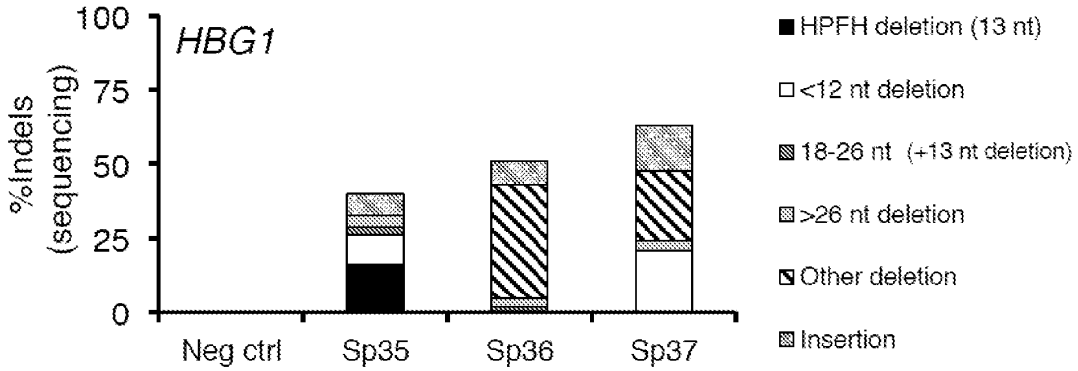

HBG1
%Indels (sequencing)
100
75
50
25
0
Neg ctrl
Sp35
Sp36
Sp37
HPFH deletion (13 nt)
<12 nt deletion
18-26 nt (+13 nt deletion)
>26 nt deletion
Other deletion
Insertion

FIG 3C

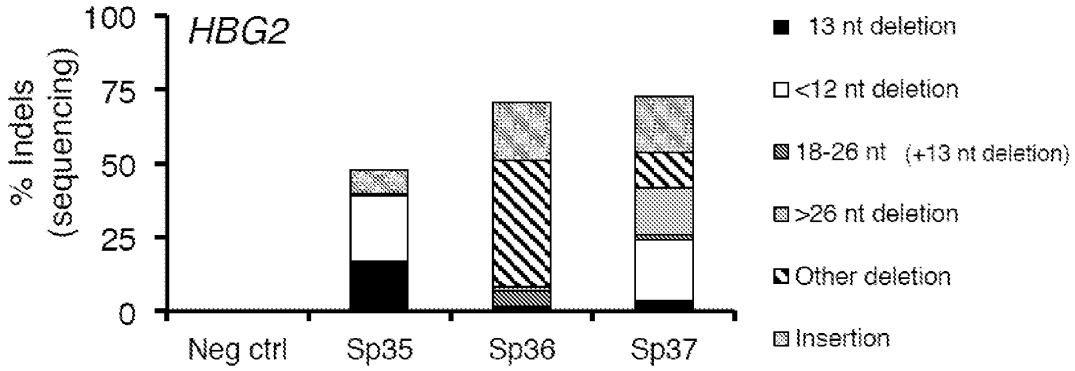

HBG2
% Indels (sequencing)
100
75
50
25
0
Neg ctrl
Sp35
Sp36
Sp37
13 nt deletion
<12 nt deletion
18-26 nt (+13 nt deletion)
>26 nt deletion
Other deletion
Insertion

FIG 4A

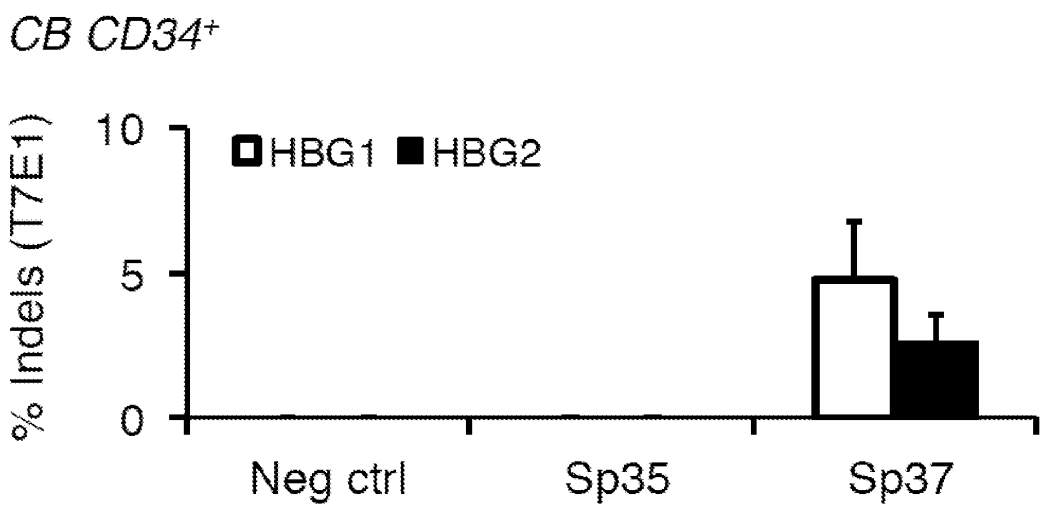

CB CD34+
% Indels (T7E1)
10
5
0
HBG1
HBG2
Neg ctrl
Sp35
Sp37

FIG 4B

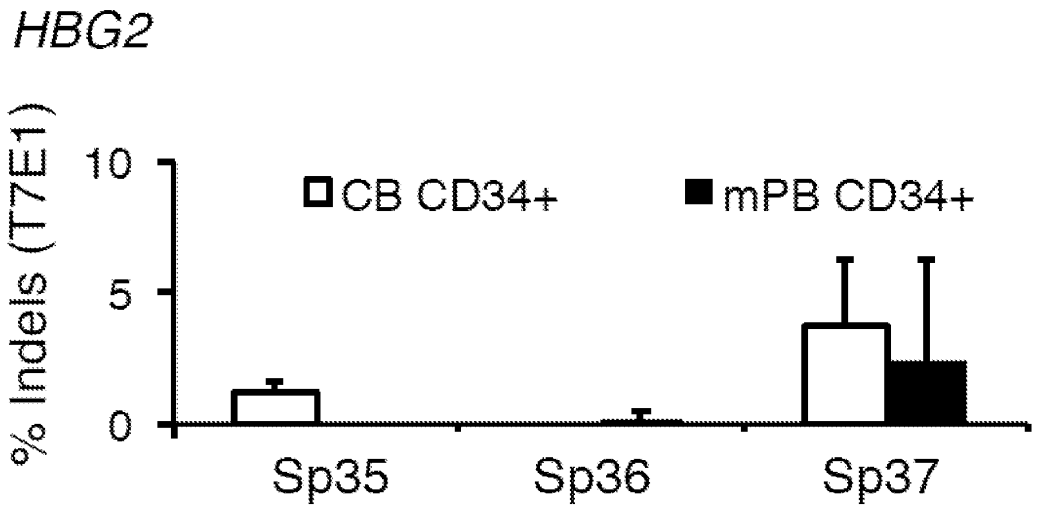

HBG2
% Indels (T7E1)
10
5
0
CB CD34+
mPB CD34+
Sp35
Sp36
Sp37

FIG 4C

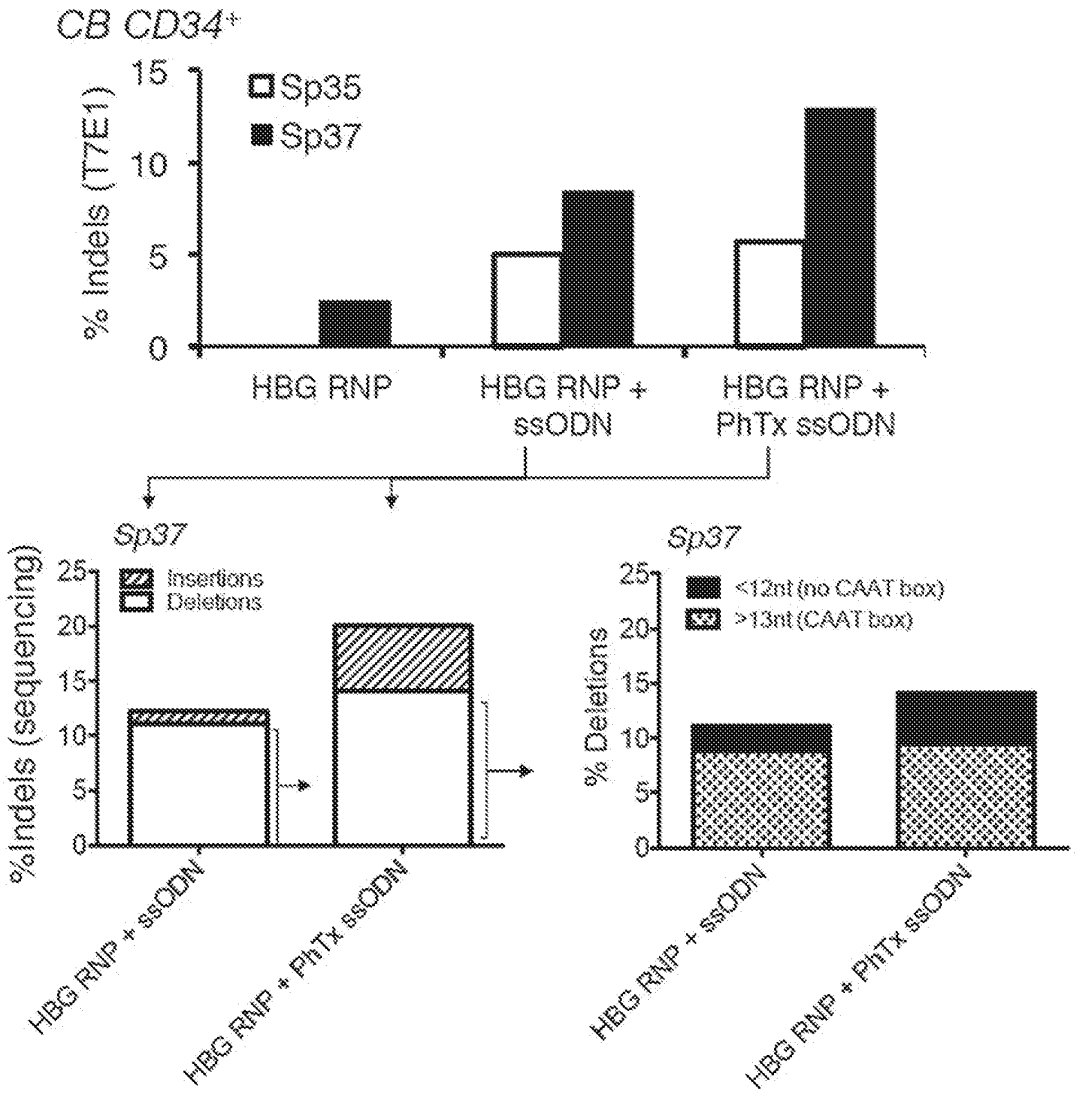

CB CD34+
% Indels (T7E1)
15
10
5
0
Sp35
Sp37
HBG RNP
HBG RNP + ssODN
HBG RNP + PhTx ssODN
Sp37
%Indels (sequencing)
25
20
15
10
5
0
Insertions
Deletions
HBG RNP + ssODN
HBG RNP + PhTx ssODN
Sp37
% Deletions
<12nt (no CAAT box)
>13nt (CAAT box)
HBG RNP + ssODN
HBG RNP + PhTx ssODN

FIG 5A

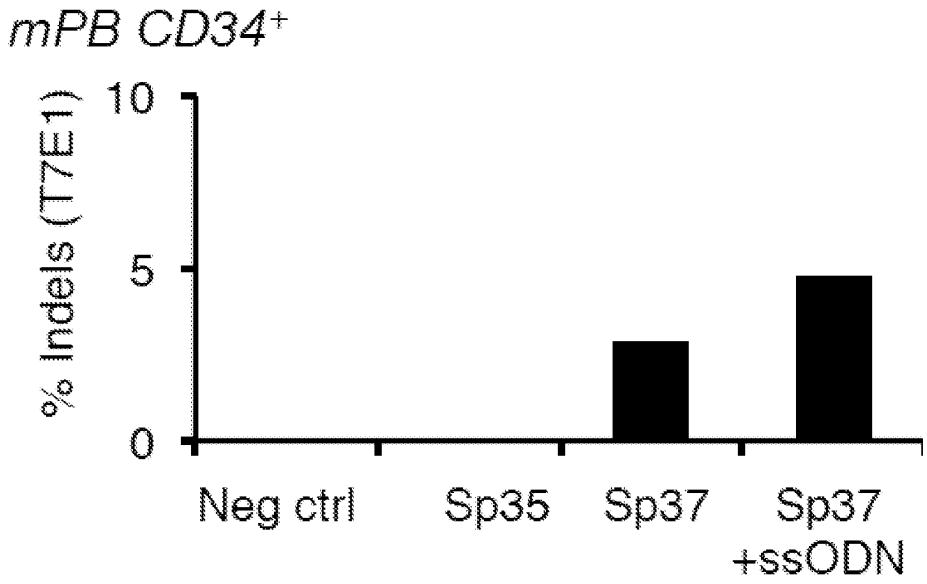

mPB CD34+
% Indels (T7E1)
10
5
0
Neg ctrl
Sp35
Sp37
Sp37
+ssODN

FIG 5B

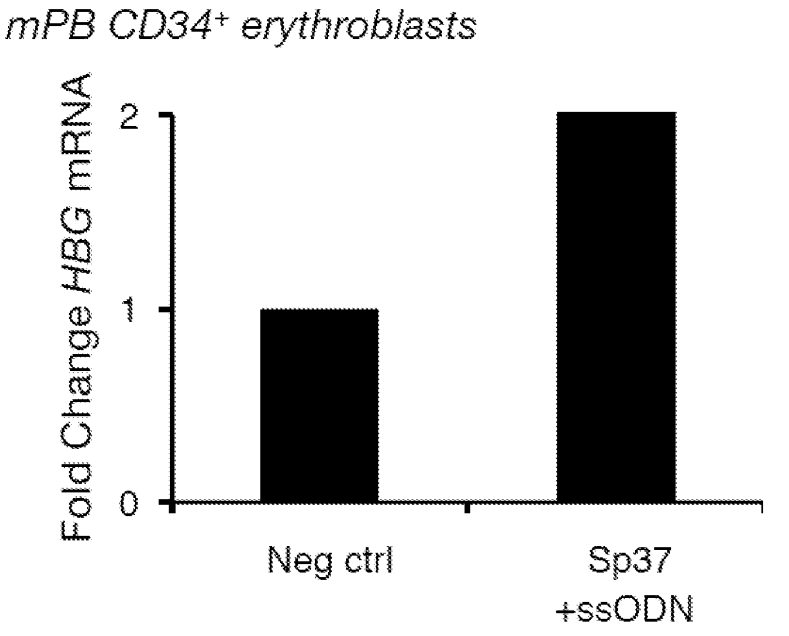

mPB CD34+ erythroblasts
Fold Change HBG mRNA
2
1
0
Neg ctrl
Sp37
+ssODN

FIG 6A

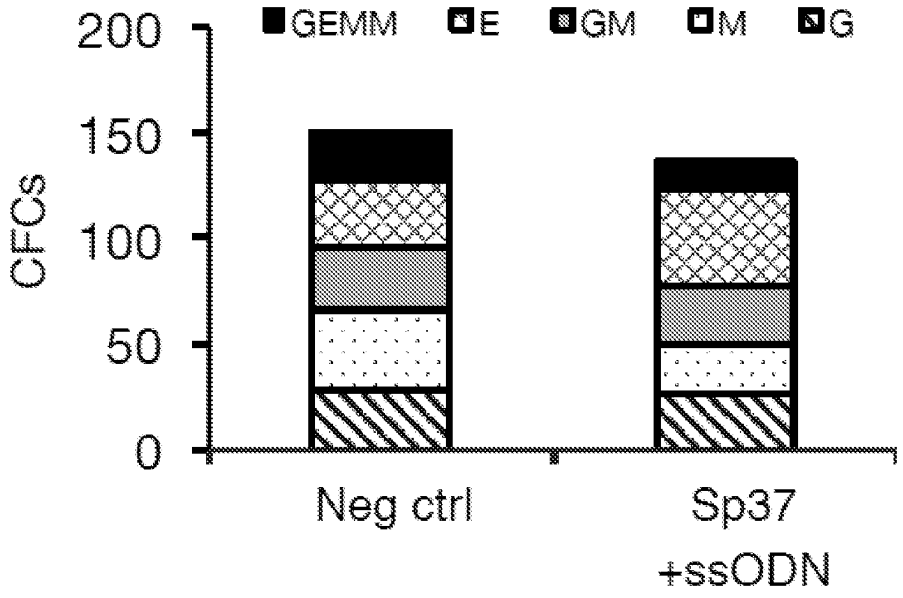

GEMM
E
GM
M
G
200
150
100
50
0
CFCs
Neg ctrl
Sp37
+ssODN

FIG 6B

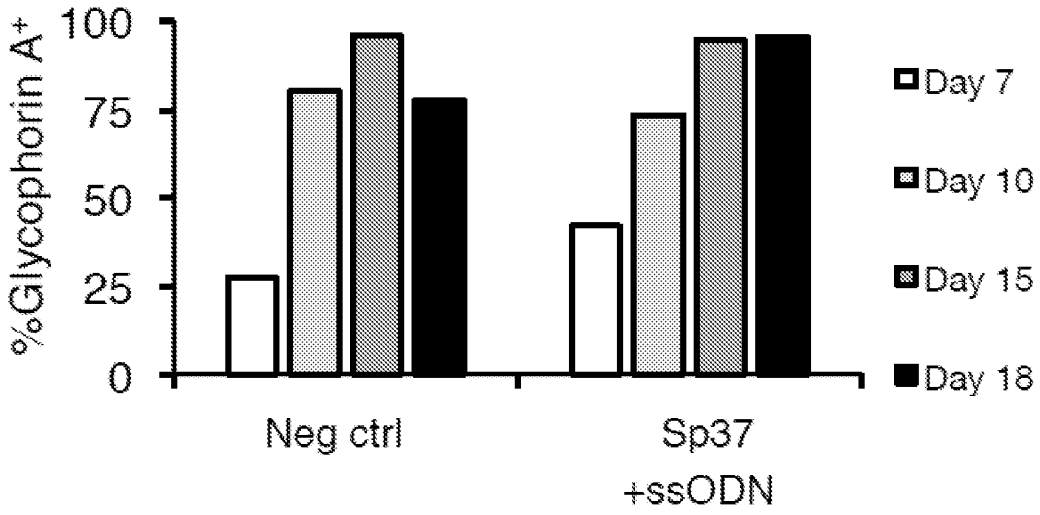

100
75
50
25
0
%Glycophorin A+
Day 7
Day 10
Day 15
Day 18
Neg ctrl
Sp37
+ssODN

FIG 7A

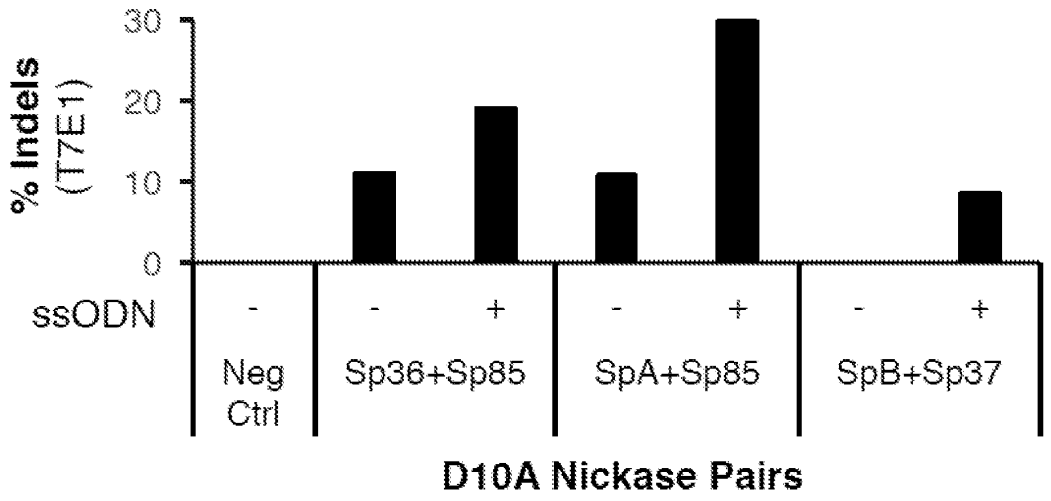

% Indels
(T7E1)
30
20
10
0
ssODN
-
-
+
-
+
-
+
Neg Ctrl
Sp36+Sp85
SpA+Sp85
SpB+Sp37
D10A Nickase Pairs

FIG 7B

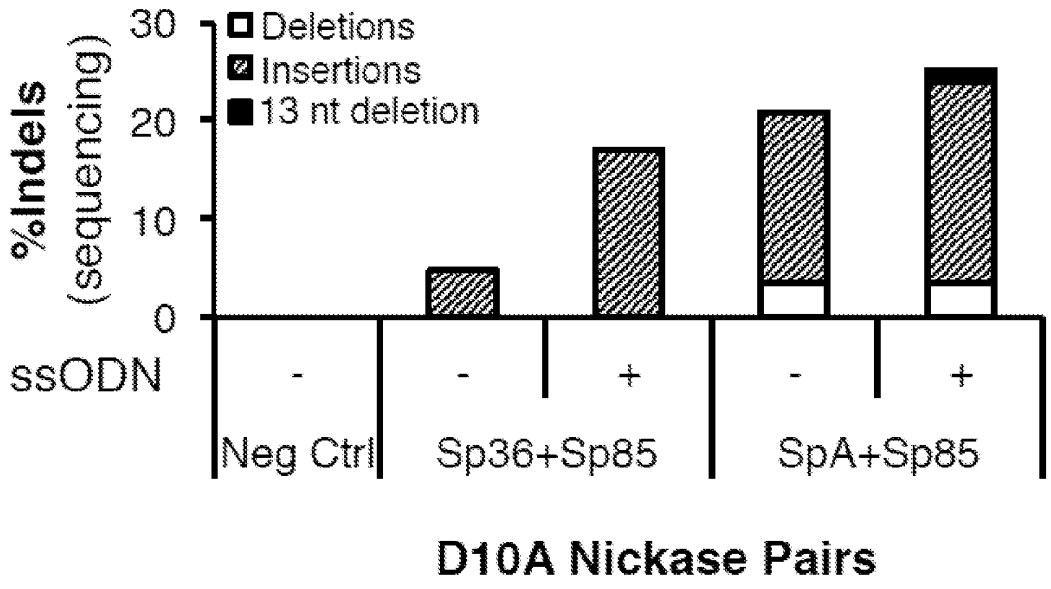

%Indels
(sequencing)
30
20
10
0
Deletions
Insertions
13 nt deletion
ssODN
-
-
+
-
+
Neg Ctrl
Sp36+Sp85
SpA+Sp85
D10A Nickase Pairs

FIG 8A

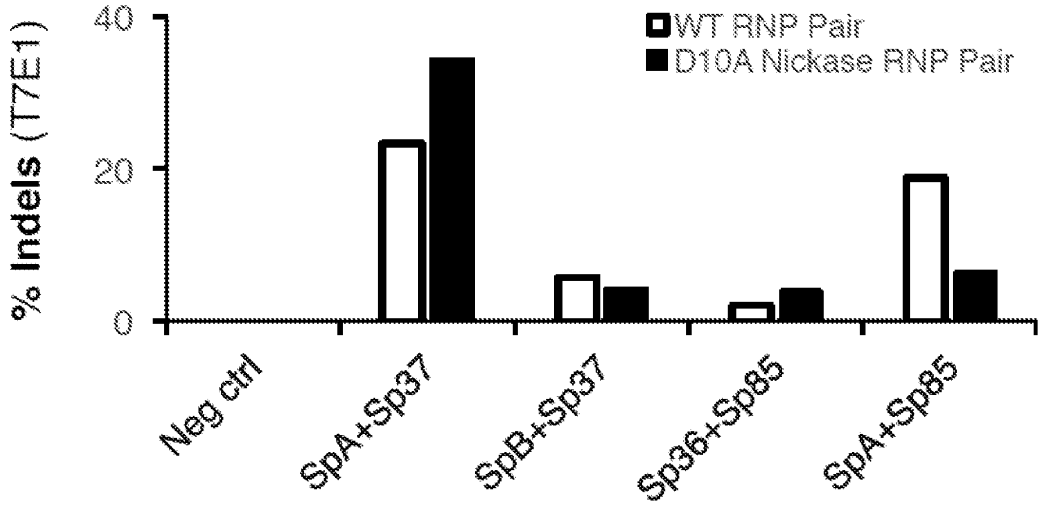

% Indels (T7E1)
40
20
0
WT RNP Pair
D10A Nickase RNP Pair
Neg ctrl
SpA+Sp37
SpB+Sp37
Sp36+Sp85
SpA+Sp85

FIG 8B

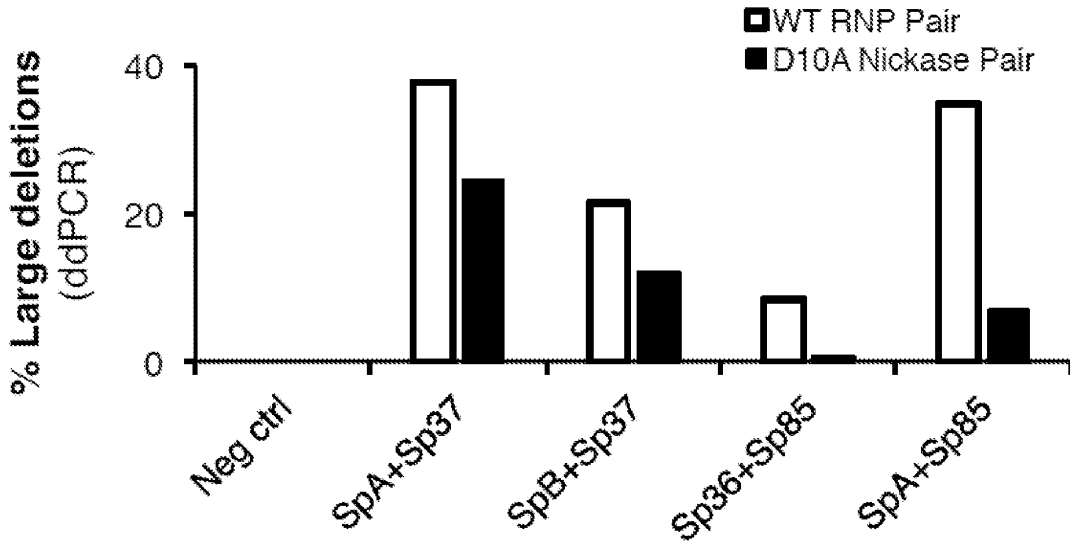

% Large deletions (ddPCR)
40
20
0
WT RNP Pair
D10A Nickase Pair
Neg ctrl
SpA+Sp37
SpB+Sp37
Sp36+Sp85
SpA+Sp85

FIG 8C

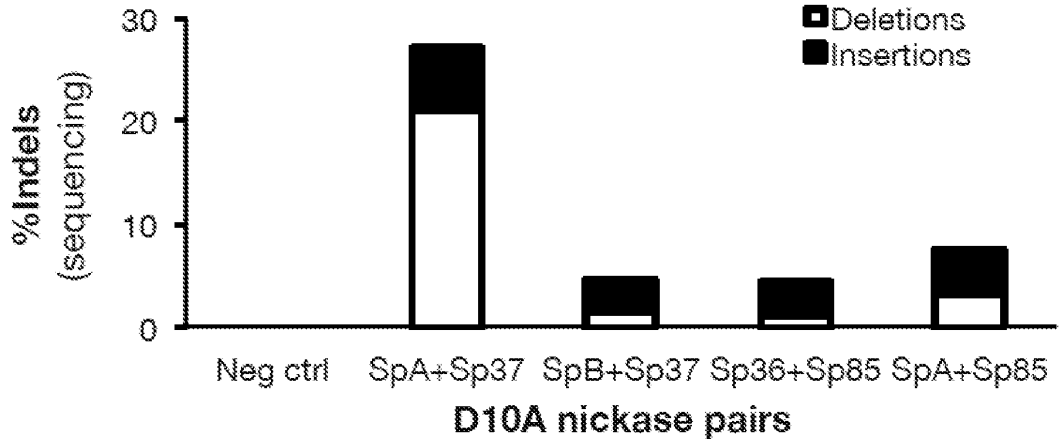

30
20
10
0
%Indels
(sequencing)
Deletions
Insertions
Neg ctrl
SpA+Sp37
SpB+Sp37
Sp36+Sp85
SpA+Sp85
D10A nickase pairs

FIG 8D

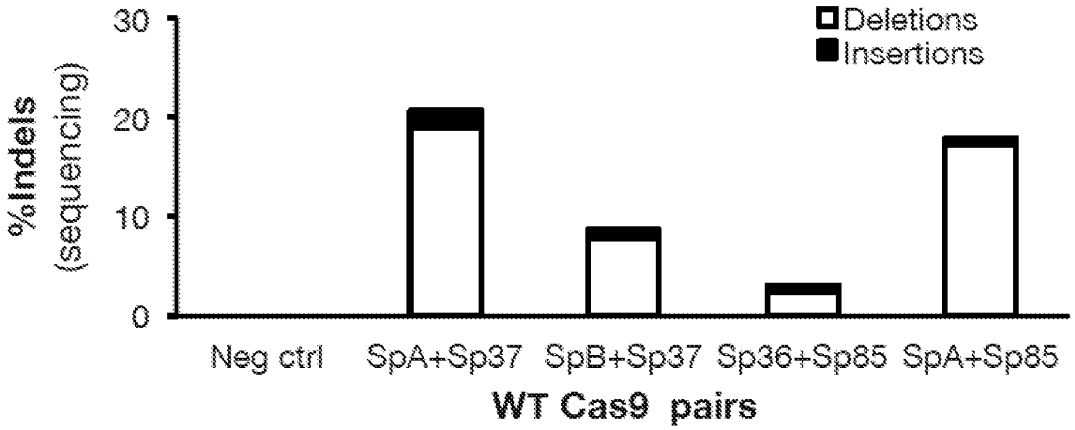

30
20
10
0
%Indels
(sequencing)
Deletions
Insertions
Neg ctrl
SpA+Sp37
SpB+Sp37
Sp36+Sp85
SpA+Sp85
WT Cas9 pairs

FIG 9

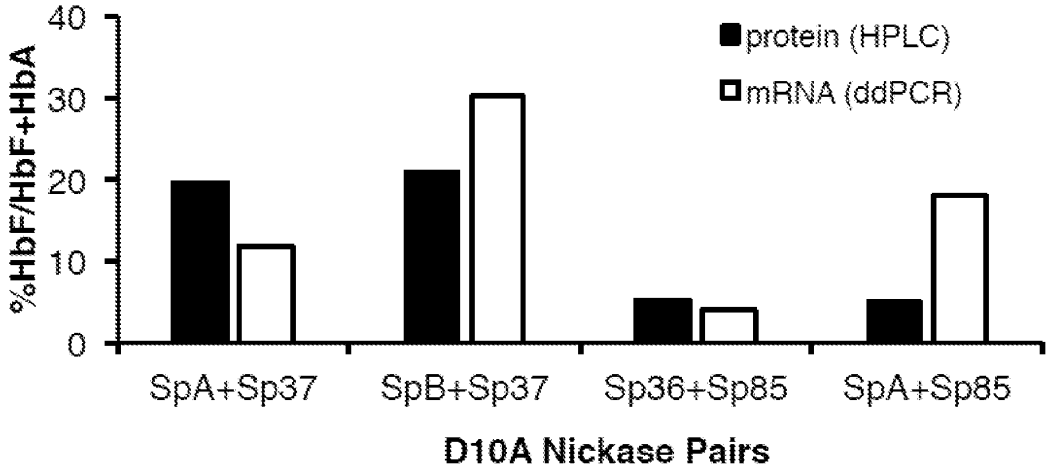

40
30
20
10
0
%HbF/HbF+HbA
protein (HPLC)
mRNA (ddPCR)
SpA+Sp37
SpB+Sp37
Sp36+Sp85
SpA+Sp85
D10A Nickase Pairs

FIG 10A

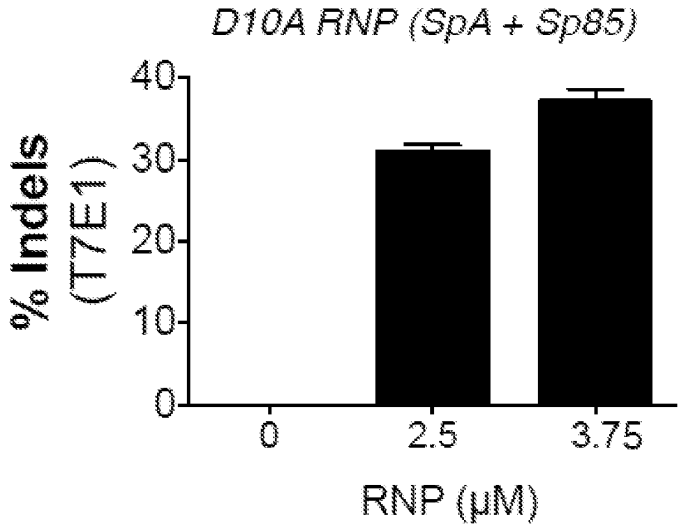


FIG 10B

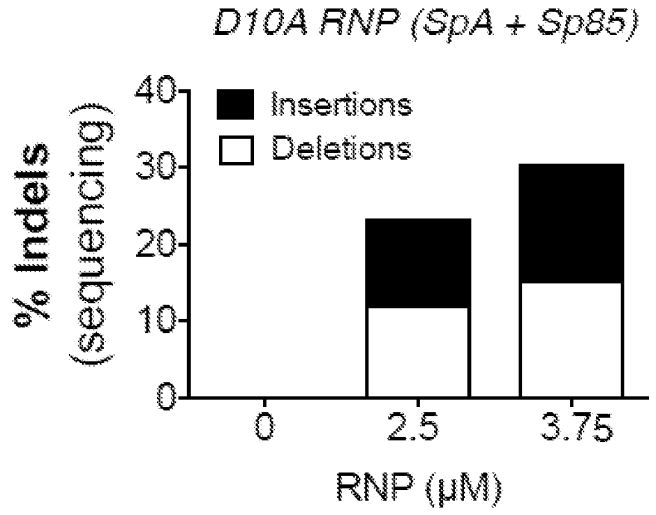


FIG 10C

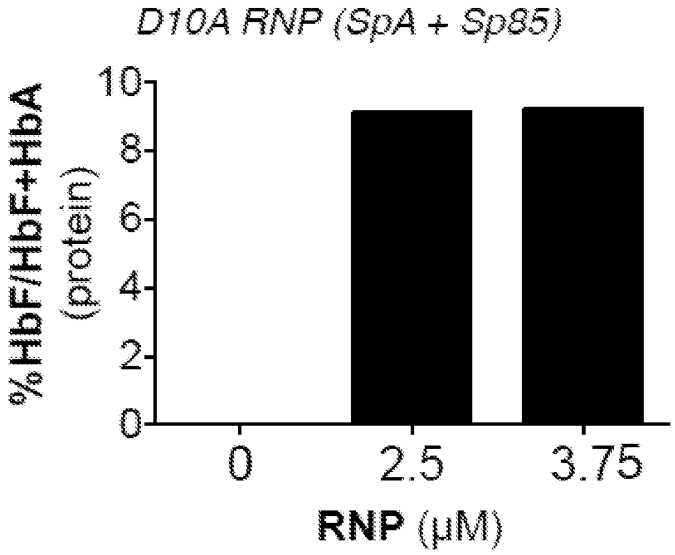


FIG 10D

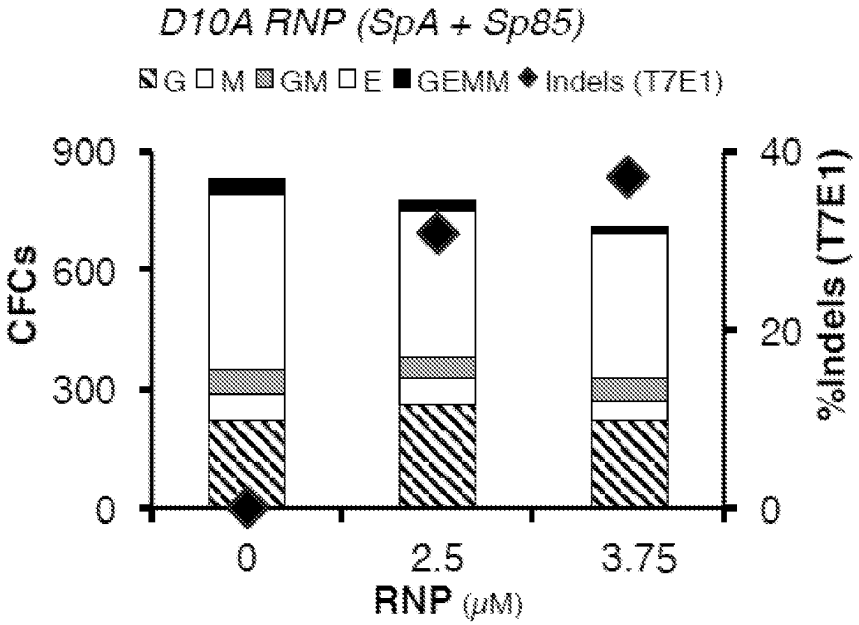


FIG 10E

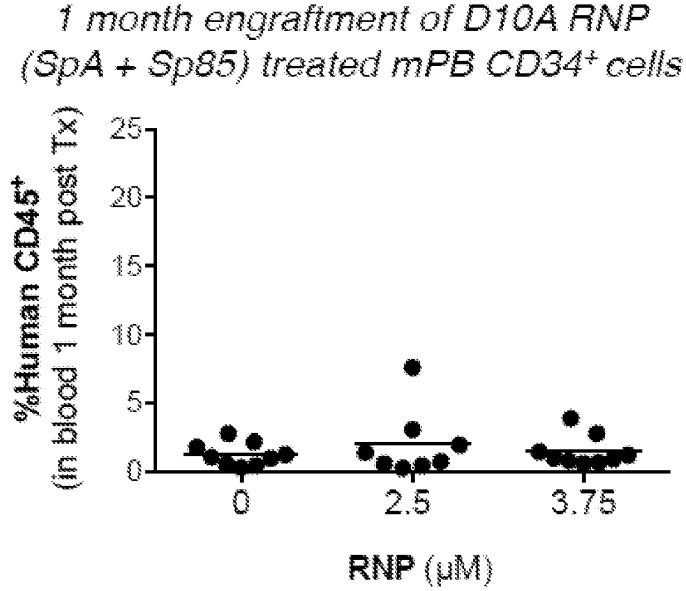


FIG 10F

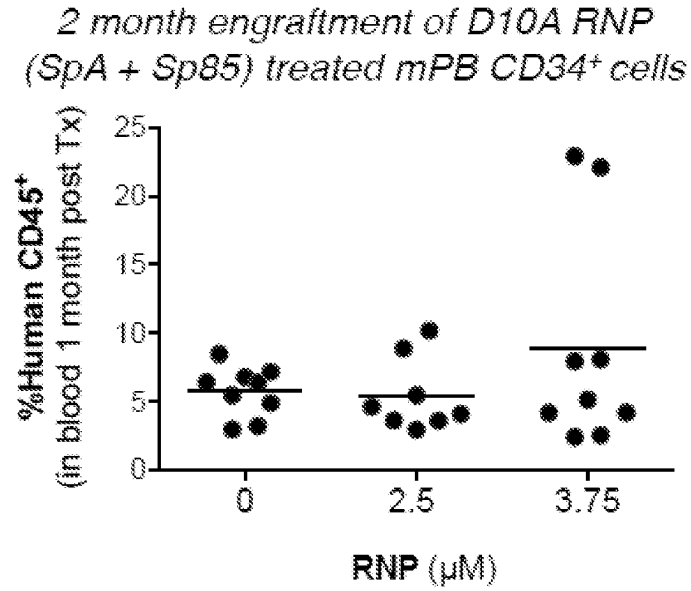


FIG 10G

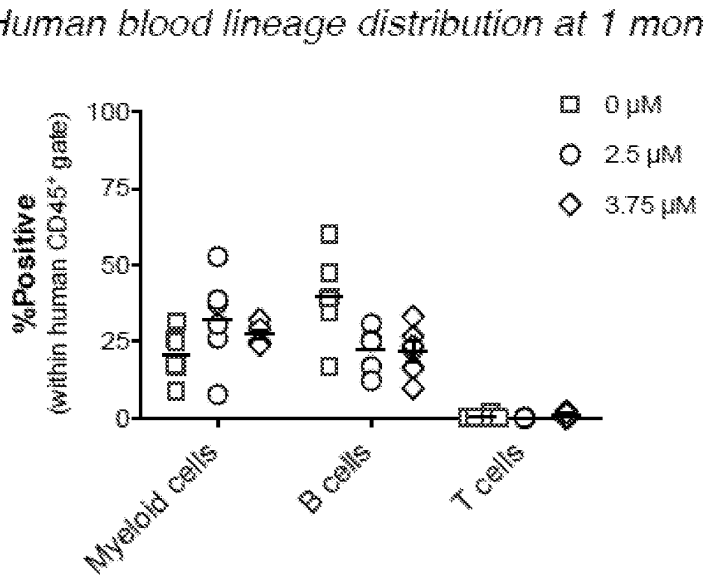


FIG 10H

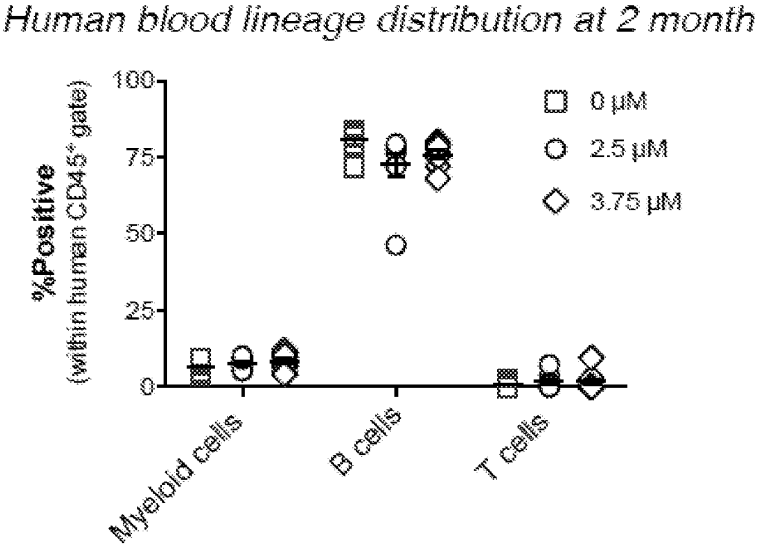


FIG 11A

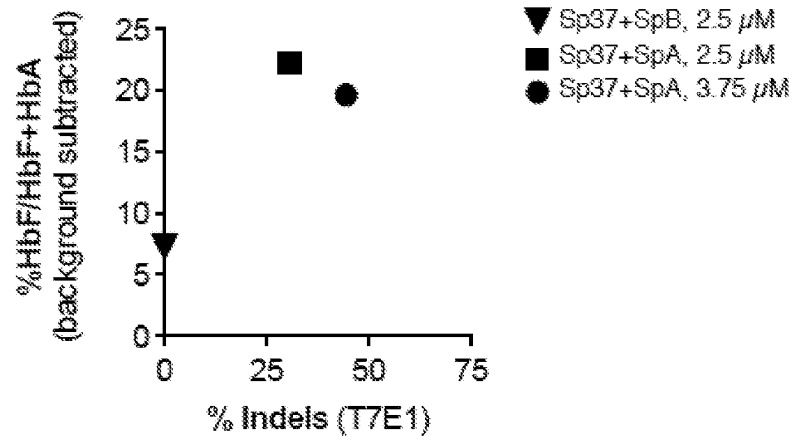


FIG 11B

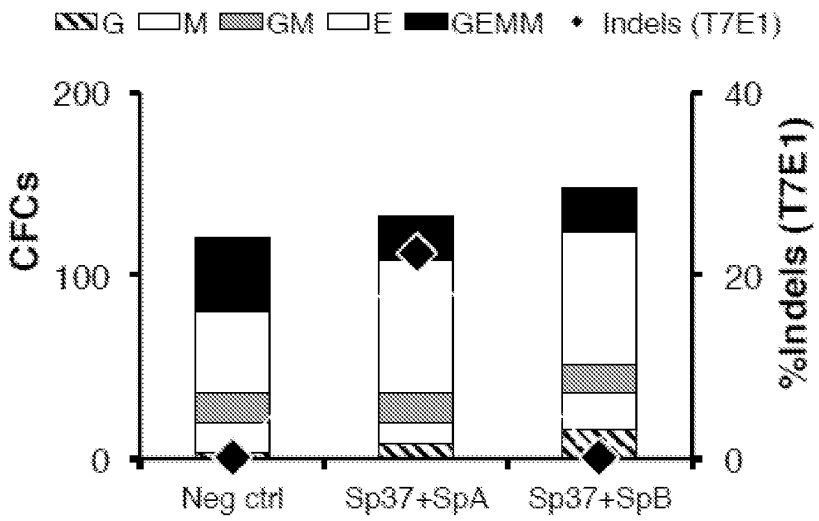


FIG 11C

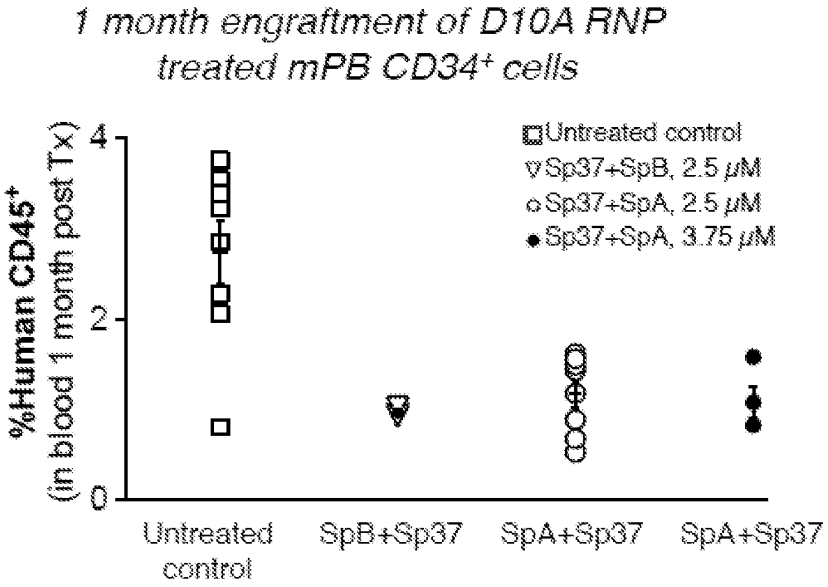
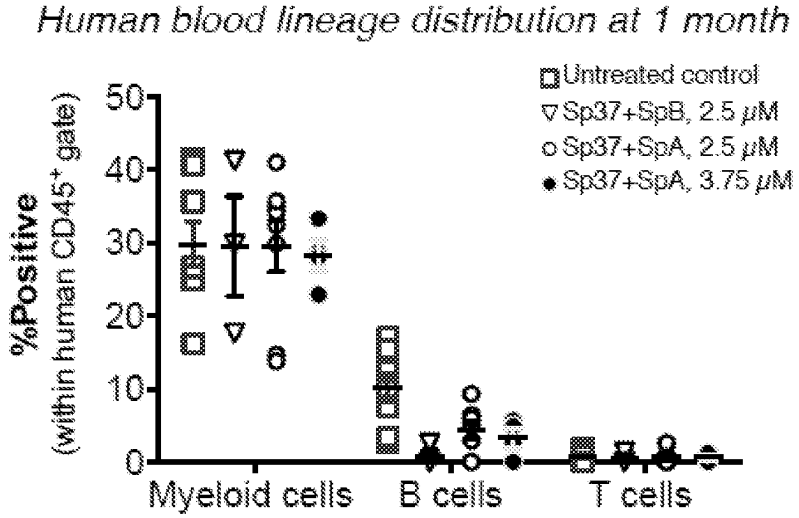


FIG 11D



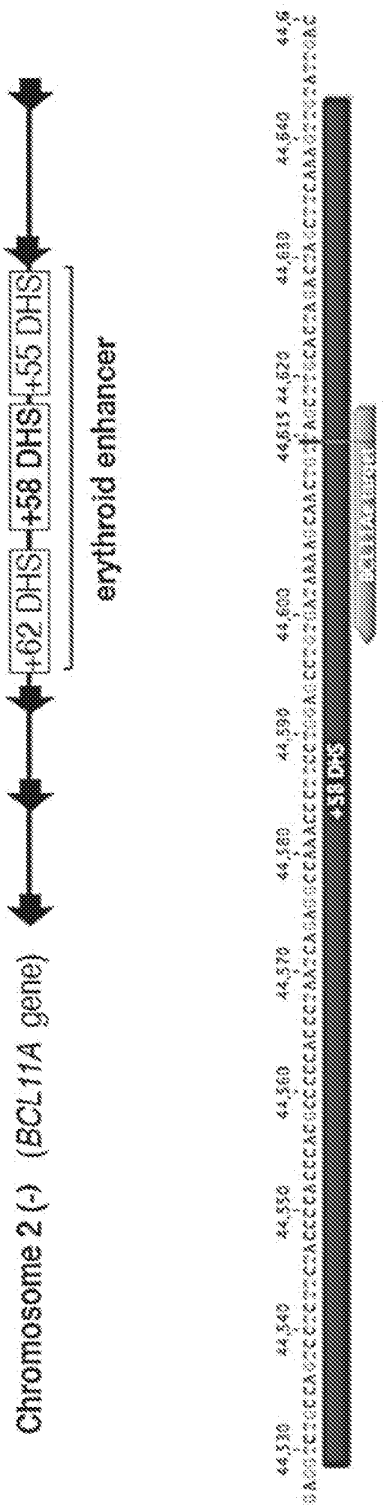

Chromosome 2 (-) (BCL11A gene)
+62 DHS
+58 DHS
+55 DHS
erythroid enhancer
44,530
44,549
44,550
44,569
44,570
44,589
44,590
44,609
44,610
44,629
44,630
44,649
44,650
BCL11A

FIG 12

Attorn

FIG 13A

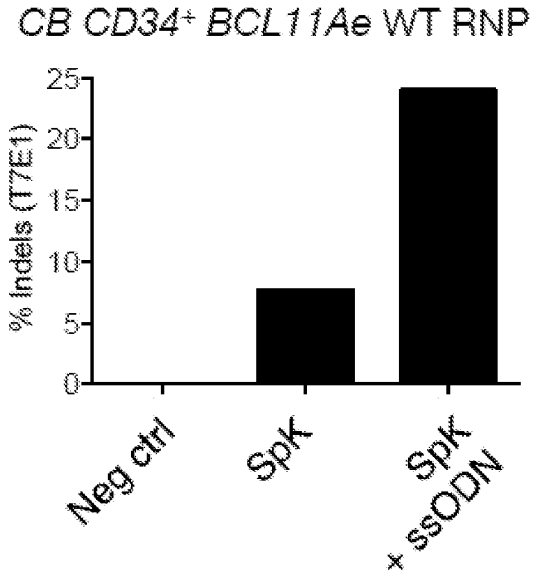
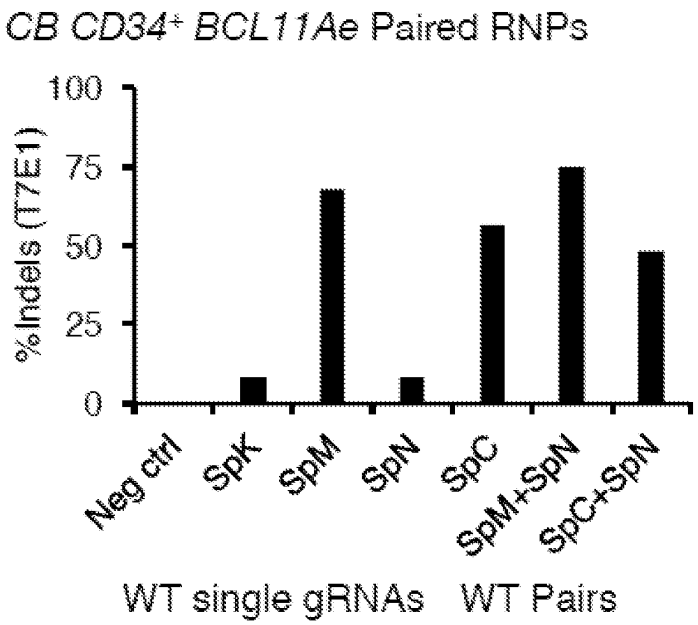


FIG 13B



Attorn

FIG 14A

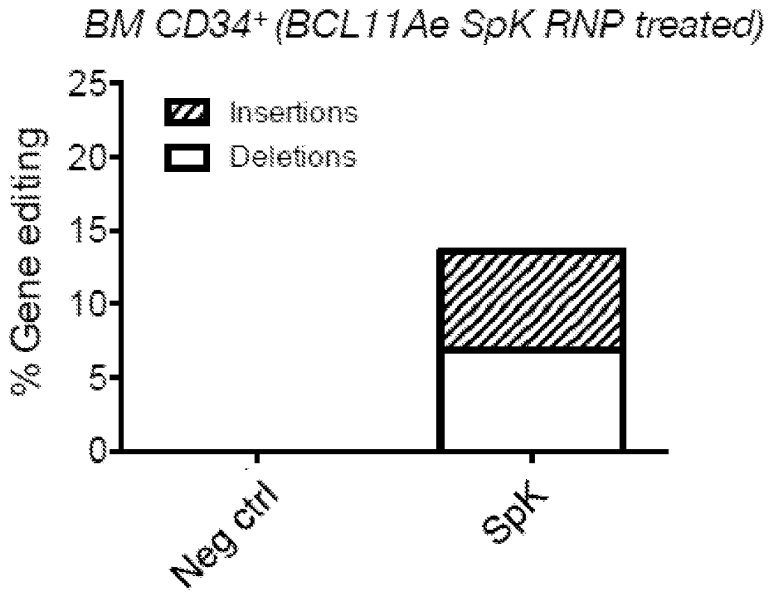
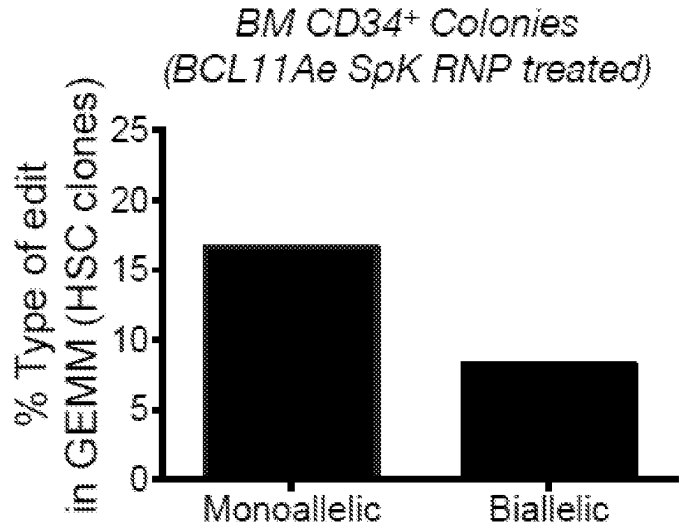


FIG 14B



Attorn

FIG 14C

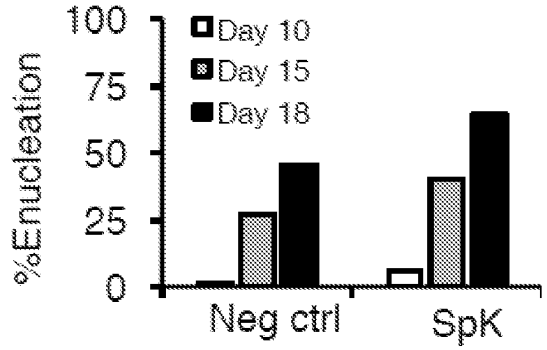


FIG 14D

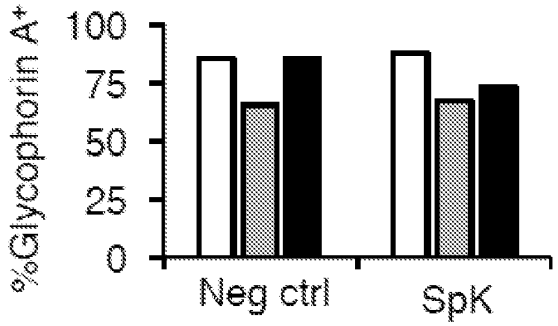
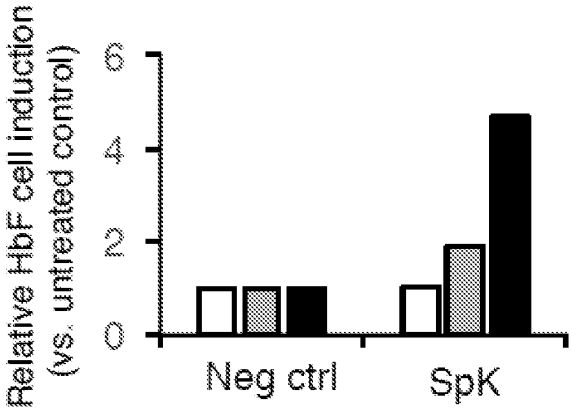


FIG 14E



Attorn

FIG 15A

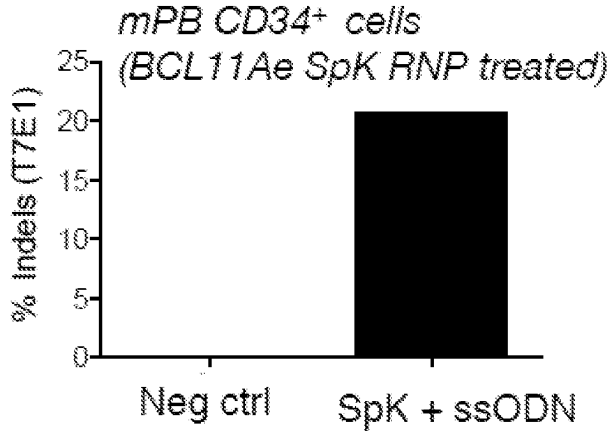


FIG 15B

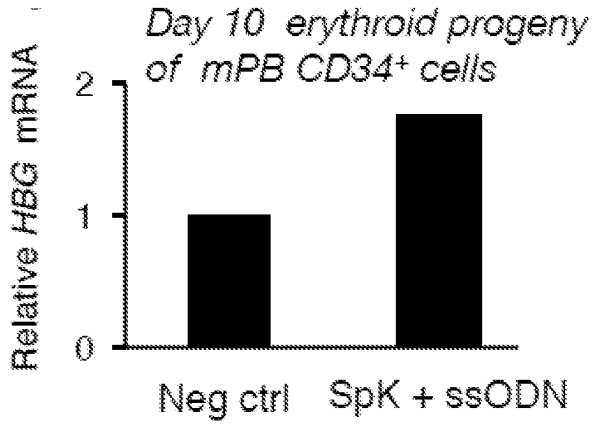


FIG 15C

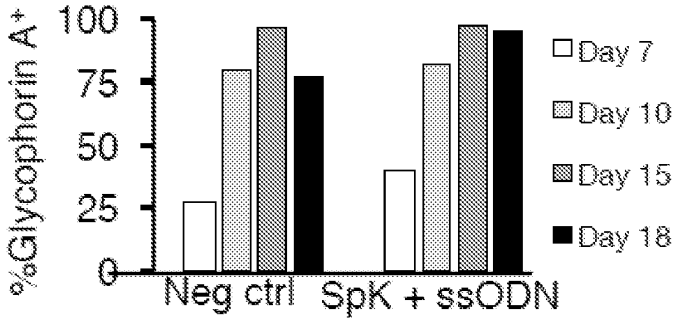


FIG 16

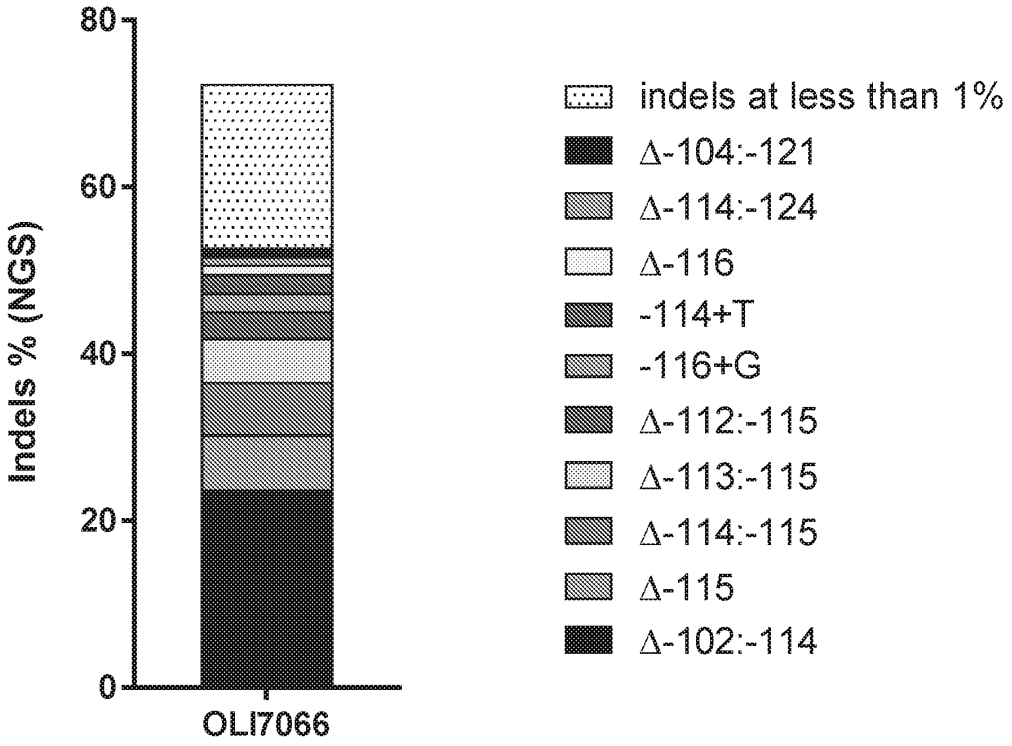

80
60
40
20
0
Indels % (NGS)
OLI7066
indels at less than 1%
Δ-104:-121
Δ-114:-124
Δ-116
-114+T
-116+G
Δ-112:-115
Δ-113:-115
Δ-114:-115
Δ-115
Δ-102:-114

FIG 17A

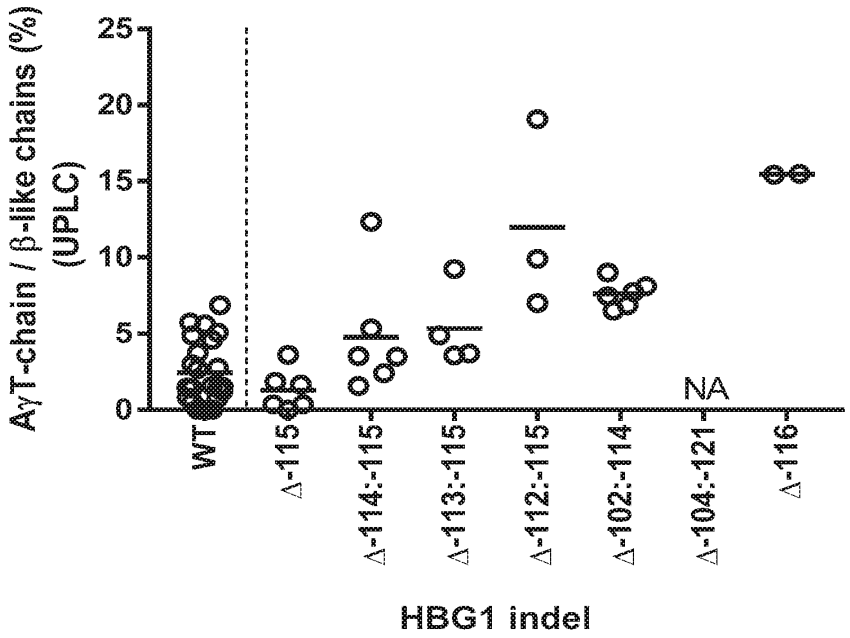


FIG 17B

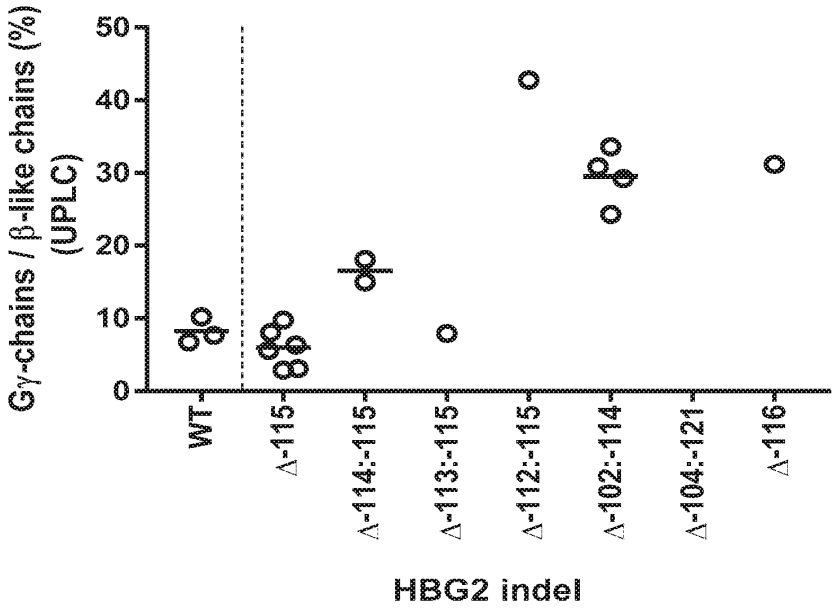


FIG 17C

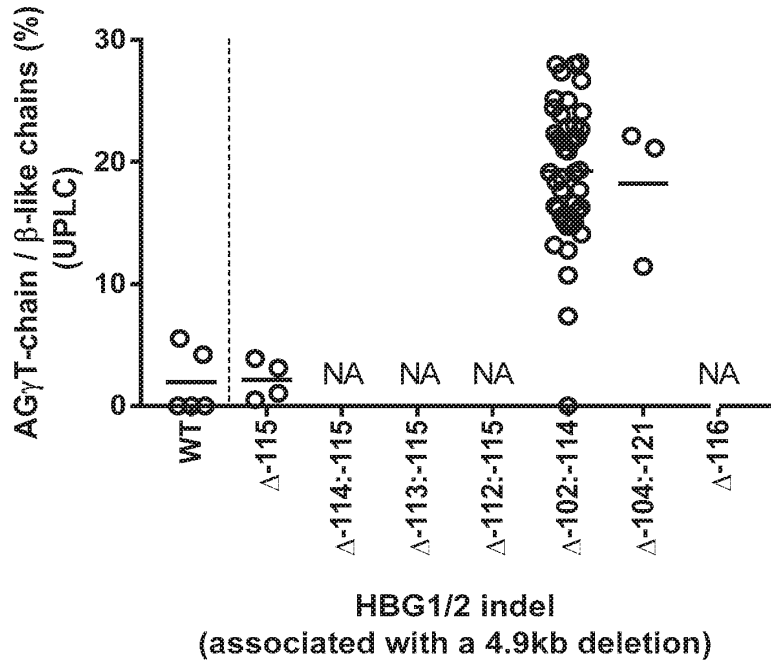
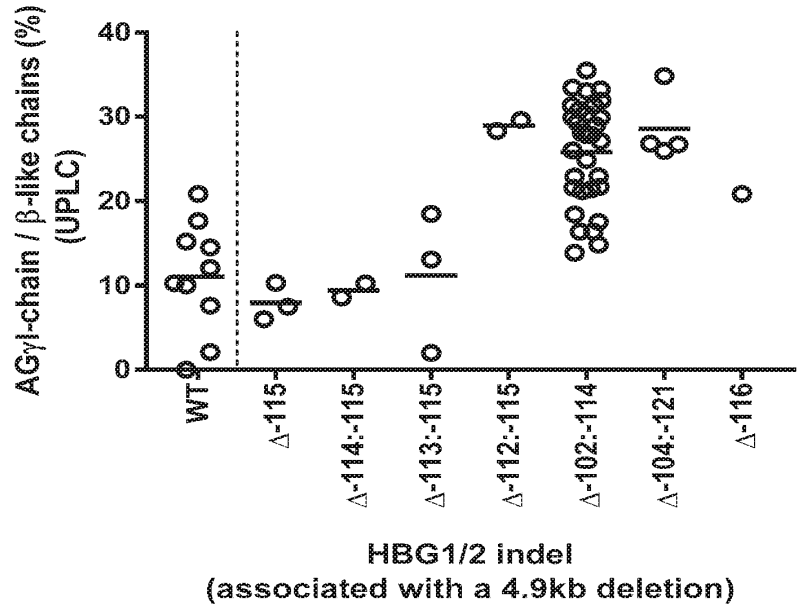


FIG 17D



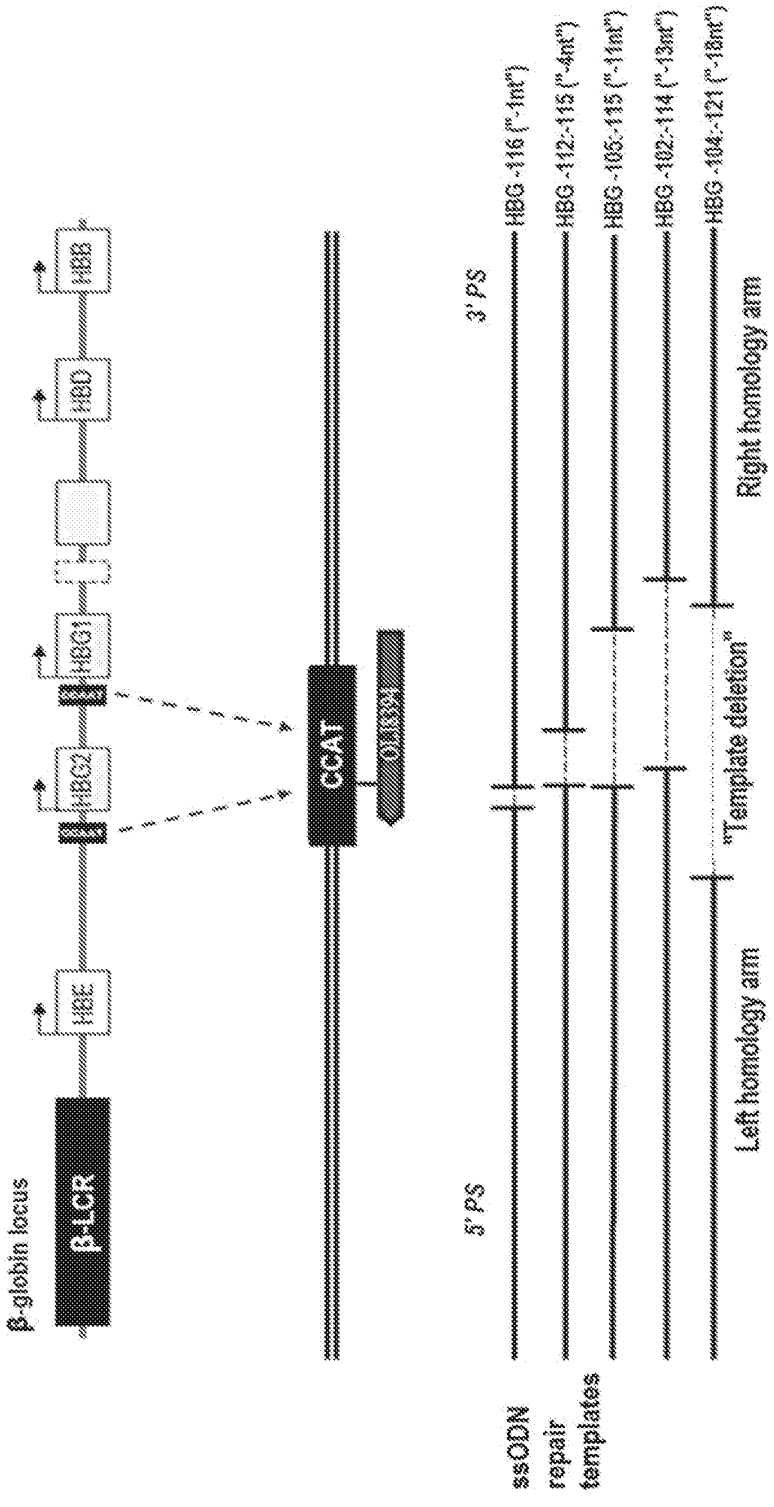

β-globin locus
β-LCR
HBE
HBG2
HBG1
HBD
HBB
CCAT
ORF95
5' PS
3' PS
HBG -116 ("-1nt")
HBG -112:-115 ("-4nt")
HBG -105:-115 ("-11nt")
HBG -102:-114 ("-13nt")
HBG -104:-121 ("-18nt")
Left homology arm
"Template deletion"
Right homology arm
ssODN repair templates

FIG 18

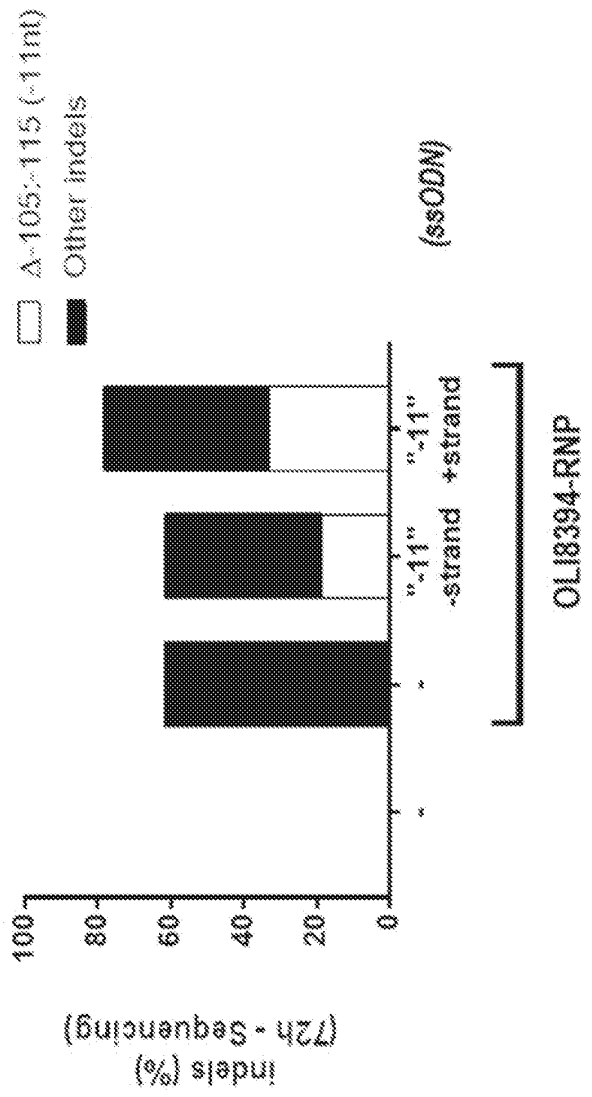

Δ-105:-115 (-11nt)
Other indels
(ssODN)
"-11" +strand
"-11" -strand
OLI8394-RNP
100
80
60
40
20
0
Indels (%)
(72h - Sequencing)

FIG 19A

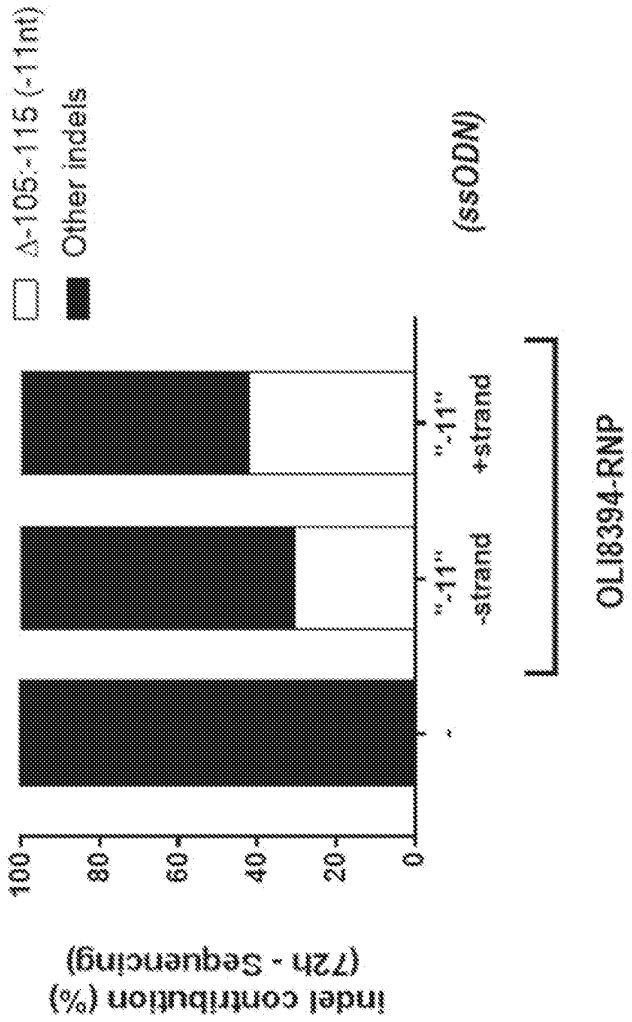

Δ-105:-115 (-11nt)
Other indels
(ssODN)
"-11" +strand
"-11" -strand
-
OLI8394-RNP
0
20
40
60
80
100
Indel contribution (%)
(72h - Sequencing)

FIG 19B

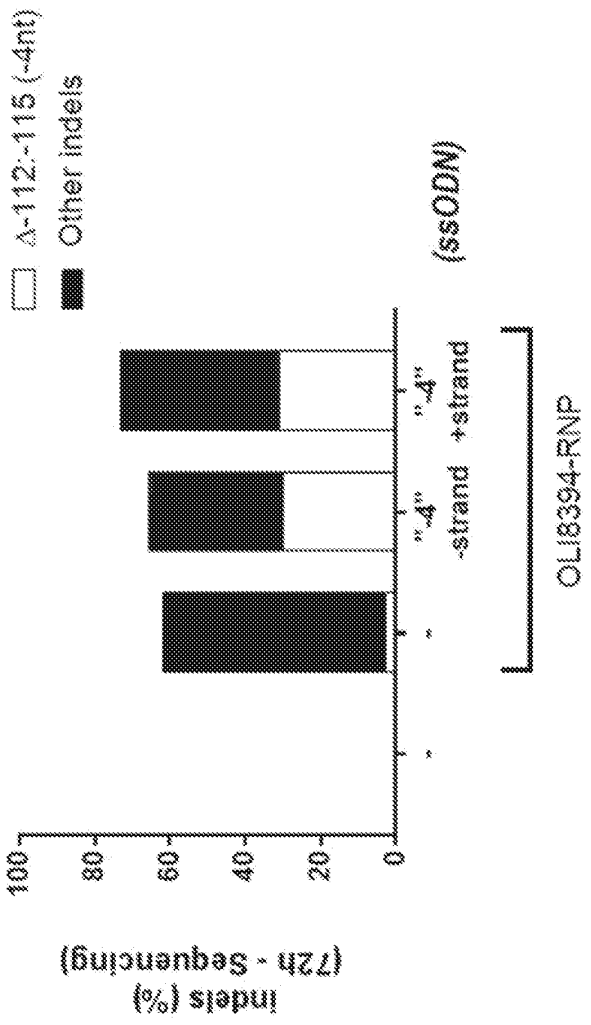
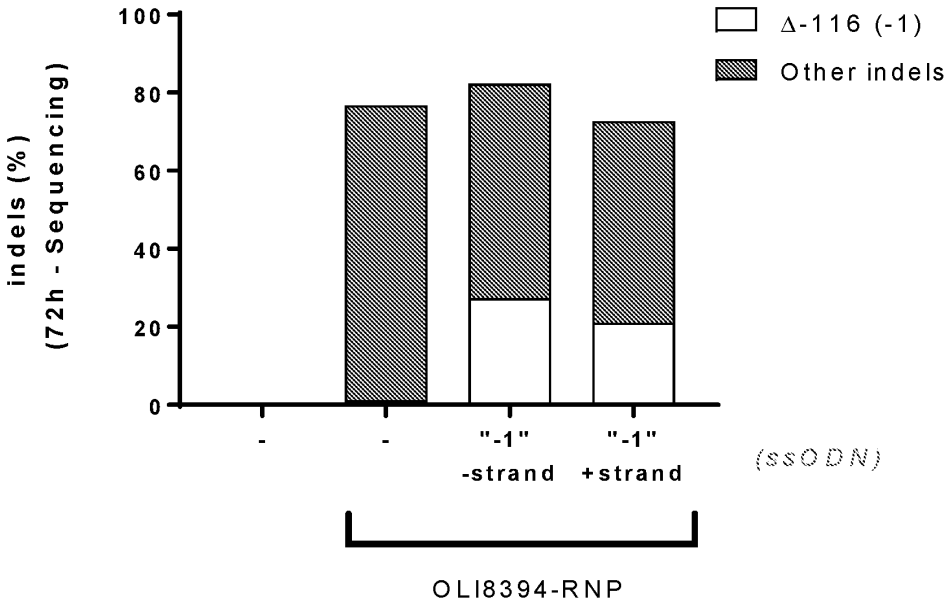

Δ-112-115 (-4nt)
Other indels
Indels (%)
(72h - Sequencing)
100
80
60
40
20
0
-
-
"-4"
-strand
"-4"
+strand
(ssODN)
OLI8394-RNP

FIG 19C

FIG 19D


100
80
60
40
20
0
indels (%)
(72h - Sequencing)
Δ-116 (-1)
Other indels
-
-
"-1"
-strand
"-1"
+strand
(ssODN)
OLI8394-RNP

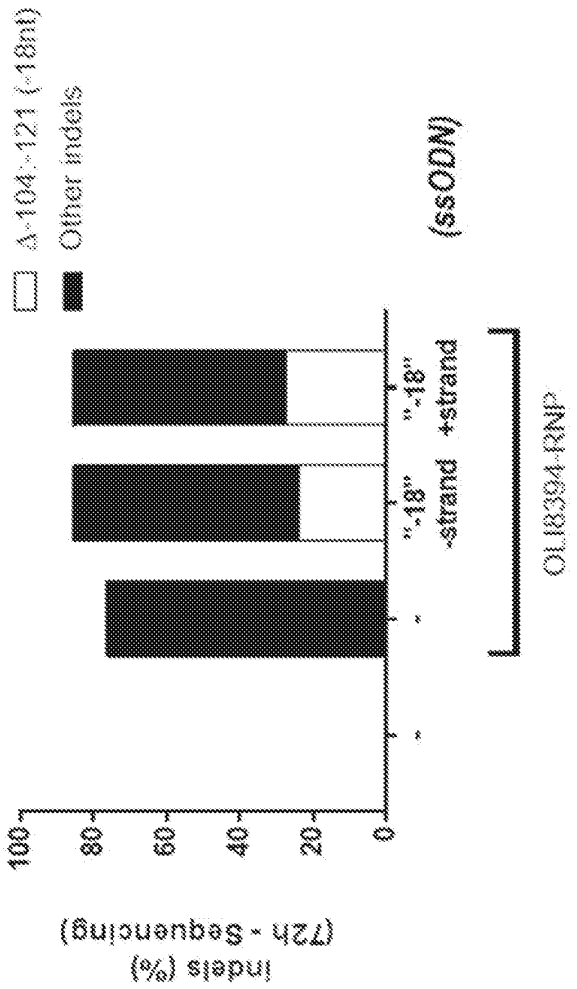

Δ-104:-121 (-18nt)
Other indels
Indels (%)
(72h - Sequencing)
100
80
60
40
20
0
-
-
"-18"
-strand
"-18"
+strand
(ssODN)
OLI8394-RNP

FIG 19E

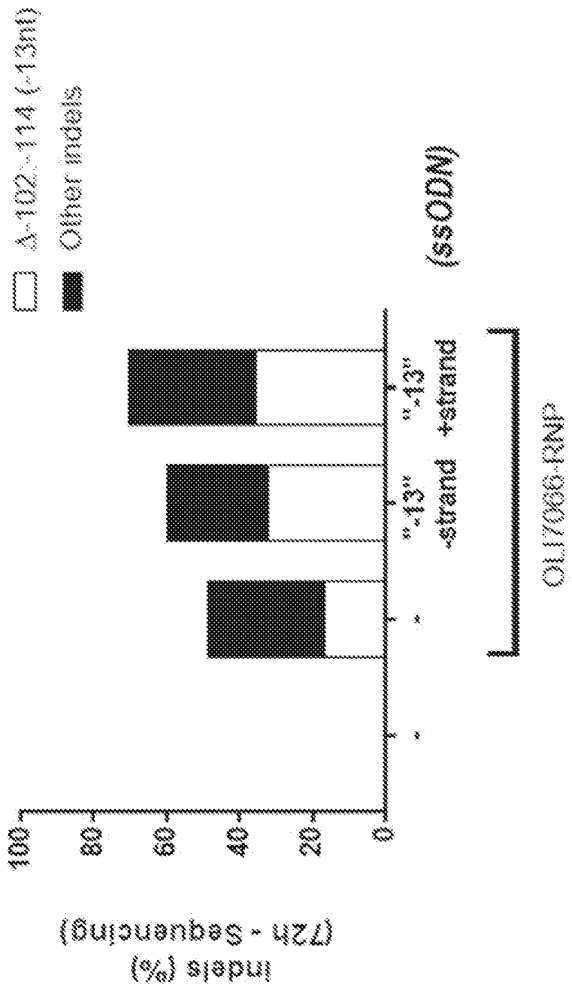

Δ-102:-114 (-13nt)
Other indels
Indels (%)
(72h - Sequencing)
100
80
60
40
20
0
"-13" +strand
"-13" -strand
(ssODN)
OLI7080-RNP

FIG 19F

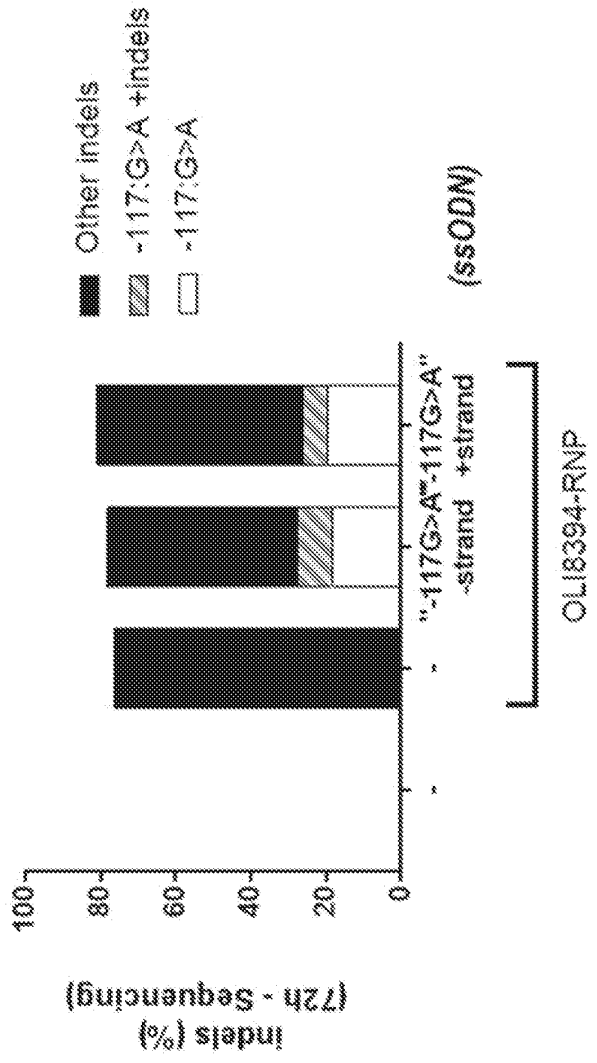

Other indels
-117:G>A +indels
-117:G>A
(ssODN)
"-117G>A"-117G>A"
-strand +strand
OLI8394-RNP
Indels (%)
(72h - Sequencing)
100
80
60
40
20
0

FIG 19G

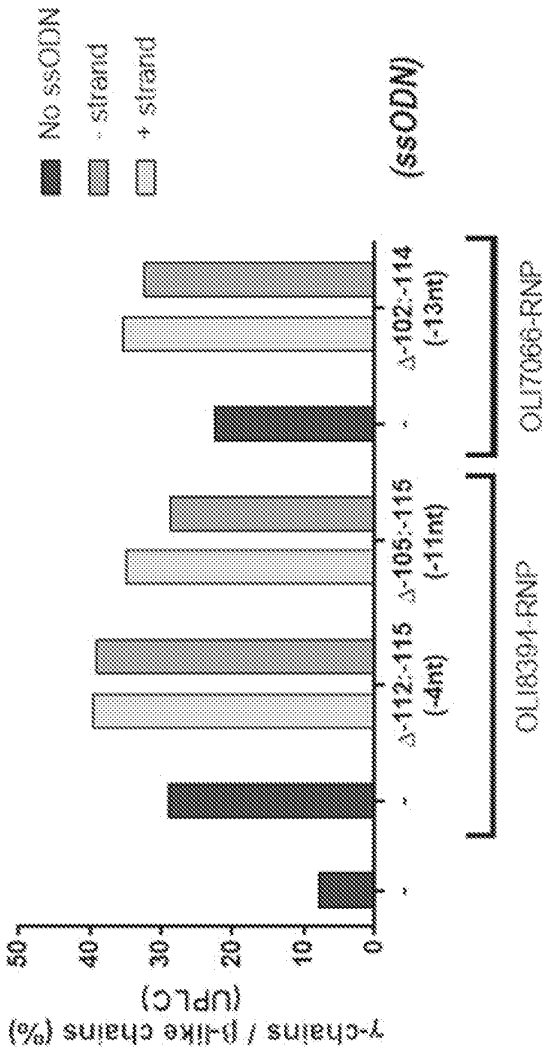

γ-chains / β-like chains (%)
(UPLC)
50
40
30
20
10
0
-
-
Δ-112,-115 (-4nt)
Δ-105,-115 (-11nt)
-
Δ-102,-114 (-13nt)
OLI8304-RNP
OLI7066-RNP
(ssODN)
No ssODN
- strand
+ strand

FIG 20A

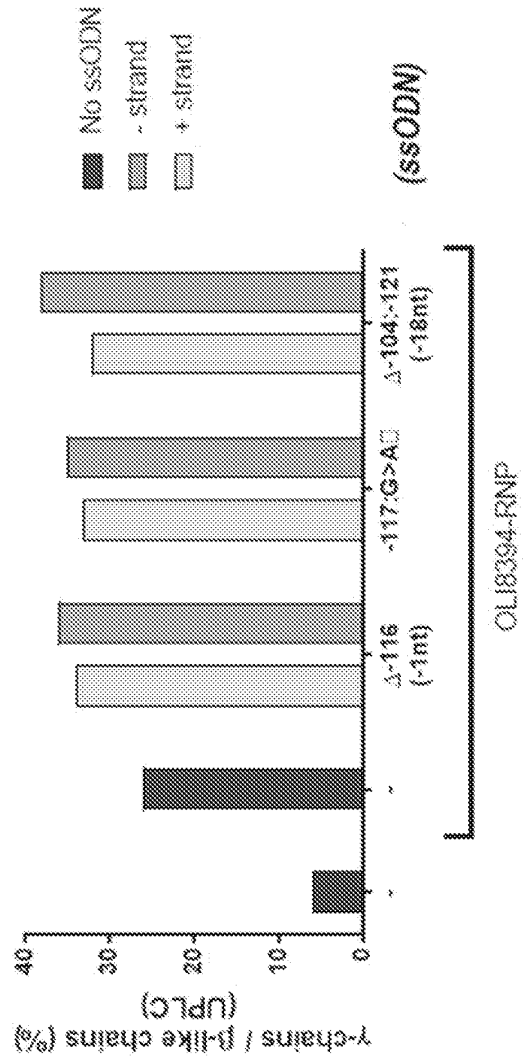

No ssODN
- strand
+ strand
(ssODN)
γ-chains / β-like chains (%)
(UPLC)
40
30
20
10
0
-
-
Δ-116 (-1nt)
-117:G>A
Δ-104:-121 (-18nt)
OLI8394-RNP

FIG 20B

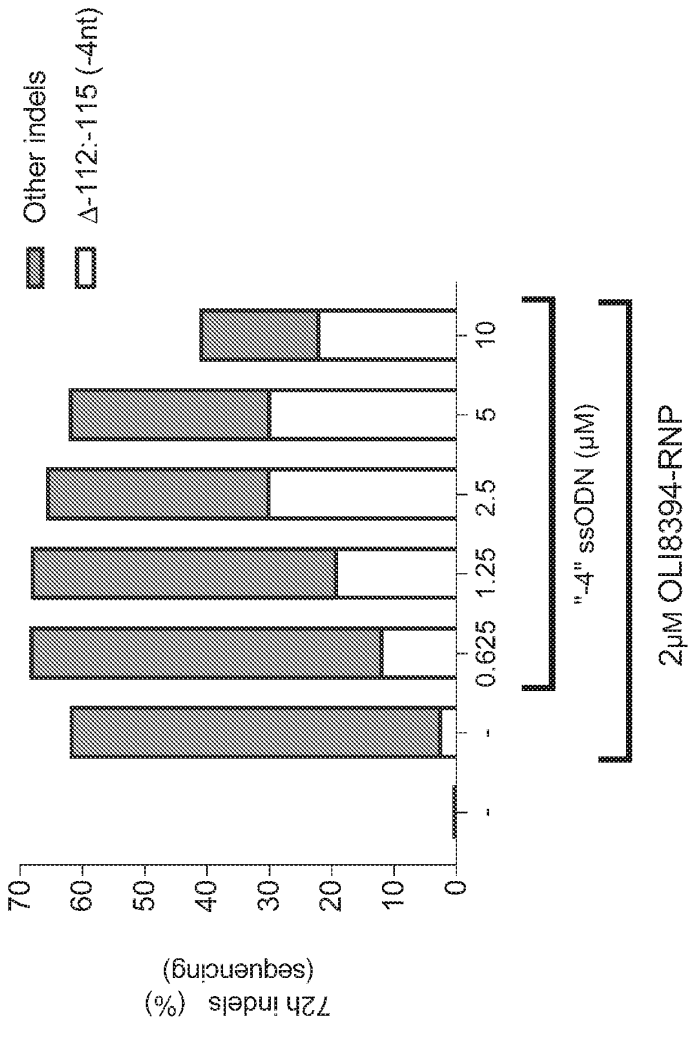

Other indels
Δ-112:-115 (-4nt)
72h indels (%)
(sequencing)
70
60
50
40
30
20
10
0
-
-
0.625
1.25
2.5
5
10
"-4" ssODN (μM)
2μM OLI8394-RNP

FIG 21A

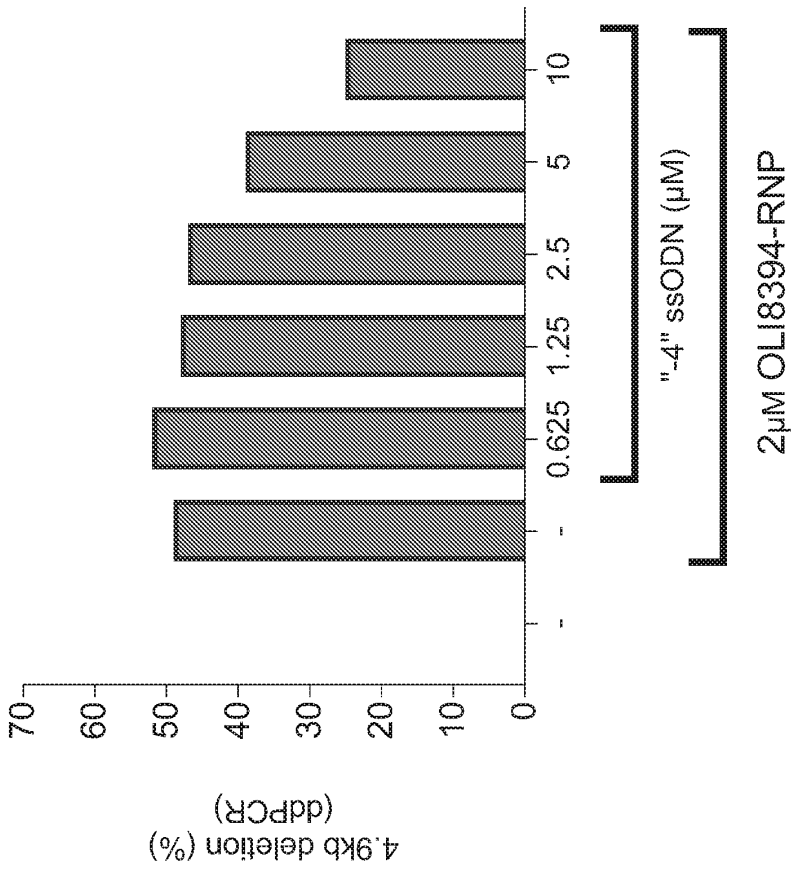

4.9kb deletion (%)
(ddPCR)
70
60
50
40
30
20
10
0
-
-
0.625
1.25
2.5
5
10
"-4" ssODN (μM)
2μM OLI8394-RNP

FIG 21B

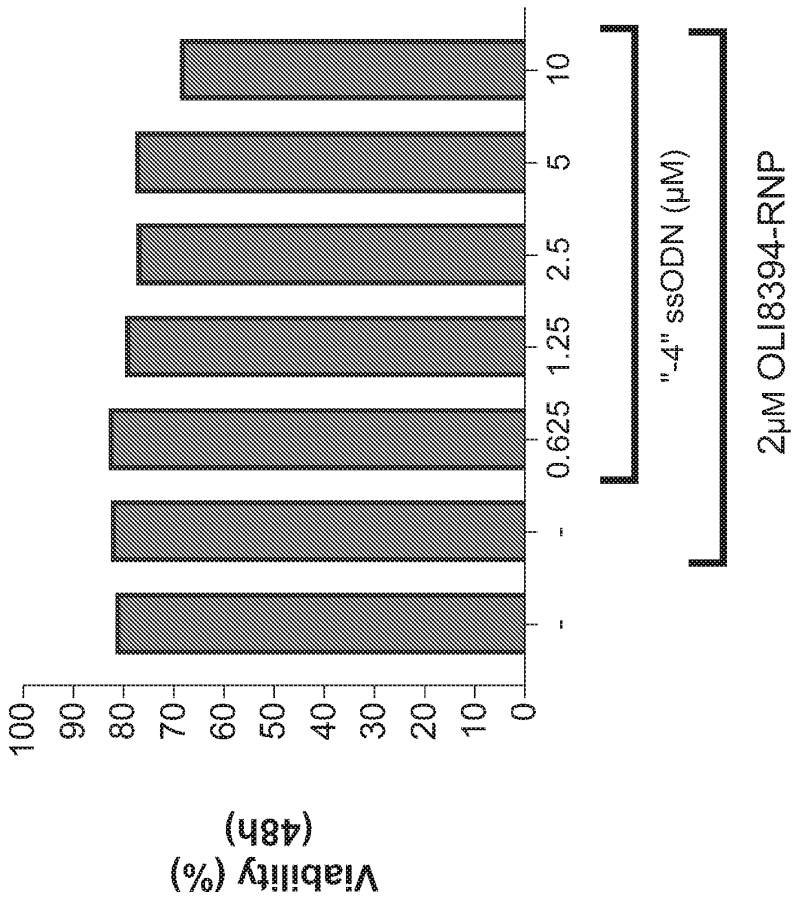

Viability (%)
(48h)
100
90
80
70
60
50
40
30
20
10
0
-
-
0.625
1.25
2.5
5
10
"-4" ssODN (μM)
2μM OLI8394-RNP

FIG 21C

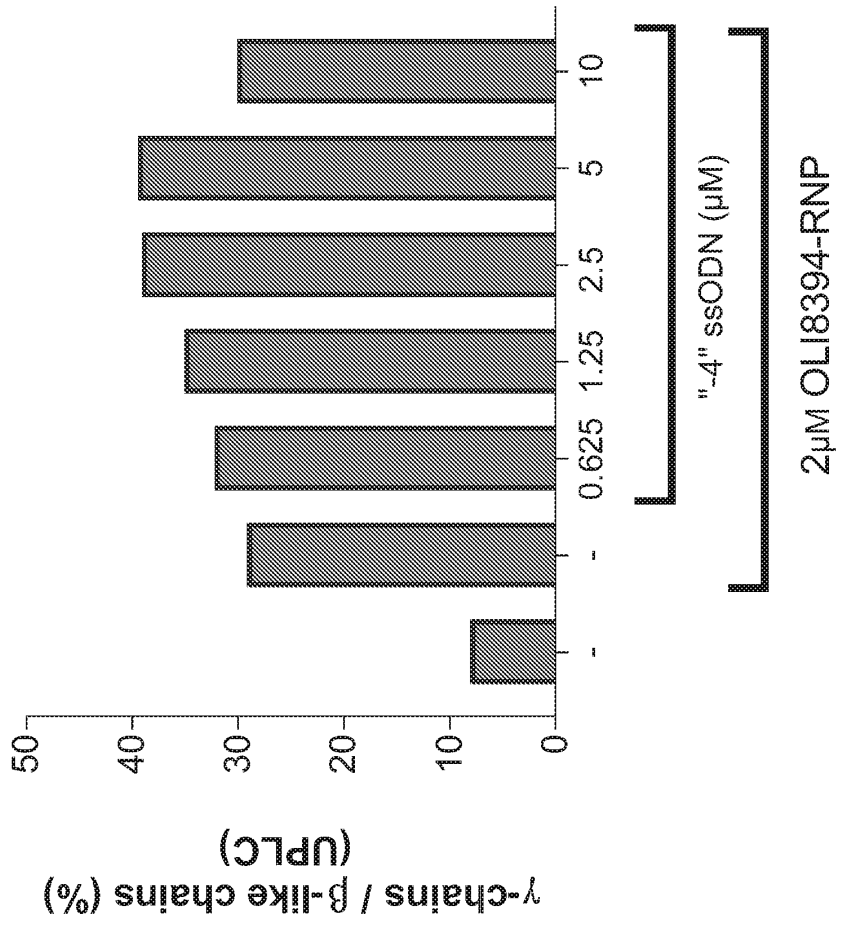

γ-chains / β-like chains (%)
(UPLC)
0
10
20
30
40
50
-
-
0.625
1.25
2.5
5
10
"-4" ssODN (μM)
2μM OLI8394-RNP

FIG 21D

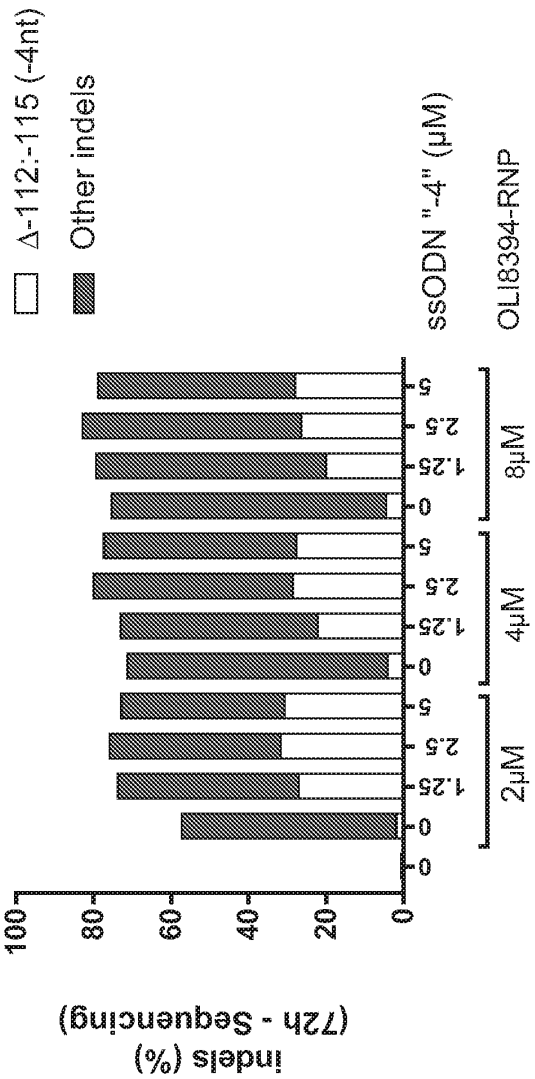

Δ-112:-115 (-4nt)
Other indels
ssODN "-4" (μM)
OLI8394-RNP
Indels (%)
(72h - Sequencing)
100
80
60
40
20
0
0
0
1.25
2.5
5
2μM
0
1.25
2.5
5
4μM
0
1.25
2.5
5
8μM

FIG 22A

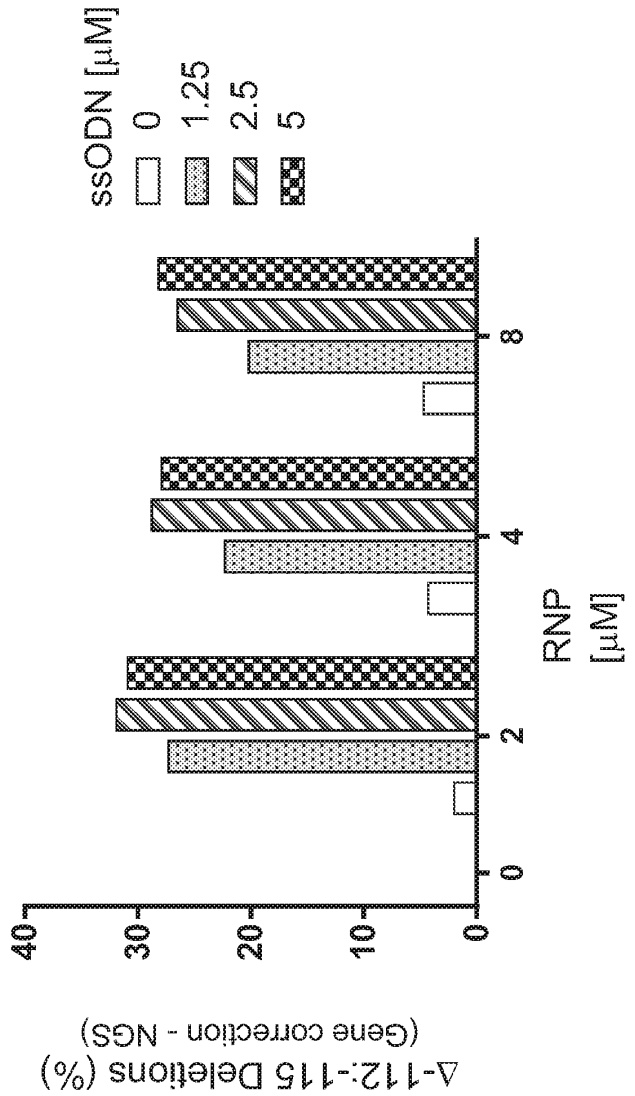

ssODN [µM]
0
1.25
2.5
5
Δ-112:-115 Deletions (%)
(Gene correction - NGS)
40
30
20
10
0
0
2
4
8
RNP
[µM]

FIG 22B

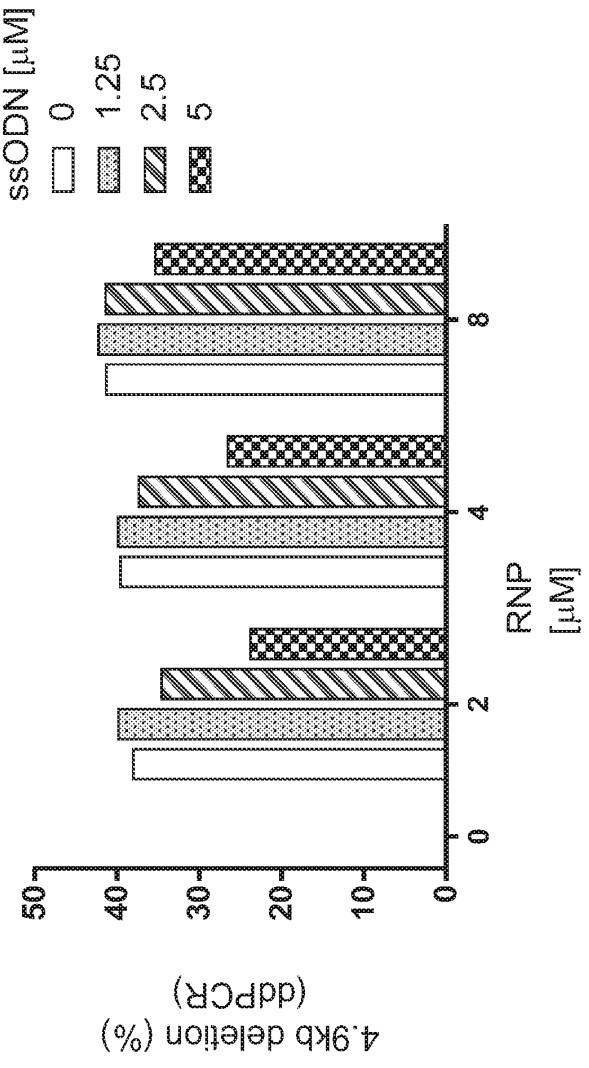

ssODN [μM]
0
1.25
2.5
5
4.9kb deletion (%)
(ddPCR)
50
40
30
20
10
0
0
2
4
8
RNP
[μM]

FIG 22C

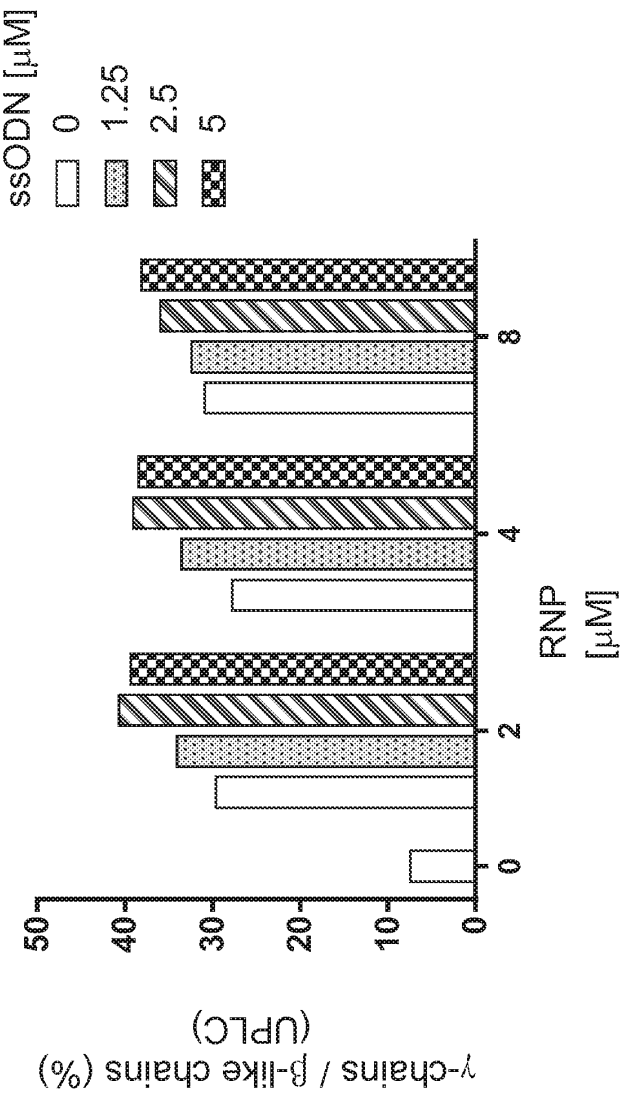

ssODN [μM]
0
1.25
2.5
5
γ-chains / β-like chains (%)
(UPLC)
50
40
30
20
10
0
0
2
4
8
RNP
[μM]

FIG 22D

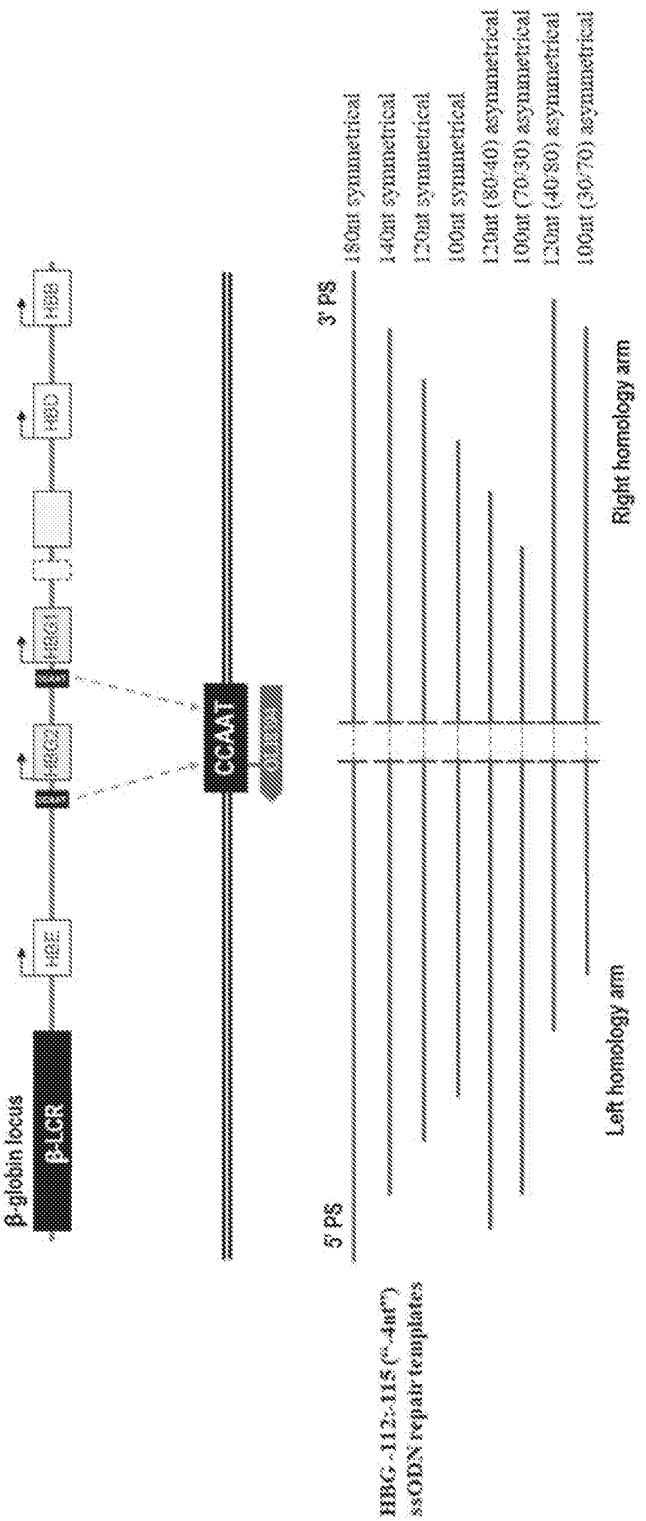

β-globin locus
LCR
HBE
HBG2
HBG1
HBD
HBB
CCAAT
5' PS
3' PS
180nt symmetrical
140nt symmetrical
120nt symmetrical
100nt symmetrical
120nt (60/40) asymmetrical
100nt (70/30) asymmetrical
120nt (40/80) asymmetrical
100nt (30/70) asymmetrical
Right homology arm
Left homology arm
HBG -112:-115 ("-4nt")
ssODN repair templates

FIG 23A

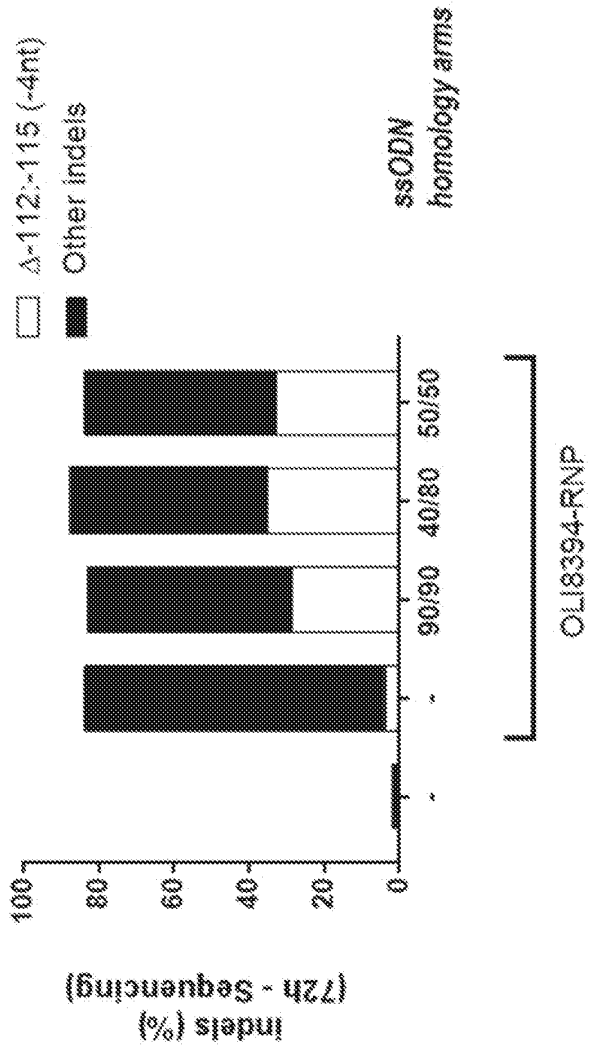

Δ-112:-115 (-4nt)
Other indels
ssODN
homology arms
Indels (%)
(72h - Sequencing)
100
80
60
40
20
0
-
-
90/90
40/80
50/50
OLI8394-RNP

FIG 23B

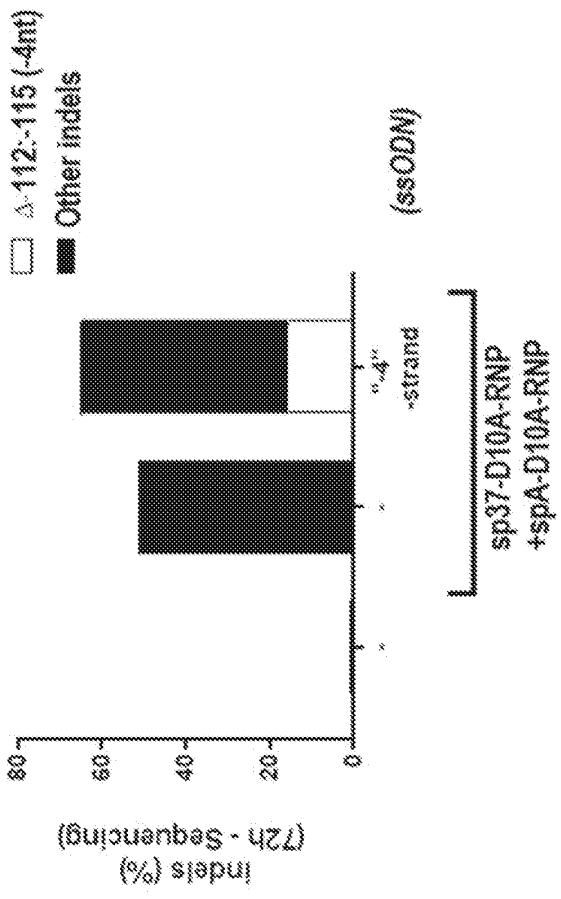
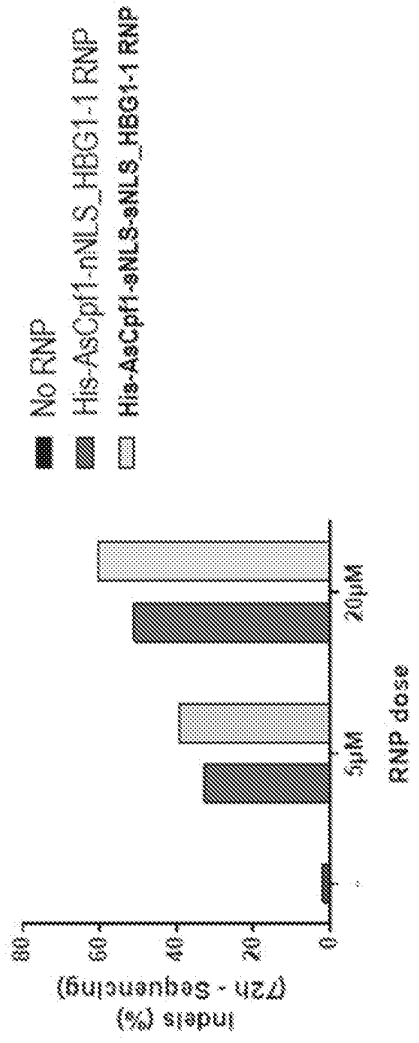

Indels (%)
(72h - Sequencing)
0
20
40
60
80
sp37-D10A-RNP
+spA-D10A-RNP
"-4"
-strand
(ssODN)
Other indels
Δ-112:-115 (-4nt)

FIG 24

FIG 25


No RNP
His-AsCpf1-nNLS_HBG1-1 RNP
His-AsCpf1-aNLS-aNLS_HBG1-1 RNP
Indels (%)
(72h - sequencing)
0
20
40
60
80
5μM
20μM
RNP dose

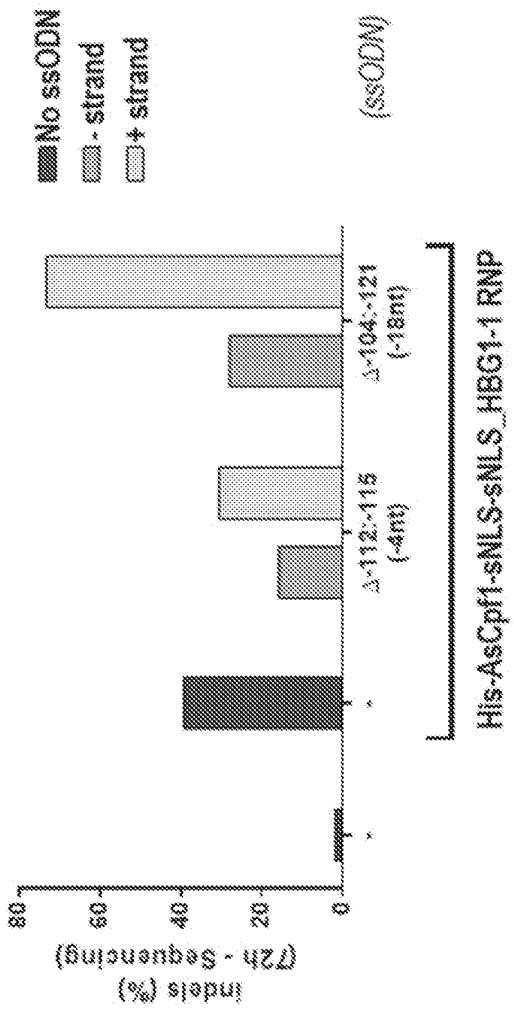

No ssODN
- strand
+ strand
(ssODN)
Δ-104:-121 (-18nt)
Δ-112:-116 (-4nt)
His-AsCpf1-sNLS-sNLS_HBG1-1 RNP
Indels (%) (72h - Sequencing)
0
20
40
60
80

FIG 26A

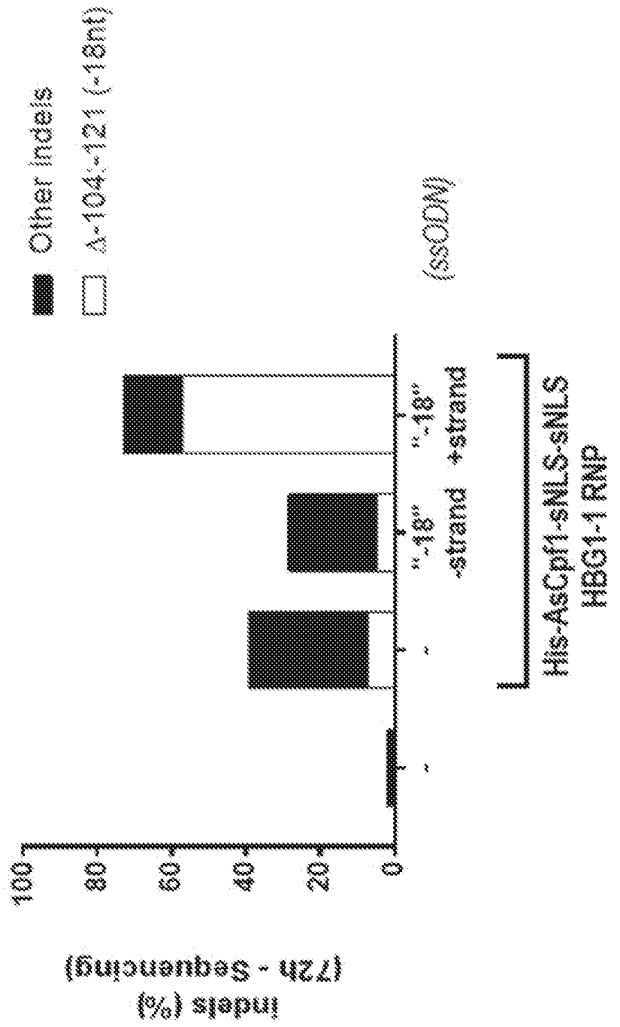

Other indels
Δ-104:-121 (-18nt)
Indels (%)
(72h - Sequencing)
100
80
60
40
20
0
-
-
"-18"
-strand
"-18"
+strand
(ssODN)
His-AsCpf1-sNLS-sNLS
HBG1-1 RNP

FIG 26B

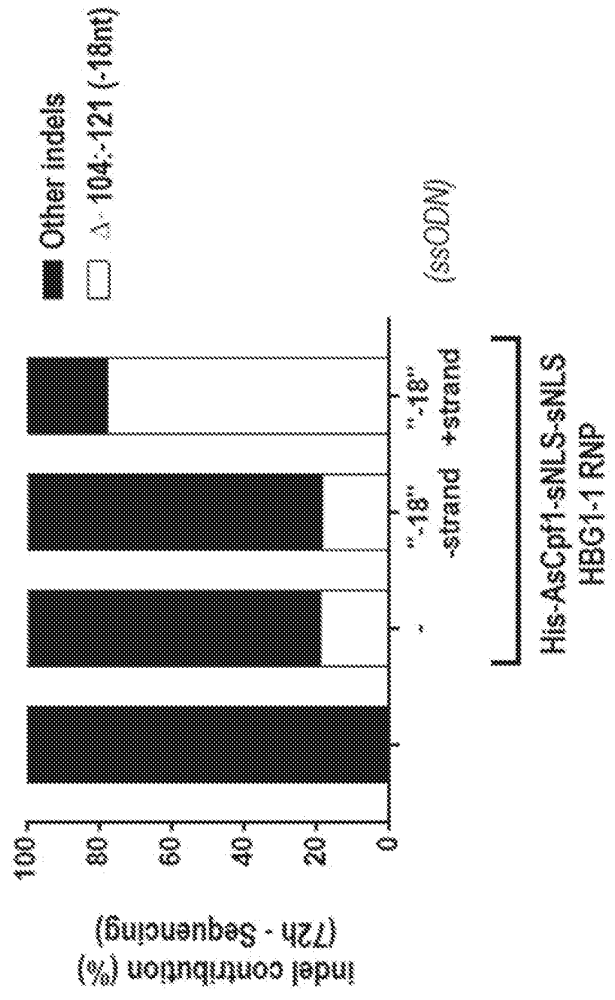

Other indels
Δ -104:-121 (-18nt)
Indel contribution (%)
(72h - Sequencing)
100
80
60
40
20
0
-
"-18" -strand
"-18" +strand
(ssODN)
His-AsCpf1-sNLS-sNLS
HBG1-1 RNP

FIG 26C

FIG 27A

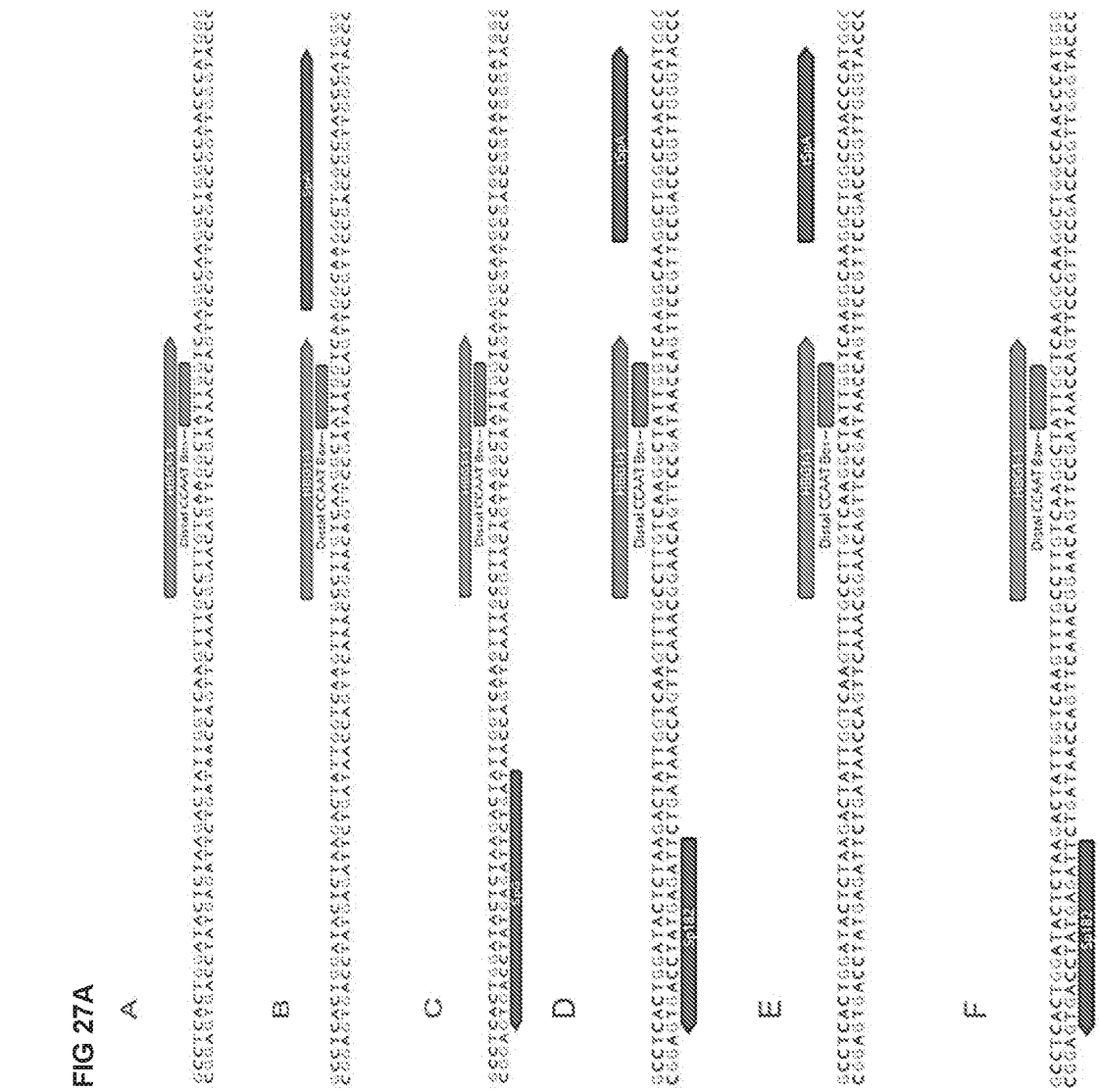


FIG 28A

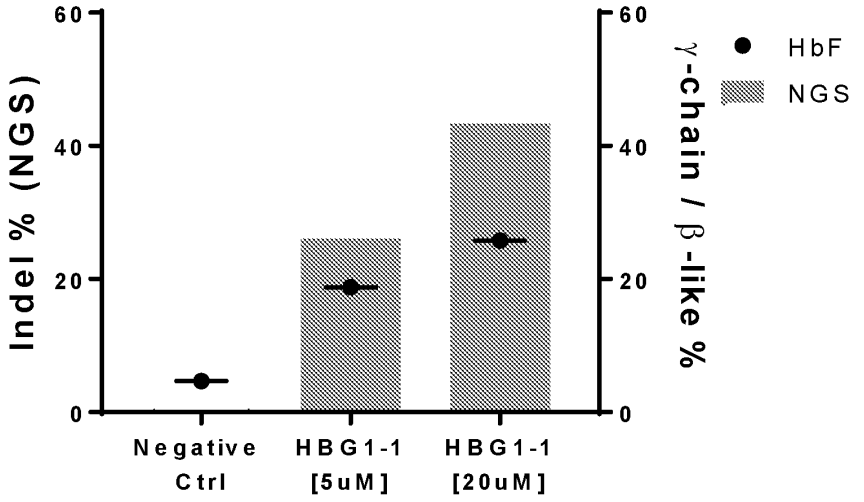
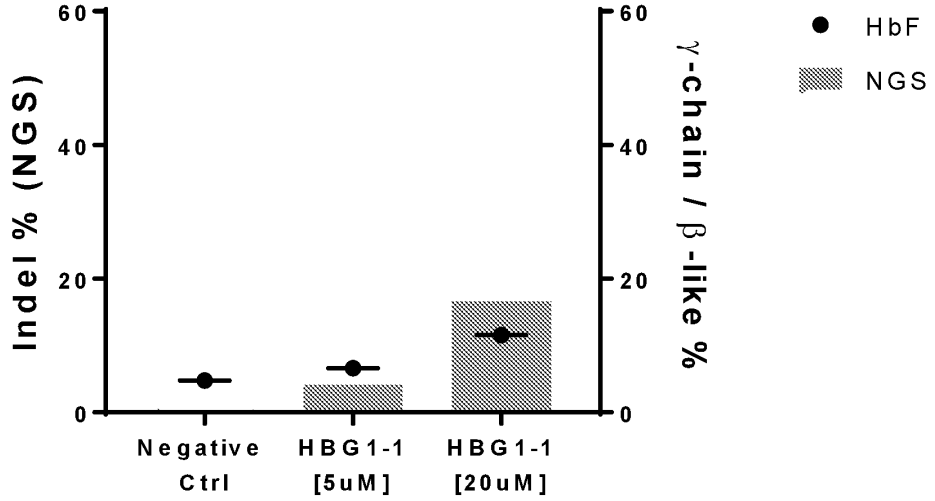


FIG 28B



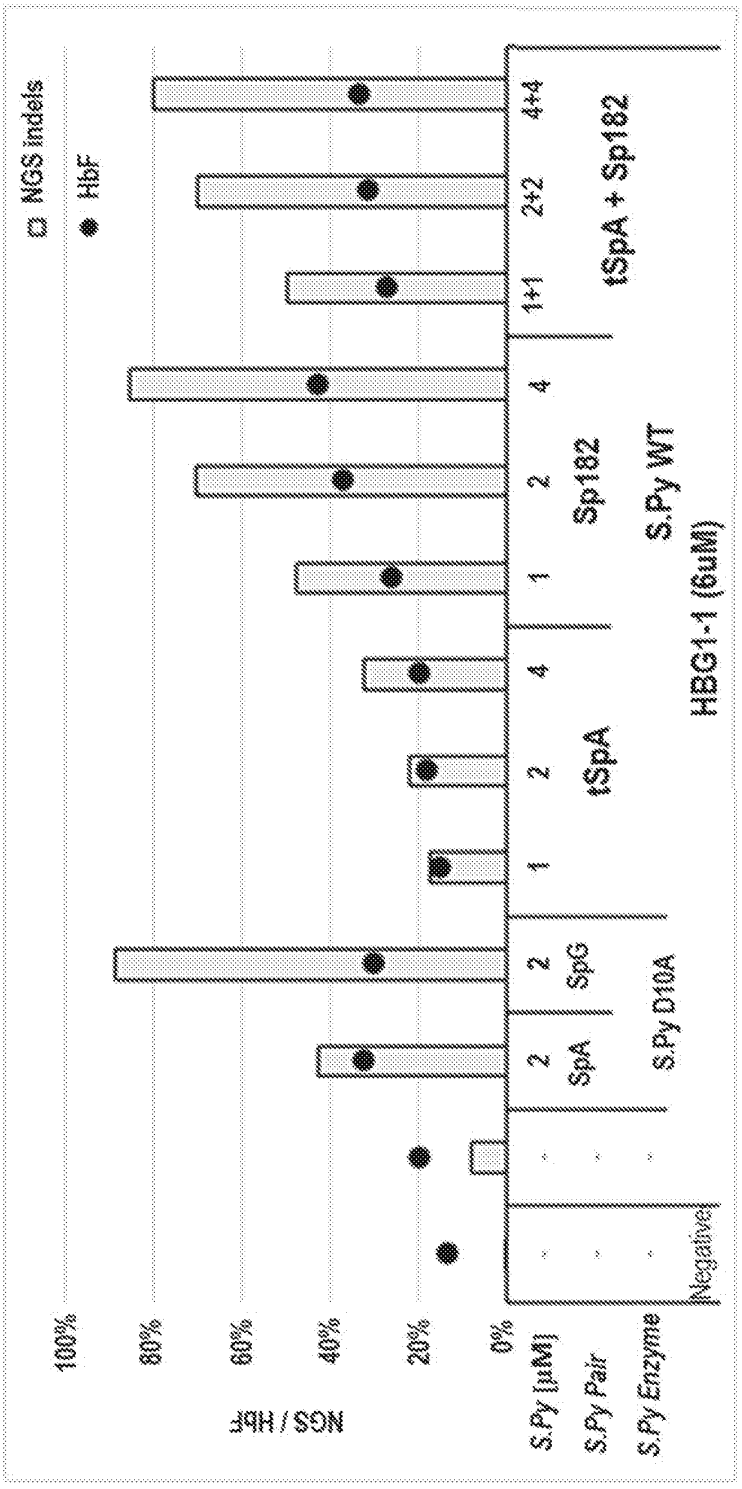

NGS indels
HbF
NGS / HbF
100%
80%
60%
40%
20%
0%
S.Py [uM]
S.Py Pair
S.Py Enzyme
Negative
2
SpA
2
SpG
S.Py D10A
1
2
4
tSpA
1
2
4
Sp182
S.Py WT
1+1
2+2
4+4
tSpA + Sp182
HBG1-1 (6uM)

FIG 29

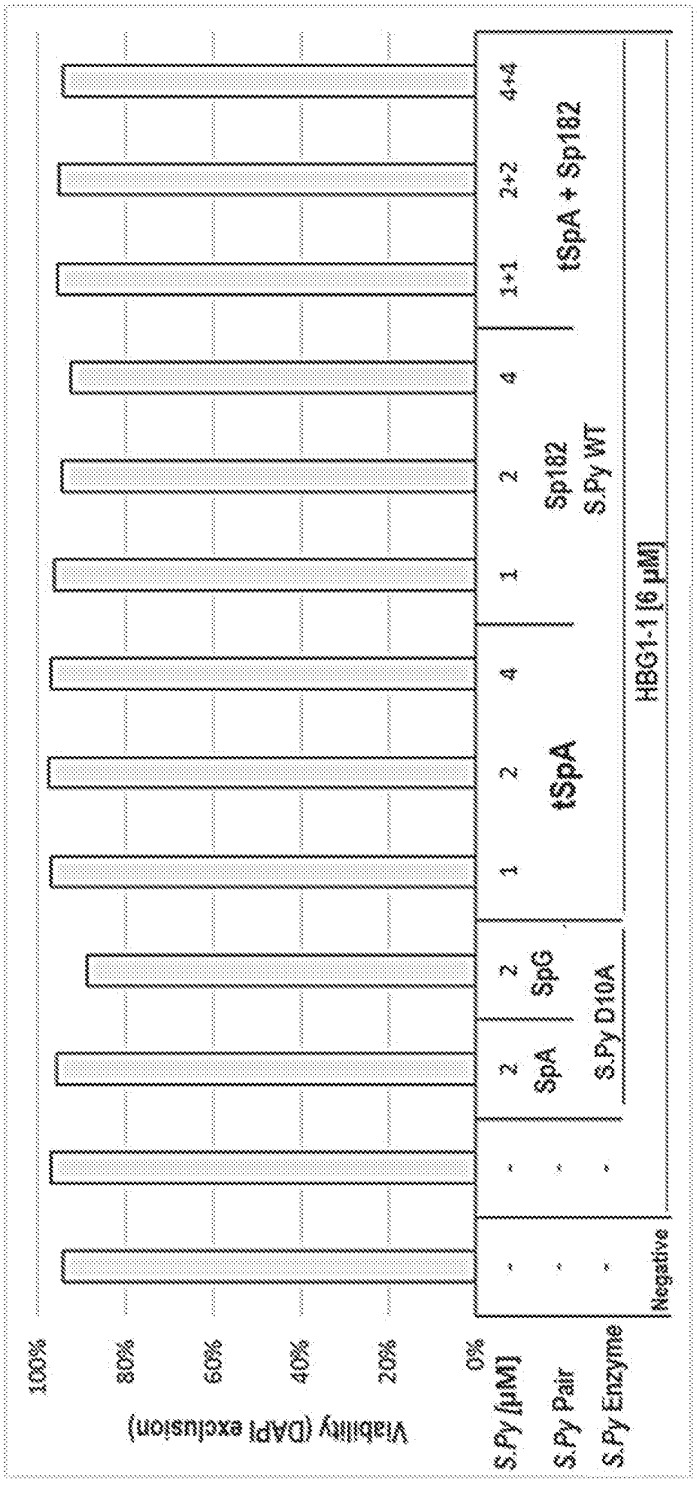

Viability (DAPI exclusion)
100%
80%
60%
40%
20%
0%
S.Py [µM]
S.Py Pair
S.Py Enzyme
-
-
-
Negative
-
-
-
2
SpA
S.Py D10A
2
SpG
1
2
4
tSpA
1
2
4
Sp182
S.Py WT
1+1
2+2
4+4
tSpA + Sp182
HBG1-1 [6 µM]

FIG 30

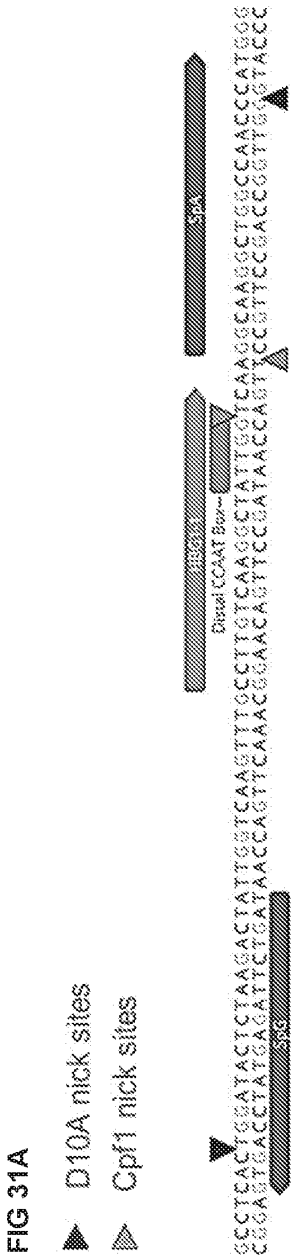


FIG 31A

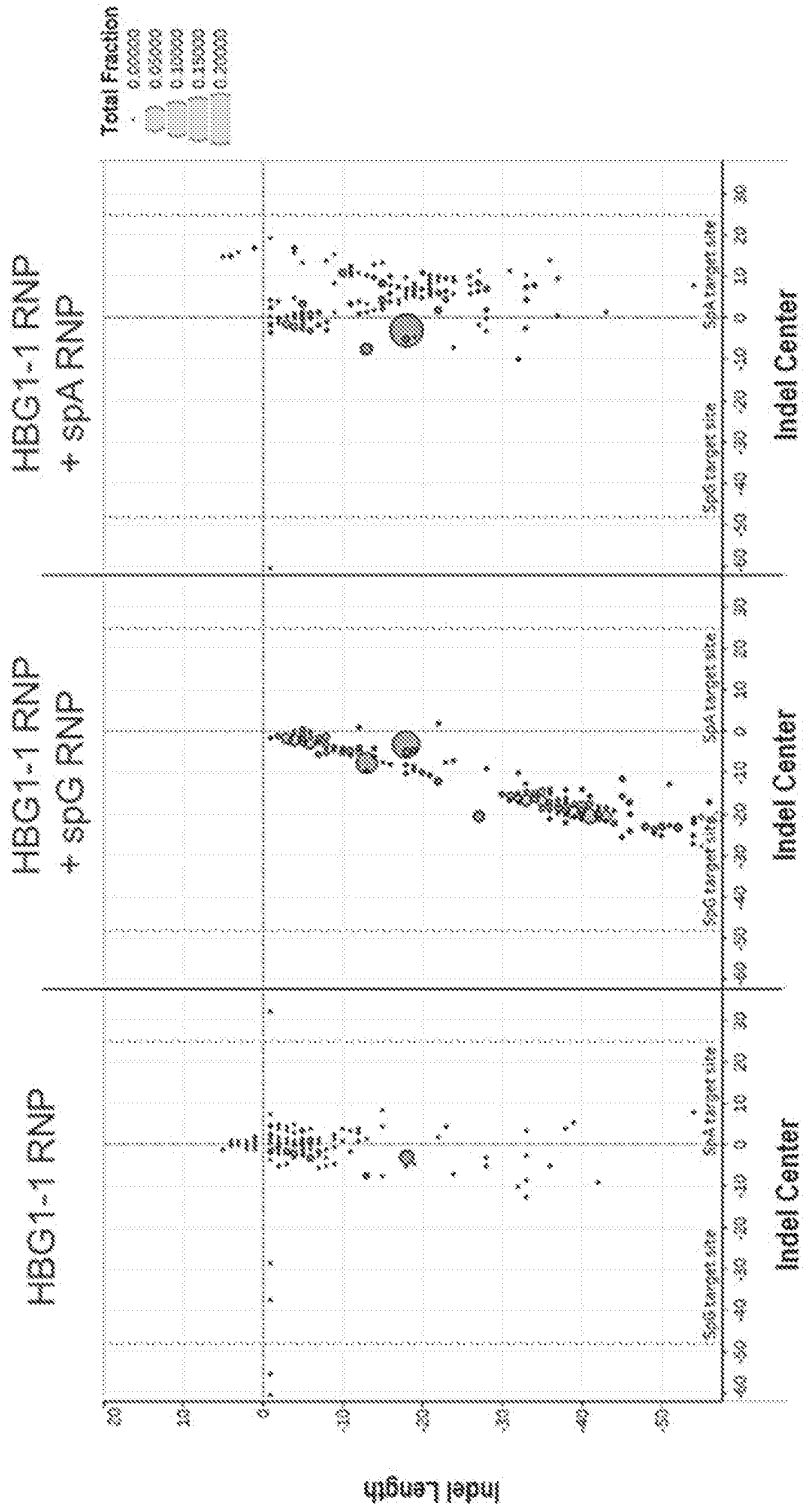

Total Fraction
0.00000
0.05000
0.10000
0.15000
0.20000
HBG1-1 RNP
HBG1-1 RNP + spG RNP
HBG1-1 RNP + spA RNP
SpG target site
SpA target site
Indel Center
Indel Length

FIG 31B

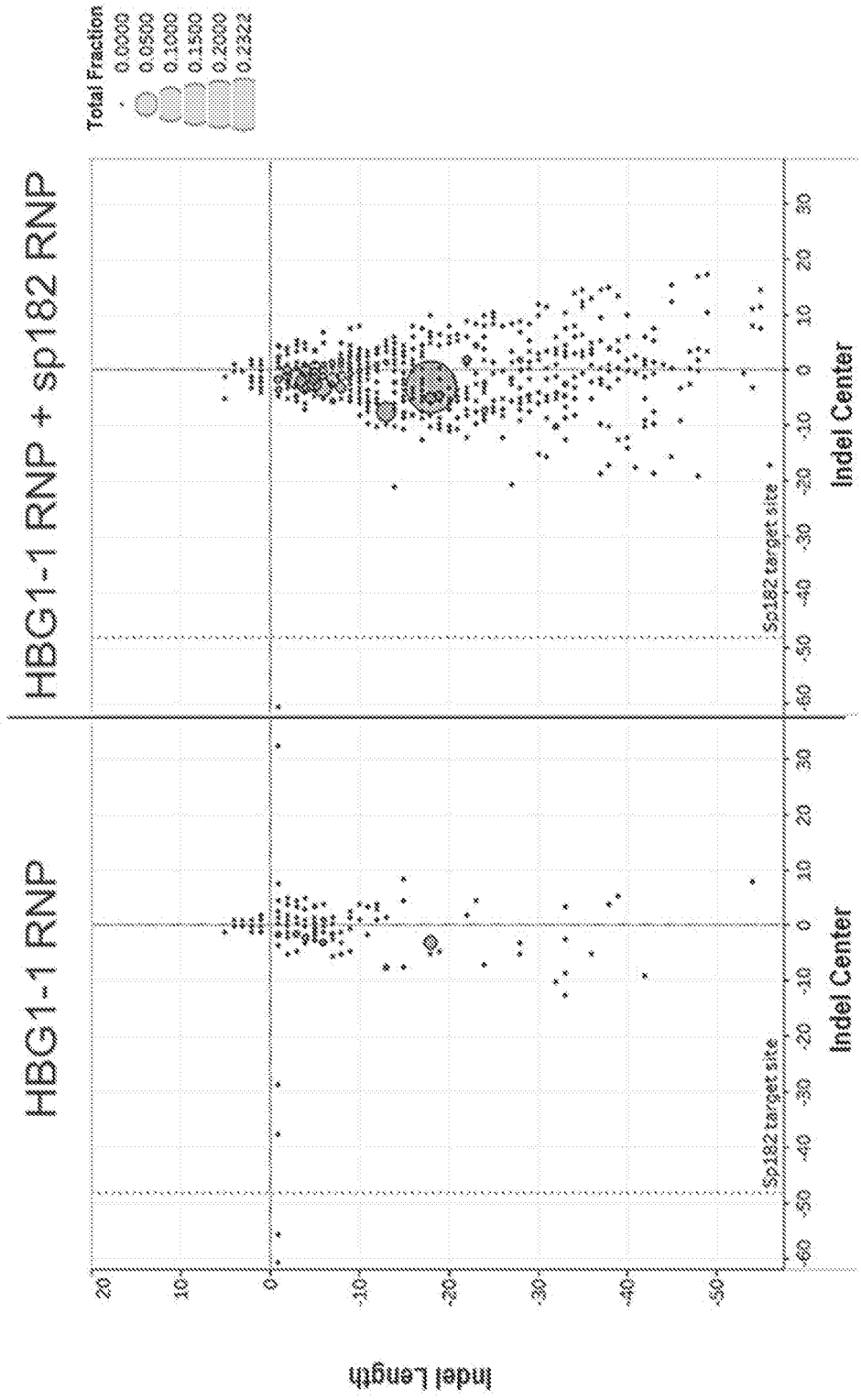
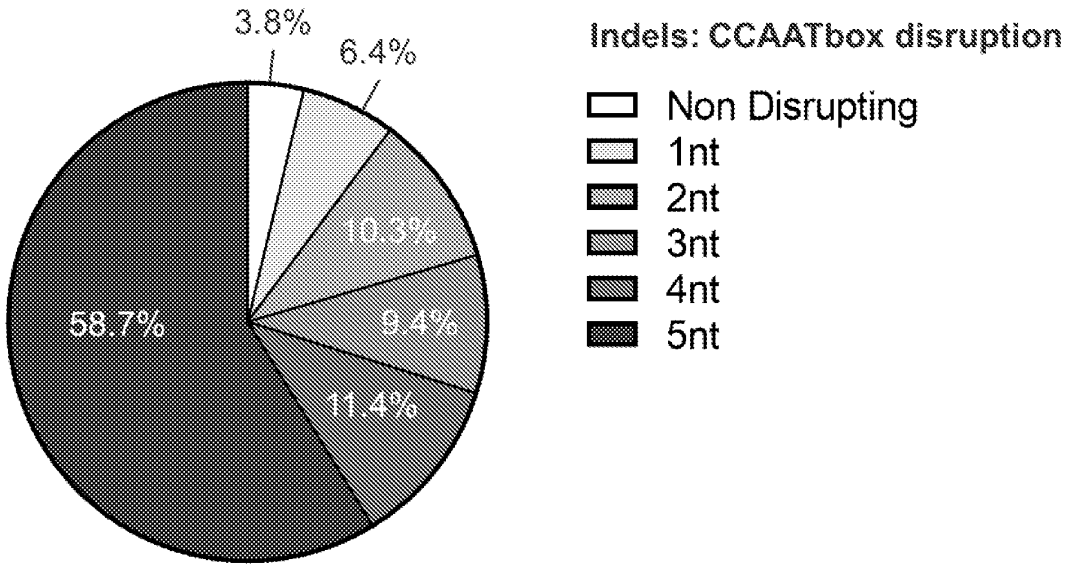

HBG1-1 RNP
HBG1-1 RNP + sp182 RNP
Total Fraction
0.0000
0.0500
0.1000
0.1500
0.2000
0.2322
Indel Length
Indel Center
Sp182 target site

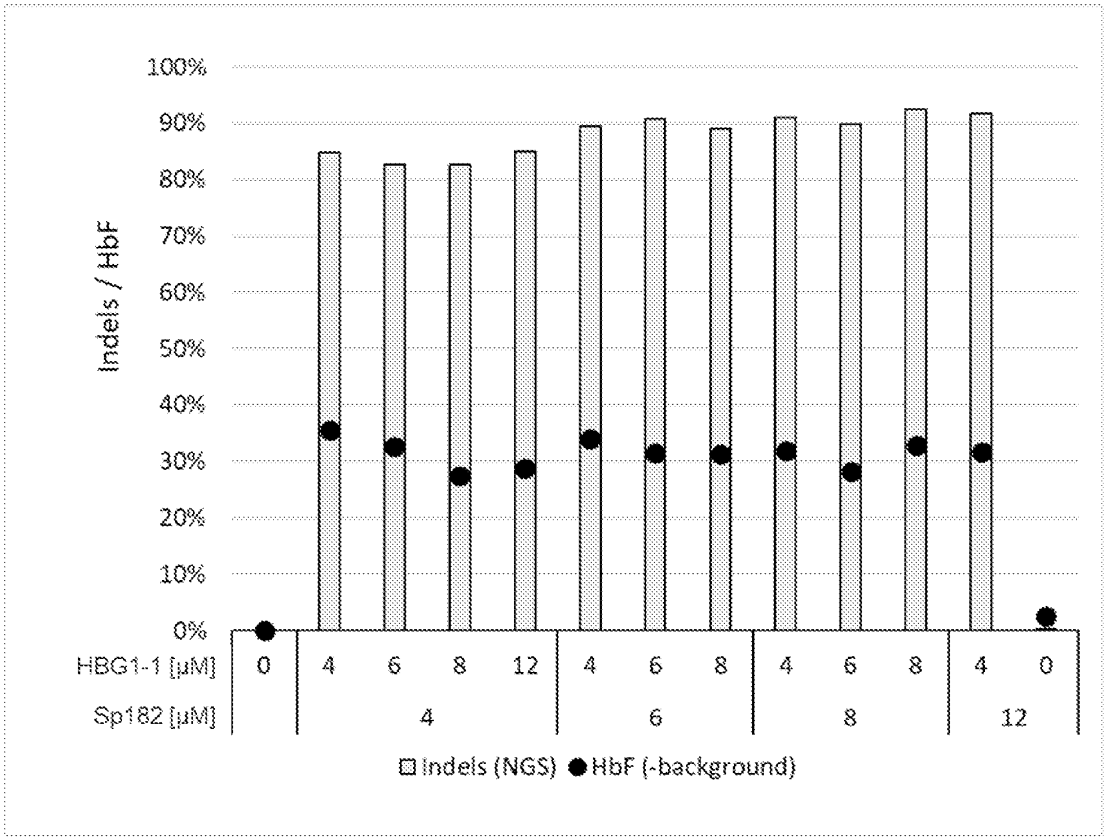
FIG 32A

FIG 32B


3.8%
6.4%
10.3%
9.4%
11.4%
58.7%
Indels: CCAATbox disruption
Non Disrupting
1nt
2nt
3nt
4nt
5nt

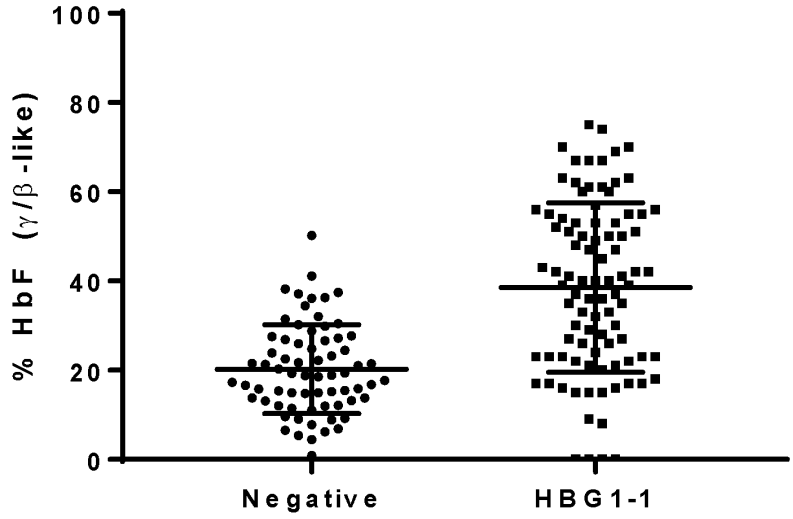
Attorn

FIG 33



Attorn

FIG 34



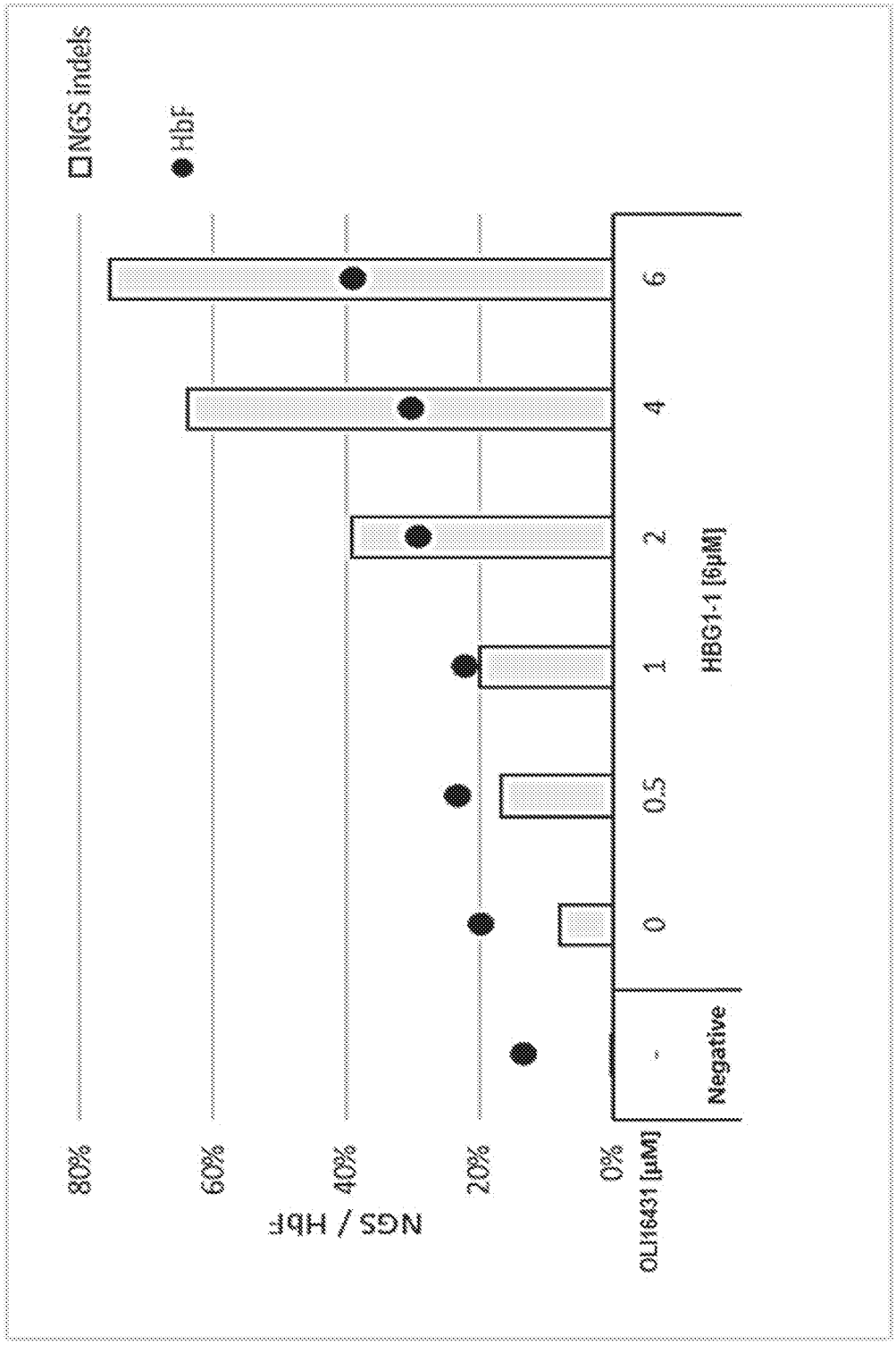

NGS indels
HbF
NGS / HbF
80%
60%
40%
20%
0%
OLI16431 [µM]
-
0
0.5
1
2
4
6
HBG1-1 [6µM]
Negative

FIG 35A

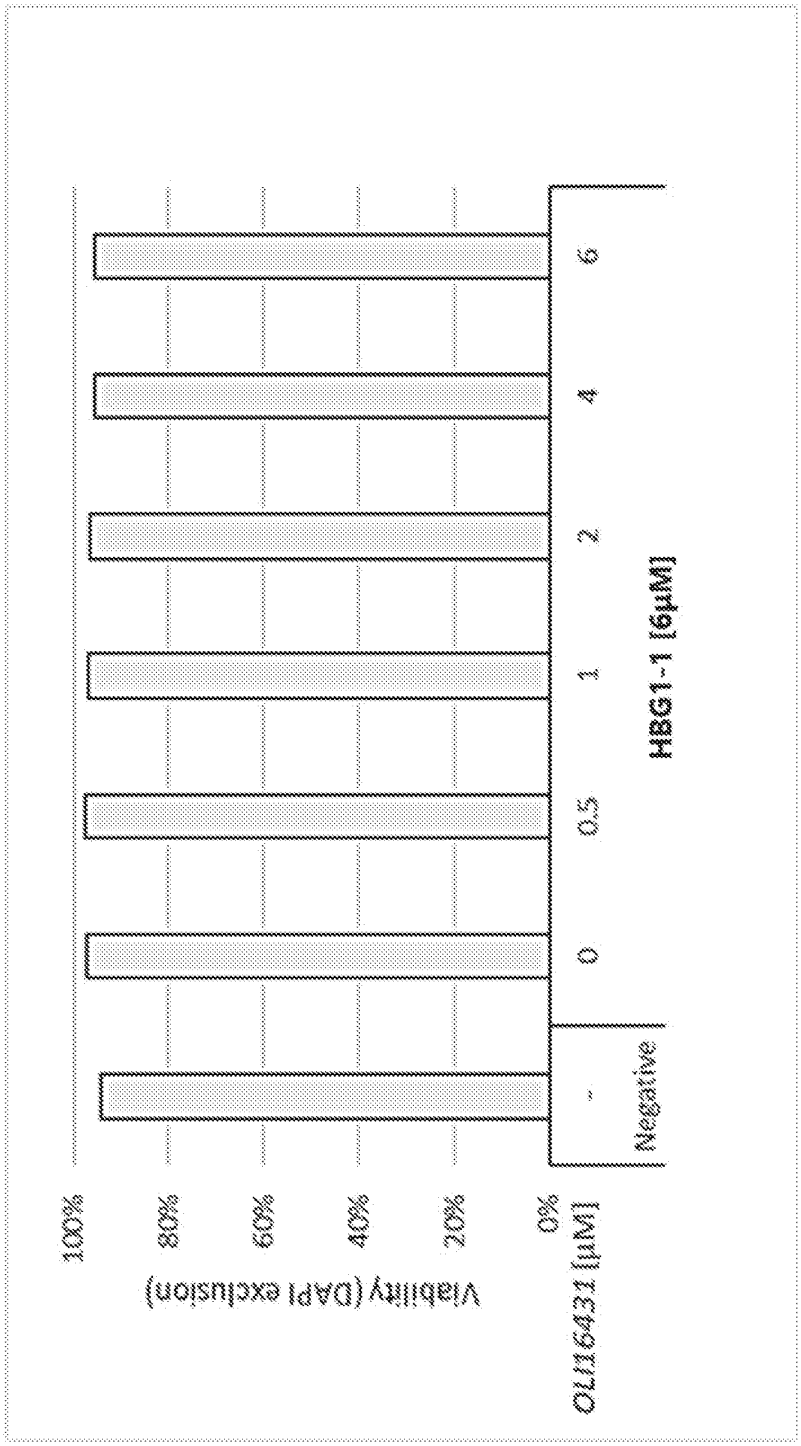

Viability (DAPI exclusion)
100%
80%
60%
40%
20%
0%
OLI16431 [µM]
-
Negative
0
0.5
1
2
4
6
HBG1-1 [6µM]

FIG 35B

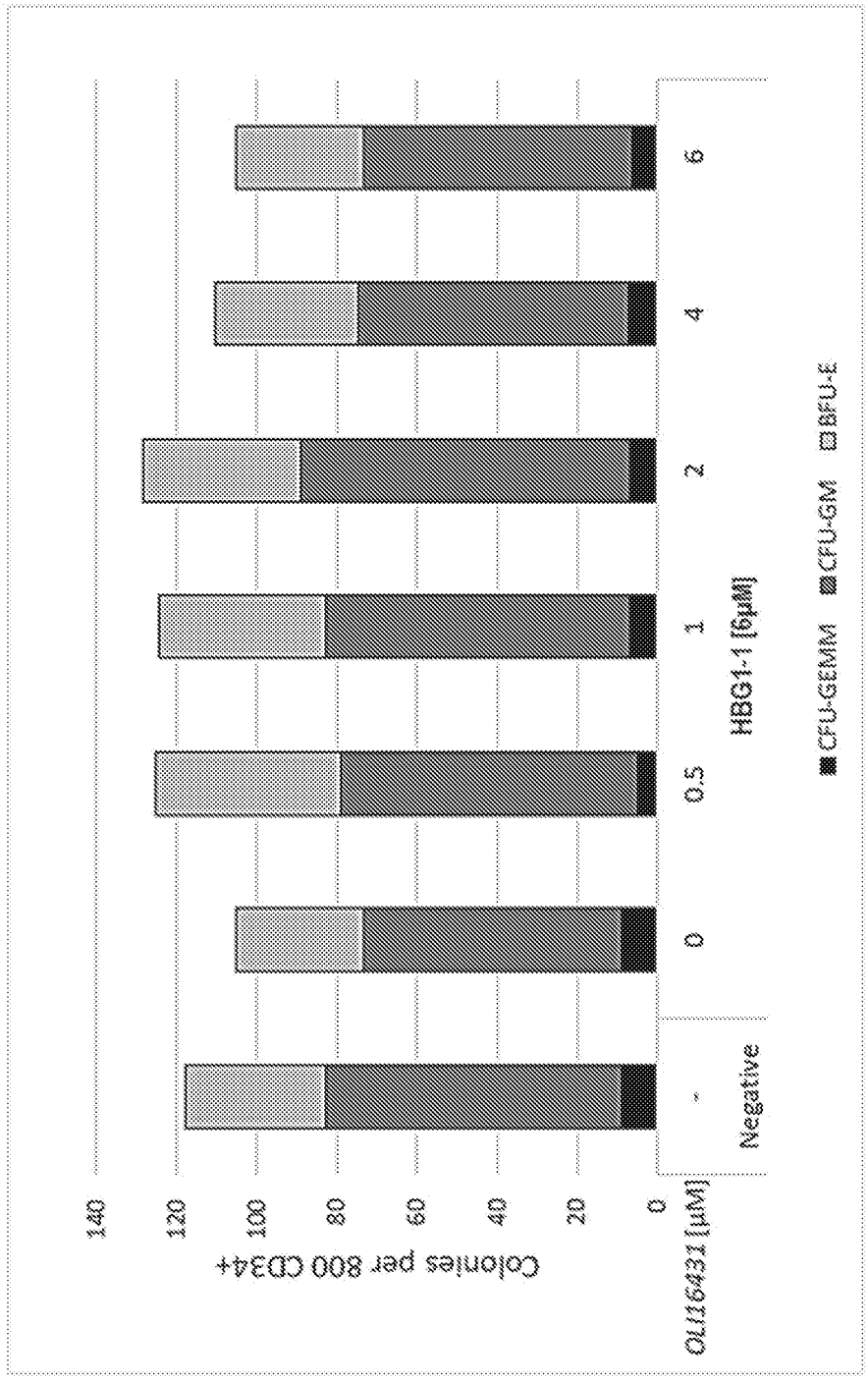
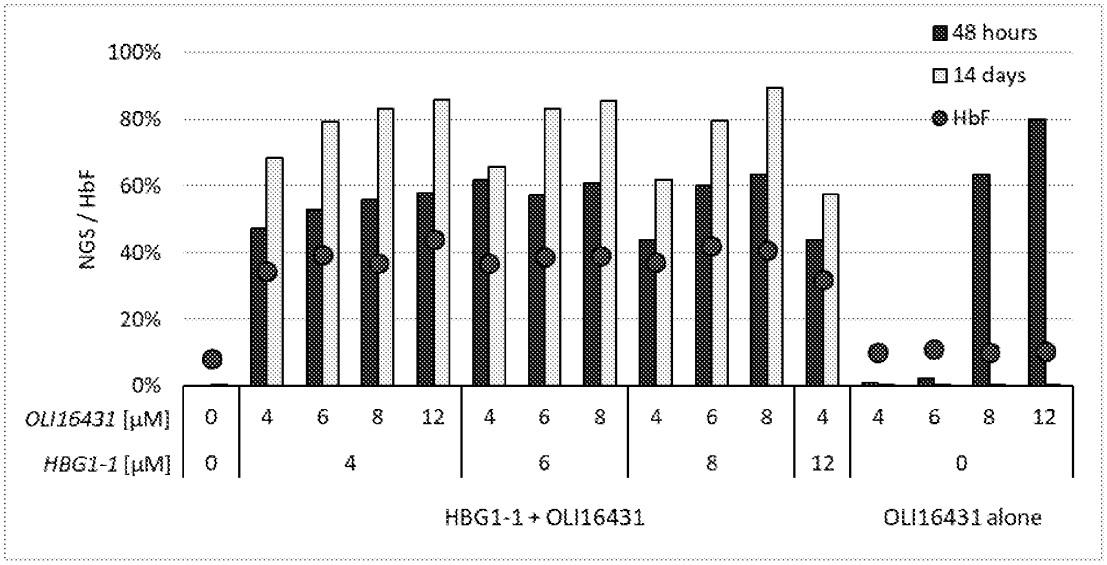

Colonies per 800 CD34+
140
120
100
80
60
40
20
0
OLI16431 [µM]
-
0
0.5
1
2
4
6
Negative
HBG1-1 [6µM]
CFU-GEMM
CFU-GM
BFU-E

FIG 35C

Attorn

FIG 36A



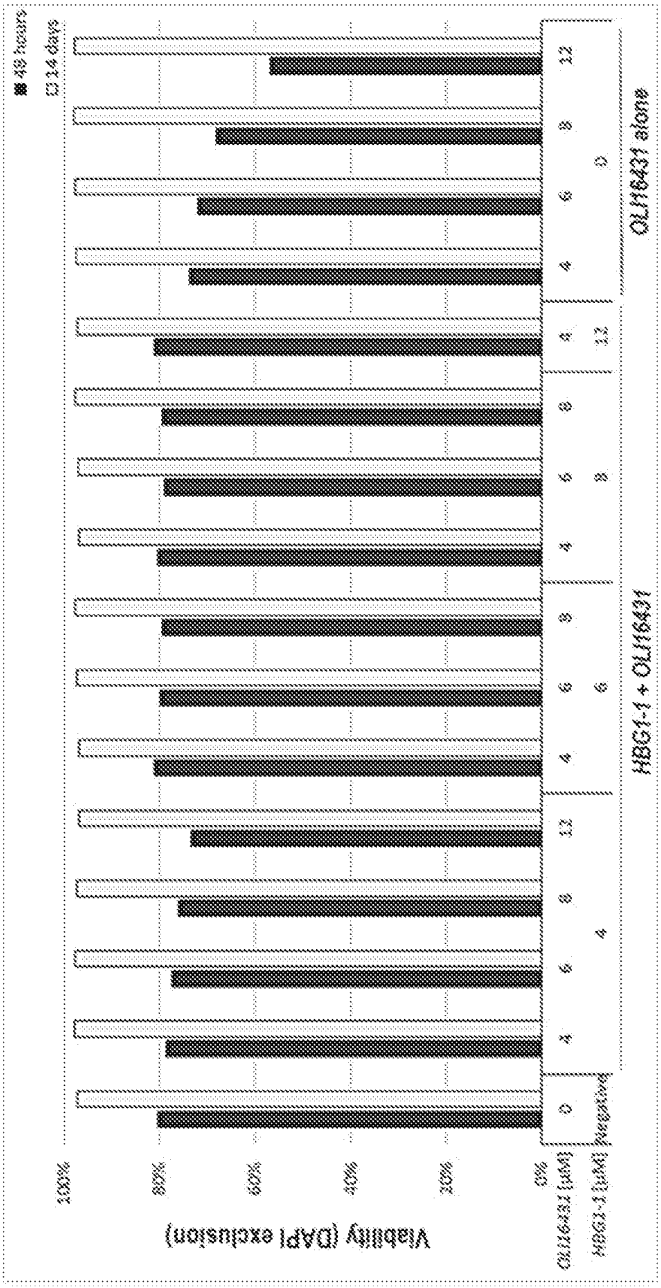
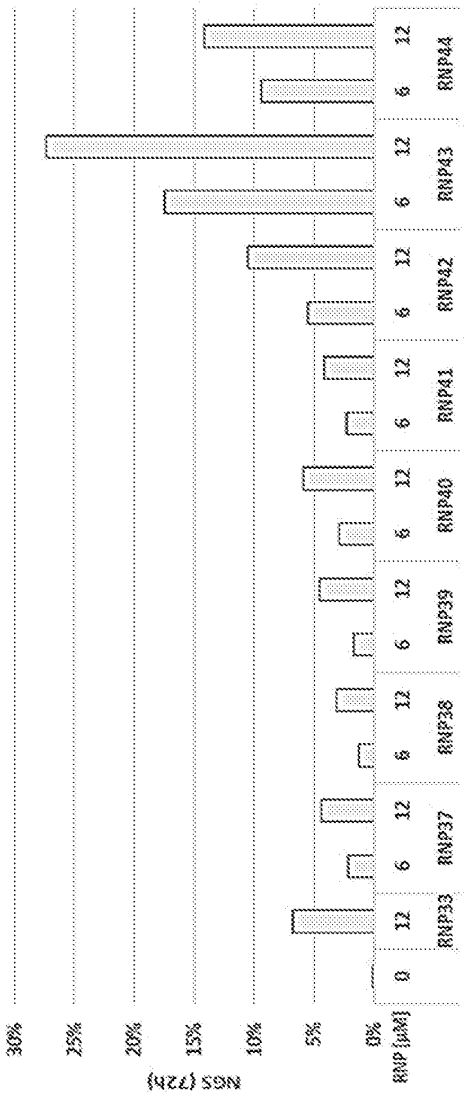

48 hours
14 days
Viability (DAPI exclusion)
100%
80%
60%
40%
20%
0%
OLI16431 [μM]
HBG1-1 [μM]
Negative
0
4
6
8
12
4
6
8
12
4
6
8
12
4
6
8
12
HBG1-1 + OLI16431
OLI16431 alone

FIG 36B

FIG 37


NGS (72h)
30%
25%
20%
15%
10%
5%
0%
RNP (μM)
0
12
6
12
RNP33
6
12
RNP37
6
12
RNP38
6
12
RNP39
6
12
RNP40
6
12
RNP41
6
12
RNP42
6
12
RNP43
6
12
RNP44

Attorn

FIG 38A

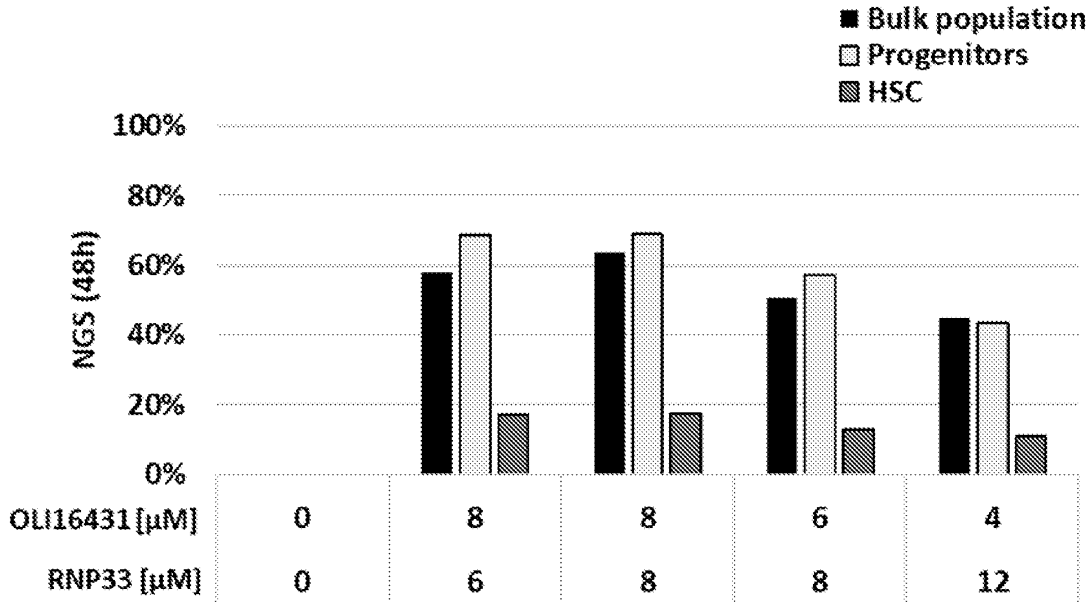


FIG 38B

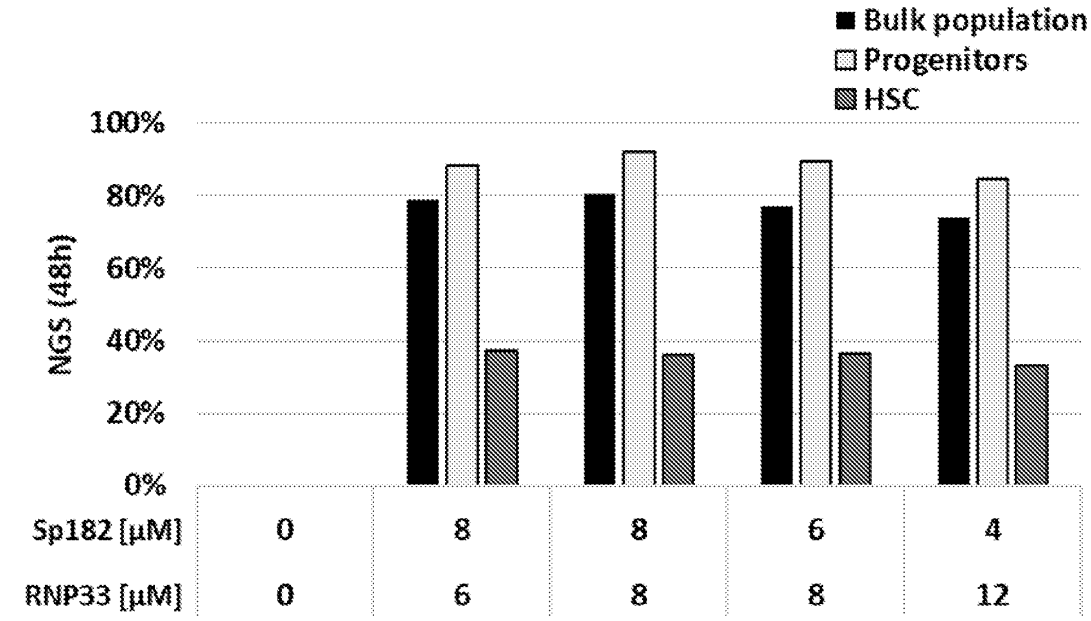
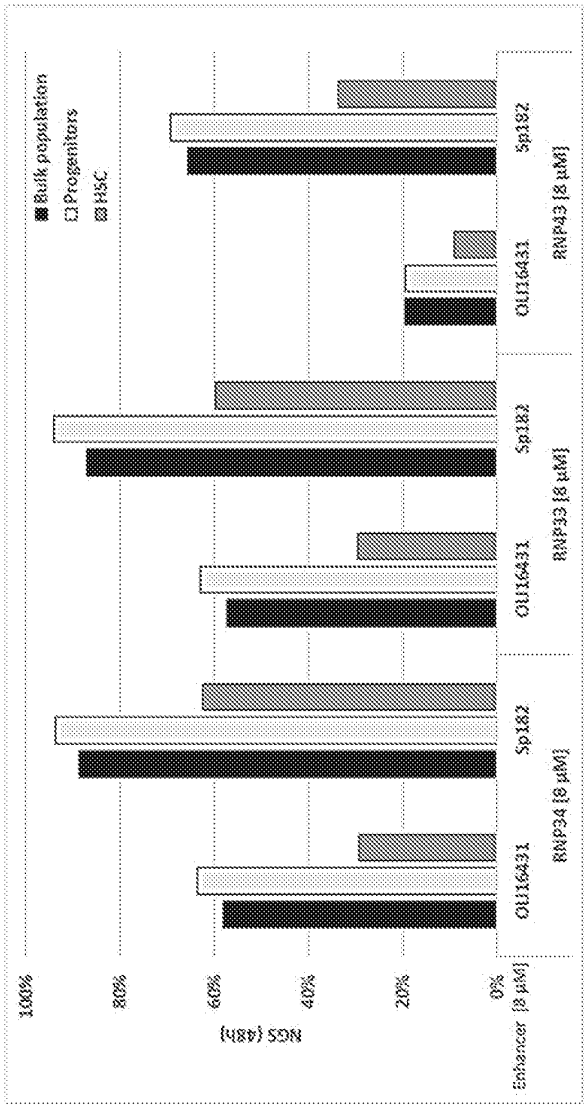


FIG 39


Bulk population
Progenitors
HSC
NGS (abs)
100%
80%
60%
40%
20%
0%
Enhancer [8 µM]
OU16431
Sp182
RNP34 [8 µM]
OU16431
Sp182
RNP33 [8 µM]
OU16431
Sp182
RNP43 [8 µM]

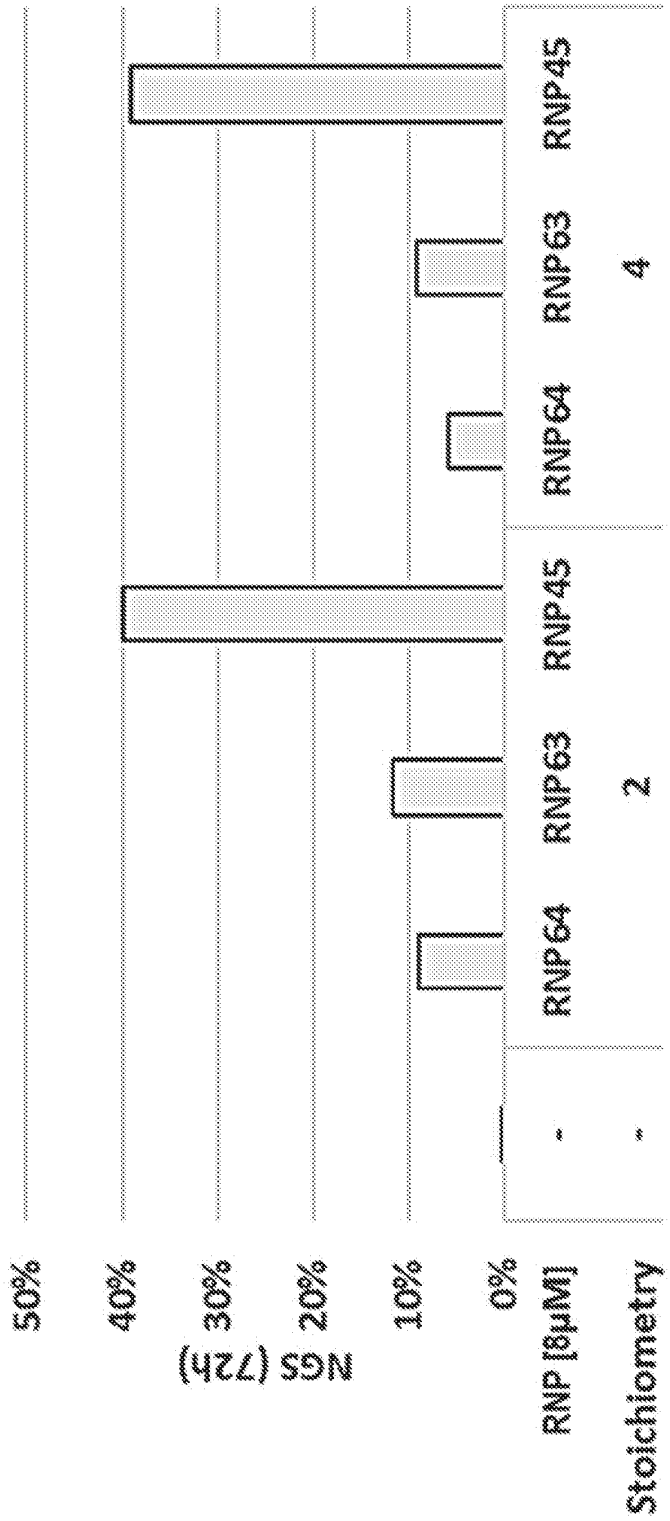
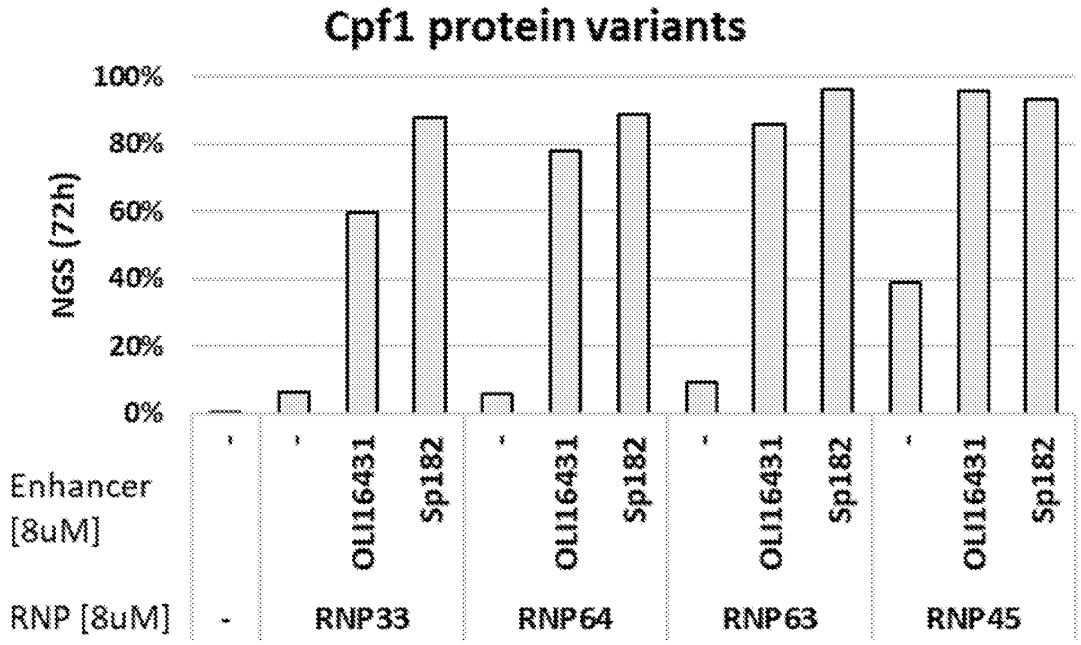

Cpf1 RNP stoichiometry comparison
NGS (72h)
50%
40%
30%
20%
10%
0%
RNP [8µM]
Stoichiometry
-
-
RNP64
RNP63
RNP45
2
RNP64
RNP63
RNP45
4

FIG 40

Attorn

FIG 41



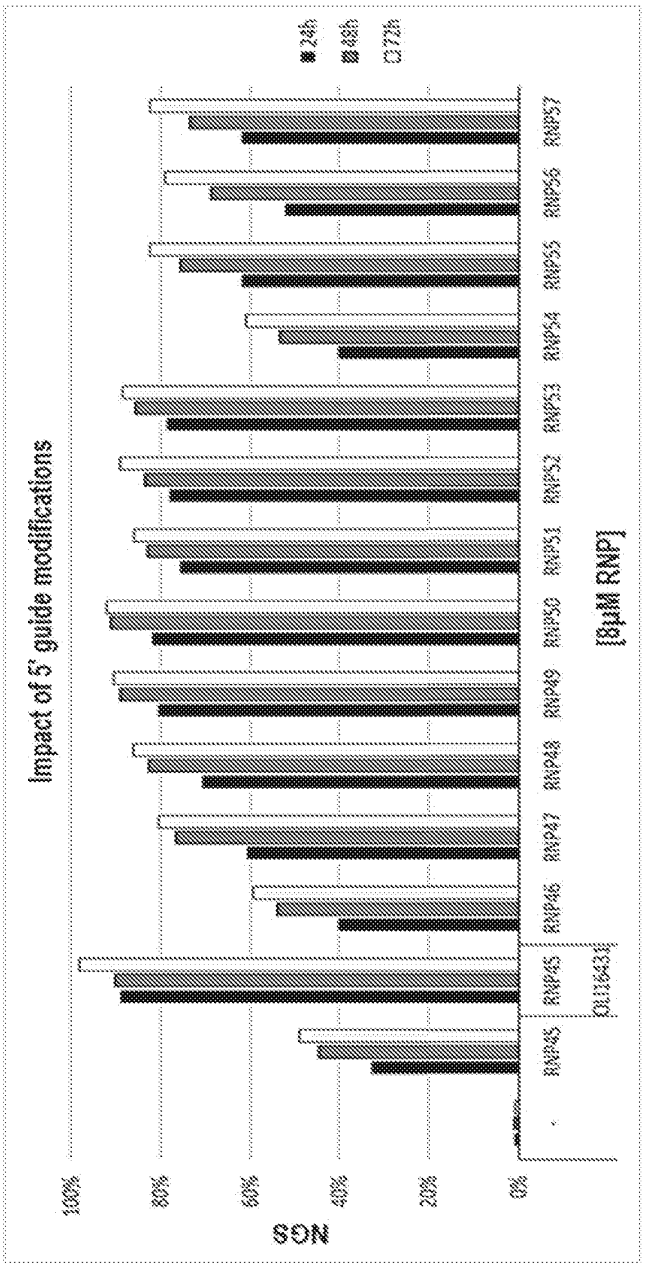
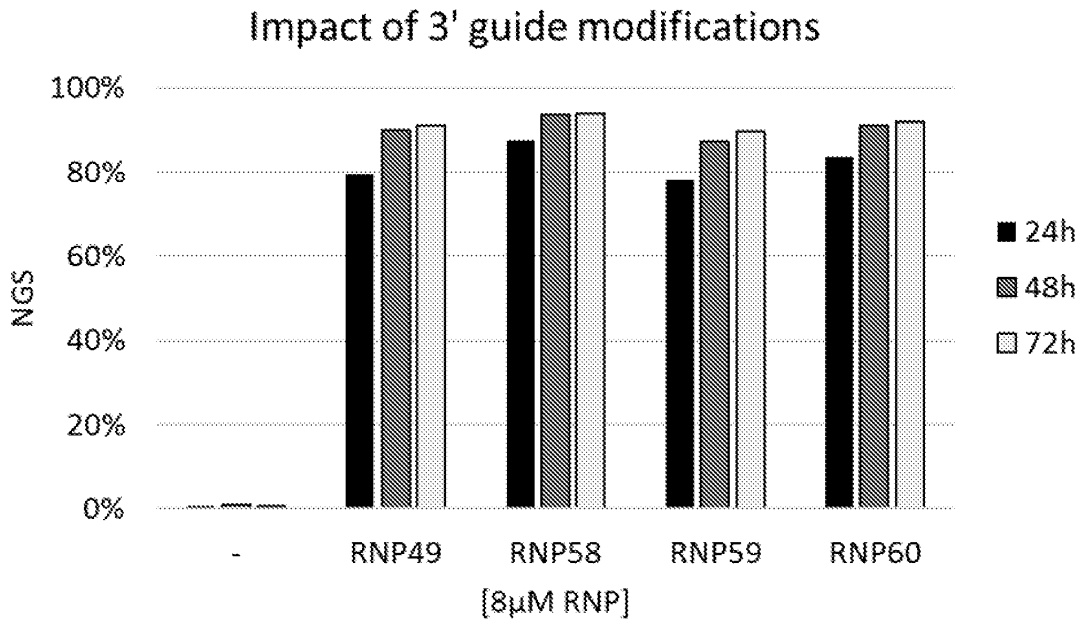

Impact of 5' guide modifications
24h
48h
72h
NGS
100%
80%
60%
40%
20%
0%
RNP57
RNP56
RNP55
RNP54
RNP53
RNP52
RNP51
RNP50
RNP49
RNP48
RNP47
RNP46
RNP45
RNP45
-
[6µM RNP]

FIG 42

Attorn

FIG 43



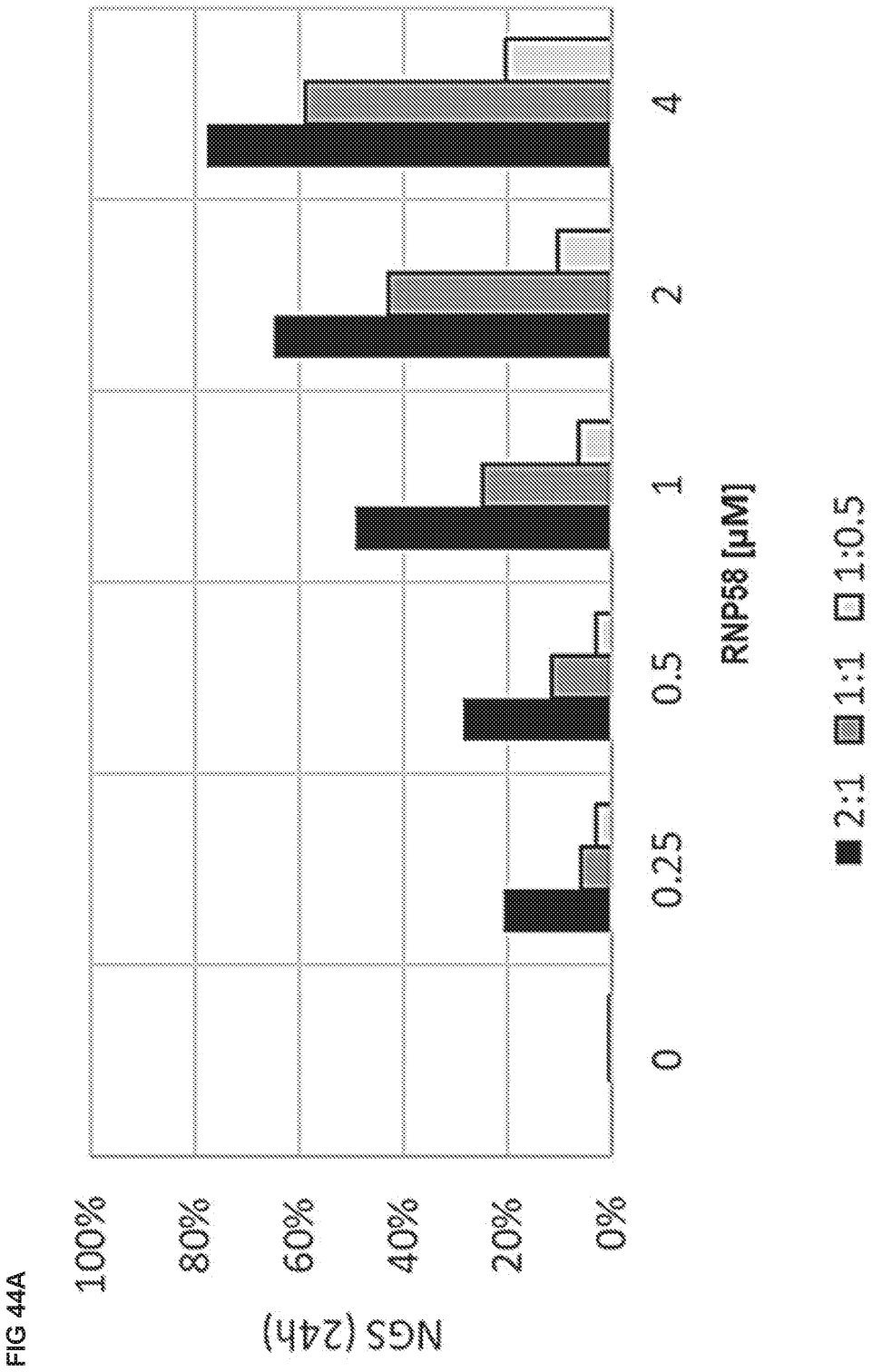


FIG 44A

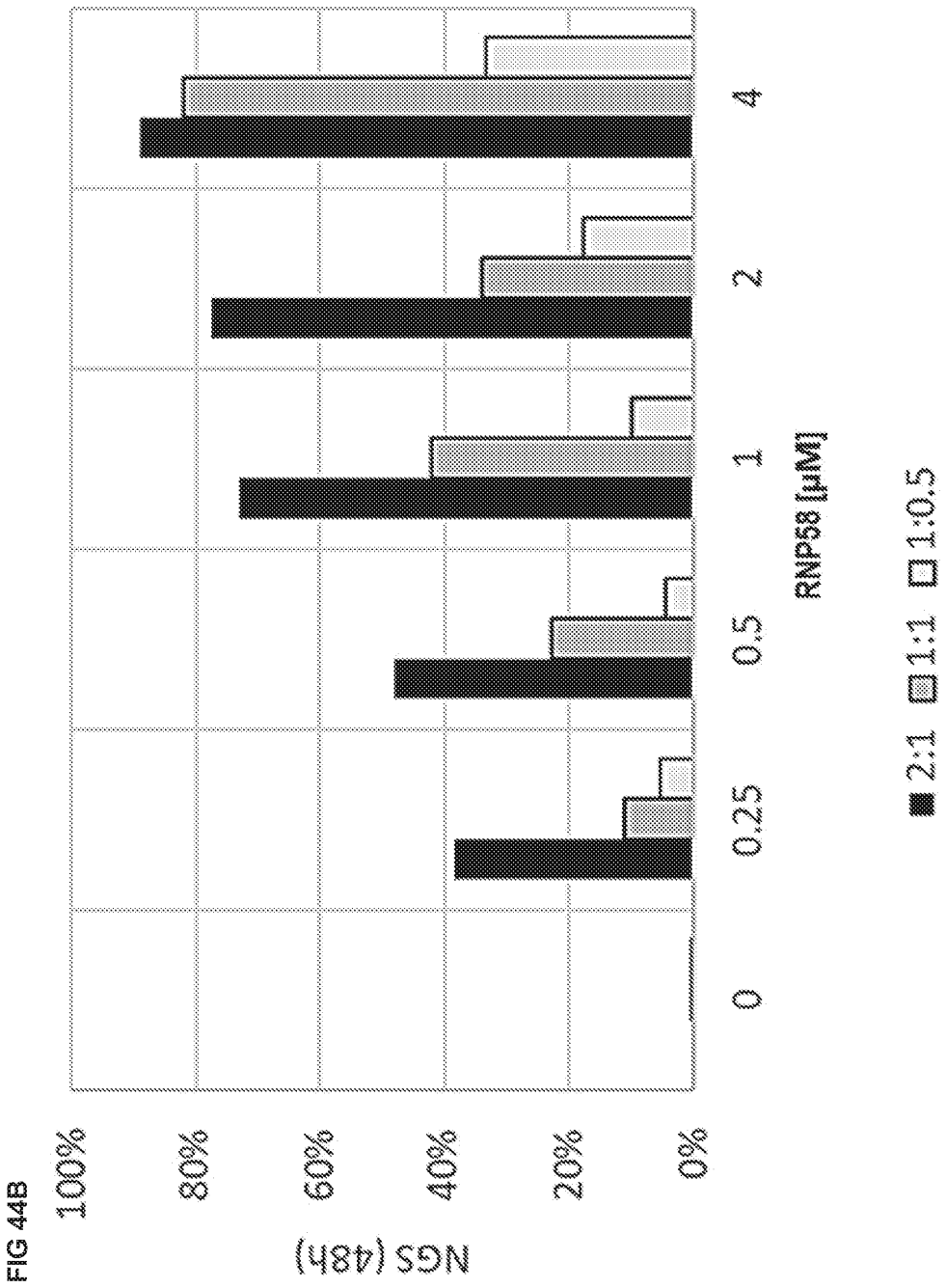


FIG 44B

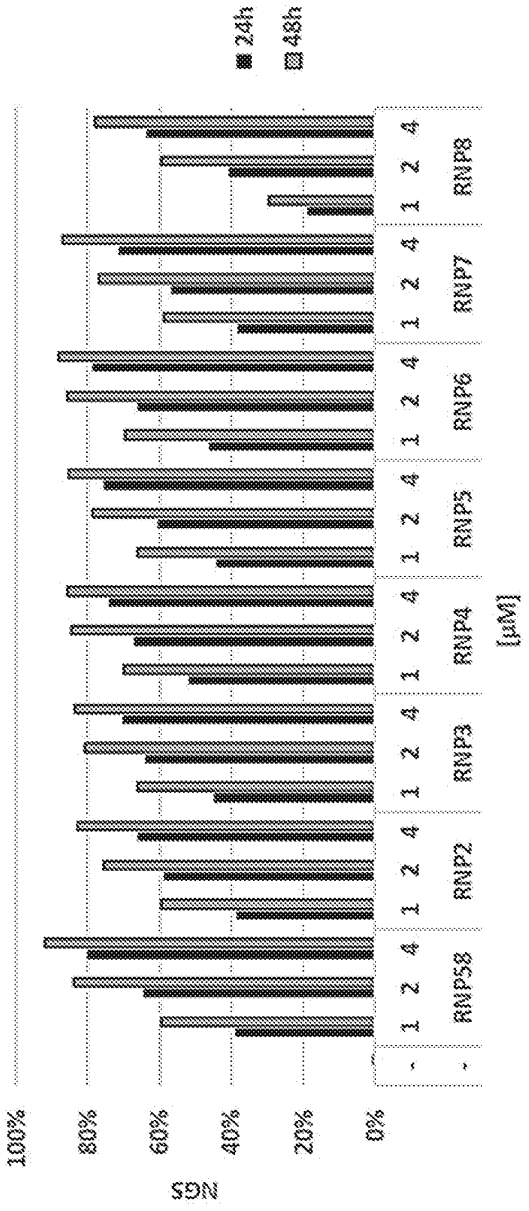

24h
48h
NGS
100%
80%
60%
40%
20%
0%
-
-
RNP58
RNP2
RNP3
RNP4
RNP5
RNP6
RNP7
RNP8
1 2 4
[μM]

FIG 45A

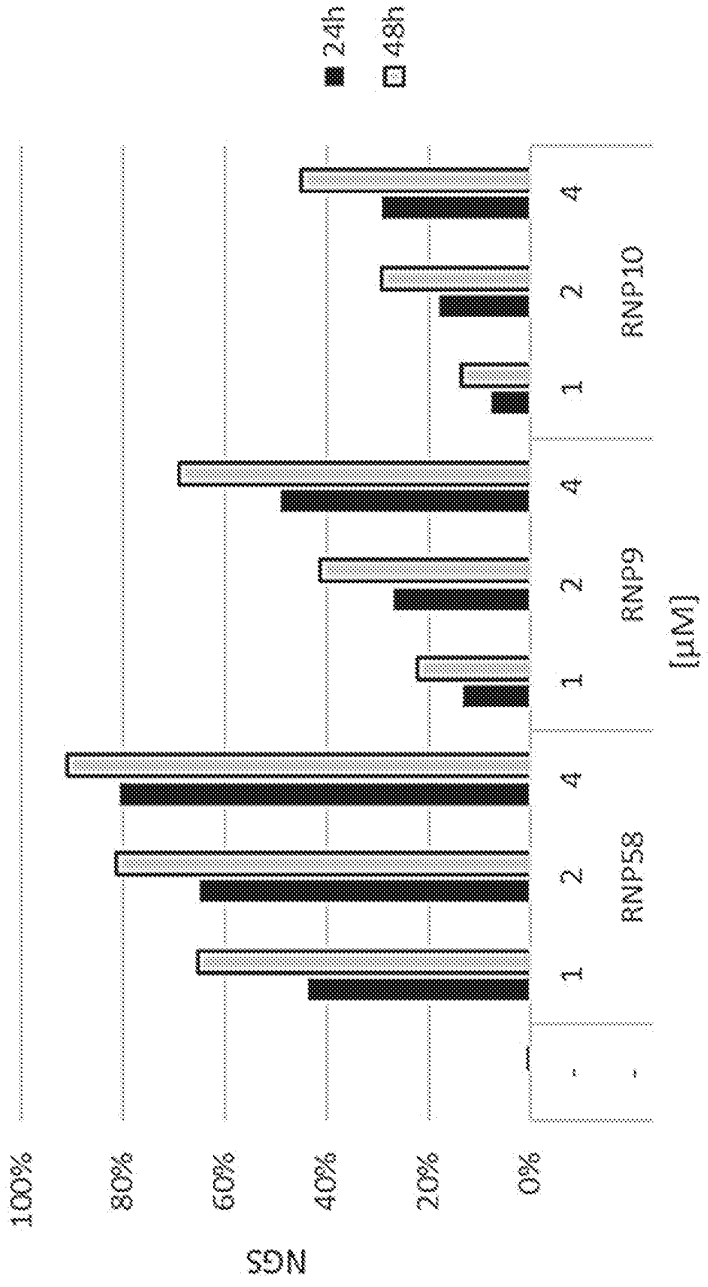


FIG 45B

Attorn

FIG 46A

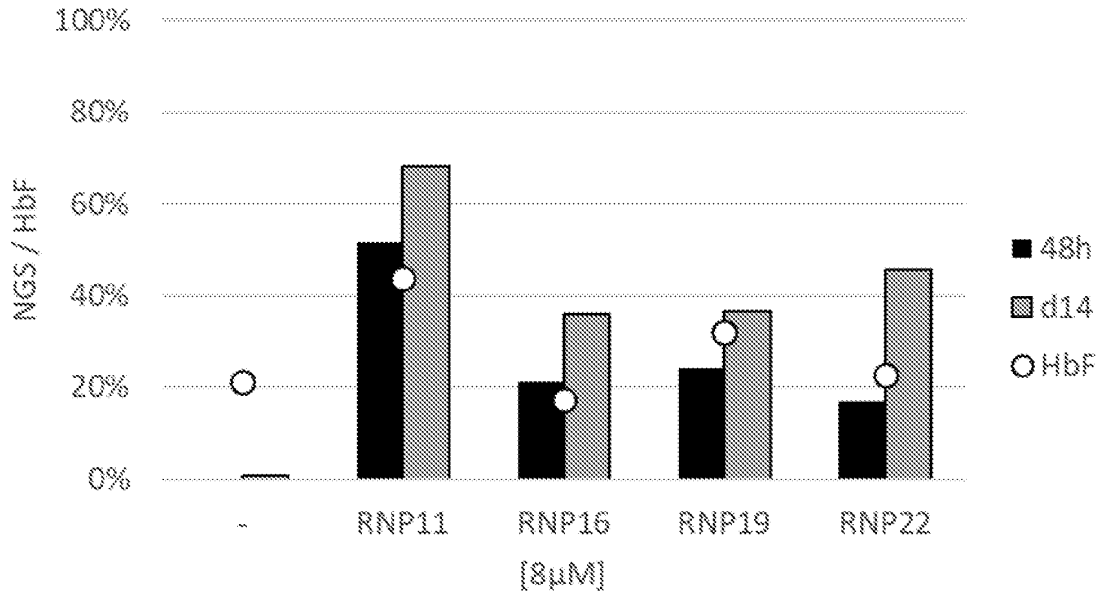
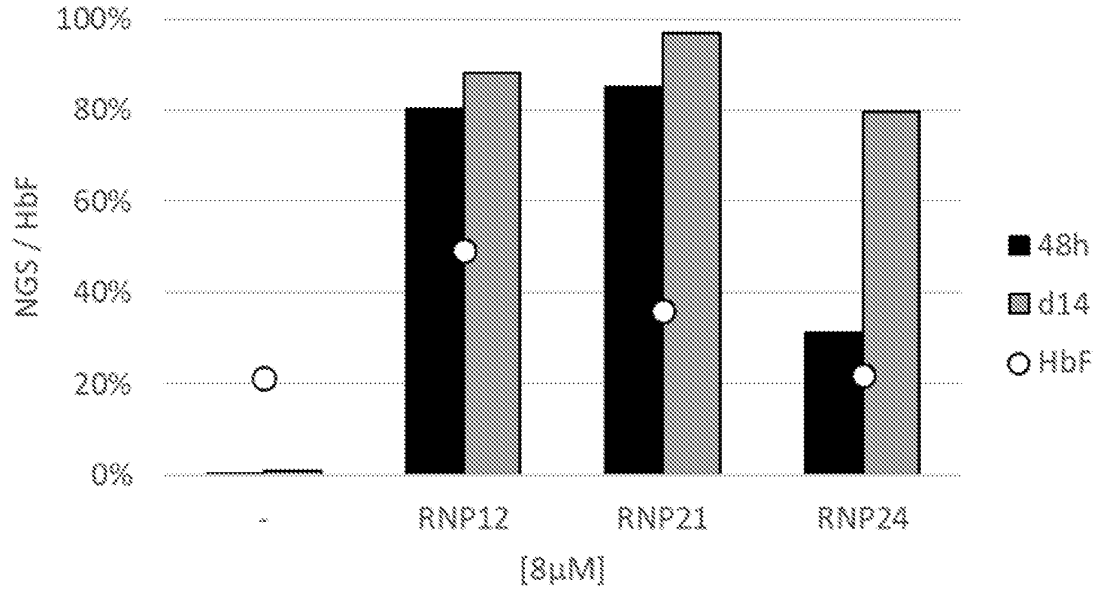
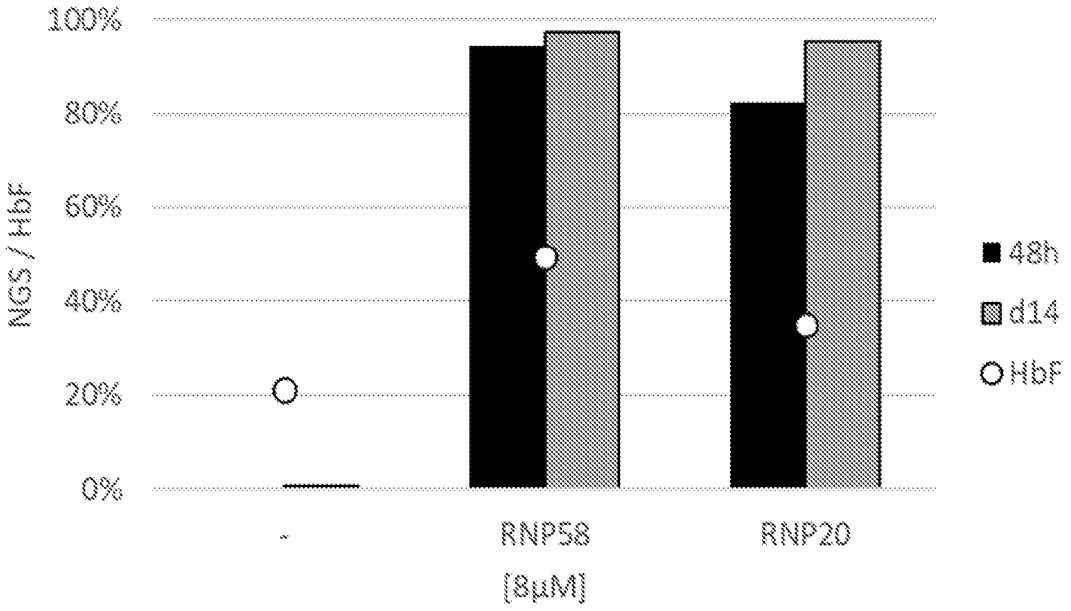


FIG 46B



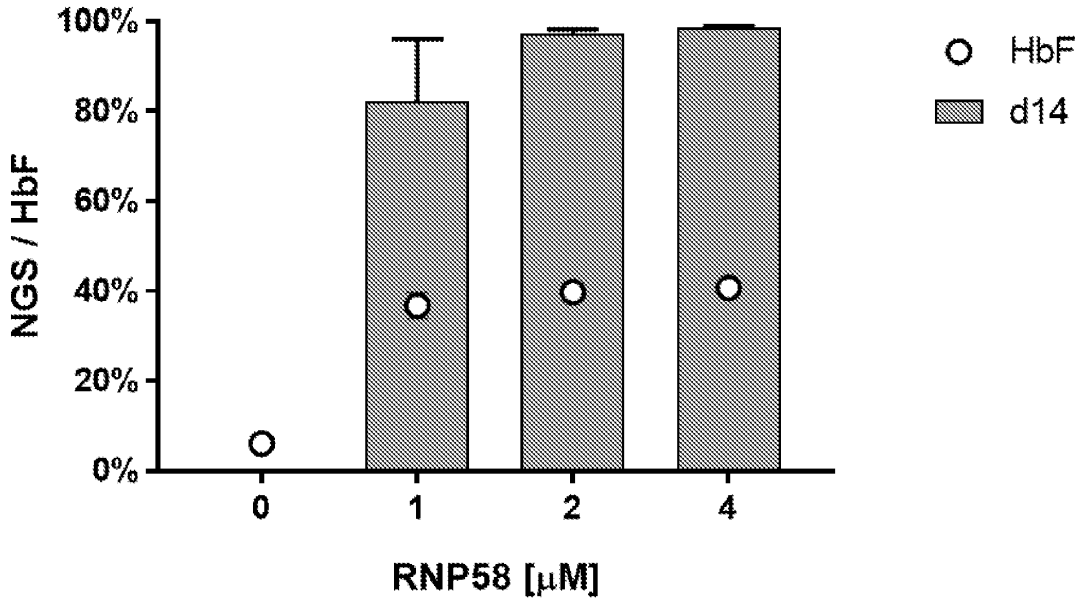
Attorn

FIG 46C



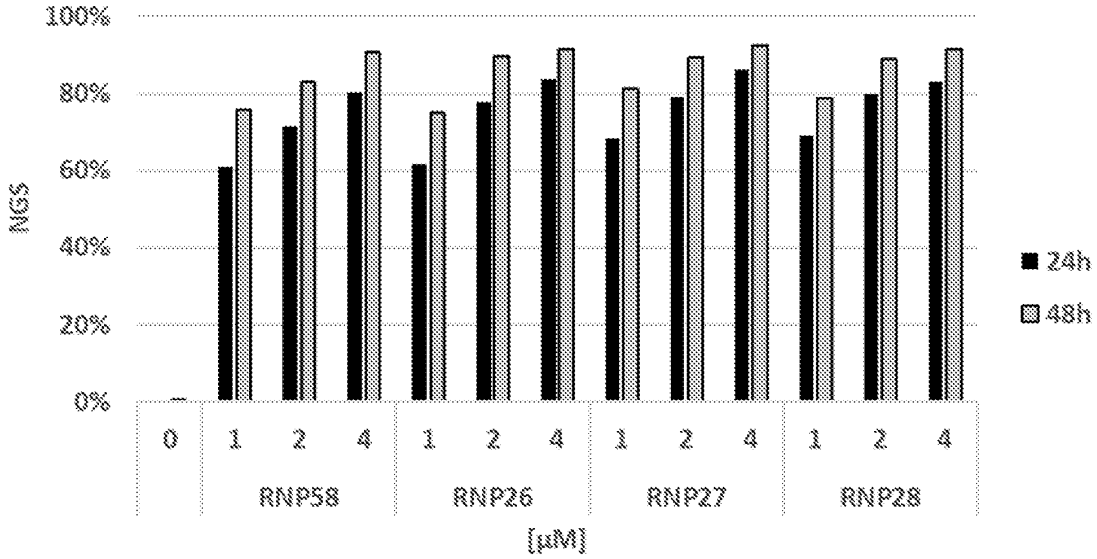
Attorn

FIG 47



Attorn

FIG 48



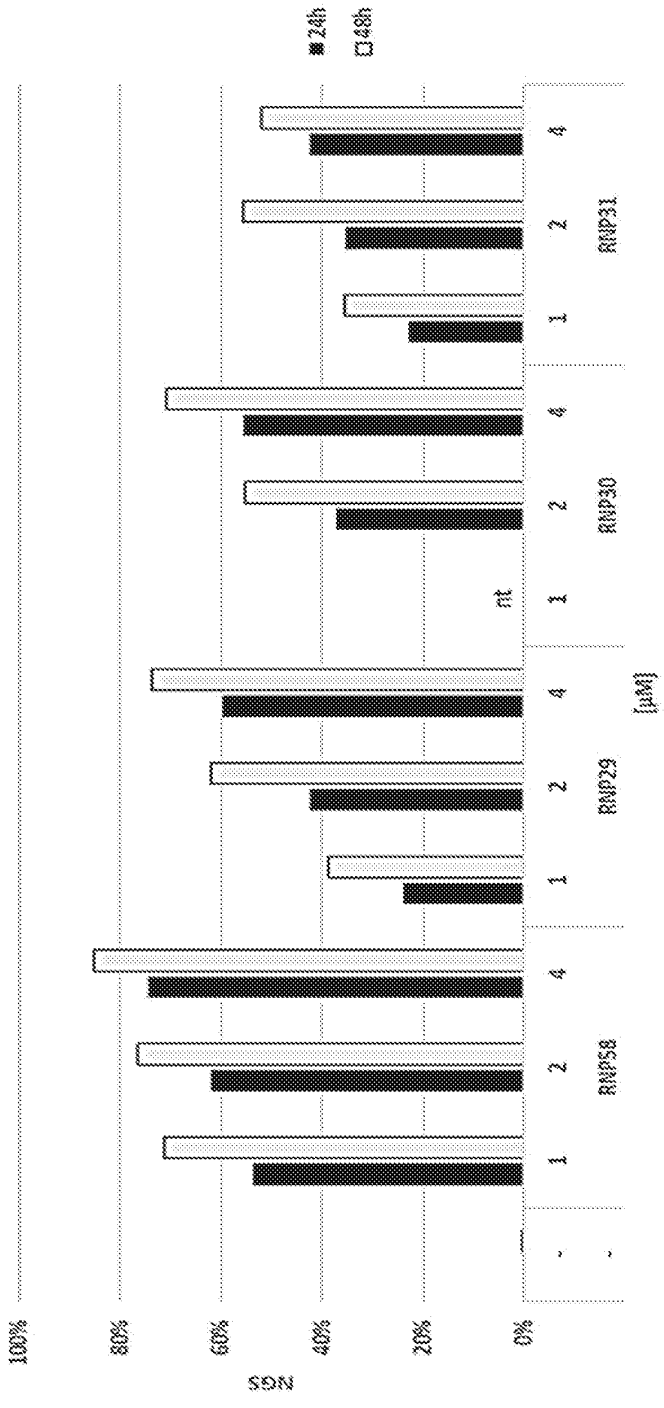
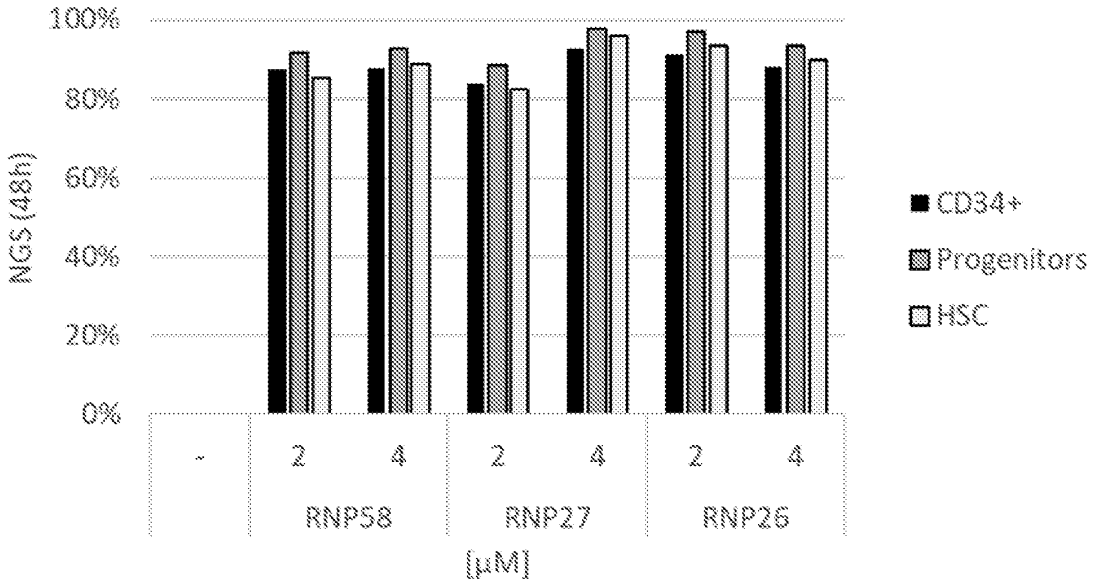

24h
48h
NGS
100%
80%
60%
40%
20%
0%
-
-
1
2
4
RNP58
1
2
4
RNP29
1
2
4
RNP30
rt
1
2
4
RNP31
[μM]

FIG 49

Attorn

FIG 50



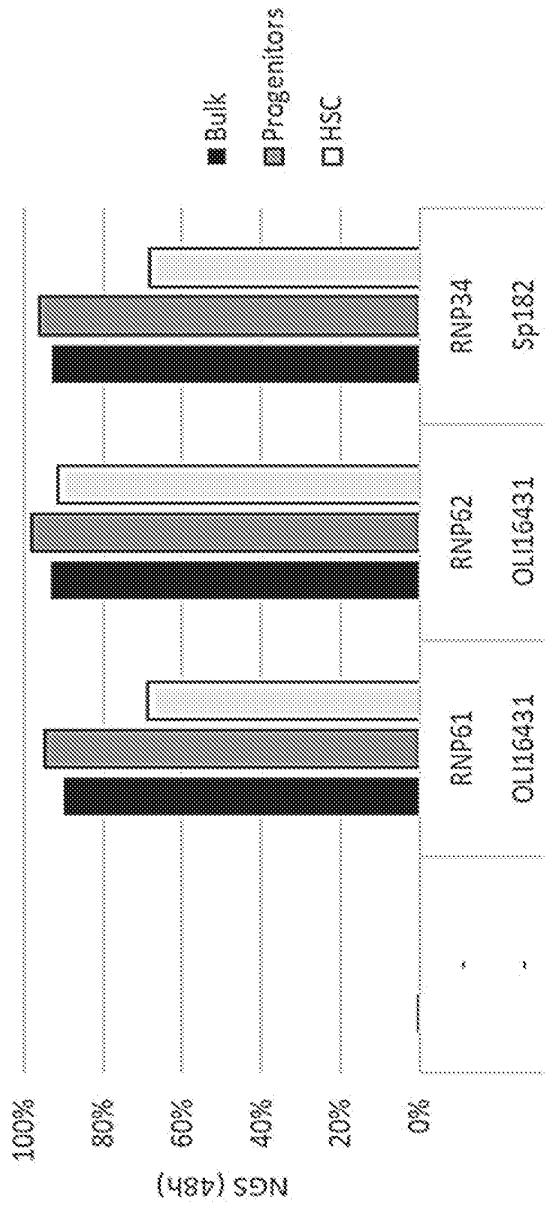

Bulk
Progenitors
HSC
RNP34
Sp182
RNP62
OLI16431
RNP61
OLI16431
100%
80%
60%
40%
20%
0%
NGS (48h)

FIG 51

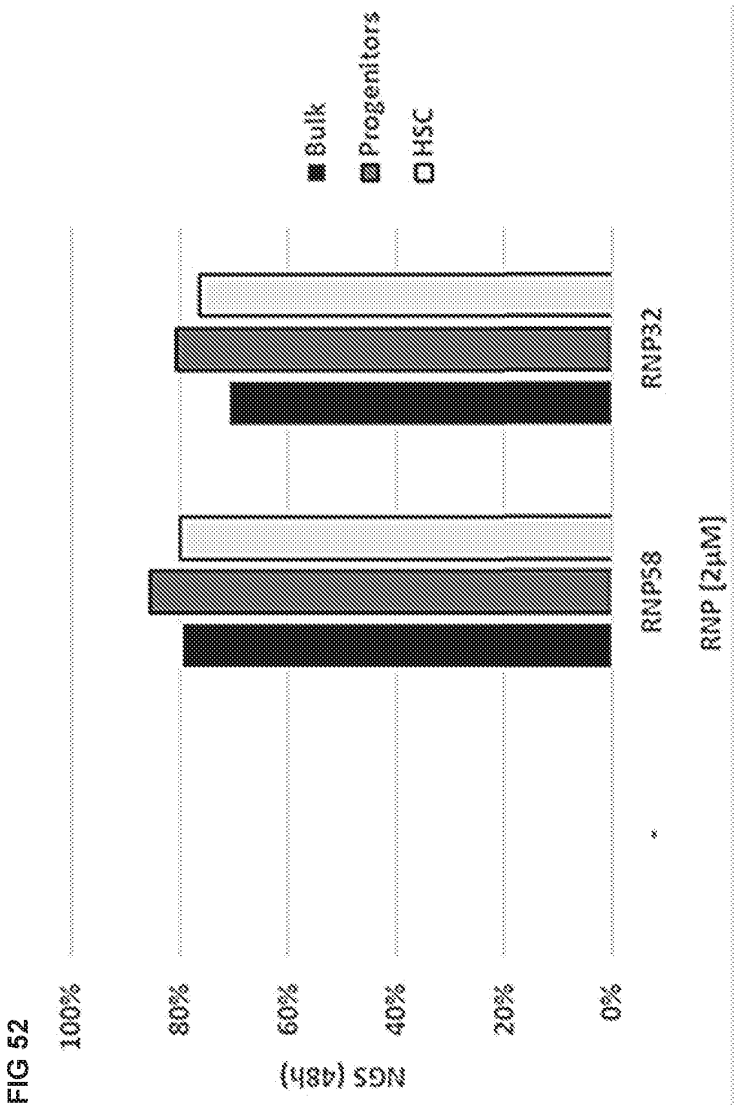
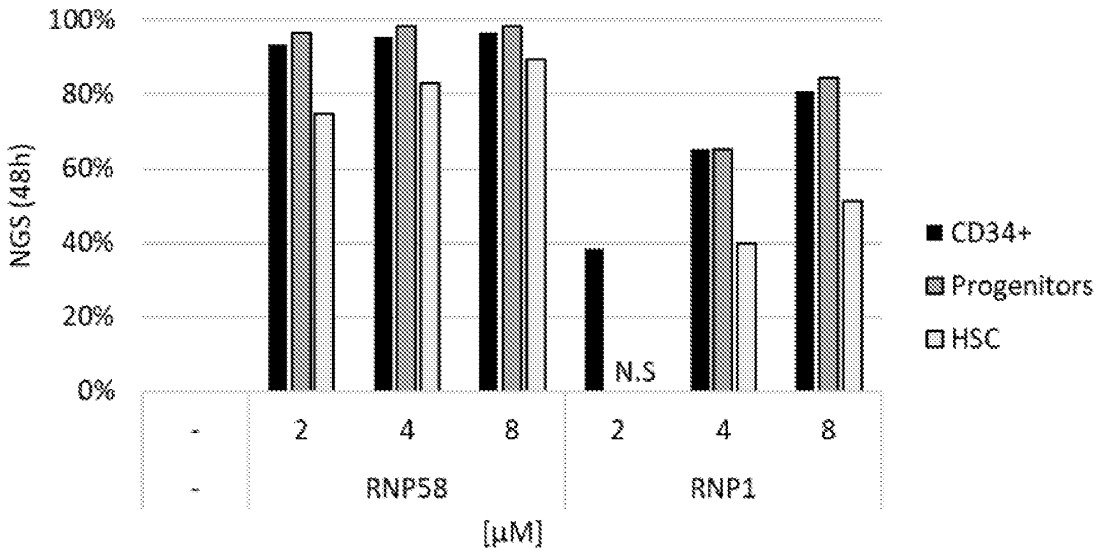

Bulk
Progenitors
HSC
100%
80%
60%
40%
20%
0%
NGS (48h)
RNP32
RNP58
-
RNP (2μM)

FIG 52

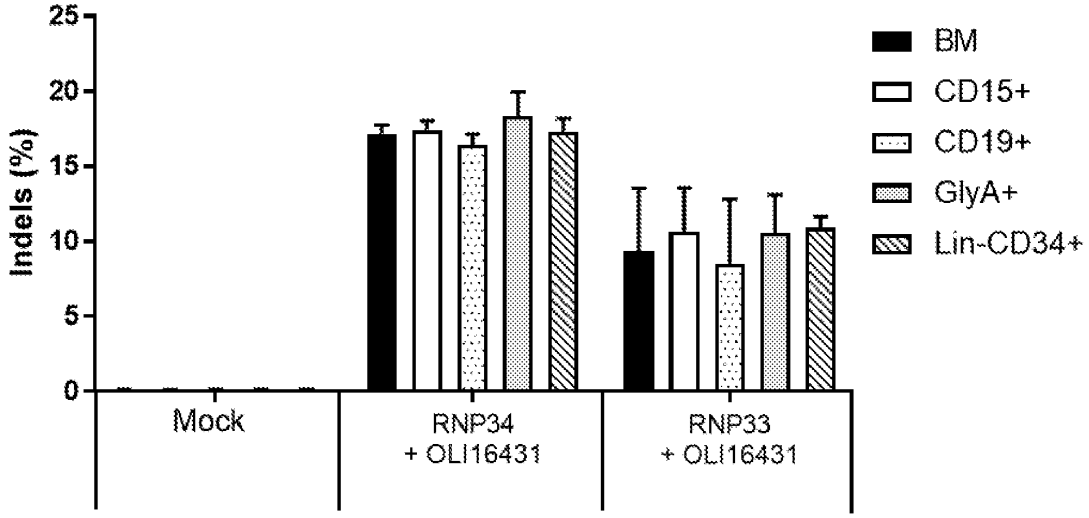
Attorn

FIG 53



Attorn

FIG 54



Attorn

FIG 55A

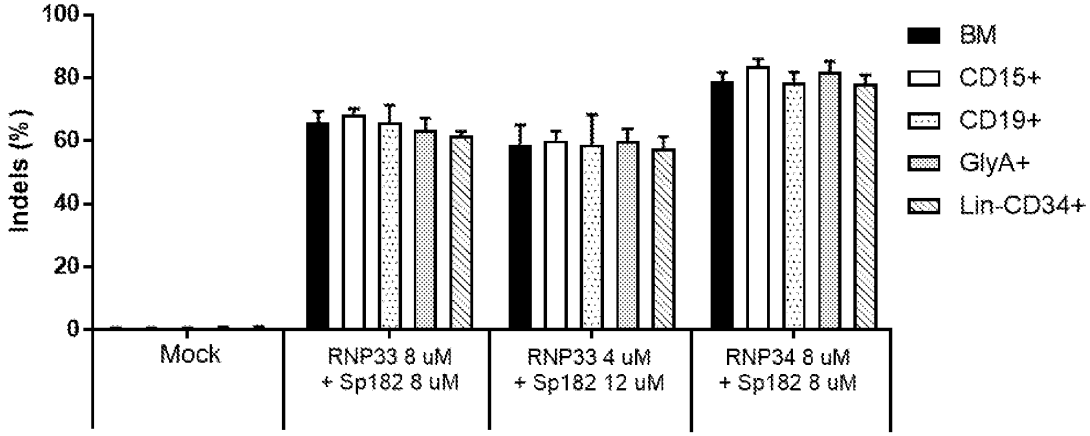
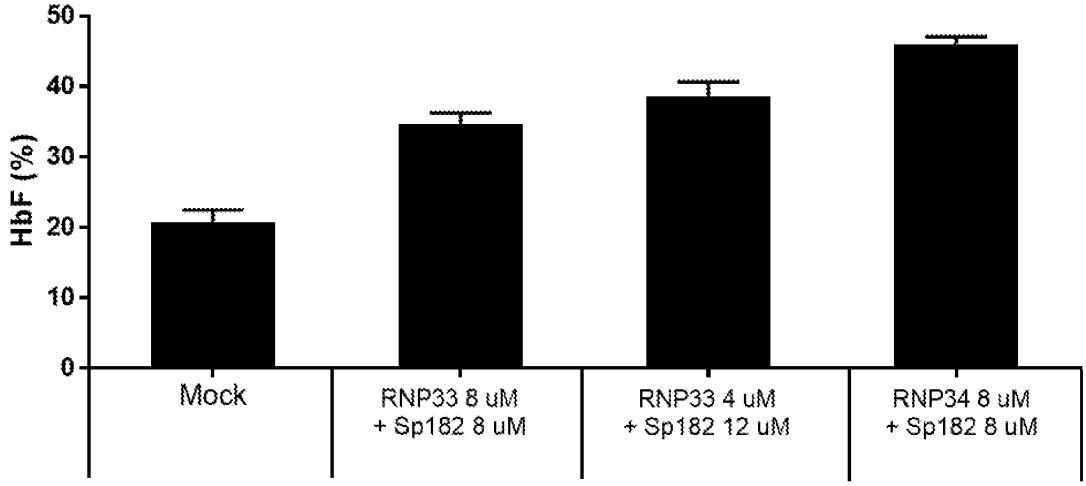


FIG 55B



Attorn

FIG 56A

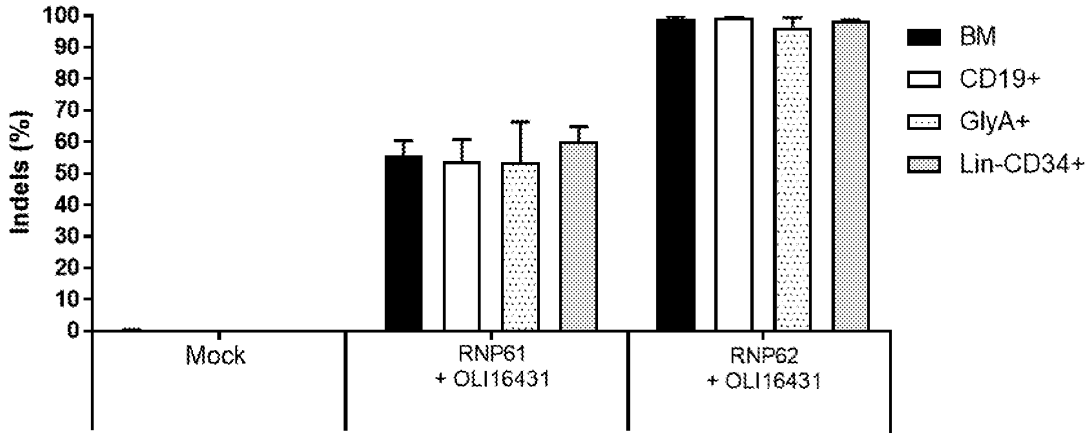
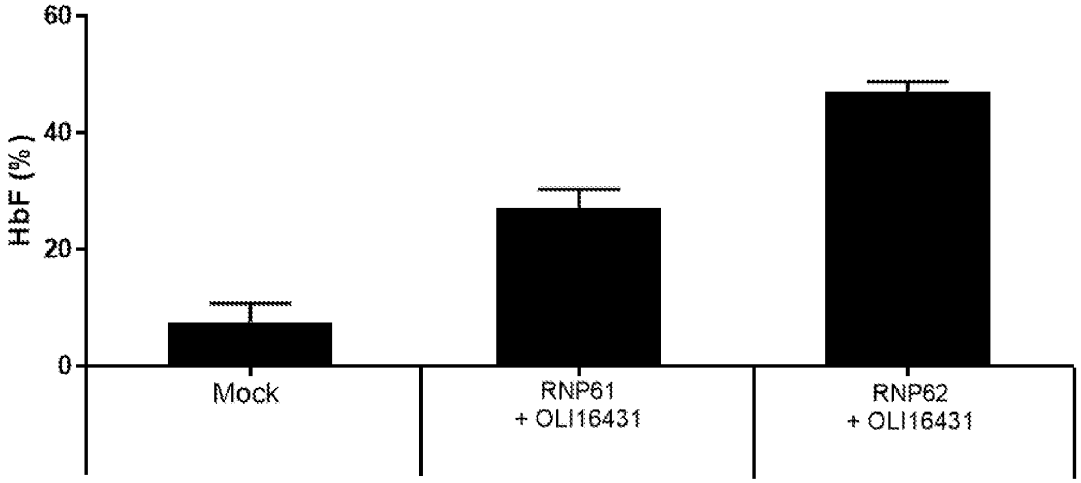
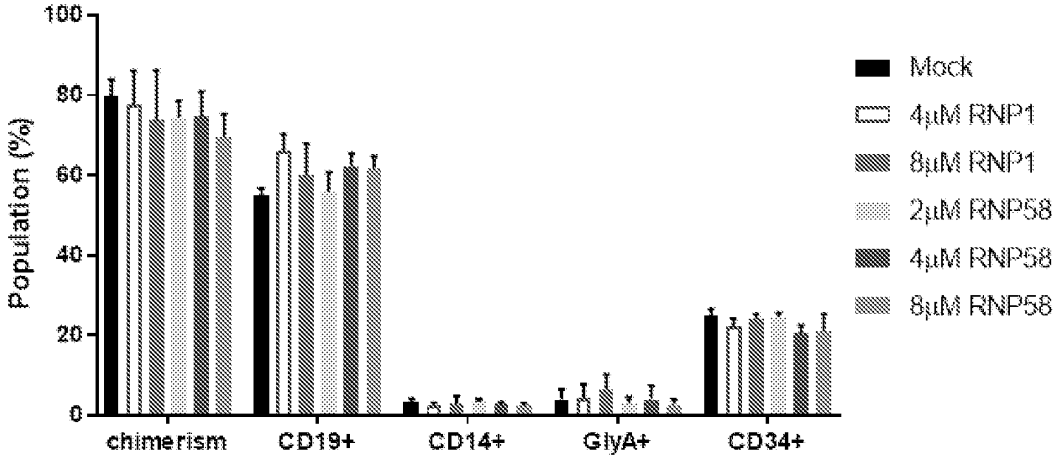


FIG 56B



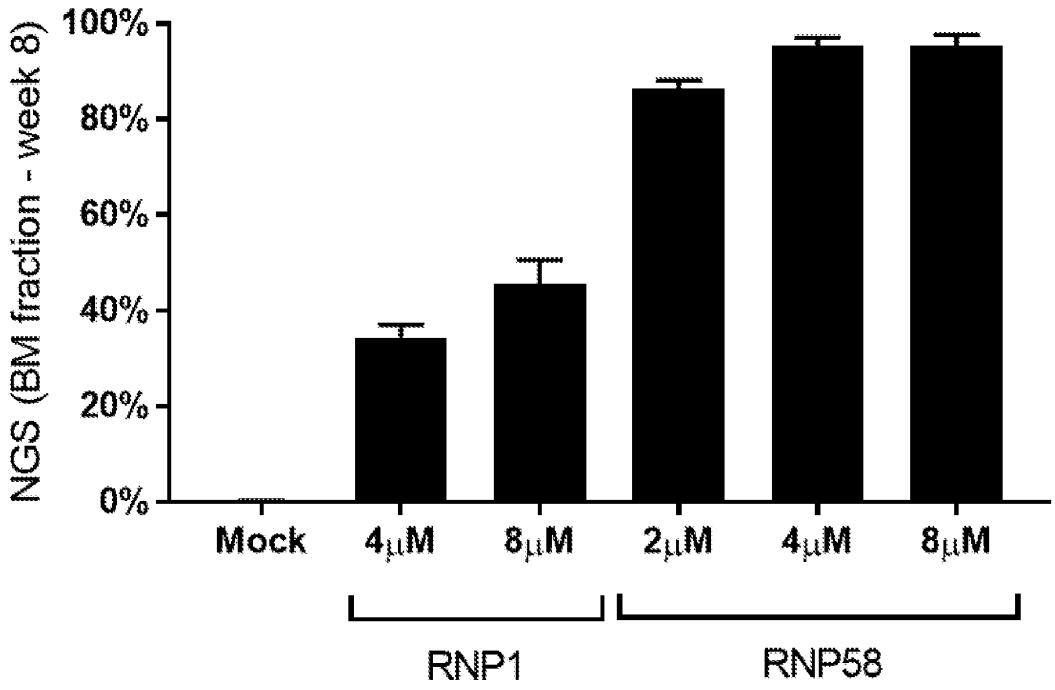
Attorn

FIG 57



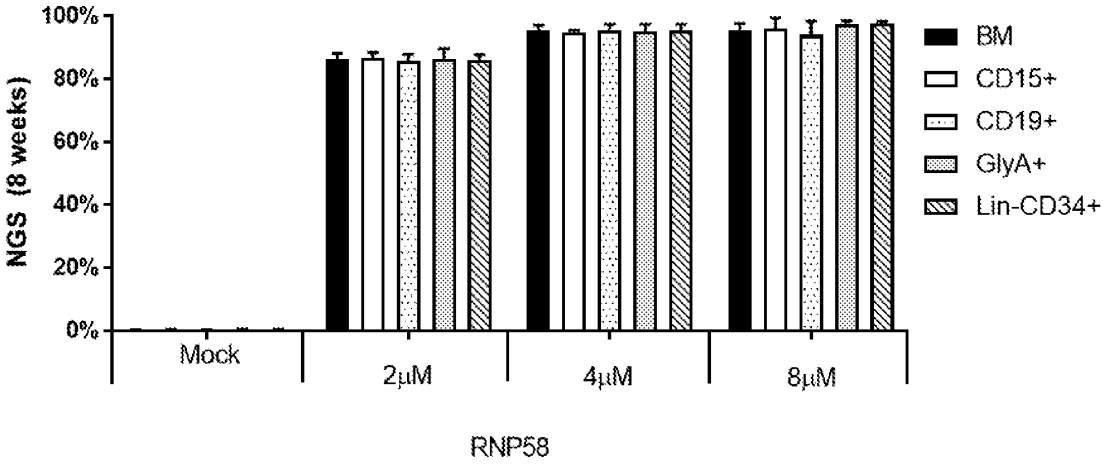
Attorn

FIG 58



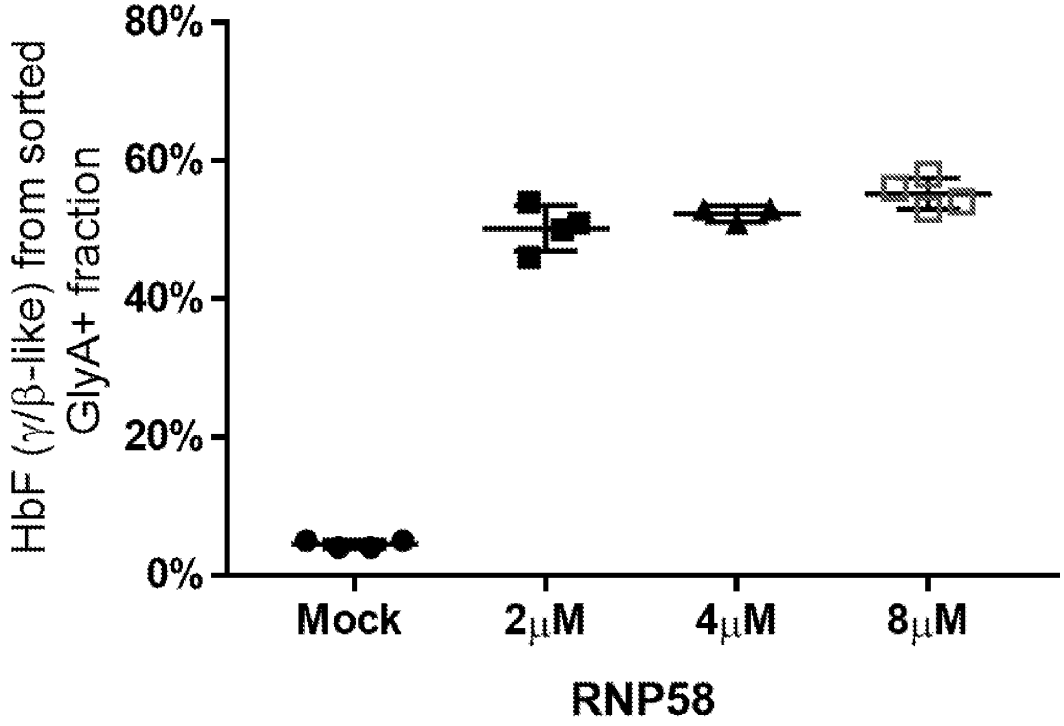
Attorn

FIG 59



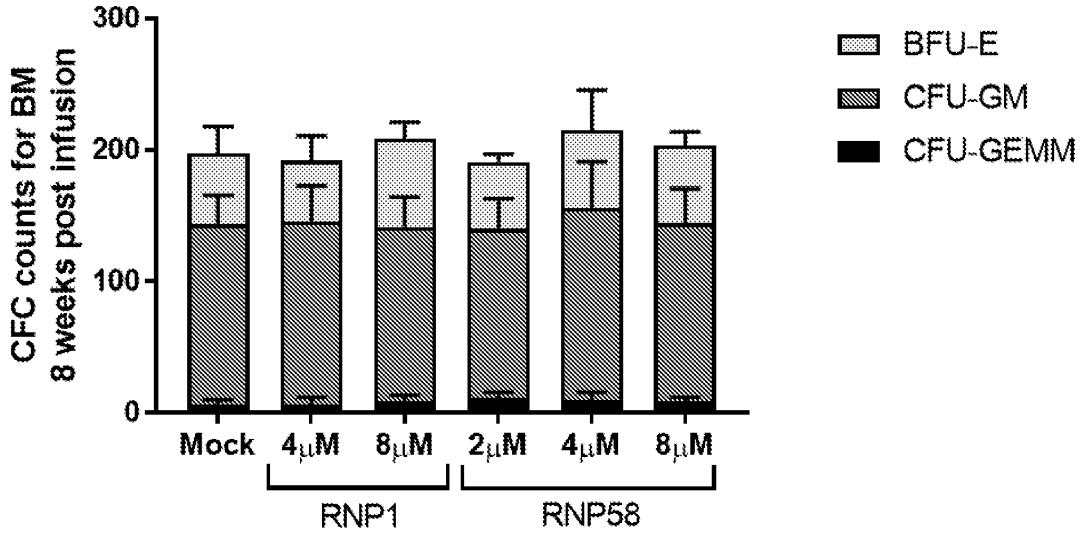
Attorn

FIG 60



Attorn

FIG 61



Attorn

FIG 62

His-AsCpf1-sNLS-sNLS H800A amino acid sequence (SEQ ID NO:1032)

MGHHHHHHGSTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYAD
QCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKA
ELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIF
TRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNL
AIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSI
DLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSE
AFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKL
EMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGR
YKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNN
PEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLL
YHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYR
PKSRMKRMAARLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKD
RRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQ
QFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGI
AEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPL
TGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQF
DAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRS
VLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQ
NGISNQDWLAYIQELRNGS**PKKKRKV**GS**PKKKRKV**

His-AsCpf1-sNLS-sNLS H800A nucleotide sequence (SEQ ID NO:1110)

ATGGGCCATCATCATCACCATCACGGTAGCACCCAGTTTGAAGGCTTTACCAATCTGTATCAGGTTAGC
AAAACCCTGCGTTTTGAACTGATTCCGCAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCAT
CGAAGAGGATAAAGCACGTAACGATCACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAA
ACGTACGCAGATCAGTGTCTGCAGCTGGTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAG
TTATCGCAAAGAGAAAACCGAAGAAACCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAAT
GCCATCCATGATTATTTCATTGGTCGTACCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGA
AATCTATAAAGGCCTGTTTAAAGCCGAACTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTA
CCACCACCGAACATGAAAATGCACTGCTGCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTT
ATGAGAATCGCAAAAACGTGTTTAGCGCAGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAG
GATAATTTCCCGAAATTCAAAGAGAACTGCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTG
CGTGAACATTTTGAAAACGTTAAAAAAGCCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTT
TAGCTTCCCGTTTTACAATCAGCTGCTGACCCAGACACAGATTGATCTGTATAACCAGCTGTTAGGTGG
TATTAGCCGTGAAGCAGGCACCGAAAAAATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAG
AAAAATGATGAAACCGCACATATTATTGCAAGCCTGCCGCATCGTTTTATTCCGCTGTTTAAGCAGATT
CTGAGCGATCGTAATACCCTGAGCTTTATTCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAG
CTTCTGCAAATACAAAACGCTGCTGCGCAATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACG
AACTGAATAGCATTGATCTGACCCACATCTTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCA
CTGTGTGATCATTGGGATACCCTGCGTAATGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAA
AATTACCAAAAGCGCGAAAGAAAAAGTTCAGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAG
ATTATTAGCGCAGCCGGTAAAGAACTGTCAGAAGCATTTAAACAGAAAACCAGCGAAATTCTGAGCC
ATGCACATGCAGCACTGGATCAGCCGCTGCCGACCACACTGAAAAAACAAGAAGAAAAAGAGATCCT
GAAAAGCCAGCTGGATAGCCTGCTGGGTCTGTATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCA
ATGAAGTTGATCCGGAATTTAGCGCACGTCTGACCGGCATTAAGCTGGAAATGGAACCGAGCCTGAGT
TTTTACAATAAAGCCCGTAATTATGCCACCAAAAAACCGTATAGCGTGGAAAAATTCAAGCTGAACTT
CCAGATGCCGACACTGGCAAGCGGTTGGGATGTTAATAAAGAAAAGAACAATGGTGCCATCCTGTTCG
TGAAAAATGGCCTGTATTATCTGGGTATCATGCCGAAACAGAAAGGTCGTTATAAAGCACTGAGCTTT
GAACCGACAGAAAAAACCAGTGAAGGCTTCGATAAAATGTACTATGATTACTTTCCGGACGCAGCCA
AAATGATTCCGAAATGCAGCACCCAGCTGAAAGCAGTTACCGCACATTTTCAGACCCATACCACACCG
ATTCTGCTGAGCAATAACTTTATTGAACCGCTGGAAATCACCAAAGAGATCTACGATCTGAATAACCC

Attorn

GGAAAAAGAGCCGAAAAAGTTTCAGACCGCCTATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGT
GAAGCGCTGTGTAAATGGATTGATTTCACCCGTGATTTTCTGAGCAAATACACCAAAACCACCAGTAT
CGATCTGAGCAGCCTGCGTCCGAGCAGCCAGTATAAAGATCTGGGCGAATATTATGCAGAACTGAATC
CGCTGCTGTATCATATTAGCTTTCAGCGTATTGCCGAGAAAGAAATAATGGATGCAGTTGAAACCGGT
AAGCTGTACCTGTTCCAGATTTATAACAAAGACTTTGCCAAAGGCCATCATGGTAAACCGAATCTGCA
TACCCTGTATTGGACCGGTCTGTTTAGCCCTGAAAATCTGGCAAAAACCAGCATTAAACTGAATGGTC
AGGCGGAACTGTTTTATCGTCCGAAAAGCCGTATGAAACGTATGGCAGCACGTCTGGGTGAAAAAATG
CTGAATAAAAAGCTGAAGGATCAGAAAACCCCGATTCCGGATACACTGTATCAAGAGCTGTATGATTA
TGTGAATCATCGTCTGAGTCATGATCTGAGTGATGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAA
AGAAGTTAGCCACGAGATCATTAAAGATCGTCGTTTTACCAGCGACAAGTTTTTTTTCCATGTTCCGAT
TACGCTGAATTATCAGGCAGCAAATAGCCCGAGCAAATTTAACCAGCGTGTGAATGCATATCTGAAAG
AACATCCGGAAACACCGATTATTGGTATTGATCGTGGTGAACGCAACCTGATTTATATCACCGTTATTG
ATAGCACCGGCAAGATCCTGGAACAGCGTAGCCTGAATACCATTCAGCAGTTTGATTATCAGAAAAAG
CTGGACAACCGTGAAAAAGAACGTGTTGCAGCCCGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGA
CCTGAAACAGGGTTATCTGAGCCAGGTTATTCATGAAATTGTGGATCTGATGATTCATTACCAGGCAG
TTGTTGTGCTGGAAAACCTGAATTTTGGCTTTAAAAGCAAACGTACCGGCATTGCGGAAAAAGCAGTT
TATCAGCAGTTCGAAAAGATGCTGATCGATAAACTGAACTGTCTGGTGCTGAAAGATTATCCGGCTGA
AAAAGTTGGCGGTGTTCTGAATCCGTATCAGCTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCA
CCCAGAGCGGATTTCTGTTTTATGTTCCGGCACCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTG
TTGATCCGTTTGTTTGGAAAACCATCAAAAACCATGAAAGCCGCAAACATTTTCTGGAAGGTTTTGATT
TCCTGCATTACGACGTTAAGACCGGTGATTTTATTCTGCACTTTAAGATGAATCGGAACCTGTCATTTC
AGCGTGGTCTGCCTGGTTTTATGCCTGCATGGGATATTGTGTTTGAGAAAAACGAAACACAGTTCGAT
GCAAAAGGCACCCCGTTTATTGCAGGTAAACGTATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGT
CGTTATCGCGATCTGTATCCGGCAAATGAACTGATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGT
GATGGCTCAAACATTCTGCCGAAACTGTTAGAGAATGATGATAGCCATGCCATTGATACCATGGTTGC
CCTGATTCGTAGCGTTCTGCAGATGCGTAATAGCAATGCAGCAACCGGTGAAGATTACATTAATAGTC
CGGTTCGTGATCTGAATGGTGTTTGTTTTGATAGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCGG
ATGCAAATGGTGCATATCATATTGCACTGAAAGGACAGCTGCTGCTGAACCACCTGAAAGAAAGTAA
AGATTTGAAGCTGCAGAATGGCATTAGCAATCAGGATTGGCTGGCATATATCCAAGAACTGCGTAACG
GTTCT**CCTAAGAAAAAGCGCAAAGTT**GGCAGT**CCGAAGAAGAAACGTAAGGTT**

Cpf1-1 amino acid sequence (SEQ ID NO:1094)

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLK
QLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSL
REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHK
KLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEIL
SHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYN
KARNYATKKPYSVEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTA
YAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEI
MDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAH
RLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFLFH
VPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDN
REKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFE
KMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWK
TIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAG
KRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNA
ATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWL
AYIQELRNGR**SSDDEATADSQHAAPPKKKRKV**GGSGGSGGSGGSGGSGGSGGSGGSLE<u>HHHHHH</u>

Cpf1-1 nucleotide sequence (SEQ ID NO:1111)

Attorn

ATGACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCG
CAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCATCGAAGAGGATAAAGCACGTAACGATC
ACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTG
GTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAAA
CCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTCATTGGTCGTA
CCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAA
CTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCT
GCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGC
AGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAGAACT
GCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAG
CCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGA
CCCAGACCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAA
ATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGC
AAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTAT
TCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCA
ATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATC
TTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGATACCCTGCGTAA
TGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTC
AGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCA
GAAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCC
GACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTG
TATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCT
GACCGGCATTAAACTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCA
AAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCCTGGCAAGCGGTTGGGAT
GTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTAT
GCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTT
GATAAAATGTACTACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAA
AGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCT
GGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCA
TATGCAAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCG
TGATTTTCTGAGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGT
ATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTG
CCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGAT
TTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAA
AATCTGGCAAAAACCTCGATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTAT
GAAACGTATGGCACATCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCG
ATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGA
TGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTC
GTTTTACCAGCGACAAATTCCTGTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGA
GCAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGAT
CGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAG
CCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAG
CACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATT
CATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTT
AAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACA
AACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAG
CTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCA
CCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAAC
CATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTC
ATCCTGCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGG
GATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACG
TATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACT
GATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGG
AAAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAAT
AGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGAT
AGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAA

Attorn

AGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAAT
CAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCA**
GATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTAAAGTTGGTGGTAGCGGTGGTTCAGGTGG
TAGTGGCGGTAGTGGTGGCTCAGGGGGTTCTGGTGGCTCTGGTGGTAGCCTCGAGCACCACCACCACC
ACCAC

Cpf1-2 amino acid sequence (SEQ ID NO:1095)

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLK
QLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSL
REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHK
KLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEIL
SHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYN
KARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTE
KTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQT
AYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEK
EIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMA
HRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFF
HVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLD
NREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQF
EKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVW
KTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIA
GKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSN
AATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDW
LAYIQELRNGR**SSDDEATADSQHAAPPKKKRKV**GGSGGSGGSGGSGGSGGSGGSLEHHHHHH

Cpf1-2 nucleotide sequence (SEQ ID NO:1112)

ATGACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCG
CAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCATCGAAGAGGATAAAGCACGTAACGATC
ACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTG
GTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAAA
CCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTCATTGGTCGTA
CCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAA
CTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCT
GCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGC
AGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAGAACT
GCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAG
CCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGA
CCCAGACCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAA
ATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGC
AAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTAT
TCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCA
ATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATC
TTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGATACCCTGCGTAA
TGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTC
AGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCA
GAAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCC
GACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTG
TATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCT
GACCGGCATTAAACTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCA
AAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGATGCCGACCCTGGCAAGCGGTTGGGAT
GTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTAT
GCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTT

Attorn

GATAAAATGTACTACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAA
AGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCT
GGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCA
TATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCG
TGATTTTCTGAGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGT
ATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTG
CCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGAT
TTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAA
AATCTGGCAAAAACCTCGATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTAT
GAAACGTATGGCACATCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCG
ATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGA
TGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTC
GTTTTACCAGCGACAAATTCTTTTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGA
GCAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGAT
CGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAG
CCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAG
CACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATT
CATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTT
AAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACA
AACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAG
CTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCA
CCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAAC
CATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTC
ATCCTGCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGG
GATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACG
TATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACT
GATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGG
AAAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAAT
AGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGAT
AGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAA
AGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAAT
CAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCA**
GATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTAAAGTTGGTGGTAGCGGTGGTTCAGGTGG
TAGTGGCGGTAGTGGTGGCTCAGGGGGTTCTGGTGGCTCTGGTGGTAGCCTCGAGCACCACCACCACC
ACCAC

Cpf1-3 amino acid sequence (SEQ ID NO:1096)

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLK
QLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSL
REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHK
KLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEIL
SHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYN
KARNYATKKPYSVEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTA
YAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEI
MDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAA
RLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFLFH
VPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDN
REKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFE
KMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWK
TIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAG
KRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNA
ATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWL
AYIQELRNGR**SSDDEATADSQHAAPPKKKRKV**GGSGGSGGSGGSGGSGGSGGSGGSLEHHHHHH

Attorn

Cpf1-3 nucleotide sequence (SEQ ID NO:1113)

ATGACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCG
CAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCATCGAAGAGGATAAAGCACGTAACGATC
ACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTG
GTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAAA
CCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTCATTGGTCGTA
CCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAA
CTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCT
GCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGC
AGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAGAACT
GCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAG
CCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGA
CCCAGACCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAA
ATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGC
AAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTAT
TCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCA
ATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATC
TTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGATACCCTGCGTAA
TGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTC
AGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCA
GAAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCC
GACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTG
TATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCT
GACCGGCATTAAACTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCA
AAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCCTGGCAAGCGGTTGGGAT
GTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTAT
GCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTT
GATAAAATGTACTACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAA
AGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCT
GGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCA
TATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCG
TGATTTTCTGAGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGT
ATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTG
CCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGAT
TTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAA
AATCTGGCAAAAACCTCGATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTAT
GAAACGTATGGCAGCTCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCG
ATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGA
TGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTC
GTTTTACCAGCGACAAATTCCTGTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGA
GCAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGAT
CGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAG
CCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAG
CACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATT
CATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTT
AAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACA
AACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAG
CTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCA
CCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAAC
CATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTC
ATCCTGCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGG
GATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACG
TATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACT
GATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGG
AAAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAAT

Attorn

AGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGAT
AGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAA
AGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAAT
CAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCA**
GATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTAAAGTTGGTGGTAGCGGTGGTTCAGGTGG
TAGTGGCGGTAGTGGTGGCTCAGGGGGTTCTGGTGGCTCTGGTGGTAGCCTCGAGCACCACCACCACC
ACCAC

Cpf1-4 amino acid sequence (SEQ ID NO:1097)

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLK
QLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSL
REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHK
KLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEIL
SHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYN
KARNYATKKPYSVEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTA
YAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEI
MDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAA
RLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFLFH
VPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDN
REKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFE
KMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWK
TIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAG
KRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNA
ATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWL
AYIQELRNGR**SSDDEATADSQHAAPPKKKRKV**

Cpf1-4 nucleotide sequence (SEQ ID NO:1114)

ATGACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCG
CAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCATCGAAGAGGATAAAGCACGTAACGATC
ACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTG
GTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAAA
CCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTCATTGGTCGTA
CCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAA
CTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCT
GCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGC
AGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAGAACT
GCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAG
CCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGA
CCCAGACCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAA
ATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGC
AAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTAT
TCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCA
ATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATC
TTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGATACCCTGCGTAA
TGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTC
AGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCA
GAAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCC
GACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTG
TATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCT
GACCGGCATTAAACTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCA
AAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCCTGGCAAGCGGTTGGGAT
GTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTAT

Attorn

GCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTT
GATAAAATGTACTACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAA
AGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCT
GGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCA
TATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCG
TGATTTTCTGAGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGT
ATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTG
CCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGAT
TTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAA
AATCTGGCAAAAACCTCGATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTAT
GAAACGTATGGCAGCTCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCG
ATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGA
TGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTC
GTTTTACCAGCGACAAATTCCTGTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGA
GCAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGAT
CGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAG
CCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAG
CACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATT
CATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTT
AAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACA
AACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAG
CTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCA
CCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAAC
CATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTC
ATCCTGCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGG
GATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACG
TATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACT
GATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGG
AAAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAAT
AGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGAT
AGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAA
AGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAAT
CAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCA**
GATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTAAAGTT

Cpf1-5 amino acid sequence (SEQ ID NO:1107)

STQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLK
QLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSL
REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHK
KLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEIL
SHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYN
KARNYATKKPYSVEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTA
YAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEI
MDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAH
RLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFLFH
VPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDN
REKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFE
KMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWK
TIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAG
KRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNA
ATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWL
AYIQELRNGR**SSDDEATADSQHAAPPKKKRKV**

Attorn

Cpf1-5 nucleotide sequence (SEQ ID NO:1115)

AGCACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCG
CAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCATCGAAGAGGATAAAGCACGTAACGATC
ACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTG
GTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAAA
CCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTCATTGGTCGTA
CCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAA
CTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCT
GCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGC
AGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAGAACT
GCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAG
CCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGA
CCCAGACCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAA
ATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGC
AAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTAT
TCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCA
ATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATC
TTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGATACCCTGCGTAA
TGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTC
AGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCA
GAAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCC
GACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTG
TATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCT
GACCGGCATTAAACTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCA
AAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCCTGGCAAGCGGTTGGGAT
GTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTAT
GCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTT
GATAAAATGTACTACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAA
AGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCT
GGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCA
TATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCG
TGATTTTCTGAGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGT
ATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTG
CCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGAT
TTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAA
AATCTGGCAAAAACCTCGATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTAT
GAAACGTATGGCACATCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCG
ATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGA
TGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTC
GTTTTACCAGCGACAAATTCCTGTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGA
GCAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGAT
CGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAG
CCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAG
CACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATT
CATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTT
AAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACA
AACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAG
CTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCA
CCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAAC
CATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTC
ATCCTGCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGG
GATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACG
TATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACT
GATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGG
AAAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAAT

Attorn

AGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGAT
AGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAA
AGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAAT
CAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCA**
GATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTAAAGTT

Cpf1-6 amino acid sequence (SEQ ID NO:1108)

STQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLD
WENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLK
QLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSL
REHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHK
KLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEIL
SHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYN
KARNYATKKPYSVEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEK
TSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTA
YAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEI
MDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAH
RLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFLFH
VPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDN
REKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFE
KMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWK
TIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAG
KRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNA
ATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWL
AYIQELRNGR**SSDDEATADSQHAAPPKKKRKV**GGSGGSGGSGGSGGSGGSGGSGGSLE<u>HHHHHH</u>

Cpf1-6 nucleotide sequence (SEQ ID NO:1116)

AGCACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCG
CAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCATCGAAGAGGATAAAGCACGTAACGATC
ACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTG
GTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAAA
CCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTCATTGGTCGTA
CCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAA
CTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCT
GCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGC
AGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAGAACT
GCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAG
CCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGA
CCCAGACCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAA
ATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGC
AAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTAT
TCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCA
ATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATC
TTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGATACCCTGCGTAA
TGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTC
AGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCA
GAAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCC
GACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTG
TATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCT
GACCGGCATTAAACTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCA
AAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCCTGGCAAGCGGTTGGGAT
GTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTAT
GCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTT
GATAAAATGTACTACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAA

Attorn

AGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCT
GGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCA
TATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCG
TGATTTTCTGAGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGT
ATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTG
CCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGAT
TTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAA
AATCTGGCAAAAACCTCGATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTAT
GAAACGTATGGCACATCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCG
ATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGA
TGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTC
GTTTTACCAGCGACAAATTCCTGTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGA
GCAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGAT
CGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAG
CCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAG
CACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATT
CATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTT
AAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACA
AACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAG
CTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCA
CCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAAC
CATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTC
ATCCTGCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGG
GATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACG
TATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACT
GATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGG
AAAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAAT
AGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGAT
AGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAA
AGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAAT
CAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCA**
GATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTAAAGTTGGTGGTAGCGGTGGTTCAGGTGG
TAGTGGCGGTAGTGGTGGCTCAGGGGGTTCTGGTGGCTCTGGTGGTAGCCTCGAGCACCACCACCACC
ACCAC

Cpf1-7 amino acid sequence (SEQ ID NO:1109)

MGRDPGKPIPNPLLGLDSTAPKKKRKVGIHGVPAATQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEE
DKARNDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFI
GRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSA
EDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLY
NQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFC
KYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKE
KVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHL
LDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEK
NNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQ
THTTPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSI
DLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLY
WTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRL
SHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGID
RGERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIV
DLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFT
SFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNR
NLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFR
DGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADA
NGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN**PKKKRKV**KLAAALEHHHHHH

Attorn

Cpf1-7 nucleotide sequence (SEQ ID NO:1117)

ATGGGTCGGGATCCAGGTAAACCGATTCCGAATCCGCTGCTGGGTCTGGATAGCACCGCACCGAAAAA
AAAACGTAAAGTTGGTATTCATGGTGTTCCGGCAGCAACCCAGTTTGAAGGTTTCACCAATCTGTATC
AGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCGCAGGGTAAAACCCTGAAACATATTCAAGAACAG
GGCTTCATCGAAGAGGATAAAGCACGTAACGATCACTACAAAGAACTGAAACCGATTATCGACCGCA
TCTATAAAACCTATGCAGATCAGTGTCTGCAGCTGGTTCAGCTGGATTGGGAAAATCTGAGCGCAGCA
ATTGATAGTTATCGCAAAGAAAAAACCGAAGAAACCCGTAATGCACTGATTGAAGAACAGGCAACCT
ATCGTAATGCCATCCATGATTATTTCATTGGTCGTACCGATAATCTGACCGATGCAATTAACAAACGTC
ACGCCGAAATCTATAAAGGCCTGTTTAAAGCCGAACTGTTTAATGGCAAAGTTCTGAAACAGCTGGGC
ACCGTTACCACCACCGAACATGAAAATGCACTGCTGCGTAGCTTTGATAAATTCACCACCTATTTCAGC
GGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGCAGAAGATATTAGCACCGCAATTCCGCATCGTAT
TGTGCAGGATAATTTCCCGAAATTCAAAGAGAACTGCCACATTTTTACCCGTCTGATTACCGCAGTTCC
GAGCCTGCGTGAACATTTTGAAAACGTTAAAAAAGCCATCGGCATCTTTGTTAGCACCAGCATTGAAG
AAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGACCCAGACCCAGATTGATCTGTATAACCAACTGC
TGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAAATCAAAGGTCTGAATGAAGTGCTGAATCTGGC
CATTCAGAAAAATGATGAAACCGCACATATTATTGCAAGCCTGCCGCATCGTTTTATTCCGCTGTTCAA
ACAAATTCTGAGCGATCGTAATACCCTGAGCTTTATTCTGGAAGAATTCAAATCCGATGAAGAGGTGA
TTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCAATGAAAATGTTCTGGAAACTGCCGAAGCACTG
TTTAACGAACTGAATAGCATTGATCTGACCCACATCTTTATCAGCCACAAAAAACTGGAAACCATTTC
AAGCGCACTGTGTGATCATTGGGATACCCTGCGTAATGCCCTGTATGAACGTCGTATTAGCGAACTGA
CCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTCAGCGCAGTCTGAAACATGAGGATATTAATCT
GCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCAGAAGCATTTAAACAGAAAACCAGCGAAATT
CTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCCGACCACCCTGAAAAAACAAGAAGAAAAAG
AAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTGTATCATCTGCTGGACTGGTTTGCAGTTGAT
GAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCTGACCGGCATTAAACTGGAAATGGAACCGA
GCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCAAAAAACCGTATAGCGTCGAAAAATTCAAA
CTGAACTTTCAGATGCCGACCCTGGCAAGCGGTTGGGATGTTAATAAAGAAAAAAACAACGGTGCCAT
CCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTATGCCGAAACAGAAAGGTCGTTATAAAGCGC
TGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTTGATAAAATGTACTACGACTATTTTCCGGAT
GCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAAAGCAGTTACCGCACATTTTCAGACCCATAC
CACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCTGGAAATCACCAAAGAGATCTACGATCTGA
ATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCATATGCAAAAAAAACCGGTGATCAGAAAG
GTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCGTGATTTTCTGAGCAAATACACCAAAACC
ACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGTATAAAGATCTGGGCGAATATTATGCAGA
ACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTGCCGAGAAAGAAATCATGGACGCAGTTG
AAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGATTTTGCCAAAGGCCATCATGGCAAACCG
AATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAAAATCTGGCAAAAACCTCGATTAAACTG
AATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTATGAAACGTATGGCACATCGTCTGGGTGA
AAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCGATCCCGGATACACTGTATCAAGAACTG
TATGATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGATGAAGCACGTGCCCTGCTGCCGAATGTT
ATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTCGTTTTACCAGCGACAAATTCTTTTTTCAT
GTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGAGCAAATTTAACCAGCGTGTTAATGCATA
TCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGATCGTGGTGAACGTAACCTGATTTATATCA
CCGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAGCCTGAATACCATTCAGCAGTTTGATTAC
CAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAGCACGTCAGGCATGGTCAGTTGTTGGTA
CAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATTCATGAAATTGTGGATCTGATGATTCAC
TATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTTAAAAGCAAACGTACCGGCATTGCAGA
AAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACAAACTGAATTGCCTGGTGCTGAAAGATT
ATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAGCTGACCGATCAGTTTACCAGCTTTGCAA
AAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCACCGTATACGAGCAAAATTGATCCGCTG
ACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAACCATGAAAGCCGCAAACATTTTCTGGA
AGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTCATCCTGCACTTTAAAATGAATCGCAA
TCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGGGATATTGTGTTTGAGAAAAACGAAAC
ACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACGTATTGTTCCGGTGATTGAAAATCATC
GTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACTGATCGCACTGCTGGAAGAGAAAGGT

Attorn

ATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGGAAAATGATGATAGCCATGCAATTGAT
ACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAATAGCAATGCAGCAACCGGTGAAGATTA
CATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGATAGCCGTTTTCAGAATCCGGAATGGCC
GATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAAAGGACAGCTGCTGCTGAACCACCTGA
AAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAATCAGGATTGGCTGGCATATATCCAAGA
ACTGCGTAAC**CCTAAAAAAAAAACGCAAAGTG**AAGCTTGCGGCCGCACTCGAGCACCACCACCACCAC
CAC