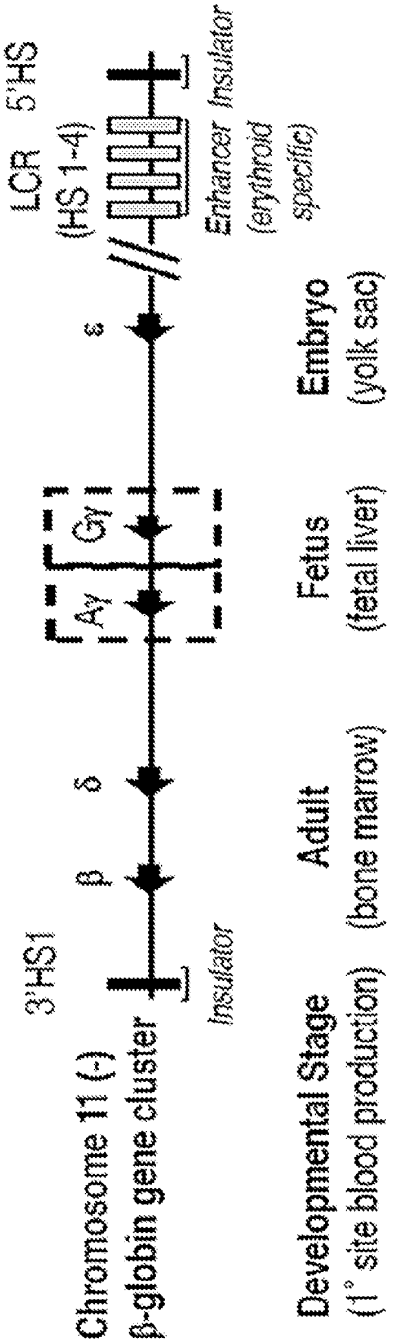
FIG 1


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

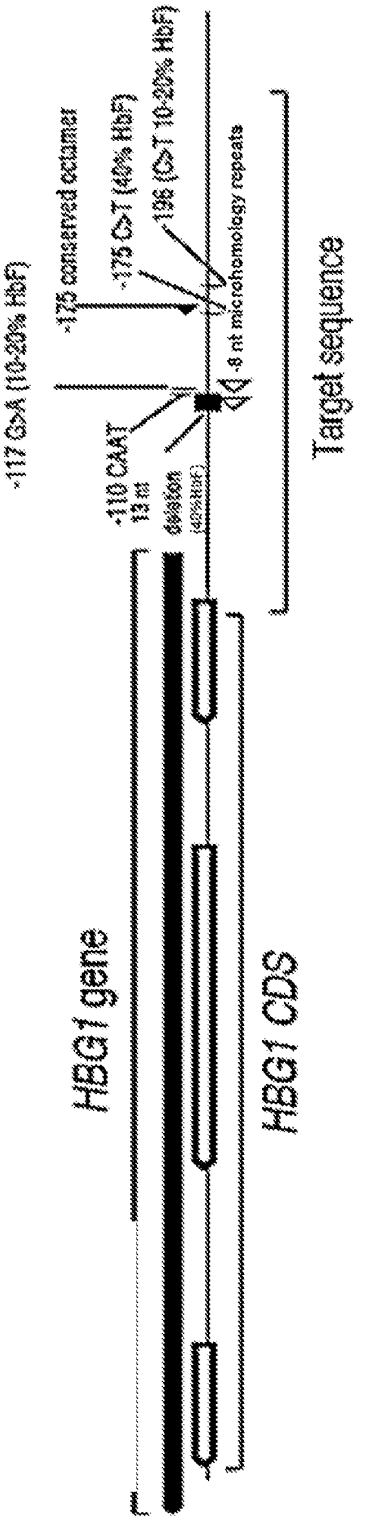

-117 C>A (10-20% HbF)
-175 conserved octamer
-175 C>T (40% HbF)
-196 (C>T 10-20% HbF)
-110 CAAT
13 nt
deletion
8 nt microhomology repeats
Target sequence
HBG1 gene
HBG1 CDS

FIG 2A

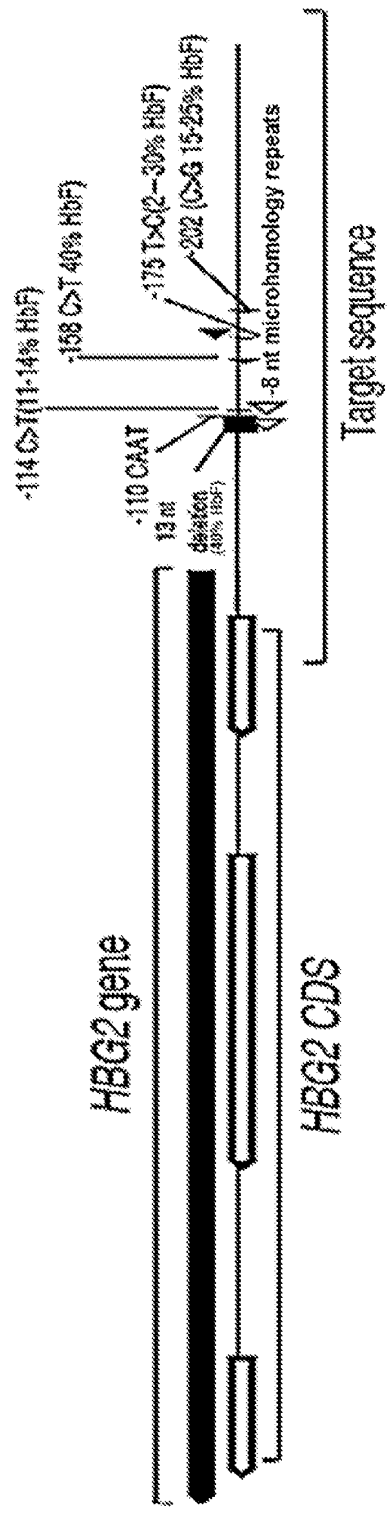

-114 C>T(11-14% HbF)
-158 C>T 40% HbF)
-175 T>C(2–30% HbF)
-202 (C>G 15-25% HbF)
-110 CAAT
13 nt
deletion
8 nt microhomology repeats
Target sequence
HBG2 gene
HBG2 CDS

FIG 2B

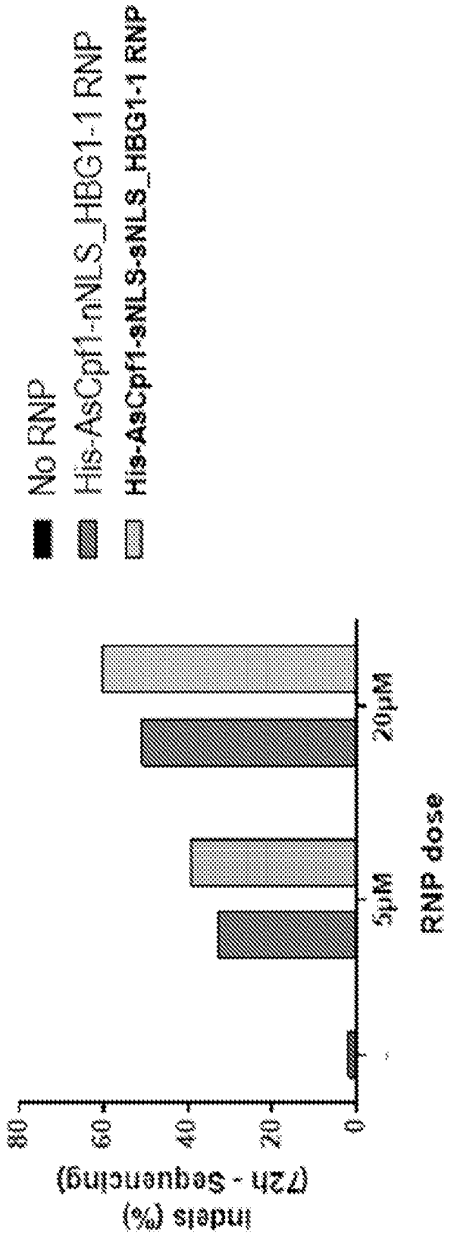

No RNP
His-AsCpf1-nNLS_HBG1-1 RNP
His-AsCpf1-sNLS-sNLS_HBG1-1 RNP
indels (%)
(72h - Sequencing)
0
20
40
60
80
5μM
20μM
RNP dose

FIG 3

FIG 4A

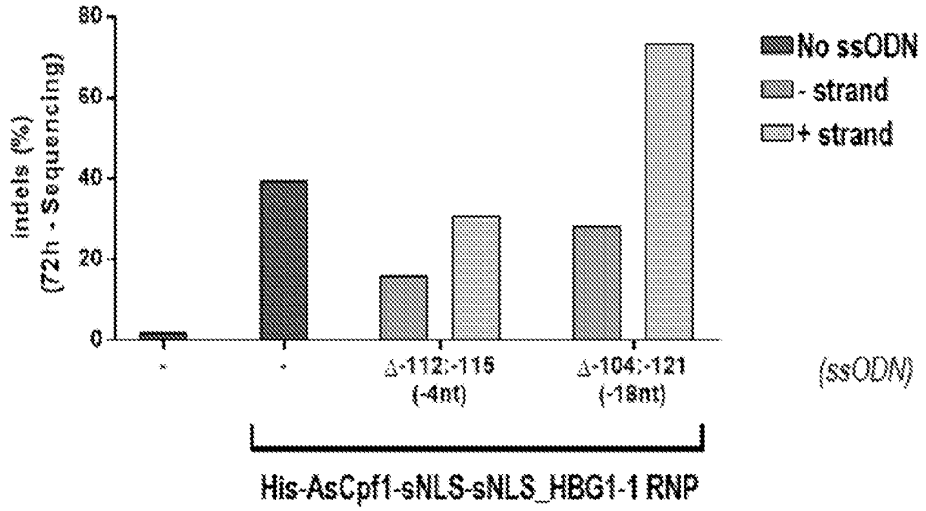

80
60
40
20
0
Indels (%)
(72h - Sequencing)
No ssODN
- strand
+ strand
-
-
Δ-112:-116
(-4nt)
Δ-104:-121
(-18nt)
(ssODN)
His-AsCpf1-sNLS-sNLS_HBG1-1 RNP

FIG 4B

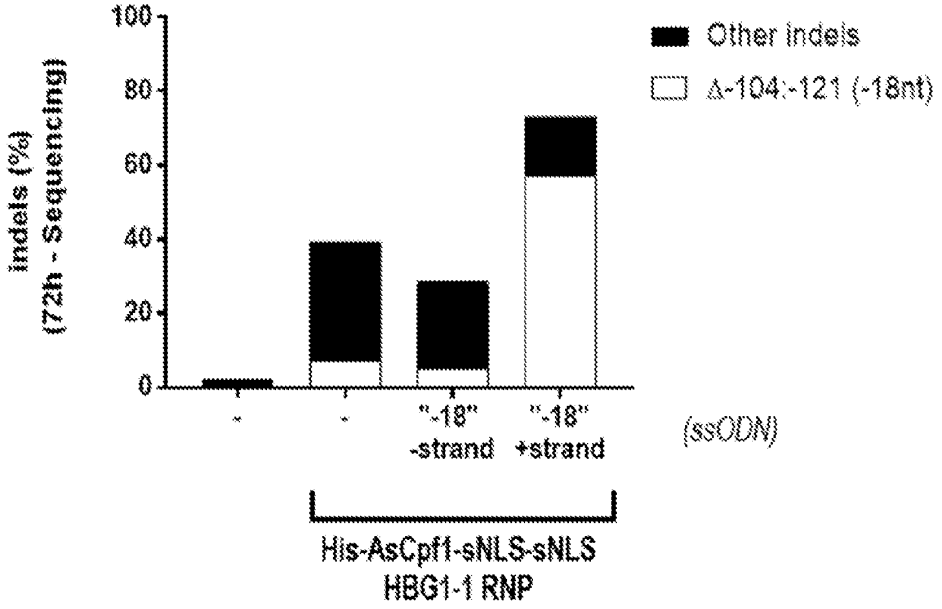

100
80
60
40
20
0
Indels (%)
(72h - Sequencing)
Other Indels
Δ-104:-121 (-18nt)
-
-
"-18"
-strand
"-18"
+strand
(ssODN)
His-AsCpf1-sNLS-sNLS
HBG1-1 RNP

FIG 4C

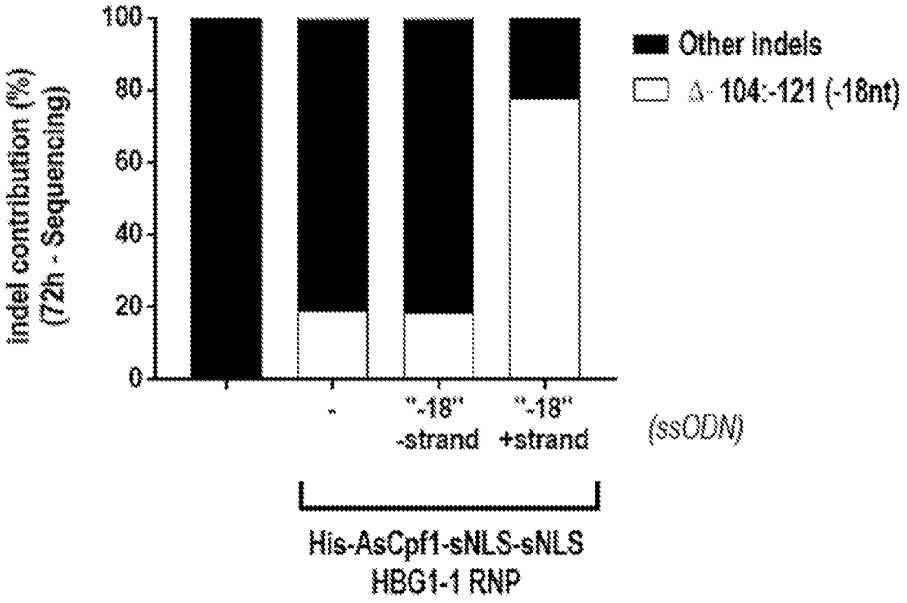

indel contribution (%)
(72h - Sequencing)
100
80
60
40
20
0
Other indels
Δ -104:-121 (-18nt)
-
"-18"
-strand
"-18"
+strand
(ssODN)
His-AsCpf1-sNLS-sNLS
HBG1-1 RNP

FIG 5

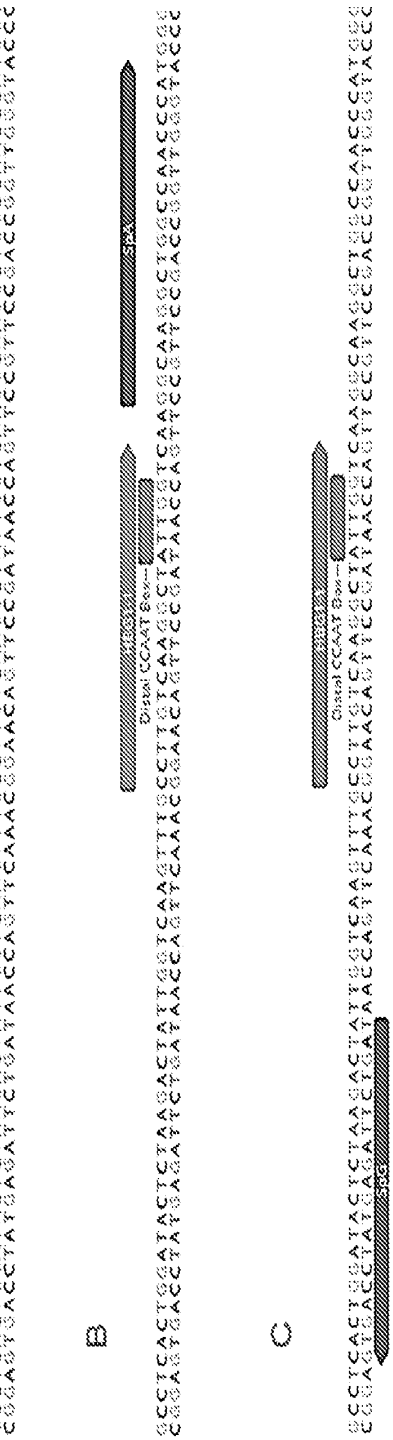


FIG 5 (cont'd)

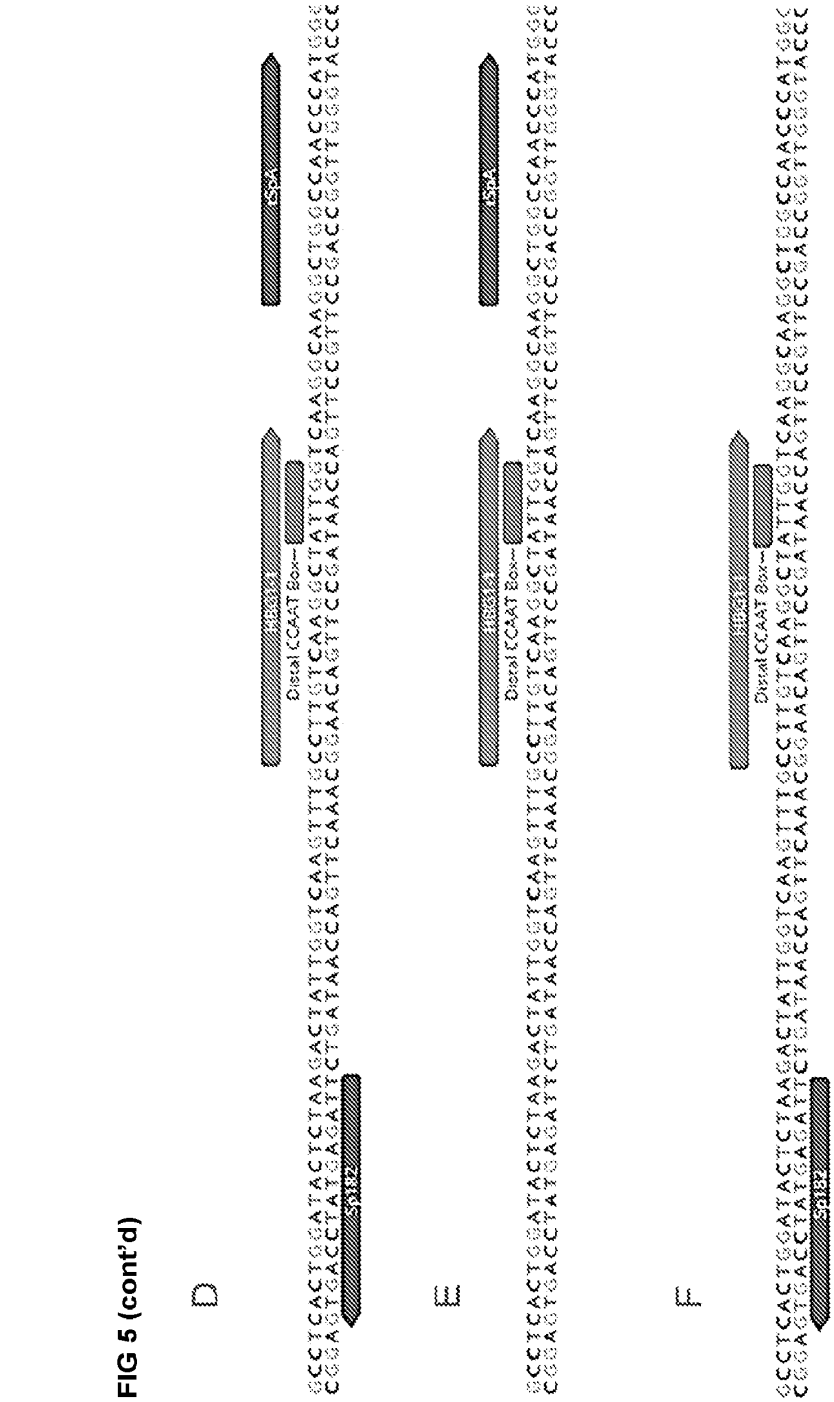


FIG 6A

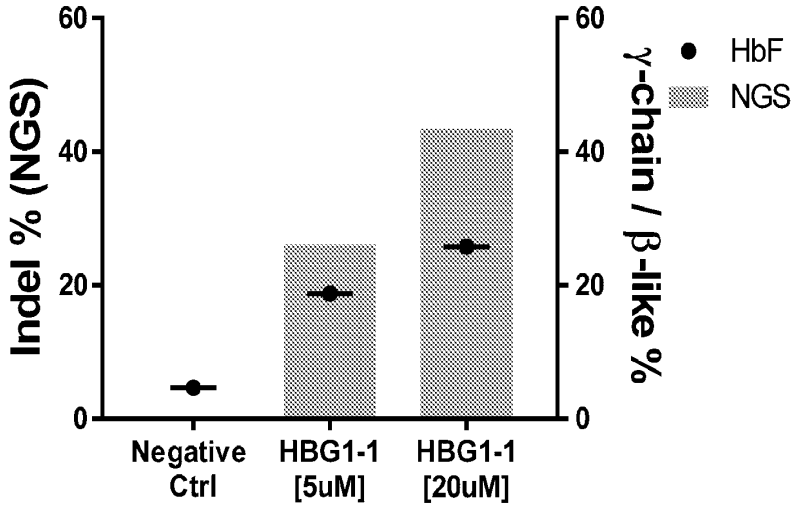
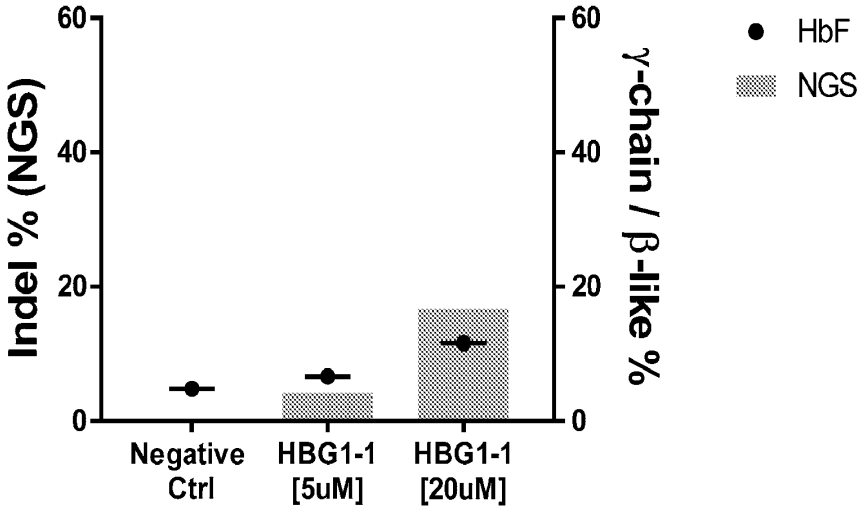

60
40
20
0
Indel % (NGS)
γ-chain / β-like %
HbF
NGS
Negative Ctrl
HBG1-1 [5uM]
HBG1-1 [20uM]

FIG 6B


60
40
20
0
Indel % (NGS)
γ-chain / β-like %
HbF
NGS
Negative Ctrl
HBG1-1 [5uM]
HBG1-1 [20uM]

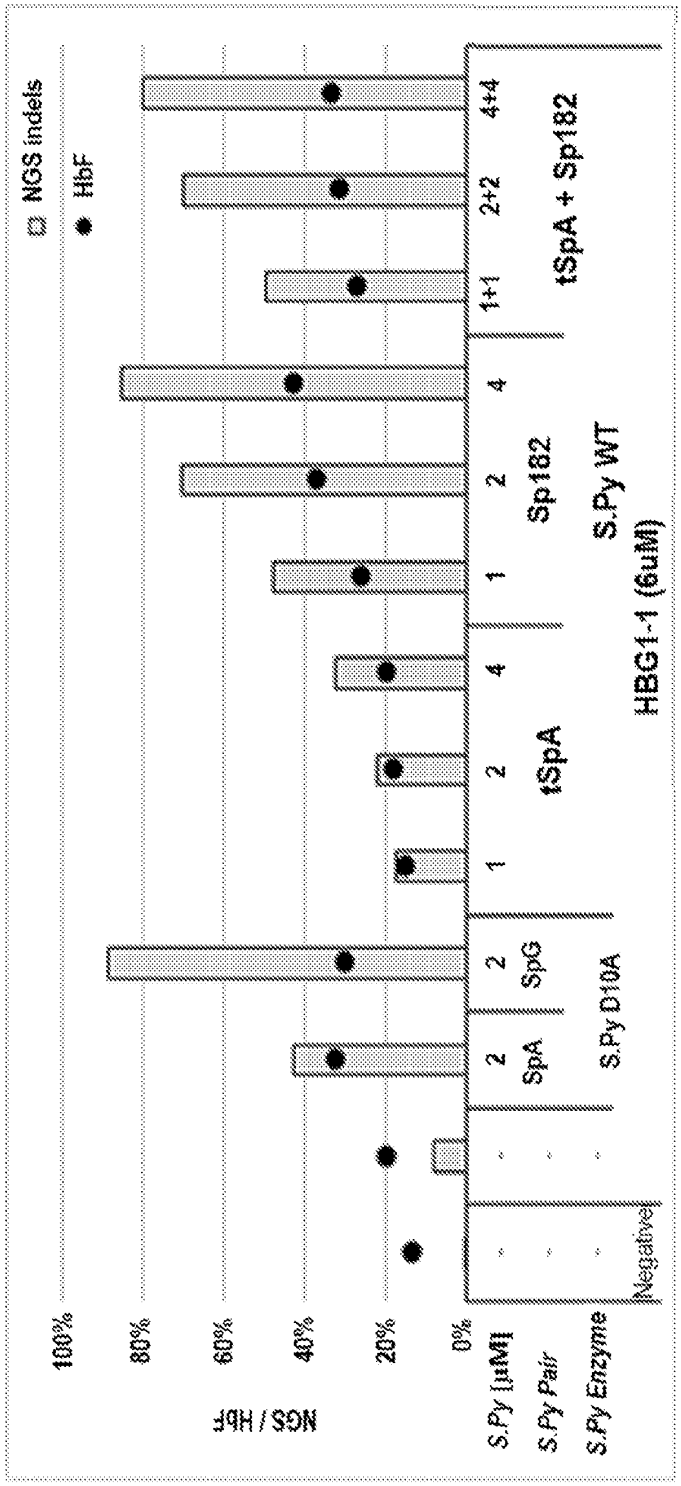

NGS indels
HbF
NGS / HbF
100%
80%
60%
40%
20%
0%
S.Py [uM]
S.Py Pair
S.Py Enzyme
Negative
SpA
SpG
S.Py D10A
tSpA
Sp182
S.Py WT
tSpA + Sp182
HBG1-1 (6uM)

FIG 7

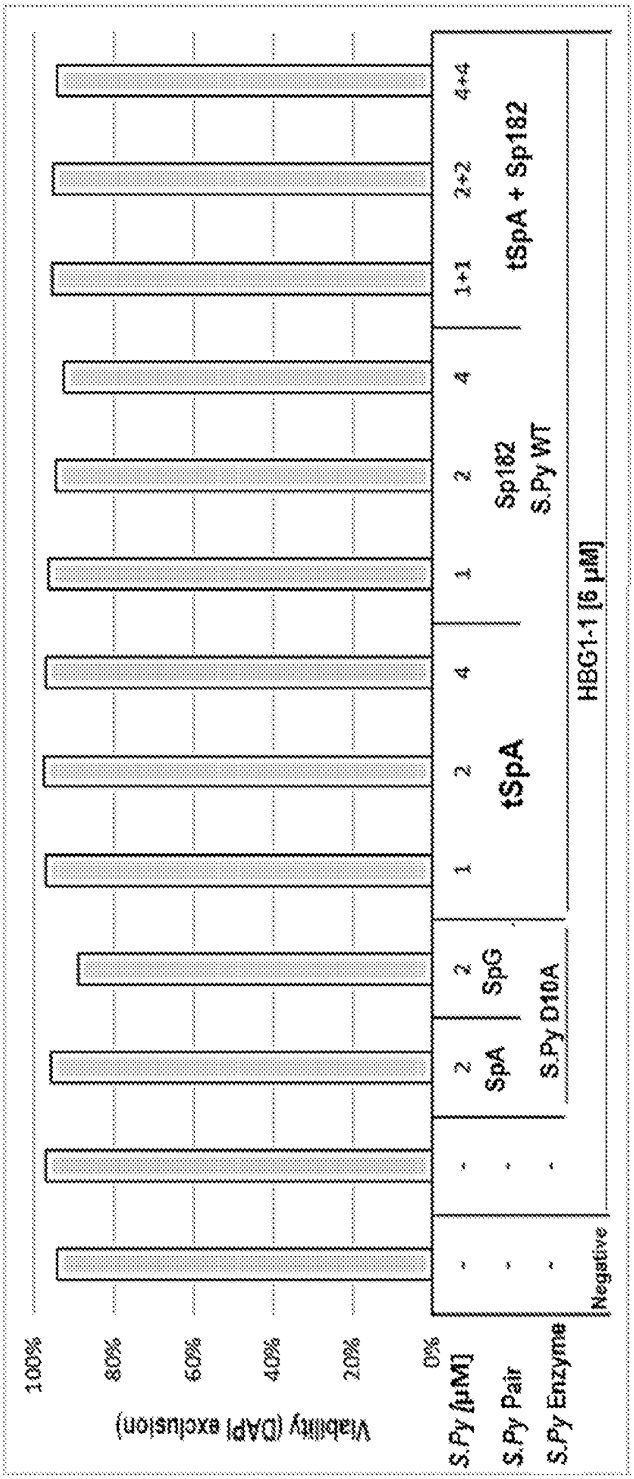
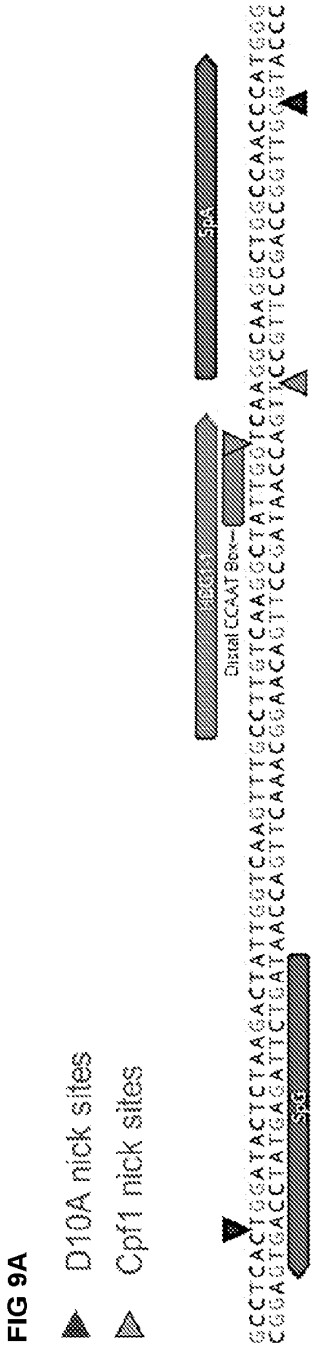

Viability (DAPI exclusion)
100%
80%
60%
40%
20%
0%
S.Py [µM]
S.Py Pair
S.Py Enzyme
-
-
-
Negative
-
-
-
2
SpA
S.Py D10A
2
SpG
1
2
4
tSpA
1
2
4
Sp182
S.Py WT
1+1
2+2
4+4
tSpA + Sp182
HBG1-1 [6 µM]

FIG 8

FIG 9A


D10A nick sites
Cpf1 nick sites
Distal CCAAT Box
SpCas9
AsCpf1

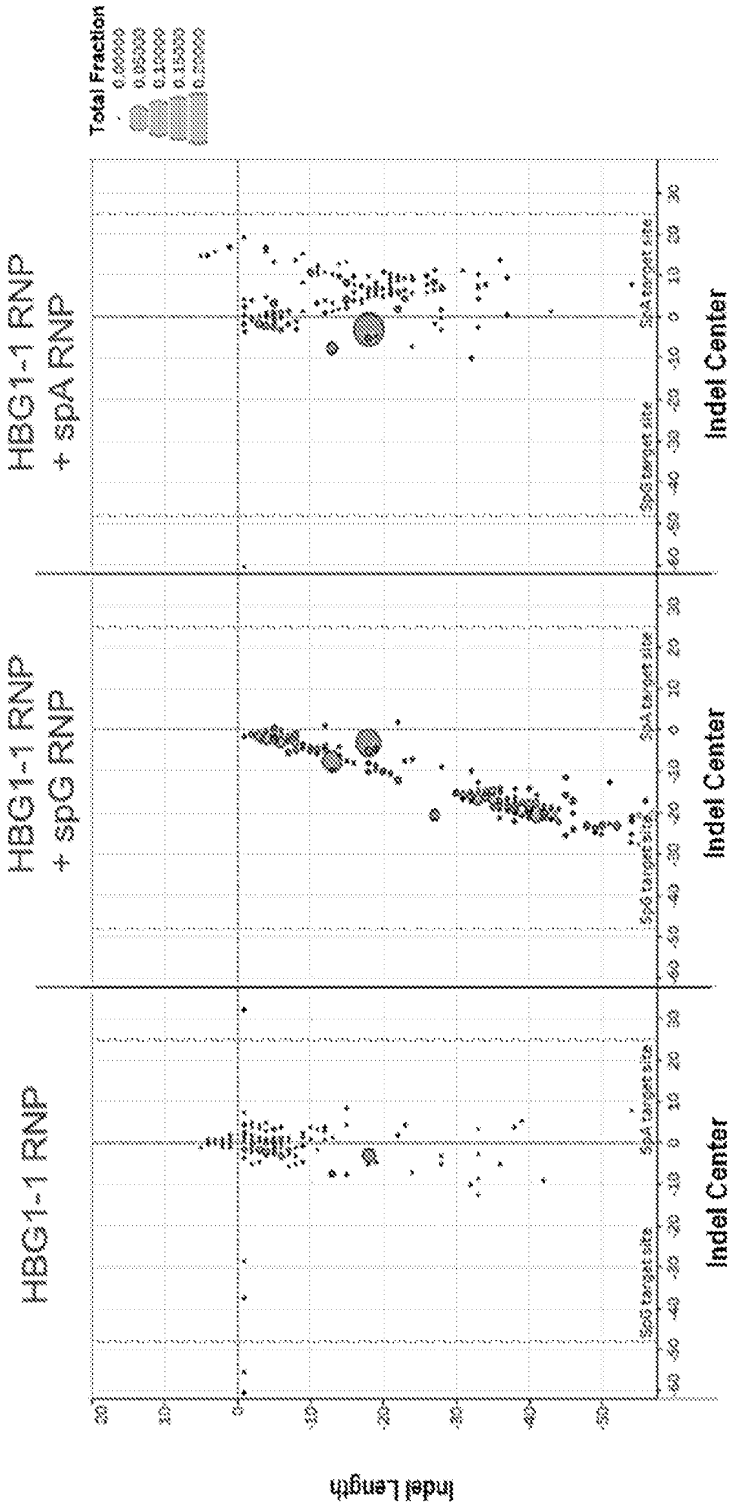

Total Fraction
0.00000
0.05000
0.10000
0.15000
0.20000
HBG1-1 RNP
HBG1-1 RNP + spG RNP
HBG1-1 RNP + spA RNP
Indel Center
SpG target site
SpA target site
Indel Length

FIG 9B

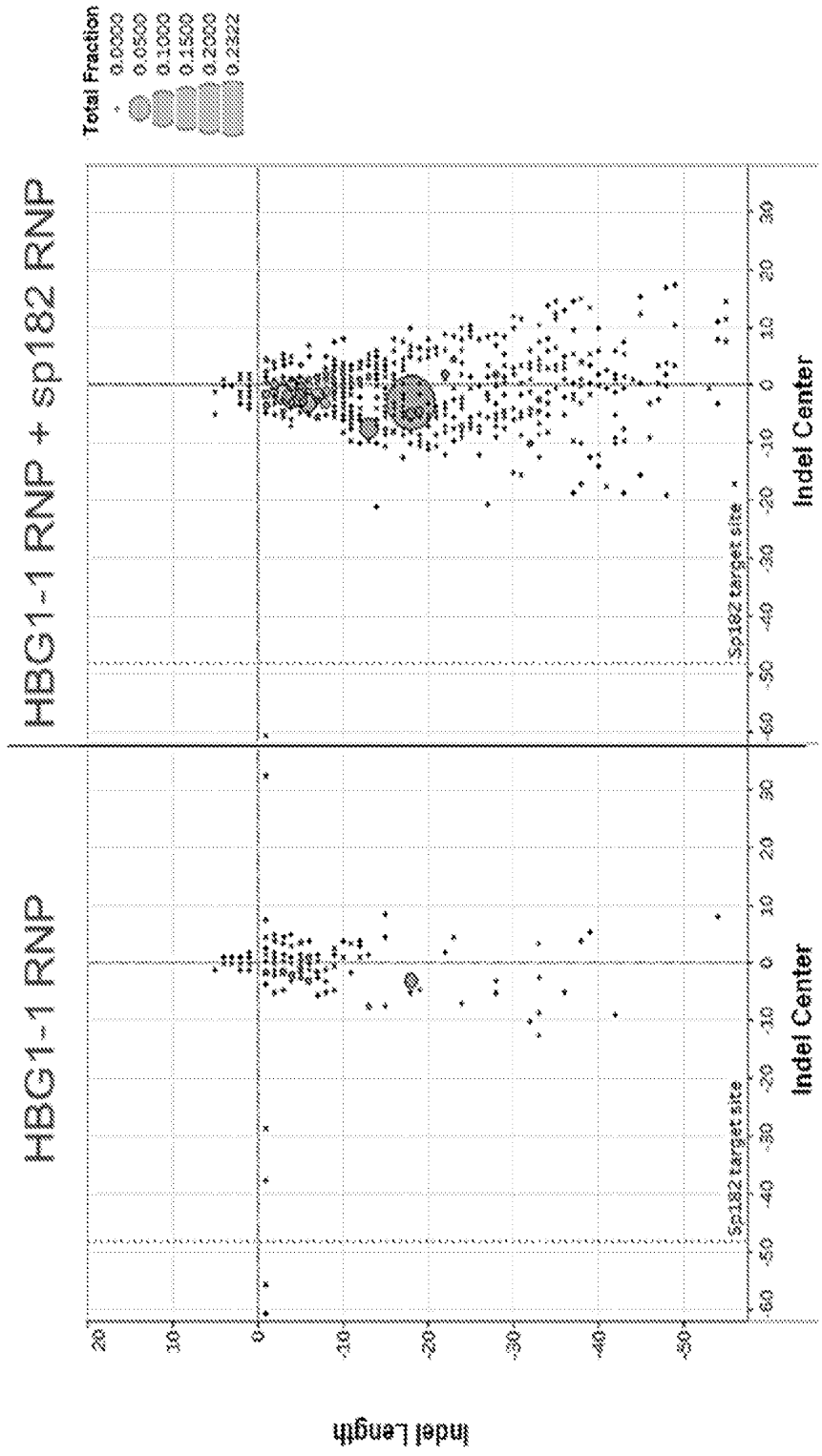

Total Fraction
0.0000
0.0500
0.1000
0.1500
0.2000
0.2322
HBG1-1 RNP + sp182 RNP
HBG1-1 RNP
Sp182 target site
Indel Center
Indel Length

FIG 10A

FIG 10B

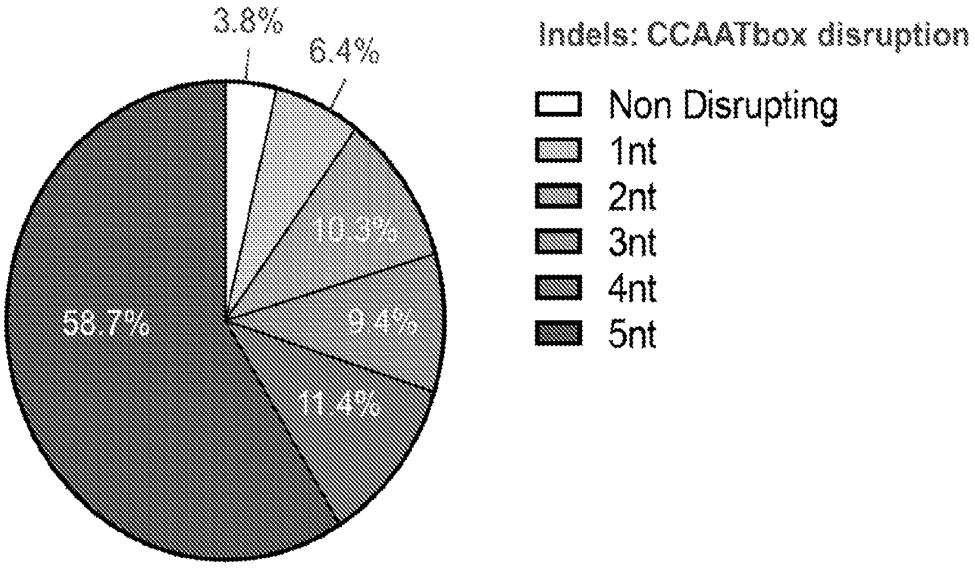

3.8%
6.4%
10.3%
9.4%
11.4%
58.7%
Indels: CCAATbox disruption
Non Disrupting
1nt
2nt
3nt
4nt
5nt

FIG 11

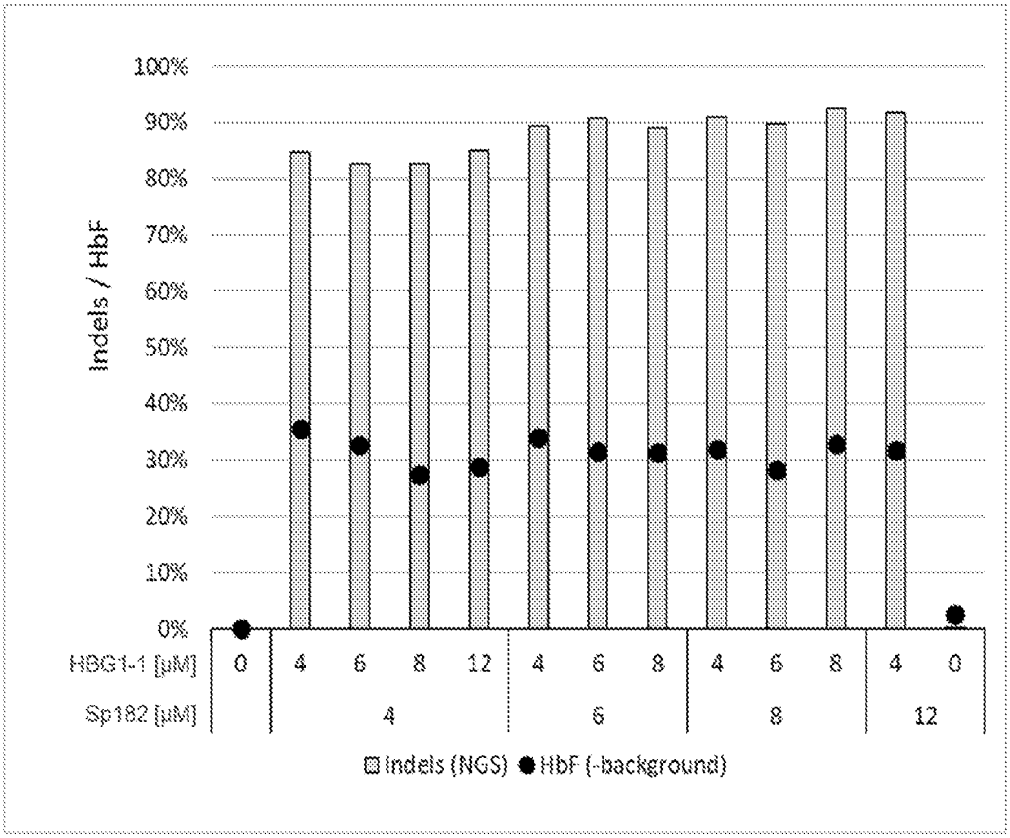
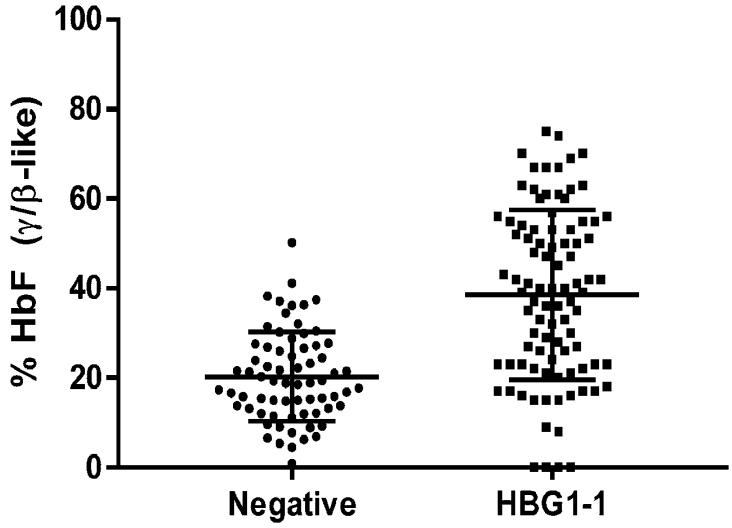

100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Indels / HbF
HBG1-1 [μM]
0 4 6 8 12 4 6 8 4 6 8 4 0
Sp182 [μM]
4 6 8 12
Indels (NGS)
HbF (-background)

FIG 12



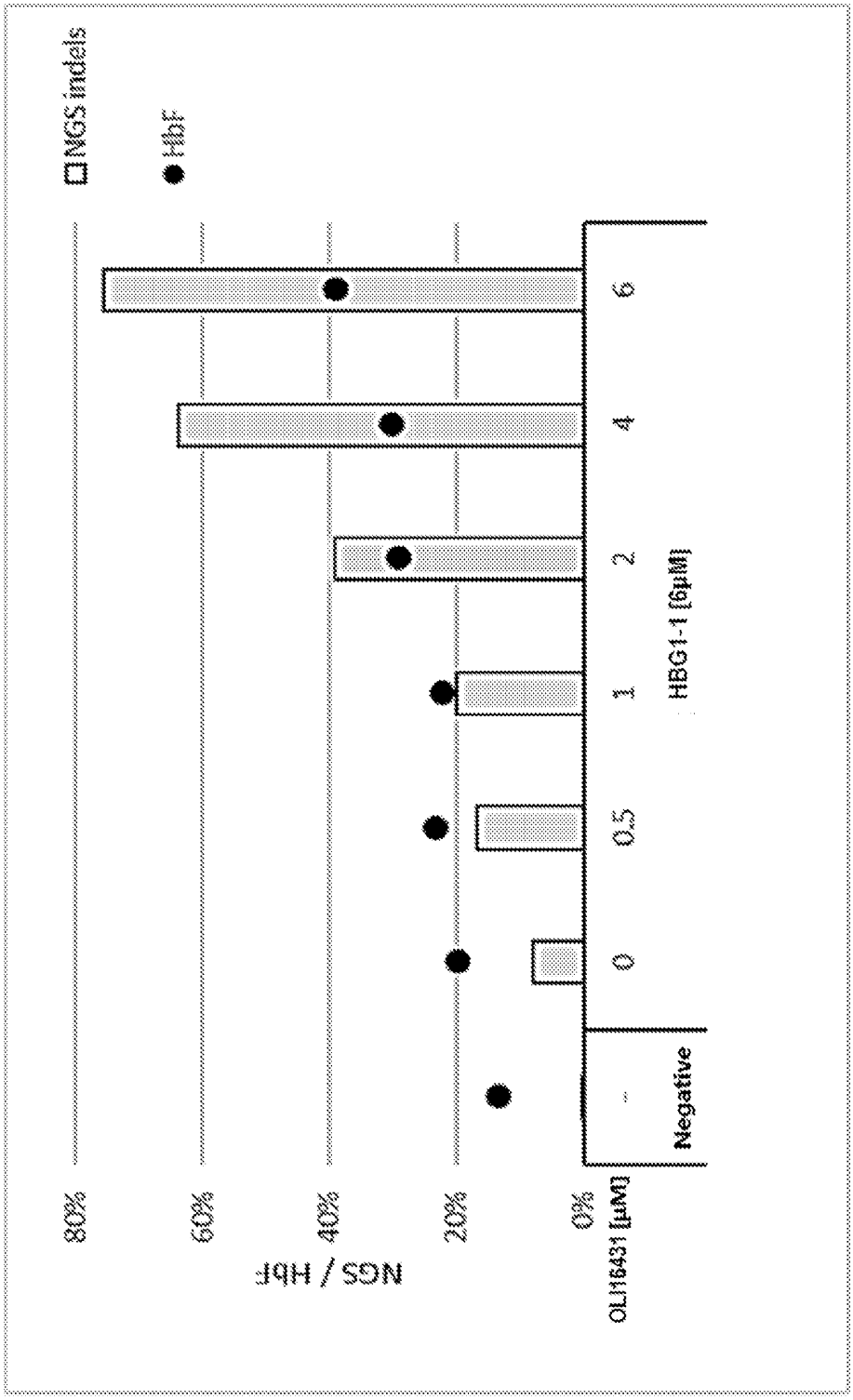

NGS indels
HbF
NGS / HbF
80%
60%
40%
20%
0%
OLI16431 [µM]
-
0
0.5
1
2
4
6
HBG1-1 [6µM]
Negative

FIG 13A

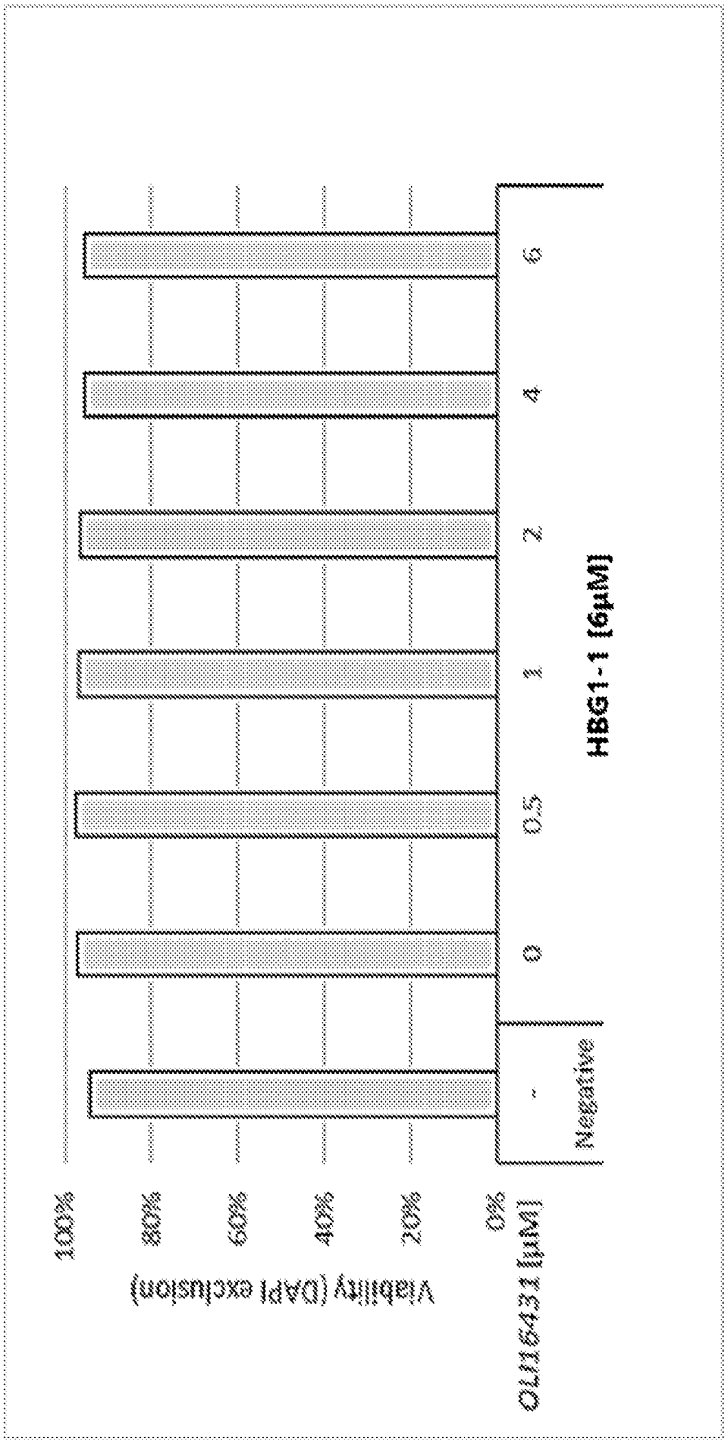

HBG1-1 [6μM]
OLI16431 [μM]
Viability (DAPI exclusion)
100%
80%
60%
40%
20%
0%
-
Negative
0
0.5
1
2
4
6

FIG 13B

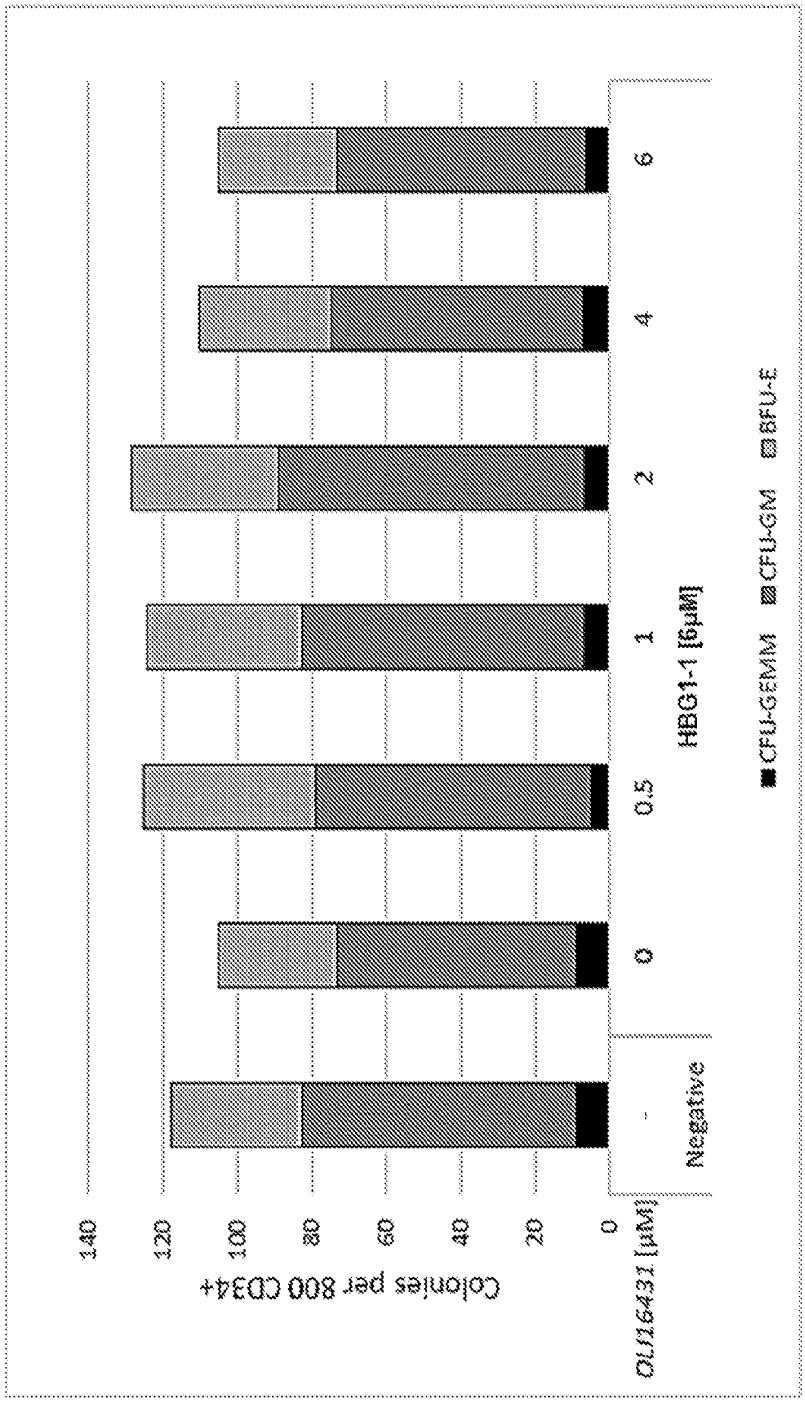

Colonies per 800 CD34+
140
120
100
80
60
40
20
0
OLI16431 [µM]
-
0
0.5
1
2
4
6
Negative
HBG1-1 [6µM]
CFU-GEMM
CFU-GM
BFU-E

FIG 13C

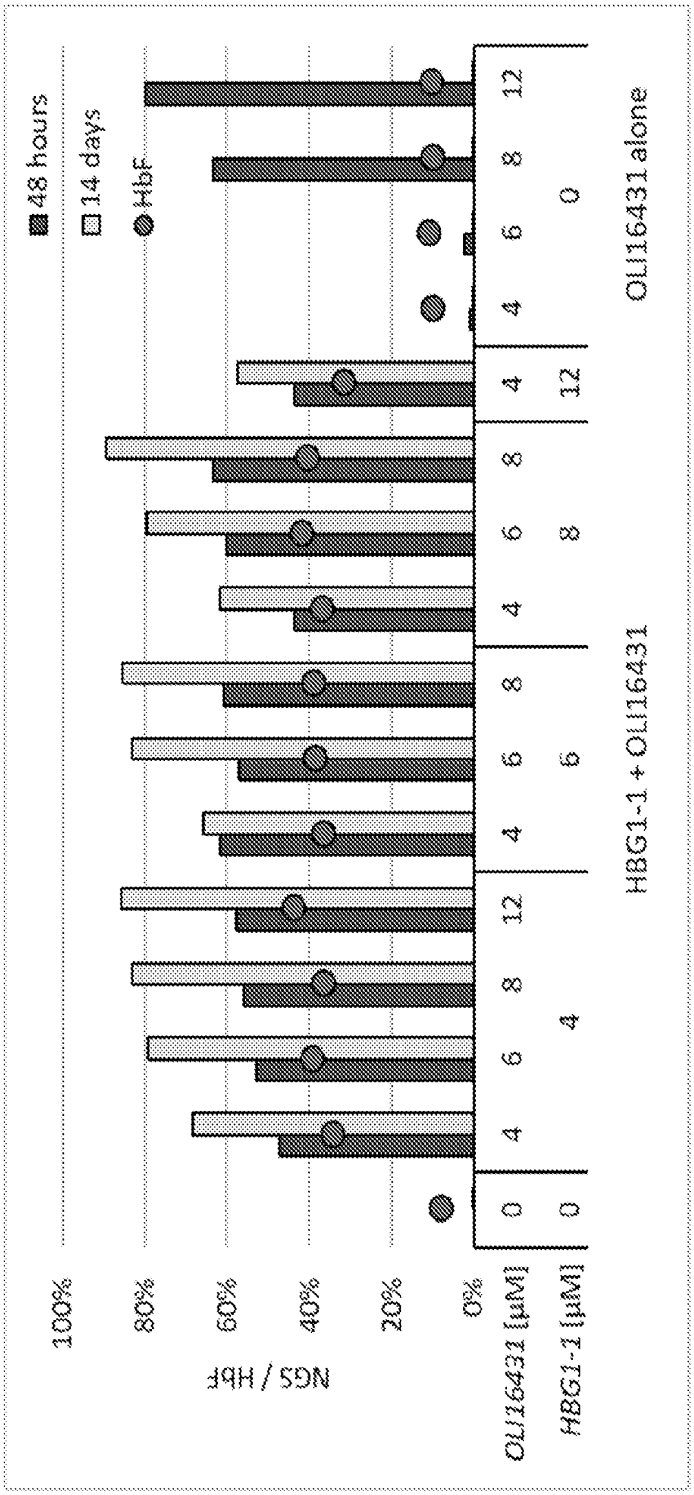

48 hours
14 days
HbF
100%
80%
60%
40%
20%
0%
NGS / HbF
OLI16431 [μM]
HBG1-1 [μM]
0
4
6
8
12
4
6
8
12
4
6
8
12
4
6
8
12
0
4
6
8
12
HBG1-1 + OLI16431
OLI16431 alone

FIG 14A

FIG 14B

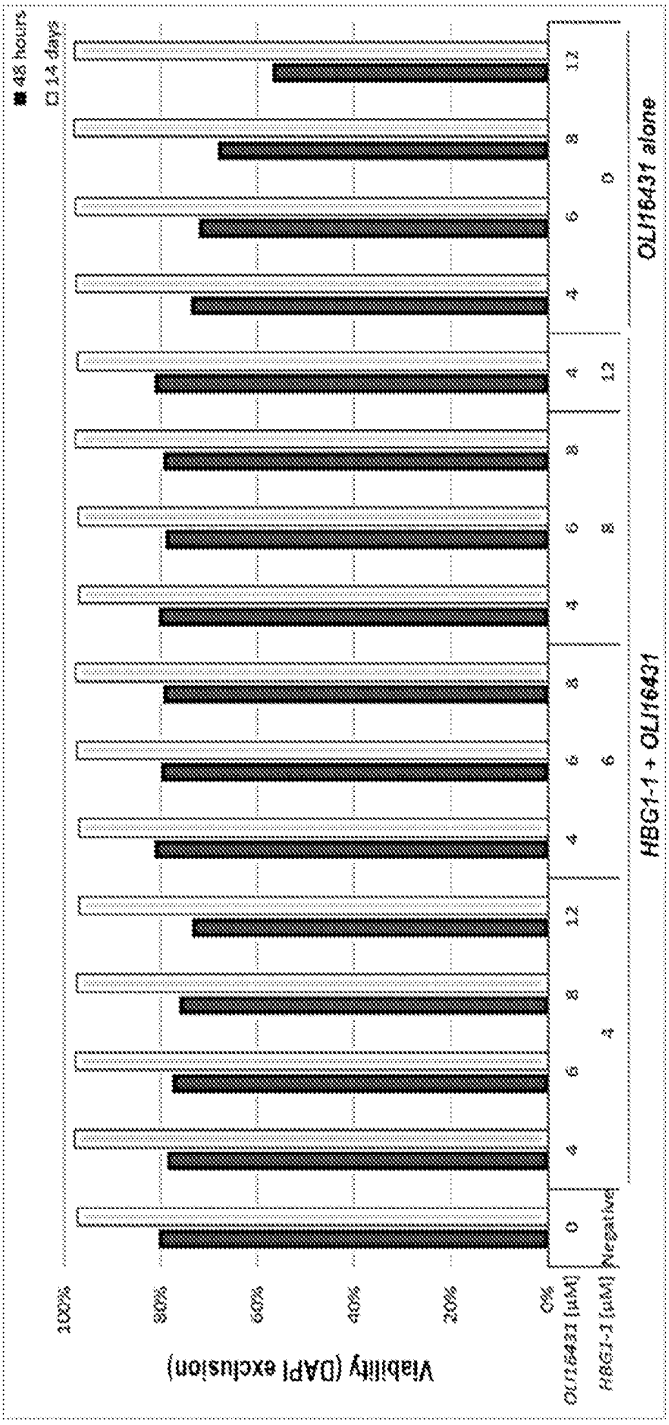

48 hours
14 days
Viability (DAPI exclusion)
100%
80%
60%
40%
20%
0%
OLI16431 (μM)
HBG1-1 (μM)
Negative
0
4
6
8
12
4
6
8
12
4
6
8
12
4
6
8
12
0
HBG1-1 + OLI16431
OLI16431 alone

FIG 15

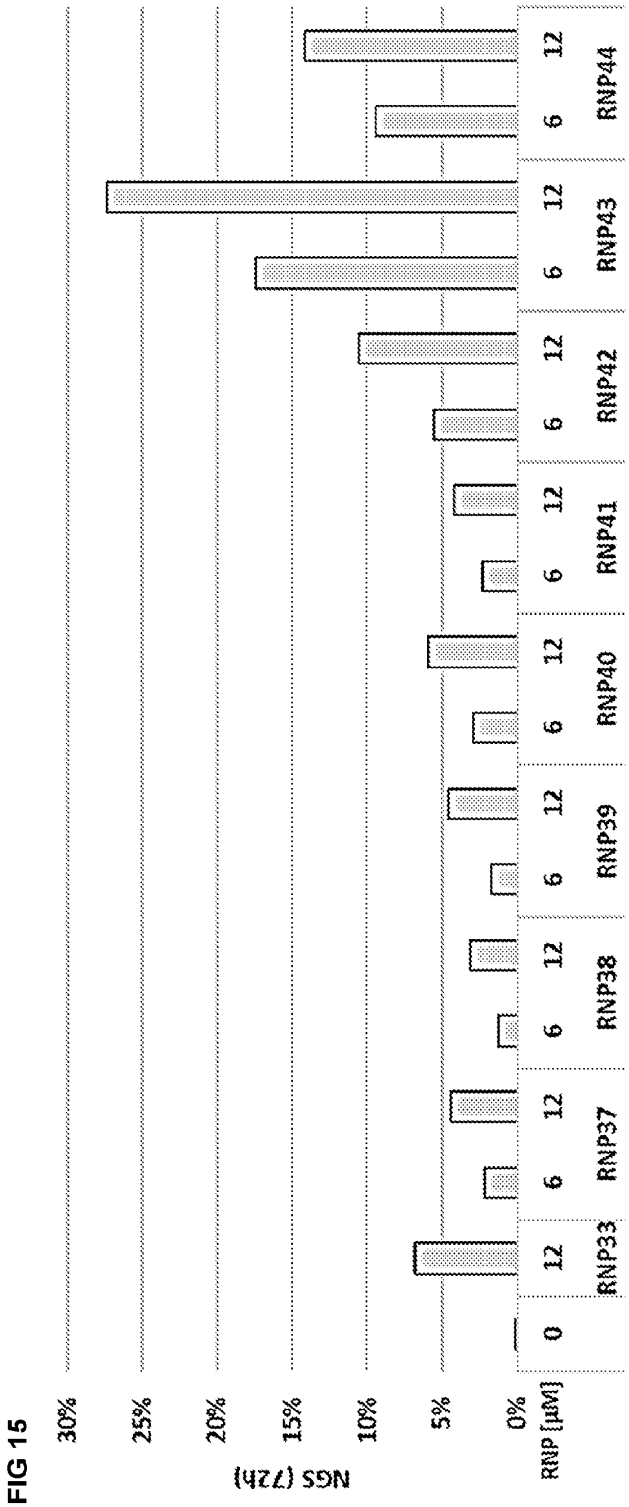

NGS (72h)
30%
25%
20%
15%
10%
5%
0%
RNP [μM]
0
12
6
12
RNP33
6
12
RNP37
6
12
RNP38
6
12
RNP39
6
12
RNP40
6
12
RNP41
6
12
RNP42
6
12
RNP43
6
12
RNP44

FIG 16A

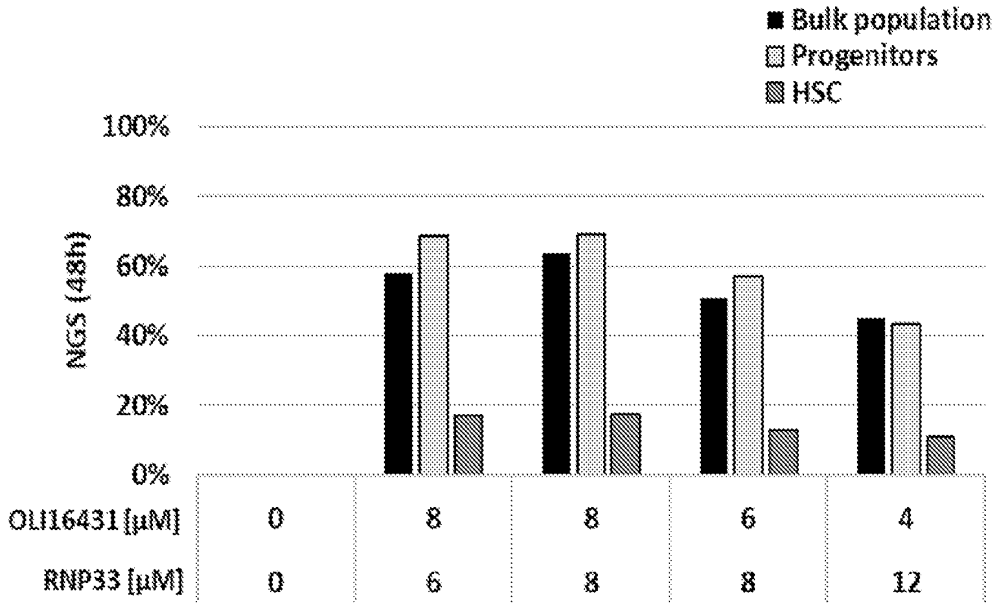


FIG 16B

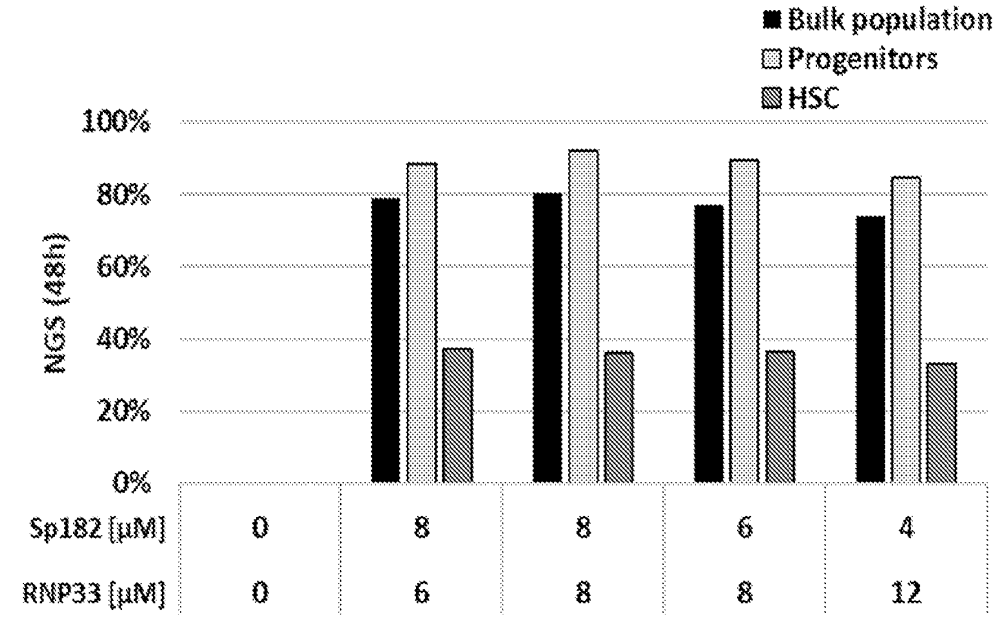
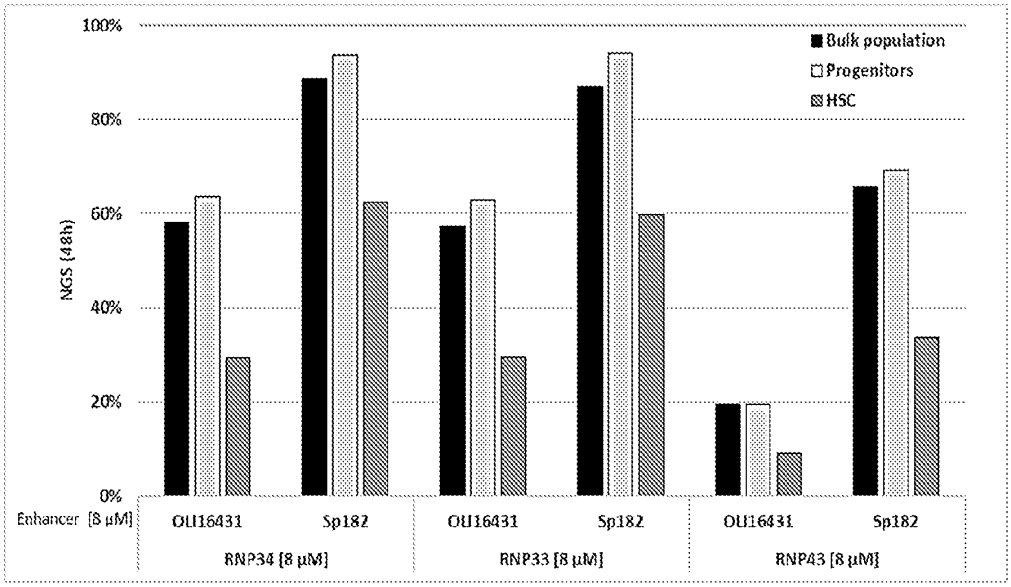


FIG 17


100%
80%
60%
40%
20%
0%
NGS (48h)
Bulk population
Progenitors
HSC
Enhancer [8 µM]
OU16431
Sp182
OU16431
Sp182
OU16431
Sp182
RNP34 [8 µM]
RNP33 [8 µM]
RNP43 [8 µM]

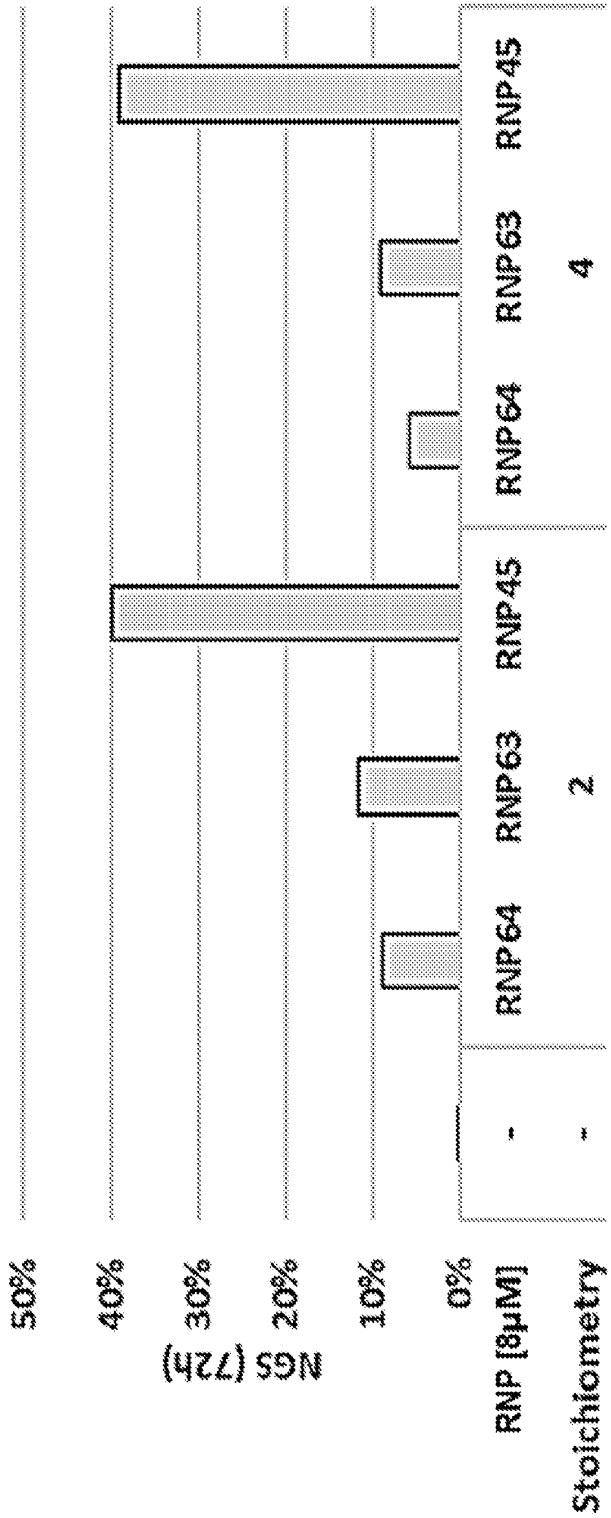
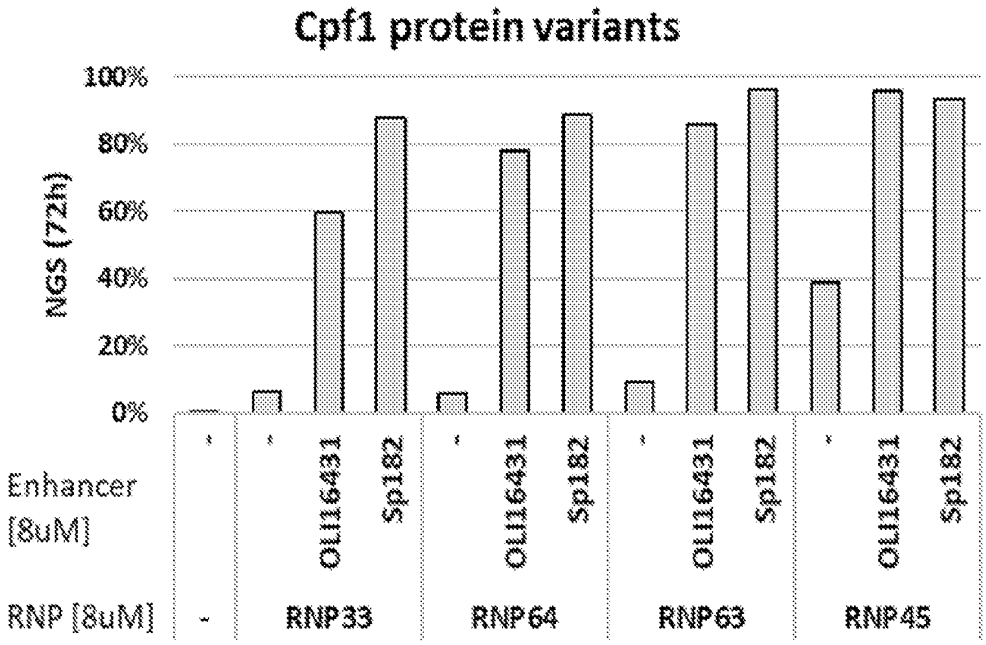

Cpf1 RNP stoichiometry comparison
NGS (72h)
50%
40%
30%
20%
10%
0%
RNP [8μM]
Stoichiometry
-
-
RNP64
RNP63
RNP45
2
RNP64
RNP63
RNP45
4

FIG 18

FIG 19


Cpf1 protein variants
100%
80%
60%
40%
20%
0%
NGS (72h)
-
-
OLI16431
Sp182
-
OLI16431
Sp182
-
OLI16431
Sp182
-
OLI16431
Sp182
Enhancer [8uM]
RNP [8uM]
-
RNP33
RNP64
RNP63
RNP45

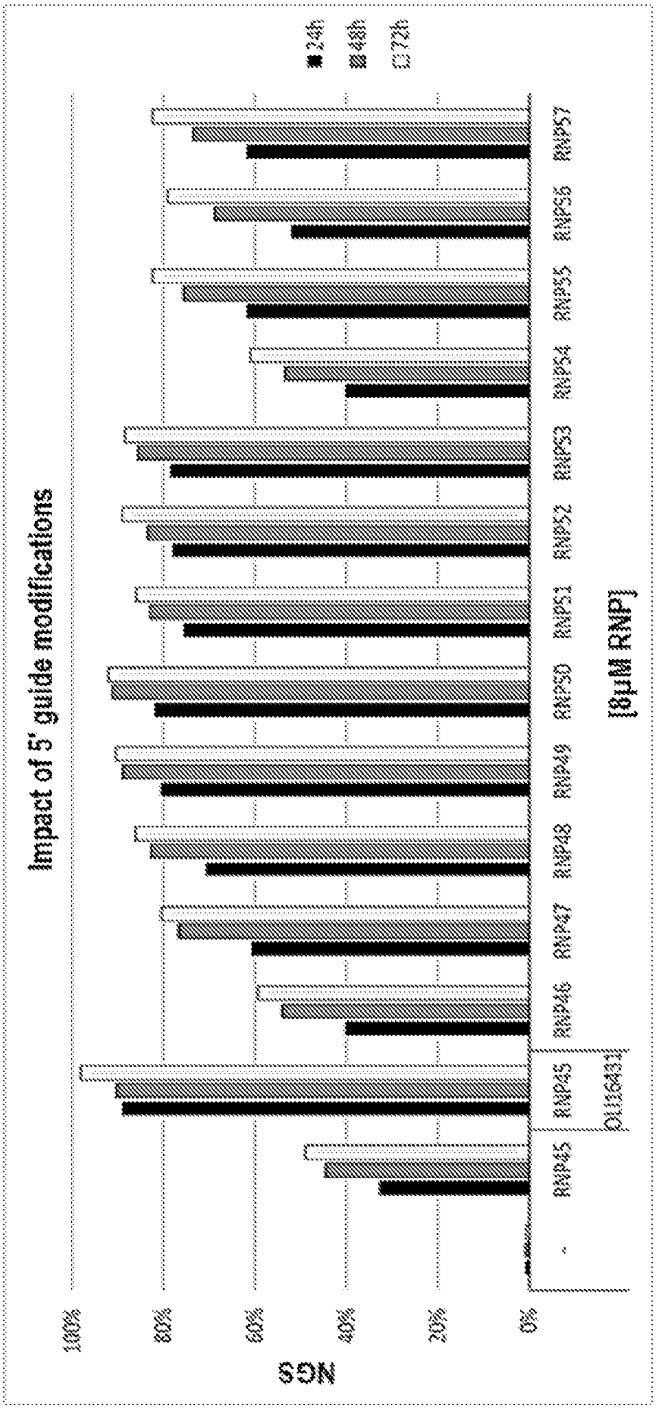
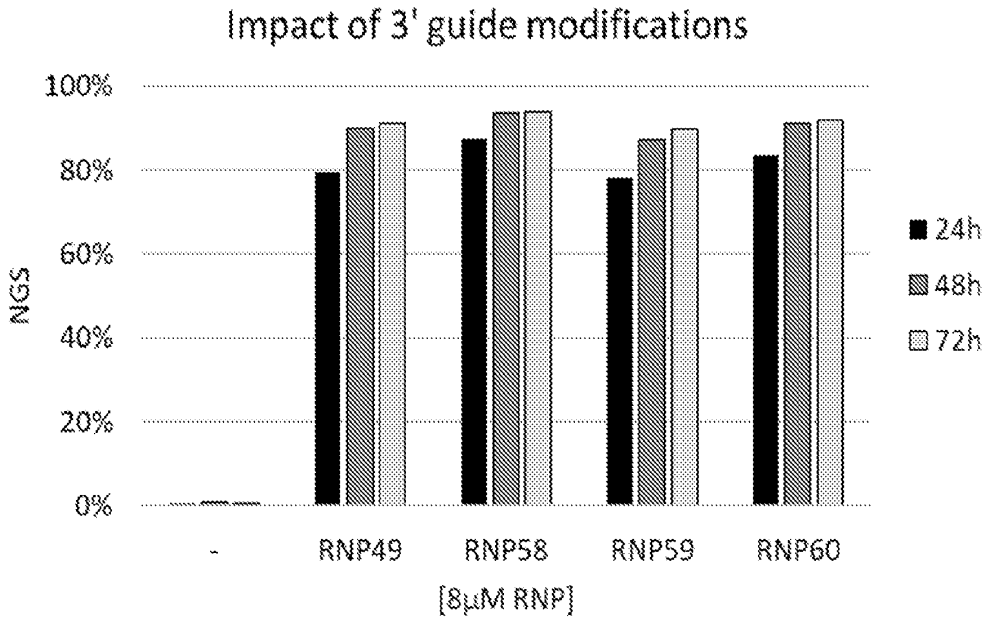

Impact of 5' guide modifications
24h
48h
72h
NGS
100%
80%
60%
40%
20%
0%
-
RNP45
RNP45
RNP46
RNP47
RNP48
RNP49
RNP50
RNP51
RNP52
RNP53
RNP54
RNP55
RNP56
RNP57
[8µM RNP]

FIG 20

FIG 21


Impact of 3' guide modifications
100%
80%
60%
40%
20%
0%
NGS
24h
48h
72h
-
RNP49
RNP58
RNP59
RNP60
[8μM RNP]

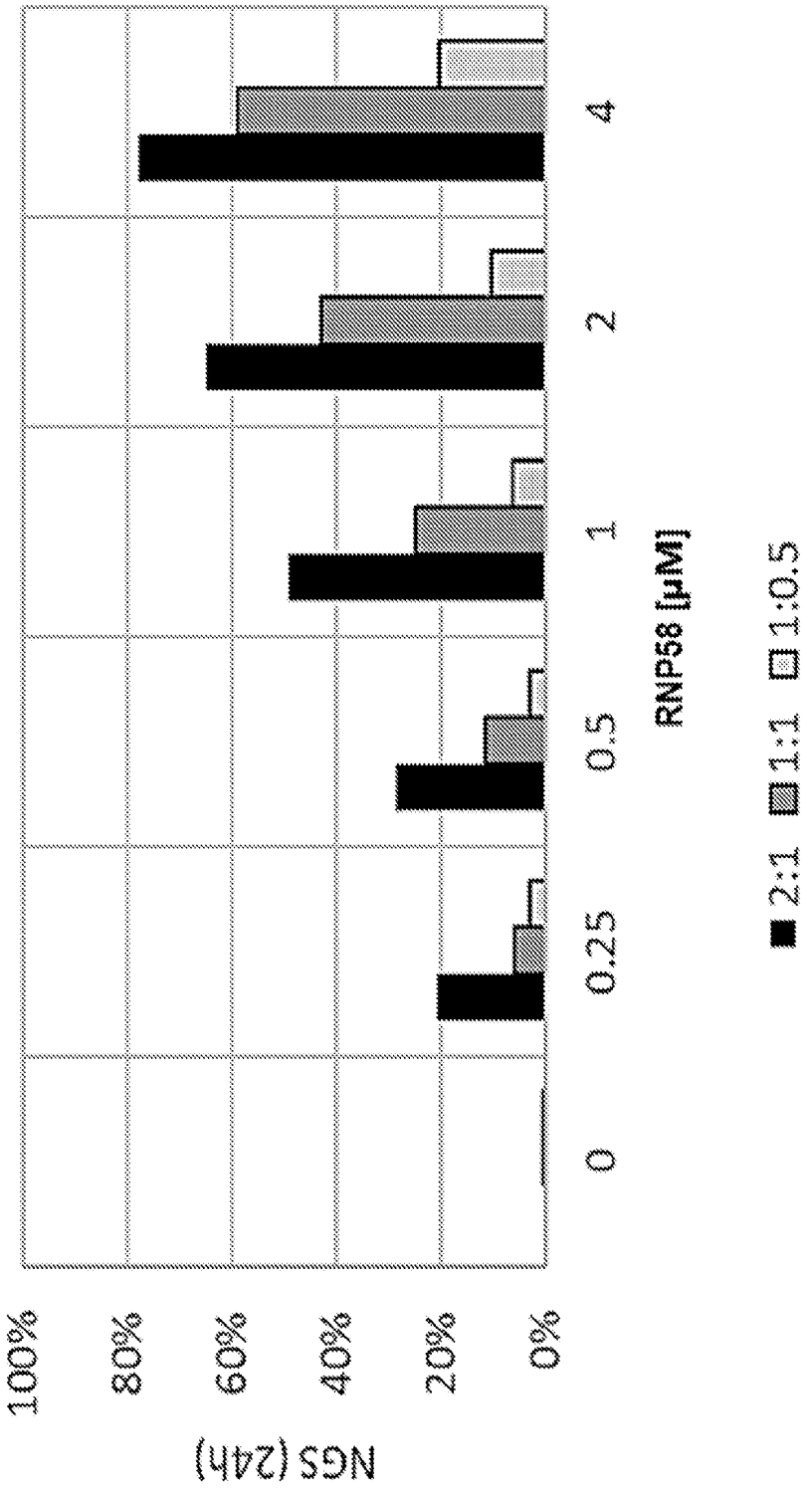


FIG 22A

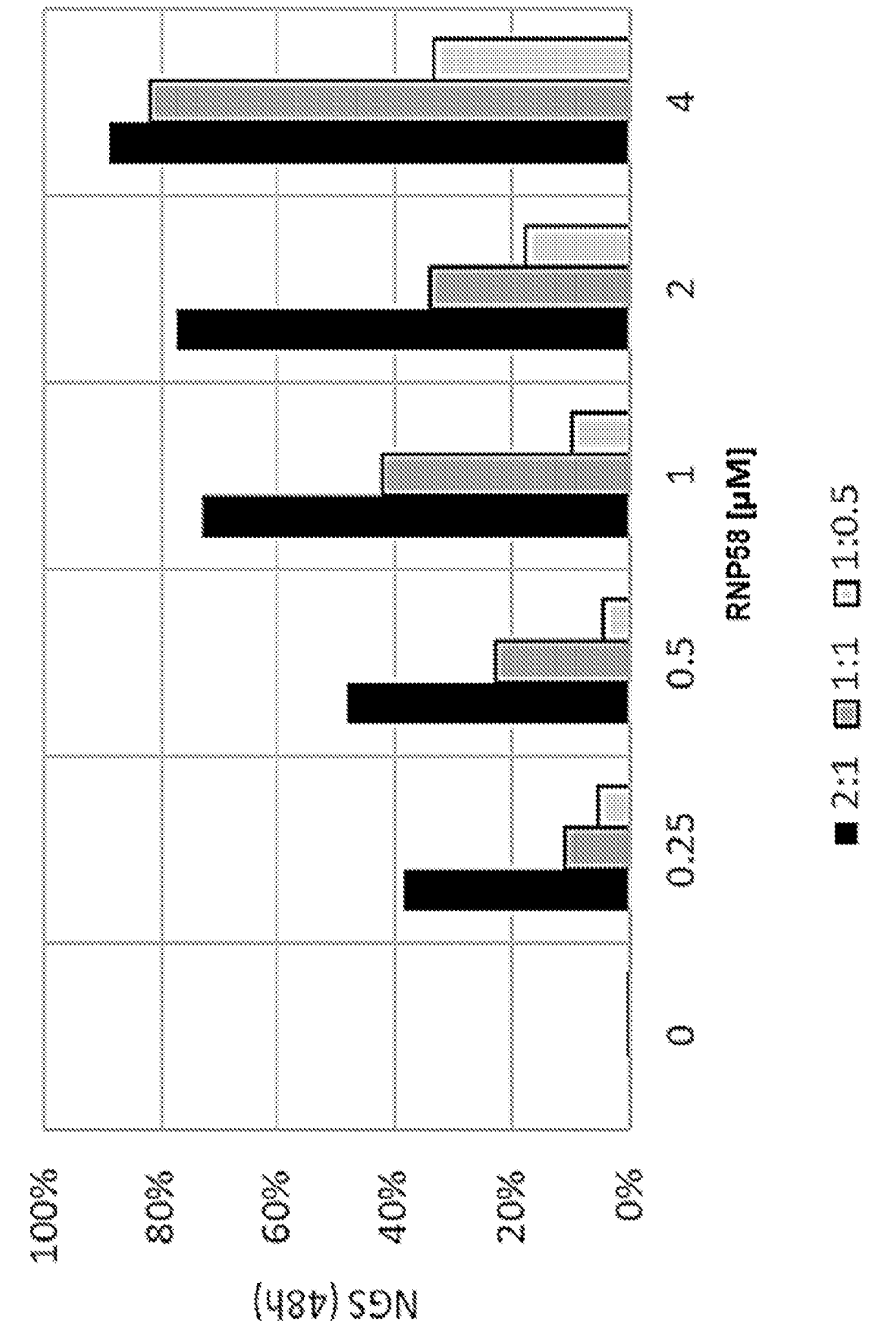


FIG 22B

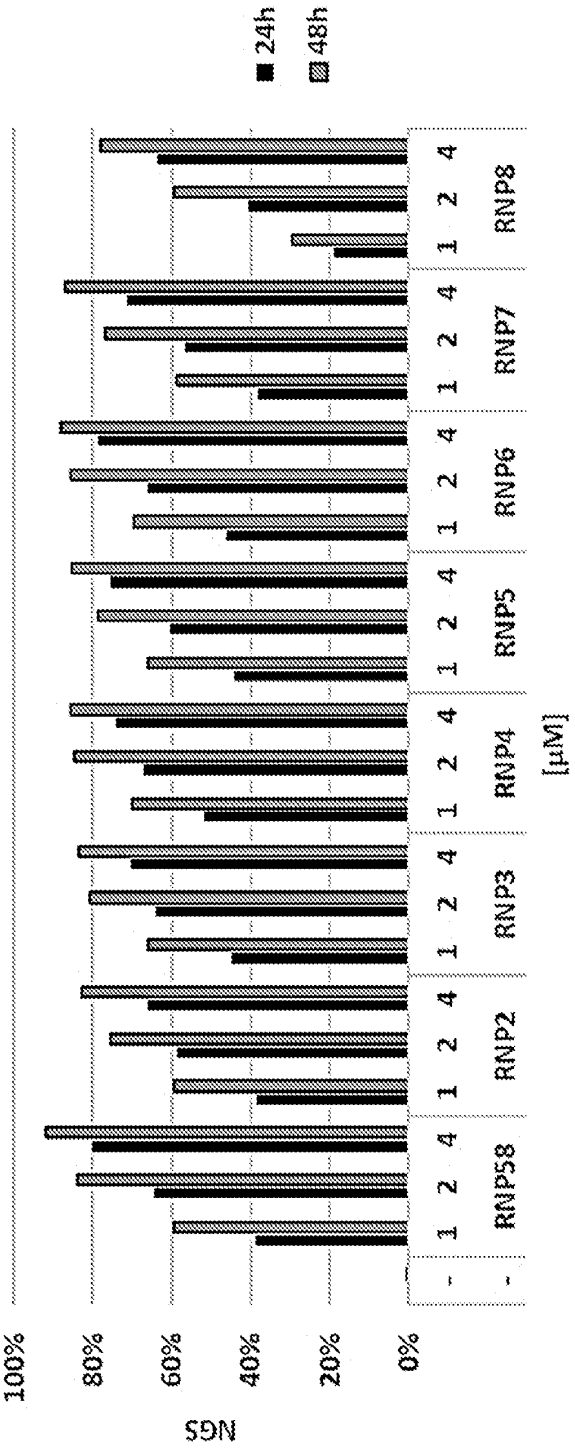

24h
48h
NGS
100%
80%
60%
40%
20%
0%
[μM]
- - RNP58 1 2 4 RNP2 1 2 4 RNP3 1 2 4 RNP4 1 2 4 RNP5 1 2 4 RNP6 1 2 4 RNP7 1 2 4 RNP8

FIG 23A

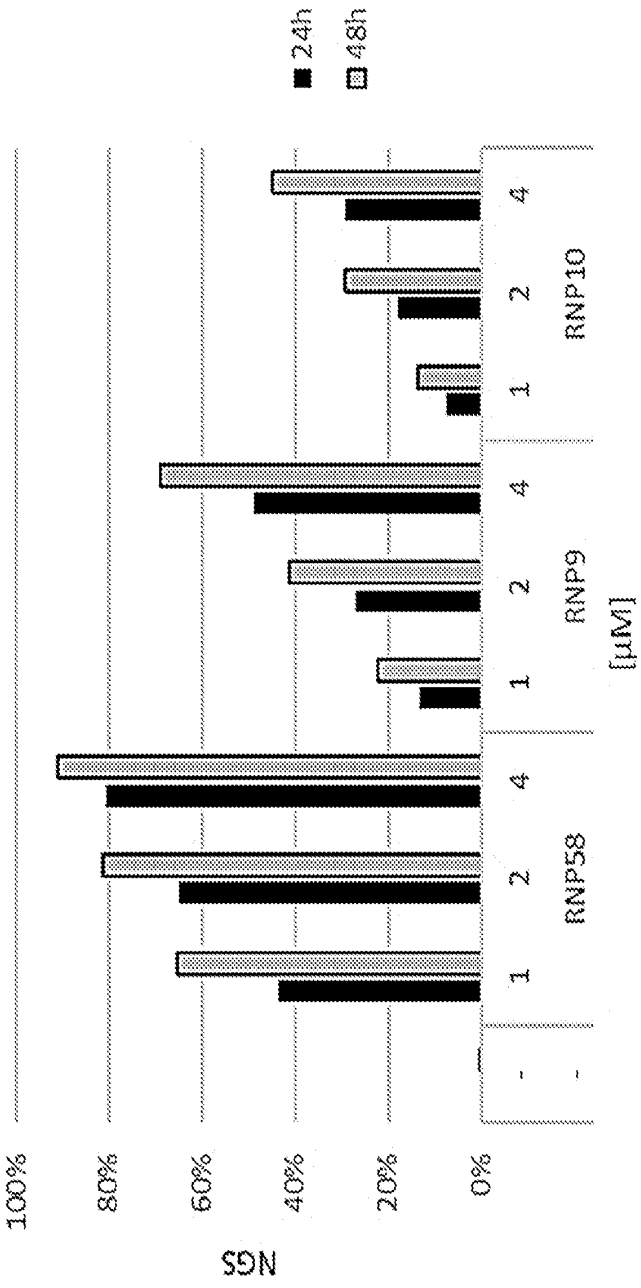


FIG 23B

FIG 24A

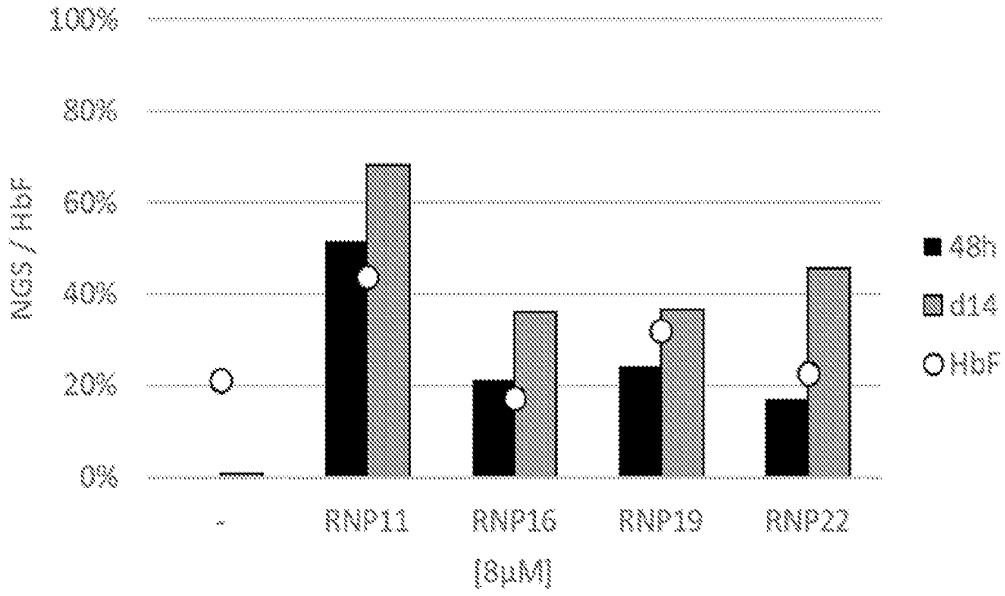


FIG 24B

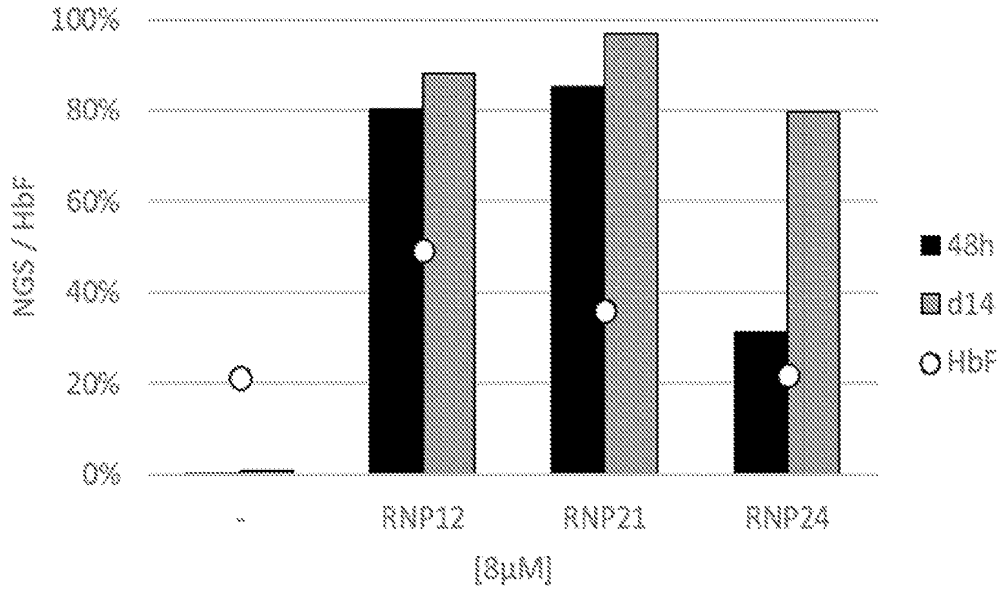


FIG 24C

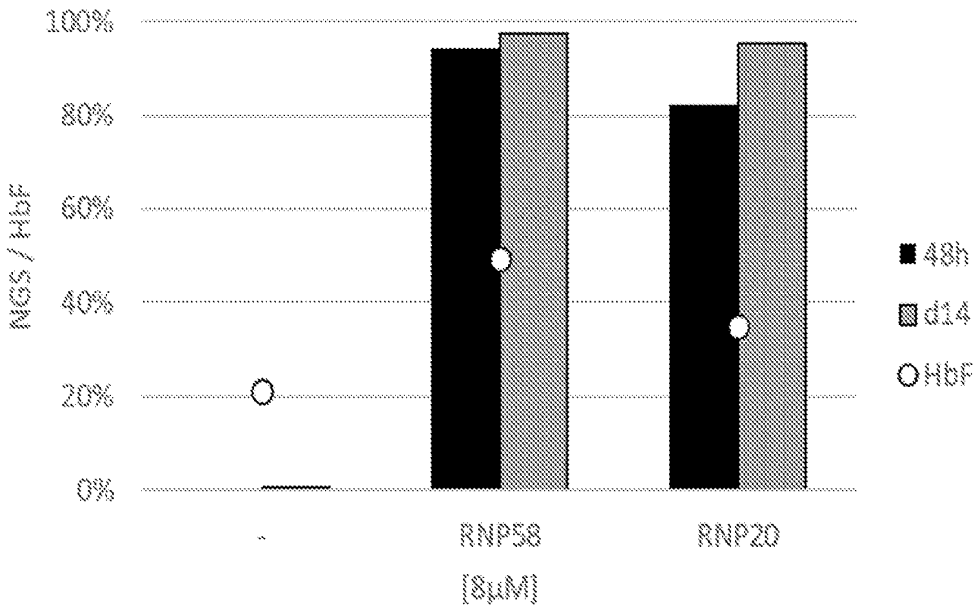


FIG 25

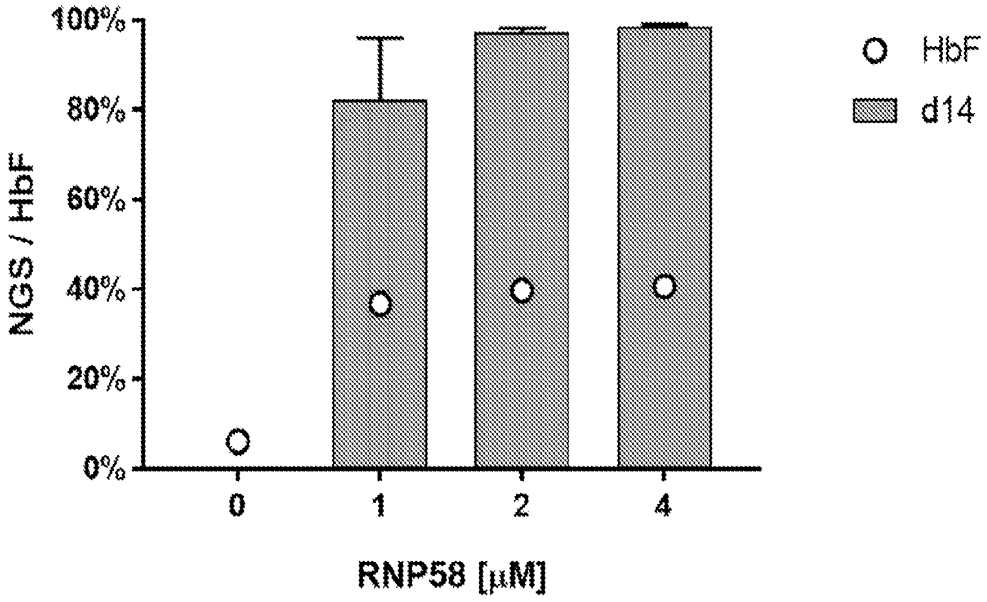


FIG 26

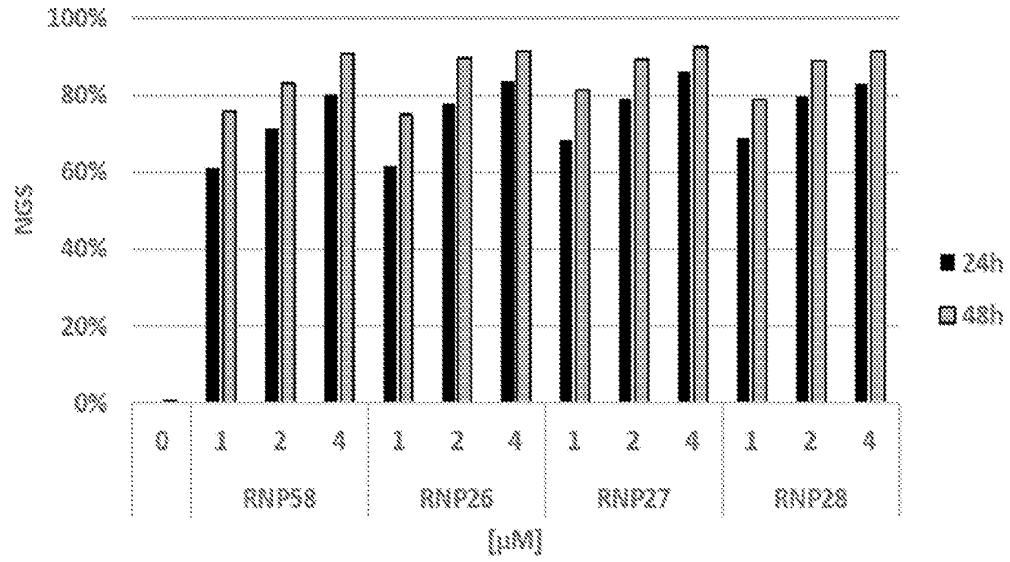

100%
80%
60%
40%
20%
0%
NGS
24h
48h
0
1
2
4
1
2
4
1
2
4
1
2
4
RNP58
RNP26
RNP27
RNP28
[μM]

FIG 27

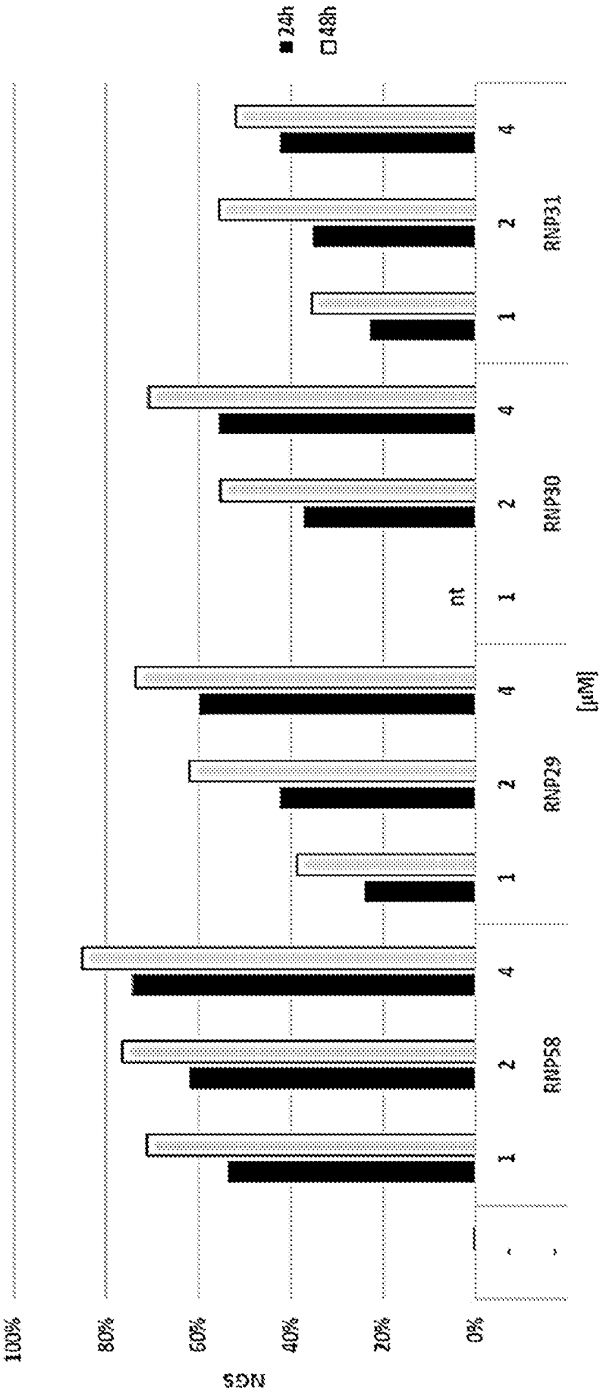


FIG 28

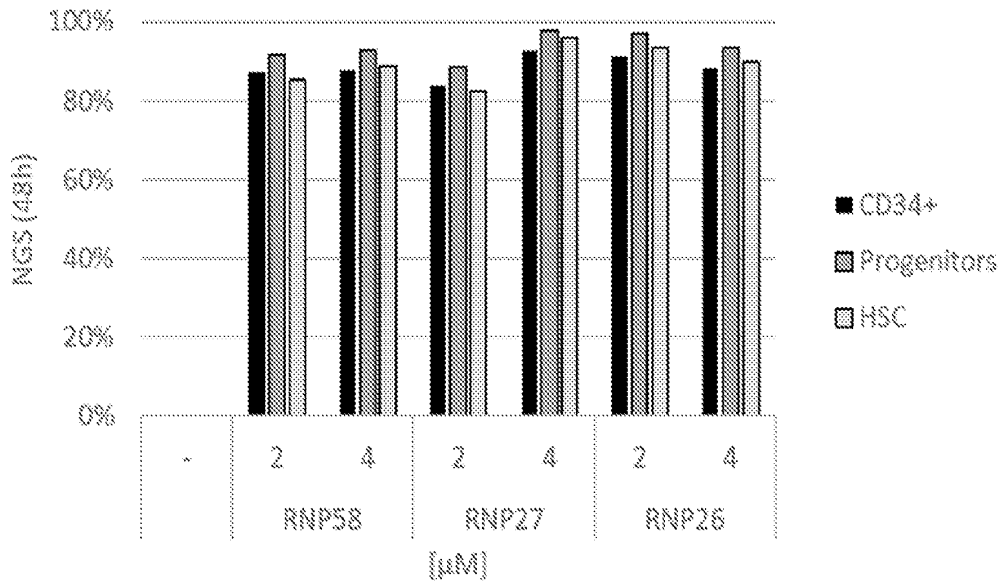

100%
80%
60%
40%
20%
0%
NGS (48h)
-
2
4
2
4
2
4
RNP58
RNP27
RNP26
[µM]
CD34+
Progenitors
HSC

FIG 29

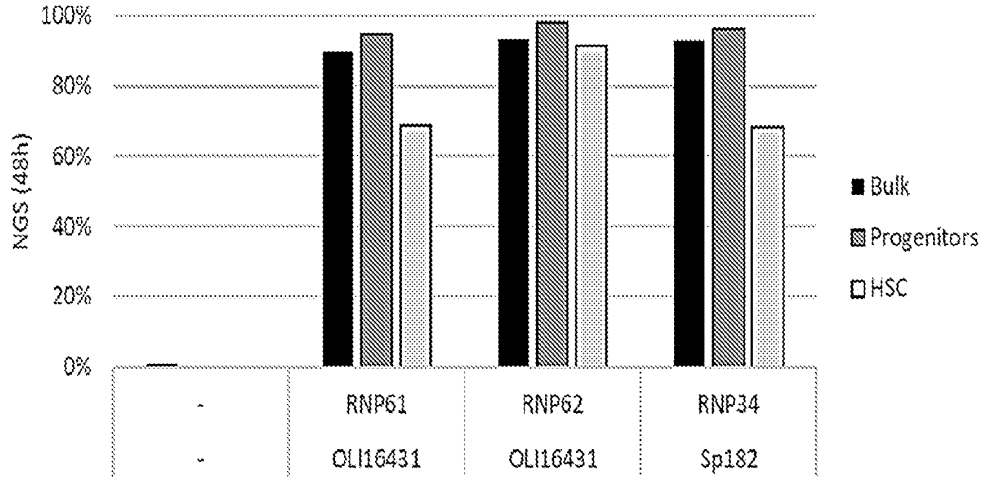

100%
80%
60%
40%
20%
0%
NGS (48h)
Bulk
Progenitors
HSC
-
-
RNP61
OLI16431
RNP62
OLI16431
RNP34
Sp182

FIG 30

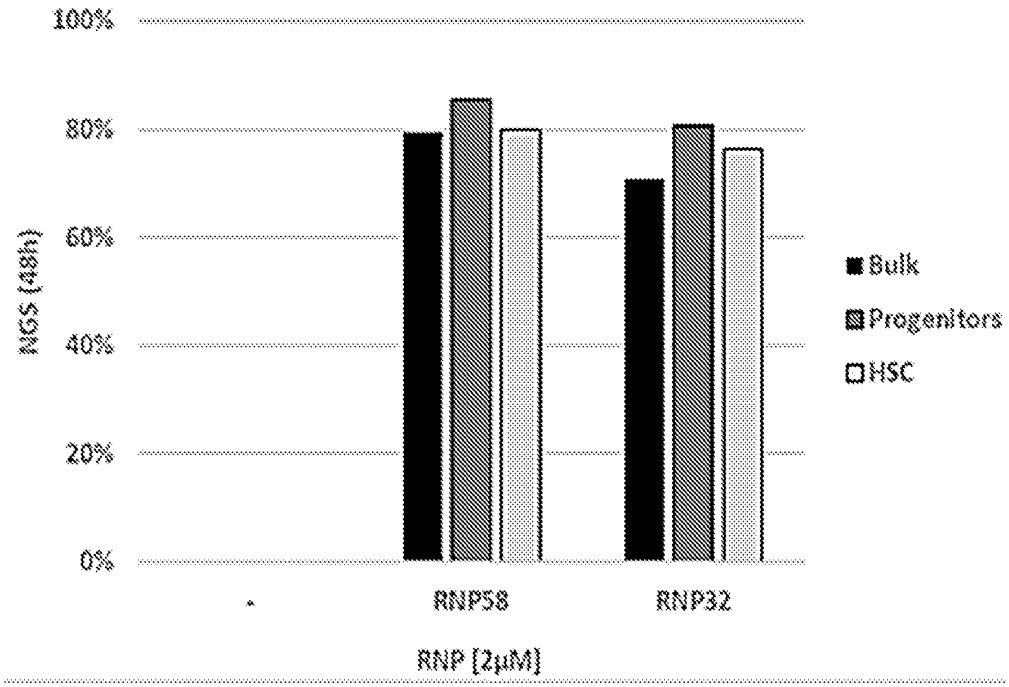

100%
80%
60%
40%
20%
0%
NGS (48h)
Bulk
Progenitors
HSC
-
RNP58
RNP32
RNP [2μM]

FIG 31

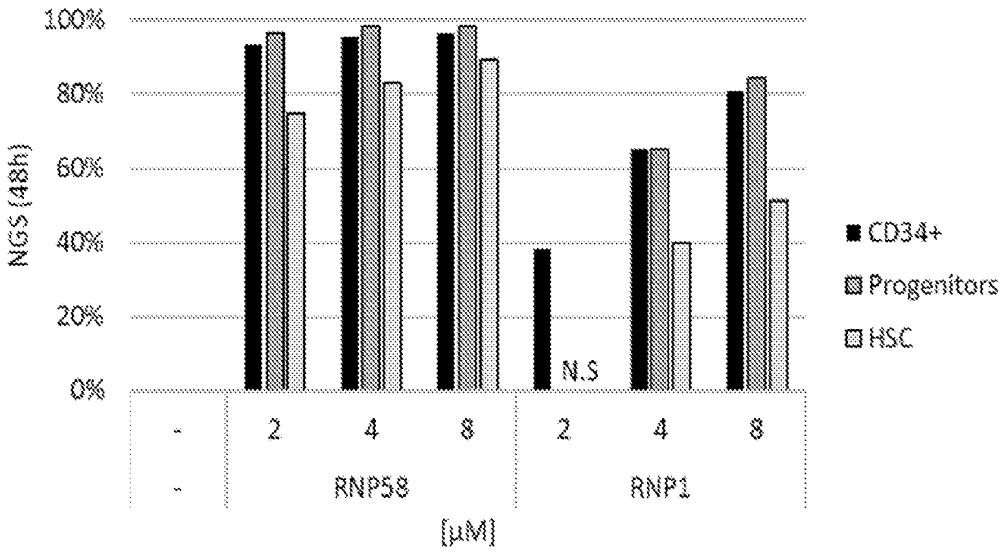

100%
80%
60%
40%
20%
0%
NGS (48h)
N.S
CD34+
Progenitors
HSC
-
2
4
8
2
4
8
-
RNP58
RNP1
[µM]

FIG 32

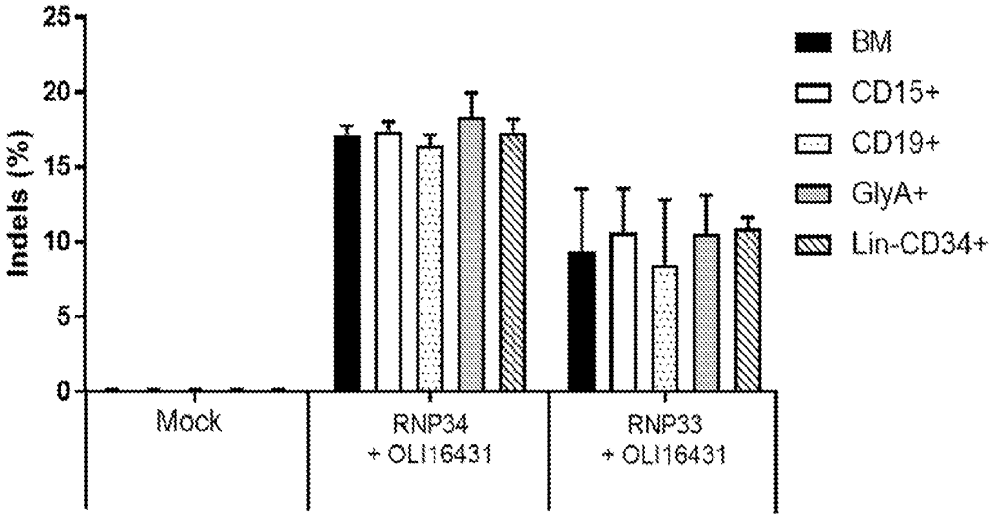

25
20
15
10
5
0
Indels (%)
Mock
RNP34
+ OLI16431
RNP33
+ OLI16431
BM
CD15+
CD19+
GlyA+
Lin-CD34+

FIG 33A

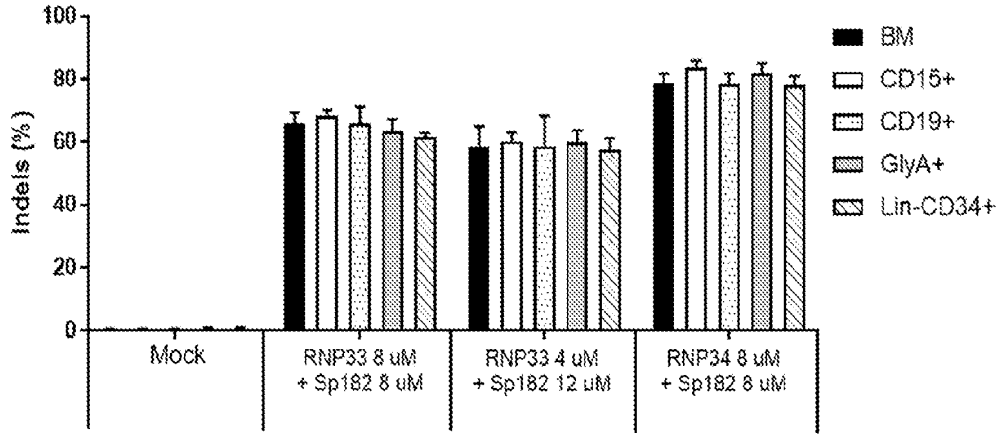


FIG 33B

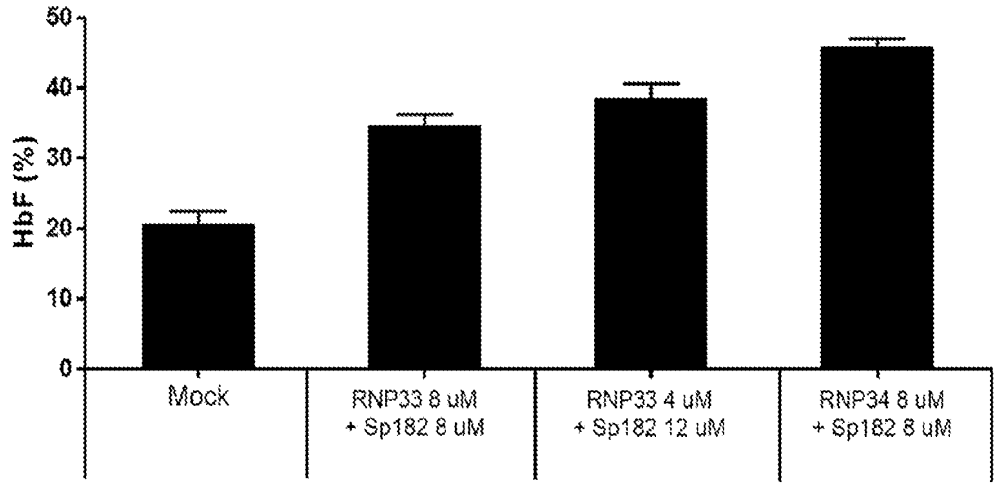


FIG 34A

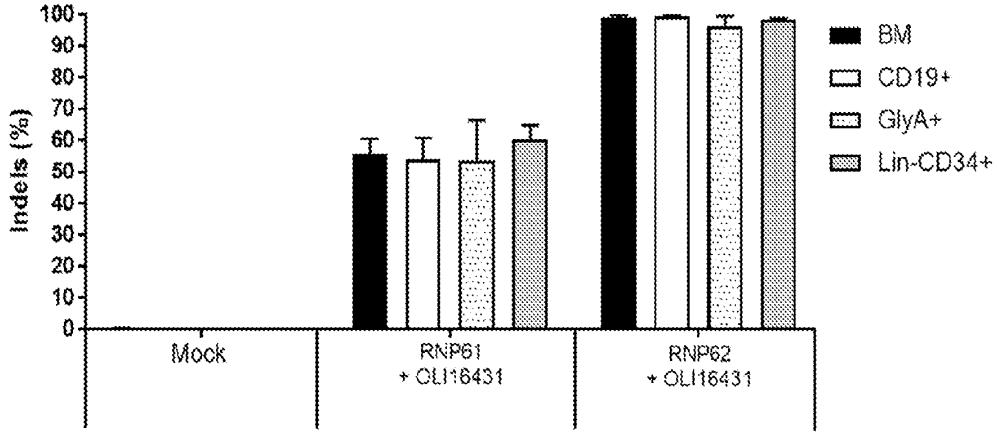


FIG 34B

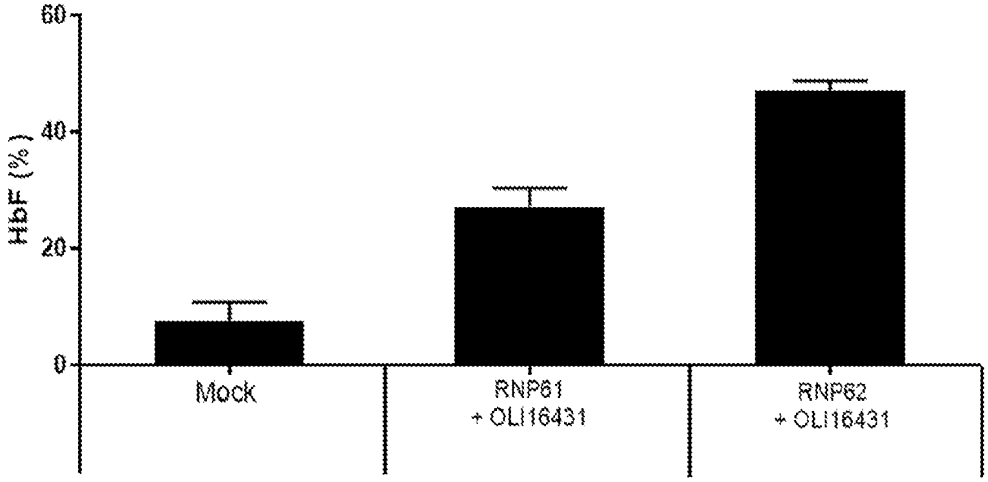


FIG 35

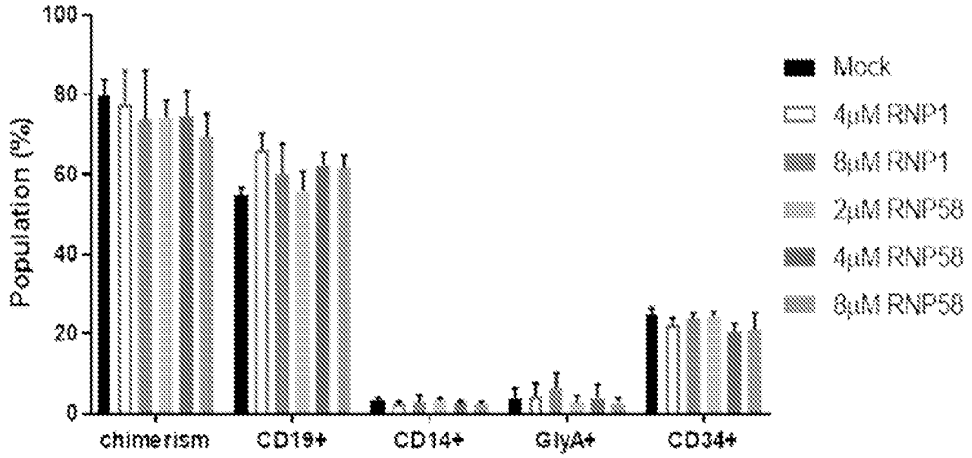

100
80
60
40
20
0
Population (%)
chimerism
CD19+
CD14+
GlyA+
CD34+
Mock
4μM RNP1
8μM RNP1
2μM RNP58
4μM RNP58
8μM RNP58

FIG 36

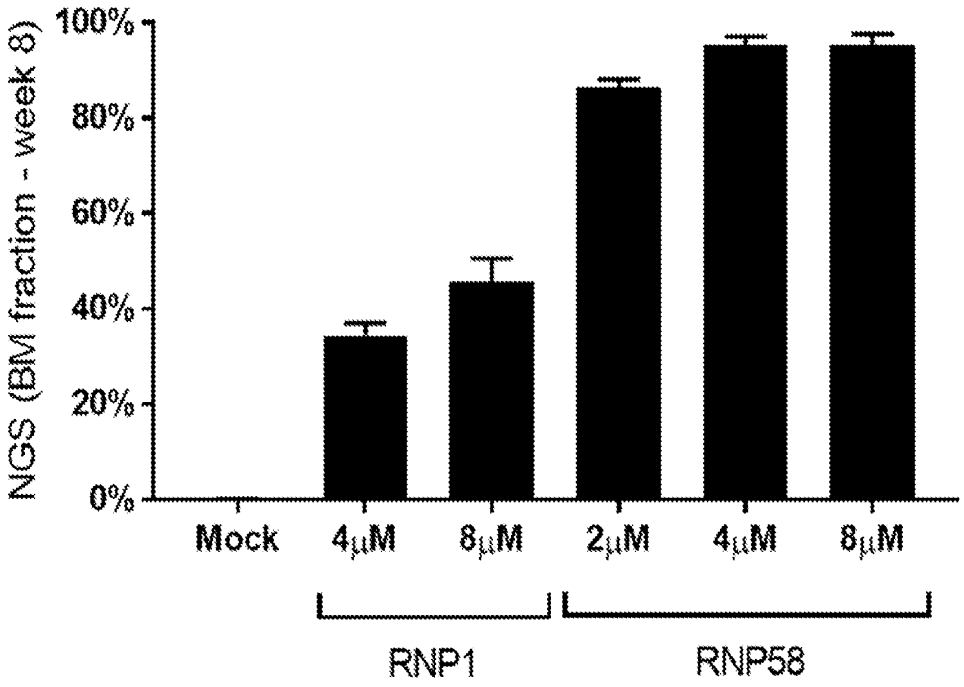

NGS (BM fraction - week 8)
100%
80%
60%
40%
20%
0%
Mock
4μM
8μM
2μM
4μM
8μM
RNP1
RNP58

FIG 37

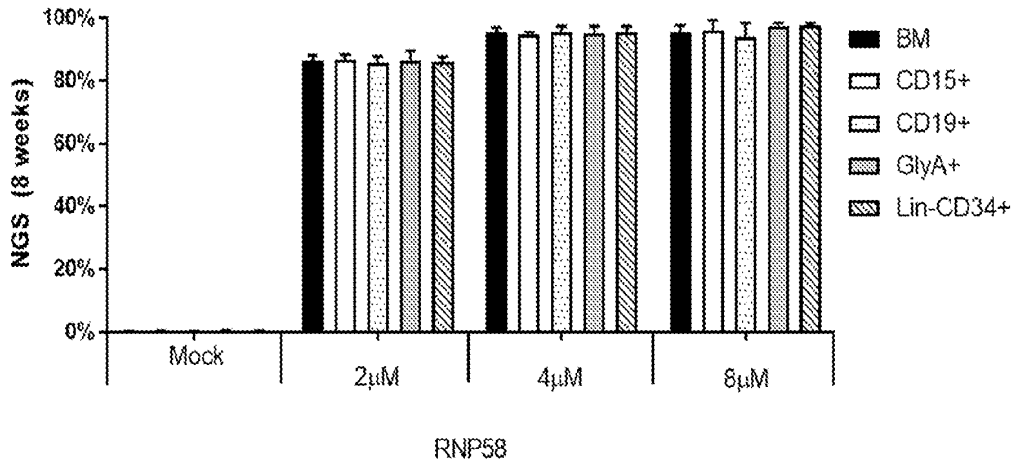

100%
80%
60%
40%
20%
0%
NGS (8 weeks)
Mock
2μM
4μM
8μM
RNP58
BM
CD15+
CD19+
GlyA+
Lin-CD34+

FIG 38

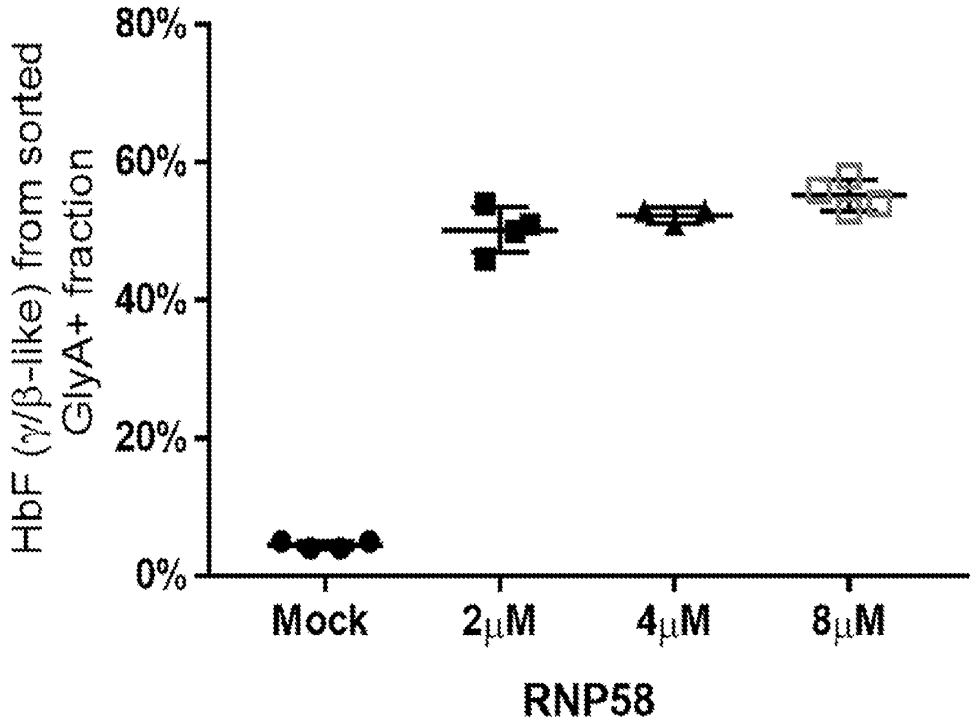

HbF (γ/β-like) from sorted GlyA+ fraction
80%
60%
40%
20%
0%
Mock
2μM
4μM
8μM
RNP58

FIG 39

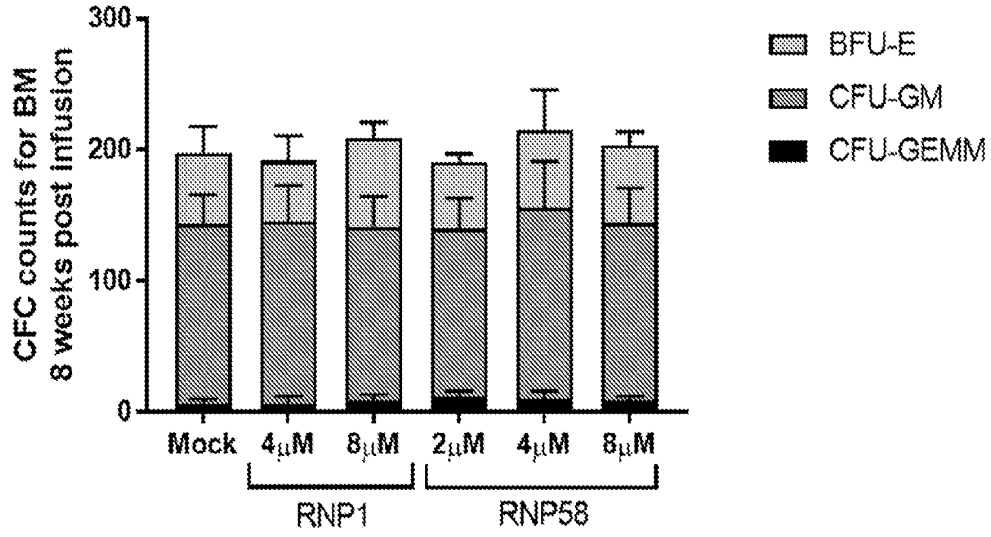

CFC counts for BM
8 weeks post infusion
300
200
100
0
Mock
4μM
8μM
2μM
4μM
8μM
RNP1
RNP58
BFU-E
CFU-GM
CFU-GEMM

FIG 40

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

MGHHHHHHGSTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHY
KELKPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAI
HDYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDK
FTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFE
NVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLN
LAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRN
ENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKIT
KSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKK
QEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKA
RNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKG
RYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSN
NFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYT
KTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYN
KDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAARL
GEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIK
DRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITV
IDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVI
HEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAE
KVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKN
HESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQ
FDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLE
NDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPM
DADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRNGS**PKKKRKV**
GS**PKKKRKV**

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

ATGGGCCATCATCATCACCATCACGGTAGCACCCAGTTTGAAGGCTTTACCAAT
CTGTATCAGGTTAGCAAAACCCTGCGTTTTGAACTGATTCCGCAGGGTAAAACC
CTGAAACATATTCAAGAACAGGGCTTCATCGAAGAGGATAAAGCACGTAACGAT
CACTACAAAGAACTGAAACCGATTATCGACCGCATCTATAAAACGTACGCAGAT
CAGTGTCTGCAGCTGGTTCAGCTGGATTGGGAAAATCTGAGCGCAGCAATTGA
TAGTTATCGCAAAGAGAAAACCGAAGAAACCCGTAATGCACTGATTGAAGAACA
GGCAACCTATCGTAATGCCATCCATGATTATTTCATTGGTCGTACCGATAATCTG
ACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGGCCTGTTTAAAGCC
GAACTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGTTACCACCACCGAA
CATGAAAATGCACTGCTGCGTAGCTTTGATAAATTCACCACCTATTTCAGCGGC
TTTTATGAGAATCGCAAAAACGTGTTTAGCGCAGAAGATATTAGCACCGCAATT
CCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAGAACTGCCACATT
TTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACATTTTGAAAACGTT
AAAAAAGCCATCGGCATCTTTGTTAGCACCAGCATTGAAGAAGTTTTTAGCTTCC
CGTTTTACAATCAGCTGCTGACCCAGACACAGATTGATCTGTATAACCAGCTGT
TAGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAAATCAAAGGTCTGAATGAA
GTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGCAAGC
CTGCCGCATCGTTTTATTCCGCTGTTTAAGCAGATT

FIG 40 (cont'd)

CTGAGCGATCGTAATACCCTGAGCTTTATTCTGGAAGAATTCAAATCCGATGAA
GAGGTGATTCAGAGCTTCTGCAAATACAAAACGCTGCTGCGCAATGAAAATGTT
CTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCAC
ATCTTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATT
GGGATACCCTGCGTAATGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGT
AAAATTACCAAAAGCGCGAAAGAAAAAGTTCAGCGCAGTCTGAAACATGAGGAT
ATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCAGAAGCATTT
AAACAGAAAACCAGCGAAATTCTGAGCCATGCACATGCAGCACTGGATCAGCC
GCTGCCGACCACACTGAAAAAACAAGAAGAAAAAGAGATCCTGAAAAGCCAGC
TGGATAGCCTGCTGGGTCTGTATCATCTGCTGGACTGGTTTGCAGTTGATGAAA
GCAATGAAGTTGATCCGGAATTTAGCGCACGTCTGACCGGCATTAAGCTGGAA
ATGGAACCGAGCCTGAGTTTTTACAATAAAGCCCGTAATTATGCCACCAAAAAA
CCGTATAGCGTGGAAAAATTCAAGCTGAACTTCCAGATGCCGACACTGGCAAG
CGGTTGGGATGTTAATAAAGAAAAGAACAATGGTGCCATCCTGTTCGTGAAAAA
TGGCCTGTATTATCTGGGTATCATGCCGAAACAGAAAGGTCGTTATAAAGCACT
GAGCTTTGAACCGACAGAAAAAACCAGTGAAGGCTTCGATAAAATGTACTATGA
TTACTTTCCGGACGCAGCCAAAATGATTCCGAAATGCAGCACCCAGCTGAAAGC
AGTTACCGCACATTTTCAGACCCATACCACACCGATTCTGCTGAGCAATAACTTT
ATTGAACCGCTGGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAA
GAGCCGAAAAAGTTTCAGACCGCCTATGCAAAAAAAACCGGTGATCAGAAAGG
TTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCGTGATTTTCTGAGCAAA
TACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGTA
TAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAG
CTTTCAGCGTATTGCCGAGAAAGAAATAATGGATGCAGTTGAAACCGGTAAGCT
GTACCTGTTCCAGATTTATAACAAAGACTTTGCCAAAGGCCATCATGGTAAACC
GAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAAAATCTGGCAAA
AACCAGCATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCG
TATGAAACGTATGGCAGCACGTCTGGGTGAAAAAATGCTGAATAAAAAGCTGAA
GGATCAGAAAACCCCGATTCCGGATACACTGTATCAAGAGCTGTATGATTATGT
GAATCATCGTCTGAGTCATGATCTGAGTGATGAAGCACGTGCCCTGCTGCCGA
ATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTCGTTTTACCAG
CGACAAGTTTTTTTTCCATGTTCCGATTACGCTGAATTATCAGGCAGCAAATAGC
CCGAGCAAATTTAACCAGCGTGTGAATGCATATCTGAAAGAACATCCGGAAACA
CCGATTATTGGTATTGATCGTGGTGAACGCAACCTGATTTATATCACCGTTATTG
ATAGCACCGGCAAGATCCTGGAACAGCGTAGCCTGAATACCATTCAGCAGTTT
GATTATCAGAAAAAGCTGGACAACCGTGAAAAAGAACGTGTTGCAGCCCGTCA
GGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCA
GGTTATTCATGAAATTGTGGATCTGATGATTCATTACCAGGCAGTTGTTGTGCTG
GAAAACCTGAATTTTGGCTTTAAAAGCAAACGTACCGGCATTGCGGAAAAAGCA
GTTTATCAGCAGTTCGAAAAGATGCTGATCGATAAACTGAACTGTCTGGTGCTG
AAAGATTATCCGGCTGAAAAAGTTGGCGGTGTTCTGAATCCGTATCAGCTGACC
GATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTAT
GTTCCGGCACCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCG
TTTGTTTGGAAAACCATCAAAAACCATGAAAGCCGCAAACATTTTCTGGAAGGTT
TTGATTTCCTGCATTACGACGTTAAGACCGGTGATTTTATTCTGCACTTTAAGAT
GAATCGGAACCTGTCATTTCAGCGTGGTCTGCCTGGTT

FIG 40 (cont'd)

TTATGCCTGCATGGGATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAA
AAGGCACCCCGTTTATTGCAGGTAAACGTATTGTTCCGGTGATTGAAAATCATC
GTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACTGATCGCACTGC
TGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGT
TAGAGAATGATGATAGCCATGCCATTGATACCATGGTTGCCCTGATTCGTAGCG
TTCTGCAGATGCGTAATAGCAATGCAGCAACCGGTGAAGATTACATTAATAGTC
CGGTTCGTGATCTGAATGGTGTTTGTTTTGATAGCCGTTTTCAGAATCCGGAAT
GGCCGATGGATGCGGATGCAAATGGTGCATATCATATTGCACTGAAAGGACAG
CTGCTGCTGAACCACCTGAAAGAAAGTAAAGATTTGAAGCTGCAGAATGGCATT
AGCAATCAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTTCT**CCTAAG**
AAAAAGCGCAAAGTTGGCAGT**CCGAAGAAGAAACGTAAGGTT**

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

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIY
KTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTD
NLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFY
ENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIF
VSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDET
AHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEA
LFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQ
RSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKS
QLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKP
YSVEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFE
PTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEIT
KEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLS
SLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGH
HGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNK
KLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSD
KFLFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKIL
EQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMI
HYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVL
NPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKH
FLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGT
PFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAI
DTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANG
AYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRNGR**SSDDEATADSQHA**
APPKKKRKVGGSGGSGGSGGSGGSGGSGGSGGSLE<u>HHHHHH</u>

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

ATGACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGT
TTTGAACTGATTCCGCAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTC
ATCGAAGAGGATAAAGCACGTAACGATCACTACAAAGAACTGAAACCGATTATC
GACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTGGTTCAGCTGGAT
TGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAA
ACCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGAT
TATTTCATTGGTCG

FIG 40 (cont'd)

TACCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAAAGG
CCTGTTTAAAGCCGAACTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCACCGT
TACCACCACCGAACATGAAAATGCACTGCTGCGTAGCTTTGATAAATTCACCAC
CTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGCAGAAGATATT
AGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAAGAG
AACTGCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAACAT
TTTGAAAACGTTAAAAAAGCCATCGGCATCTTTGTTAGCACCAGCATTGAAGAA
GTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGACCCAGACCCAGATTGATCTGT
ATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAAATCAAA
GGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCACAT
ATTATTGCAAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCG
ATCGTAATACCCTGAGCTTTATTCTGGAAGAATTCAAATCCGATGAAGAGGTGA
TTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCAATGAAAATGTTCTGGAAA
CTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATCTTTAT
CAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGATAC
CCTGCGTAATGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAATTAC
CAAAAGCGCGAAAGAAAAAGTTCAGCGCAGTCTGAAACATGAGGATATTAATCT
GCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCAGAAGCATTTAAACAGAA
AACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGCCGA
CCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCC
TGCTGGGTCTGTATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAG
TTGATCCGGAATTTAGCGCACGTCTGACCGGCATTAAACTGGAAATGGAACCGA
GCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCAAAAAACCGTATAGCG
TCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCCTGGCAAGCGGTTGGGAT
GTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTAT
TATCTGGGTATTATGCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAA
CCGACGGAAAAAACCAGTGAAGGTTTTGATAAAATGTACTACGACTATTTTCCG
GATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAAAGCAGTTACCGC
ACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAACCG
CTGGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAA
AAATTCCAGACCGCATATGCAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAA
GCGCTGTGTAAATGGATTGATTTCACCCGTGATTTTCTGAGCAAATACACCAAA
ACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGTATAAAGATCT
GGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCG
TATTGCCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTT
CCAGATCTACAATAAAGATTTTGCCAAAGGCCATCATGGCAAACCGAATCTGCA
TACCCTGTATTGGACCGGTCTGTTTAGCCCTGAAAATCTGGCAAAAACCTCGAT
TAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTATGAAACG
TATGGCACATCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAA
AACCCCGATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCG
TCTGAGCCATGATCTGAGTGATGAAGCACGTGCCCTGCTGCCGAATGTTATTAC
CAAAGAAGTTAGCCACGAGATCATTAAAGATCGTCGTTTTACCAGCGACAAATT
CCTGTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGAGCAA
ATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATT
GGTATTGATCGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACC
GGCAAAATCCTGGAACAGCGTAGCCTGAATACCA

FIG 40 (cont'd)

TTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTG
CAGCACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTT
ATCTGAGCCAGGTTATTCATGAAATTGTGGATCTGATGATTCACTATCAGGCCG
TTGTTGTGCTGGAAAACCTGAATTTTGGCTTTAAAAGCAAACGTACCGGCATTG
CAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACAAACTGAATT
GCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGT
ATCAGCTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGA
TTTCTGTTTTATGTTCCGGCACCGTATACGAGCAAAATTGATCCGCTGACCGGT
TTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAACCATGAAAGCCGCAAACATT
TTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTCATCCT
GCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTAT
GCCTGCATGGGATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGG
CACCCCGTTTATTGCAGGTAAACGTATTGTTCCGGTGATTGAAAATCATCGTTTC
ACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACTGATCGCACTGCTGGA
AGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGGA
AAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCT
GCAGATGCGTAATAGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGT
TCGTGATCTGAATGGTGTTTGTTTTGATAGCCGTTTTCAGAATCCGGAATGGCC
GATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAAAGGACAGCTGCT
GCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAA
TCAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATG**
ATGAAGCAACCGCAGATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTA
AAGTTGGTGGTAGCGGTGGTTCAGGTGGTAGTGGCGGTAGTGGTGGCTCAGG
GGGTTCTGGTGGCTCTGGTGGTAGCCTCGAGCACCACCACCACCACCAC

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

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIY
KTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTD
NLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFY
ENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIF
VSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDET
AHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEA
LFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQ
RSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKS
QLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKP
YSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFE
PTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEIT
KEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLS
SLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGH
HGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNK
KLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSD
KFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKIL
EQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMI
HYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVL
NPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGF

FIG 40 (cont'd)

VDPFVWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPA
WDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVF
RDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFD
SRFQNPEWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELR
NGR**SSDDEATADSQHAAPPKKKRKV**GGSGGSGGSGGSGGSGGSGGSGGSLEH
HHHHH

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

ATGACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGT
TTTGAACTGATTCCGCAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTC
ATCGAAGAGGATAAAGCACGTAACGATCACTACAAAGAACTGAAACCGATTATC
GACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTGGTTCAGCTGGAT
TGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAA
ACCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGAT
TATTTCATTGGTCGTACCGATAATCTGACCGATGCAATTAACAAACGTCACGCC
GAAATCTATAAAGGCCTGTTTAAAGCCGAACTGTTTAATGGCAAAGTTCTGAAAC
AGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCTGCGTAGCTTT
GATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTA
GCGCAGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCC
CGAAATTCAAAGAGAACTGCCACATTTTTACCCGTCTGATTACCGCAGTTCCGA
GCCTGCGTGAACATTTTGAAAACGTTAAAAAAGCCATCGGCATCTTTGTTAGCA
CCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGACCCAGA
CCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGC
ACCGAAAAAATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAAT
GATGAAACCGCACATATTATTGCAAGCCTGCCGCATCGTTTTATTCCGCTGTTC
AAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTATTCTGGAAGAATTCAAAT
CCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCAATG
AAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATC
TGACCCACATCTTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGT
GTGATCATTGGGATACCCTGCGTAATGCCCTGTATGAACGTCGTATTAGCGAAC
TGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTCAGCGCAGTCTGAAAC
ATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCAG
AAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGG
ATCAGCCGCTGCCGACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAA
GCCAGCTGGATAGCCTGCTGGGTCTGTATCATCTGCTGGACTGGTTTGCAGTT
GATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCTGACCGGCATTAAA
CTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACC
AAAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGATGCCGACCCTG
GCAAGCGGTTGGGATGTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTG
AAAAATGGCCTGTATTATCTGGGTATTATGCCGAAACAGAAAGGTCGTTATAAA
GCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTTGATAAAATGTAC
TACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTG
AAAGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAAT
AACTTTATTGAACCGCTGGAAATCACCAAAGAGATCTACGATCTGAATAACCCG
GAAAAAGAGCCGAAAAAATTCCAGACCGCATATGCAAAAAAAACCGGTGATCAG
AAAGGTTATCGTGAAGCGCT

FIG 40 (cont'd)

GTGTAAATGGATTGATTTCACCCGTGATTTTCTGAGCAAATACACCAAAACCACC
AGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGTATAAAGATCTGGGCGA
ATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTGCC
GAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATC
TACAATAAAGATTTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTG
TATTGGACCGGTCTGTTTAGCCCTGAAAATCTGGCAAAAACCTCGATTAAACTG
AATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTATGAAACGTATGGCA
CATCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCG
ATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGC
CATGATCTGAGTGATGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAA
GTTAGCCACGAGATCATTAAAGATCGTCGTTTTACCAGCGACAAATTCTTTTTTC
ATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGAGCAAATTTAACC
AGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTG
ATCGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAA
TCCTGGAACAGCGTAGCCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAAC
TGGATAATCGCGAGAAAGAACGTGTTGCAGCACGTCAGGCATGGTCAGTTGTT
GGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATTCATGAAATT
GTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTT
GGCTTTAAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTC
GAGAAAATGCTGATTGACAAACTGAATTGCCTGGTGCTGAAAGATTATCCGGCT
GAAAAAGTTGGTGGTGTTCTGAATCCGTATCAGCTGACCGATCAGTTTACCAGC
TTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCACCGTAT
ACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACC
ATCAAAAACCATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATT
ACGACGTTAAAACGGGTGATTTCATCCTGCACTTTAAAATGAATCGCAATCTGA
GTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGGGATATTGTGTTTGAGA
AAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACGTA
TTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATC
CGGCAAATGAACTGATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATG
GCTCAAACATTCTGCCGAAACTGCTGGAAAATGATGATAGCCATGCAATTGATA
CCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAATAGCAATGCAGCAA
CCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGA
TAGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCAT
ATCATATTGCACTGAAAGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAG
ATCTGAAACTGCAAAACGGCATTAGCAATCAGGATTGGCTGGCATATATCCAAG
AACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCAGATAGCCAGCA**
TGCAGCACCGCCTAAAAAGAAACGTAAAGTTGGTGGTAGCGGTGGTTCAGGT
GGTAGTGGCGGTAGTGGTGGCTCAGGGGGTTCTGGTGGCTCTGGTGGTAGCC
TCGAGCACCACCACCACCACCAC

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

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIY
KTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTD
NLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFY
ENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIF
VSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDET
AHIIASLPHRFIPL

FIG 40 (cont'd)

FKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIF
ISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEI
SAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLD
WFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQRP
TLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGFDKMY
YDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEKEPK
KFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQYKDLGE
YYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLYWT
GLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAARLGEKMLNKKLKDQKTPIPDTL
YQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFLFHVPITLNYQ
AANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQRSLNTIQQFD
YQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQAVVVLENLN
FGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPYQLTDQFTSFA
KMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDFLHYDVK
TGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPVIEN
HRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSVLQ
MRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAYHIALKGQLLLN
HLKESKDLKLQNGISNQDWLAYIQELRNGR**SSDDEATADSQHAAPPKKKRKV**GG
SGGSGGSGGSGGSGGSGGSGGSLE<u>HHHHHH</u>

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

ATGACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGT
TTTGAACTGATTCCGCAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTC
ATCGAAGAGGATAAAGCACGTAACGATCACTACAAAGAACTGAAACCGATTATC
GACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTGGTTCAGCTGGAT
TGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAA
ACCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGAT
TATTTCATTGGTCGTACCGATAATCTGACCGATGCAATTAACAAACGTCACGCC
GAAATCTATAAAGGCCTGTTTAAAGCCGAACTGTTTAATGGCAAAGTTCTGAAAC
AGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCTGCGTAGCTTT
GATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTA
GCGCAGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCC
CGAAATTCAAAGAGAACTGCCACATTTTTACCCGTCTGATTACCGCAGTTCCGA
GCCTGCGTGAACATTTTGAAAACGTTAAAAAAGCCATCGGCATCTTTGTTAGCA
CCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGACCCAGA
CCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGC
ACCGAAAAAATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAAT
GATGAAACCGCACATATTATTGCAAGCCTGCCGCATCGTTTTATTCCGCTGTTC
AAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTATTCTGGAAGAATTCAAAT
CCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCAATG
AAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATC
TGACCCACATCTTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGT
GTGATCATTGGGATACCCTGCGTAATGCCCTGTATGAACGTCGTATTAGCGAAC
TGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTCAGCGCAGTCTGAAAC
ATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCAG
AAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGG
ATCAGCCGCTGCCGACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAA

FIG 40 (cont'd)

AAGCCAGCTGGATAGCCTGCTGGGTCTGTATCATCTGCTGGACTGGTTTGCAG
TTGATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCTGACCGGCATTA
AACTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCA
CCAAAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCC
TGGCAAGCGGTTGGGATGTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCG
TGAAAAATGGCCTGTATTATCTGGGTATTATGCCGAAACAGAAAGGTCGTTATA
AAGCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTTGATAAAATGT
ACTACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGC
TGAAAGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCA
ATAACTTTATTGAACCGCTGGAAATCACCAAAGAGATCTACGATCTGAATAACCC
GGAAAAAGAGCCGAAAAAATTCCAGACCGCATATGCAAAAAAAACCGGTGATCA
GAAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCGTGATTTTCT
GAGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCA
GCCAGTATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATC
ATATTAGCTTTCAGCGTATTGCCGAGAAAGAAATCATGGACGCAGTTGAAACCG
GTAAACTGTACCTGTTCCAGATCTACAATAAAGATTTTGCCAAAGGCCATCATG
GCAAACCGAATCTGCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAAAATC
TGGCAAAAACCTCGATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGA
AAAGCCGTATGAAACGTATGGCAGCTCGTCTGGGTGAAAAAATGCTGAACAAAA
AACTGAAAGACCAGAAAACCCCGATCCCGGATACACTGTATCAAGAACTGTATG
ATTATGTGAACCATCGTCTGAGCCATGATCTGAGTGATGAAGCACGTGCCCTGC
TGCCGAATGTTATTACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTCGTTT
TACCAGCGACAAATTCCTGTTTCATGTGCCGATTACCCTGAATTATCAGGCAGC
AAATAGCCCGAGCAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCC
AGAAACGCCGATTATTGGTATTGATCGTGGTGAACGTAACCTGATTTATATCAC
CGTTATTGATAGCACCGGCAAAATCCTGGAACAGCGTAGCCTGAATACCATTCA
GCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAGC
ACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCT
GAGCCAGGTTATTCATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGT
TGTGCTGGAAAACCTGAATTTTGGCTTTAAAAGCAAACGTACCGGCATTGCAGA
AAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACAAACTGAATTGCCT
GGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCA
GCTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCT
GTTTTATGTTCCGGCACCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGT
TGATCCGTTTGTTTGGAAAACCATCAAAAACCATGAAAGCCGCAAACATTTTCTG
GAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTCATCCTGCACT
TTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTG
CATGGGATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCC
CGTTTATTGCAGGTAAACGTATTGTTCCGGTGATTGAAAATCATCGTTTCACCG
GTCGTTATCGCGATCTGTATCCGGCAAATGAACTGATCGCACTGCTGGAAGAG
AAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGGAAAAT
GATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAG
ATGCGTAATAGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGT
GATCTGAATGGTGTTTGTTTTGATAGCCGTTTTCAGAATCCGGAATGGCCGATG
GATGCAGATGCAAATGGTGCATATCATATTGCACTGAAAGGACAGCTGCTGCTG
AACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAATCAG
GATTGGCTGGCATATATCCAAGAACTGCGTAACGGT

FIG 40 (cont'd)

CGT**AGCAGTGATGATGAAGCAACCGCAGATAGCCAGCATGCAGCACCGCCTA AAAAGAAACGTAAAGTT**GGTGGTAGCGGTGGTTCAGGTGGTAGTGGCGGTAG TGGTGGCTCAGGGGGTTCTGGTGGCTCTGGTGGTAGCCTCGAGCACCACCAC CACCACCAC

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

MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIY
KTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTD
NLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFY
ENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIF
VSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDET
AHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEA
LFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQ
RSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKS
QLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKP
YSVEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFE
PTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEIT
KEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLS
SLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGH
HGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAARLGEKMLNK
KLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSD
KFLFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKIL
EQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMI
HYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVL
NPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKH
FLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGT
PFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAI
DTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANG
AYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRNGR**SSDDEATADSQHA APPKKKRKV**

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

ATGACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGT
TTTGAACTGATTCCGCAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTC
ATCGAAGAGGATAAAGCACGTAACGATCACTACAAAGAACTGAAACCGATTATC
GACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTGGTTCAGCTGGAT
TGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAA
ACCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGAT
TATTTCATTGGTCGTACCGATAATCTGACCGATGCAATTAACAAACGTCACGCC
GAAATCTATAAAGGCCTGTTTAAAGCCGAACTGTTTAATGGCAAAGTTCTGAAAC
AGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCTGCGTAGCTTT
GATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTA
GCGCAGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCC
CGAAATTCAAAGAGAACTGCCACATTTTTACCCGTCTGATTACCGCAGTTCCGA
GCCTGCGTGAACATTTTGAAAACGTTAAAAAAGCCATCGGCATCTTTGTTAGCA
CCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGACCCAGA
CCCAGATTGATCTGTATAACCAACTGCTGGG

FIG 40 (cont'd)

TGGTATTAGCCGTGAAGCAGGCACCGAAAAAATCAAAGGTCTGAATGAAGTGCT
GAATCTGGCCATTCAGAAAAATGATGAAACCGCACATATTATTGCAAGCCTGCC
GCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAGCGATCGTAATACCCTGAG
CTTTATTCTGGAAGAATTCAAATCCGATGAAGAGGTGATTCAGAGCTTTTGCAAA
TACAAAACGCTGCTGCGCAATGAAAATGTTCTGGAAACTGCCGAAGCACTGTTT
AACGAACTGAATAGCATTGATCTGACCCACATCTTTATCAGCCACAAAAAACTG
GAAACCATTTCAAGCGCACTGTGTGATCATTGGGATACCCTGCGTAATGCCCTG
TATGAACGTCGTATTAGCGAACTGACCGGTAAAATTACCAAAAGCGCGAAAGAA
AAAGTTCAGCGCAGTCTGAAACATGAGGATATTAATCTGCAAGAGATTATTAGC
GCAGCCGGTAAAGAACTGTCAGAAGCATTTAAACAGAAAACCAGCGAAATTCTG
TCACATGCACATGCAGCACTGGATCAGCCGCTGCCGACCACCCTGAAAAAACA
AGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATAGCCTGCTGGGTCTGTATCA
TCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATGAAGTTGATCCGGAATTTAG
CGCACGTCTGACCGGCATTAAACTGGAAATGGAACCGAGCCTGAGCTTTTATAA
CAAAGCCCGTAATTATGCCACCAAAAAACCGTATAGCGTCGAAAAATTCAAACT
GAACTTTCAGCGTCCGACCCTGGCAAGCGGTTGGGATGTTAATAAAGAAAAAAA
CAACGGTGCCATCCTGTTCGTGAAAAATGGCCTGTATTATCTGGGTATTATGCC
GAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTTGAACCGACGGAAAAAACCA
GTGAAGGTTTTGATAAAATGTACTACGACTATTTTCCGGATGCAGCCAAAATGAT
TCCGAAATGTAGCACCCAGCTGAAAGCAGTTACCGCACATTTTCAGACCCATAC
CACCCCGATTCTGCTGAGCAATAACTTTATTGAACCGCTGGAAATCACCAAAGA
GATCTACGATCTGAATAACCCGGAAAAAGAGCCGAAAAAATTCCAGACCGCATA
TGCAAAAAAAACCGGTGATCAGAAAGGTTATCGTGAAGCGCTGTGTAAATGGAT
TGATTTCACCCGTGATTTTCTGAGCAAATACACCAAAACCACCAGTATCGATCTG
AGCAGCCTGCGTCCGAGCAGCCAGTATAAAGATCTGGGCGAATATTATGCAGA
ACTGAATCCGCTGCTGTATCATATTAGCTTTCAGCGTATTGCCGAGAAAGAAAT
CATGGACGCAGTTGAAACCGGTAAACTGTACCTGTTCCAGATCTACAATAAAGA
TTTTGCCAAAGGCCATCATGGCAAACCGAATCTGCATACCCTGTATTGGACCGG
TCTGTTTAGCCCTGAAAATCTGGCAAAAACCTCGATTAAACTGAATGGTCAGGC
GGAACTGTTTTATCGTCCGAAAAGCCGTATGAAACGTATGGCAGCTCGTCTGG
GTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCGATCCCGGATA
CACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTGA
GTGATGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCAC
GAGATCATTAAAGATCGTCGTTTTACCAGCGACAAATTCCTGTTTCATGTGCCG
ATTACCCTGAATTATCAGGCAGCAAATAGCCCGAGCAAATTTAACCAGCGTGTT
AATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGATCGTGGT
GAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAA
CAGCGTAGCCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAAT
CGCGAGAAAGAACGTGTTGCAGCACGTCAGGCATGGTCAGTTGTTGGTACAAT
TAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATTCATGAAATTGTGGATCT
GATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTTAAA
AGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATG
CTGATTGACAAACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTT
GGTGGTGTTCTGAATCCGTATCAGCTGACCGATCAGTTTACCAGCTTTGCAAAA
ATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCACCGTATACGAGCAAA
ATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAACC
ATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTT

FIG 40 (cont'd)

TCTGCATTACGACGTTAAAACGGGTGATTTCATCCTGCACTTTAAAATGAATCGC
AATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGGGATATTGTG
TTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGT
AAACGTATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGCGAT
CTGTATCCGGCAAATGAACTGATCGCACTGCTGGAAGAGAAAGGTATTGTTTTT
CGTGATGGCTCAAACATTCTGCCGAAACTGCTGGAAAATGATGATAGCCATGCA
ATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAATAGCAAT
GCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGTGTT
TGTTTTGATAGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCAAAT
GGTGCATATCATATTGCACTGAAAGGACAGCTGCTGCTGAACCACCTGAAAGAA
AGCAAAGATCTGAAACTGCAAAACGGCATTAGCAATCAGGATTGGCTGGCATAT
ATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATGATGAAGCAACCGCAGATA**
GCCAGCATGCAGCACCGCCTAAAAAGAAACGTAAAGTT

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

STQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYK
TYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDN
LTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYE
NRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFV
STSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEAL
FNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQR
SLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQ
LDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYS
VEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPT
EKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEI
YDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLR
PSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHG
KPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLK
DQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFL
FHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQR
SLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQ
AVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPY
QLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLE
GFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFI
AGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDT
MVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPEWPMDADANGAY
HIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRNGR**SSDDEATADSQHAAP**
PKKKRKV

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

AGCACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGT
TTTGAACTGATTCCGCAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTC
ATCGAAGAGGATAAAGCACGTAACGATCACTACAAAGAACTGAAACCGATTATC
GACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTGGTTCAGCTGGAT
TGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAA
ACCCGTAATG

FIG 40 (cont'd)

CACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTCATTGG
TCGTACCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATCTATAA
AGGCCTGTTTAAAGCCGAACTGTTTAATGGCAAAGTTCTGAAACAGCTGGGCAC
CGTTACCACCACCGAACATGAAAATGCACTGCTGCGTAGCTTTGATAAATTCAC
CACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGCAGAAGA
TATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAATTCAAA
GAGAACTGCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTGCGTGAA
CATTTTGAAAACGTTAAAAAAGCCATCGGCATCTTTGTTAGCACCAGCATTGAAG
AAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGACCCAGACCCAGATTGATCT
GTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAAAAAATCA
AAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAAACCGCAC
ATATTATTGCAAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAATTCTGAG
CGATCGTAATACCCTGAGCTTTATTCTGGAAGAATTCAAATCCGATGAAGAGGT
GATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCAATGAAAATGTTCTGGA
AACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCACATCTT
TATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCATTGGGA
TACCCTGCGTAATGCCCTGTATGAACGTCGTATTAGCGAACTGACCGGTAAAAT
TACCAAAAGCGCGAAAGAAAAAGTTCAGCGCAGTCTGAAACATGAGGATATTAA
TCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCAGAAGCATTTAAACA
GAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCCGCTGC
CGACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGCTGGATA
GCCTGCTGGGTCTGTATCATCTGCTGGACTGGTTTGCAGTTGATGAAAGCAATG
AAGTTGATCCGGAATTTAGCGCACGTCTGACCGGCATTAAACTGGAAATGGAAC
CGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCAAAAAACCGTATA
GCGTCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCCTGGCAAGCGGTTGG
GATGTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCTG
TATTATCTGGGTATTATGCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTTT
GAACCGACGGAAAAAACCAGTGAAGGTTTTGATAAAATGTACTACGACTATTTT
CCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAAAGCAGTTAC
CGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAA
CCGCTGGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCC
GAAAAAATTCCAGACCGCATATGCAAAAAAAACCGGTGATCAGAAAGGTTATCG
TGAAGCGCTGTGTAAATGGATTGATTTCACCCGTGATTTTCTGAGCAAATACAC
CAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGTATAAAG
ATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCA
GCGTATTGCCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCT
GTTCCAGATCTACAATAAAGATTTTGCCAAAGGCCATCATGGCAAACCGAATCT
GCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAAAATCTGGCAAAAACCTC
GATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTATGAA
ACGTATGGCACATCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCA
GAAAACCCCGATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCA
TCGTCTGAGCCATGATCTGAGTGATGAAGCACGTGCCCTGCTGCCGAATGTTAT
TACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTCGTTTTACCAGCGACAA
ATTCCTGTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGAG
CAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGAT
TATTGGTATTGATCGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGC
ACCGGCAAAATCCTGGAACAGCGTAGCCTGAATACCA

FIG 40 (cont'd)

TTCAGCAGTTTGATTACCAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTG
CAGCACGTCAGGCATGGTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTT
ATCTGAGCCAGGTTATTCATGAAATTGTGGATCTGATGATTCACTATCAGGCCG
TTGTTGTGCTGGAAAACCTGAATTTTGGCTTTAAAAGCAAACGTACCGGCATTG
CAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATGCTGATTGACAAACTGAATT
GCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGT
ATCAGCTGACCGATCAGTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGA
TTTCTGTTTTATGTTCCGGCACCGTATACGAGCAAAATTGATCCGCTGACCGGT
TTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAACCATGAAAGCCGCAAACATT
TTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAAAACGGGTGATTTCATCCT
GCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTAT
GCCTGCATGGGATATTGTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGG
CACCCCGTTTATTGCAGGTAAACGTATTGTTCCGGTGATTGAAAATCATCGTTTC
ACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAACTGATCGCACTGCTGGA
AGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGGA
AAATGATGATAGCCATGCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCT
GCAGATGCGTAATAGCAATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGT
TCGTGATCTGAATGGTGTTTGTTTTGATAGCCGTTTTCAGAATCCGGAATGGCC
GATGGATGCAGATGCAAATGGTGCATATCATATTGCACTGAAAGGACAGCTGCT
GCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAA
TCAGGATTGGCTGGCATATATCCAAGAACTGCGTAACGGTCGT**AGCAGTGATG**
ATGAAGCAACCGCAGATAGCCAGCATGCAGCACCGCCTAAAAAGAAACGTA
AAGTT

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

STQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYK
TYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDN
LTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYE
NRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFV
STSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETA
HIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEAL
FNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQR
SLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQ
LDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYS
VEKFKLNFQRPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPT
EKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEI
YDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLSKYTKTTSIDLSSLR
PSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHG
KPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMAHRLGEKMLNKKLK
DQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSHEIIKDRRFTSDKFL
FHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLIYITVIDSTGKILEQR
SLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLMIHYQ
AVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPY
QLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLE
GFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFI
AGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDT
MVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNP

FIG 40 (cont'd)

EWPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRNGR**SSD DEATADSQHAAPPKKKRKV**GGSGGSGGSGGSGGSGGSGGSGGSLEHHHHHH

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

AGCACCCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGT
TTTGAACTGATTCCGCAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTC
ATCGAAGAGGATAAAGCACGTAACGATCACTACAAAGAACTGAAACCGATTATC
GACCGCATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTGGTTCAGCTGGAT
TGGGAAAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAA
ACCCGTAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGAT
TATTTCATTGGTCGTACCGATAATCTGACCGATGCAATTAACAAACGTCACGCC
GAAATCTATAAAGGCCTGTTTAAAGCCGAACTGTTTAATGGCAAAGTTCTGAAAC
AGCTGGGCACCGTTACCACCACCGAACATGAAAATGCACTGCTGCGTAGCTTT
GATAAATTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTA
GCGCAGAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCC
CGAAATTCAAAGAGAACTGCCACATTTTTACCCGTCTGATTACCGCAGTTCCGA
GCCTGCGTGAACATTTTGAAAACGTTAAAAAAGCCATCGGCATCTTTGTTAGCA
CCAGCATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGACCCAGA
CCCAGATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGC
ACCGAAAAAATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAAT
GATGAAACCGCACATATTATTGCAAGCCTGCCGCATCGTTTTATTCCGCTGTTC
AAACAAATTCTGAGCGATCGTAATACCCTGAGCTTTATTCTGGAAGAATTCAAAT
CCGATGAAGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCAATG
AAAATGTTCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATC
TGACCCACATCTTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGT
GTGATCATTGGGATACCCTGCGTAATGCCCTGTATGAACGTCGTATTAGCGAAC
TGACCGGTAAAATTACCAAAAGCGCGAAAGAAAAAGTTCAGCGCAGTCTGAAAC
ATGAGGATATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCAG
AAGCATTTAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGG
ATCAGCCGCTGCCGACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAA
GCCAGCTGGATAGCCTGCTGGGTCTGTATCATCTGCTGGACTGGTTTGCAGTT
GATGAAAGCAATGAAGTTGATCCGGAATTTAGCGCACGTCTGACCGGCATTAAA
CTGGAAATGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACC
AAAAAACCGTATAGCGTCGAAAAATTCAAACTGAACTTTCAGCGTCCGACCCTG
GCAAGCGGTTGGGATGTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTG
AAAAATGGCCTGTATTATCTGGGTATTATGCCGAAACAGAAAGGTCGTTATAAA
GCGCTGAGCTTTGAACCGACGGAAAAAACCAGTGAAGGTTTTGATAAAATGTAC
TACGACTATTTTCCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTG
AAAGCAGTTACCGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAAT
AACTTTATTGAACCGCTGGAAATCACCAAAGAGATCTACGATCTGAATAACCCG
GAAAAAGAGCCGAAAAAATTCCAGACCGCATATGCAAAAAAAACCGGTGATCAG
AAAGGTTATCGTGAAGCGCTGTGTAAATGGATTGATTTCACCCGTGATTTTCTG
AGCAAATACACCAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAG
CCAGTATAAAGATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCA
TATTAGCTTTCAGCGTATTGCCGAGAAAGAAATCATGGACGCAGTTGAAACCGG
TAAACTGTACCTGTTCCAGATCTACAATAAAGATTTTGCCAAAGGCCATCATGGC
AAACCGAATCTGCATACCCTGTATTGGACC

FIG 40 (cont'd)

GGTCTGTTTAGCCCTGAAAATCTGGCAAAAACCTCGATTAAACTGAATGGTCAG
GCGGAACTGTTTTATCGTCCGAAAAGCCGTATGAAACGTATGGCACATCGTCTG
GGTGAAAAAATGCTGAACAAAAAACTGAAAGACCAGAAAACCCCGATCCCGGAT
ACACTGTATCAAGAACTGTATGATTATGTGAACCATCGTCTGAGCCATGATCTG
AGTGATGAAGCACGTGCCCTGCTGCCGAATGTTATTACCAAAGAAGTTAGCCAC
GAGATCATTAAAGATCGTCGTTTTACCAGCGACAAATTCCTGTTTCATGTGCCG
ATTACCCTGAATTATCAGGCAGCAAATAGCCCGAGCAAATTTAACCAGCGTGTT
AATGCATATCTGAAAGAACATCCAGAAACGCCGATTATTGGTATTGATCGTGGT
GAACGTAACCTGATTTATATCACCGTTATTGATAGCACCGGCAAAATCCTGGAA
CAGCGTAGCCTGAATACCATTCAGCAGTTTGATTACCAGAAAAAACTGGATAAT
CGCGAGAAAGAACGTGTTGCAGCACGTCAGGCATGGTCAGTTGTTGGTACAAT
TAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTATTCATGAAATTGTGGATCT
GATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAACCTGAATTTTGGCTTTAAA
AGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTATCAGCAGTTCGAGAAAATG
CTGATTGACAAACTGAATTGCCTGGTGCTGAAAGATTATCCGGCTGAAAAAGTT
GGTGGTGTTCTGAATCCGTATCAGCTGACCGATCAGTTTACCAGCTTTGCAAAA
ATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCCGGCACCGTATACGAGCAAA
ATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGTTTGGAAAACCATCAAAAACC
ATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGATTTTCTGCATTACGACGTTAA
AACGGGTGATTTCATCCTGCACTTTAAAATGAATCGCAATCTGAGTTTTCAGCGT
GGCCTGCCTGGTTTTATGCCTGCATGGGATATTGTGTTTGAGAAAAACGAAACA
CAGTTCGATGCAAAAGGCACCCCGTTTATTGCAGGTAAACGTATTGTTCCGGTG
ATTGAAAATCATCGTTTCACCGGTCGTTATCGCGATCTGTATCCGGCAAATGAA
CTGATCGCACTGCTGGAAGAGAAAGGTATTGTTTTTCGTGATGGCTCAAACATT
CTGCCGAAACTGCTGGAAAATGATGATAGCCATGCAATTGATACCATGGTTGCA
CTGATTCGTAGCGTTCTGCAGATGCGTAATAGCAATGCAGCAACCGGTGAAGAT
TACATTAATAGTCCGGTTCGTGATCTGAATGGTGTTTGTTTTGATAGCCGTTTTC
AGAATCCGGAATGGCCGATGGATGCAGATGCAAATGGTGCATATCATATTGCAC
TGAAAGGACAGCTGCTGCTGAACCACCTGAAAGAAAGCAAAGATCTGAAACTG
CAAAACGGCATTAGCAATCAGGATTGGCTGGCATATATCCAAGAACTGCGTAAC
GGTCGT**AGCAGTGATGATGAAGCAACCGCAGATAGCCAGCATGCAGCACCG**
CCTAAAAAGAAACGTAAAGTTGGTGGTAGCGGTGGTTCAGGTGGTAGTGGCG
GTAGTGGTGGCTCAGGGGGTTCTGGTGGCTCTGGTGGTAGCCTCGAGCACCA
CCACCACCACCAC

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

MGRDPGKPIPNPLLGLDSTAPKKKRKVGIHGVPAATQFEGFTNLYQVSKTLRFELIP
QGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRIYKTYADQCLQLVQLDWENLSAAID
SYRKEKTEETRNALIEEQATYRNAIHDYFIGRTDNLTDAINKRHAEIYKGLFKAELFN
GKVLKQLGTVTTTEHENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIVQD
NFPKFKENCHIFTRLITAVPSLREHFENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQID
LYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTL
SFILEEFKSDEEVIQSFCKYKTLLRNENVLETAEALFNELNSIDLTHIFISHKKLETISS
ALCDHWDTLRNALYERRISELTGKITKSAKEKVQRSLKHEDINLQEIISAAGKELSEA
FKQKTSEILSHAHAALDQPLPTTLKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNE
VDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEKFKLNFQMPTLASGWDVN
KEKNNGAILFVKNGLYYLGIMPK

FIG 40 (cont'd)

QKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPIL
LSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTRDFLS
KYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLF
QIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRMKRMA
HRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVITKEVSH
EIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDRGERNLI
YITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKDLKQGYL
SQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKD
YPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWK
TIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEK
NETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIVFRDGSNILP
KLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVCFDSRFQNPE
WPMDADANGAYHIALKGQLLLNHLKESKDLKLQNGISNQDWLAYIQELRN**PKKKR**
KVKLAAALE<u>HHHHHH</u>

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

ATGGGTCGGGATCCAGGTAAACCGATTCCGAATCCGCTGCTGGGTCTGGATAG
CACCGCACCGAAAAAAAAACGTAAAGTTGGTATTCATGGTGTTCCGGCAGCAAC
CCAGTTTGAAGGTTTCACCAATCTGTATCAGGTTAGCAAAACCCTGCGTTTTGA
ACTGATTCCGCAGGGTAAAACCCTGAAACATATTCAAGAACAGGGCTTCATCGA
AGAGGATAAAGCACGTAACGATCACTACAAAGAACTGAAACCGATTATCGACCG
CATCTATAAAACCTATGCAGATCAGTGTCTGCAGCTGGTTCAGCTGGATTGGGA
AAATCTGAGCGCAGCAATTGATAGTTATCGCAAAGAAAAAACCGAAGAAACCCG
TAATGCACTGATTGAAGAACAGGCAACCTATCGTAATGCCATCCATGATTATTTC
ATTGGTCGTACCGATAATCTGACCGATGCAATTAACAAACGTCACGCCGAAATC
TATAAAGGCCTGTTTAAAGCCGAACTGTTTAATGGCAAAGTTCTGAAACAGCTG
GGCACCGTTACCACCACCGAACATGAAAATGCACTGCTGCGTAGCTTTGATAAA
TTCACCACCTATTTCAGCGGCTTTTATGAGAATCGCAAAAACGTGTTTAGCGCA
GAAGATATTAGCACCGCAATTCCGCATCGTATTGTGCAGGATAATTTCCCGAAA
TTCAAAGAGAACTGCCACATTTTTACCCGTCTGATTACCGCAGTTCCGAGCCTG
CGTGAACATTTTGAAAACGTTAAAAAAGCCATCGGCATCTTTGTTAGCACCAGC
ATTGAAGAAGTTTTTAGCTTCCCGTTTTACAATCAGCTGCTGACCCAGACCCAG
ATTGATCTGTATAACCAACTGCTGGGTGGTATTAGCCGTGAAGCAGGCACCGAA
AAAATCAAAGGTCTGAATGAAGTGCTGAATCTGGCCATTCAGAAAAATGATGAA
ACCGCACATATTATTGCAAGCCTGCCGCATCGTTTTATTCCGCTGTTCAAACAAA
TTCTGAGCGATCGTAATACCCTGAGCTTTATTCTGGAAGAATTCAAATCCGATGA
AGAGGTGATTCAGAGCTTTTGCAAATACAAAACGCTGCTGCGCAATGAAAATGT
TCTGGAAACTGCCGAAGCACTGTTTAACGAACTGAATAGCATTGATCTGACCCA
CATCTTTATCAGCCACAAAAAACTGGAAACCATTTCAAGCGCACTGTGTGATCAT
TGGGATACCCTGCGTAATGCCCTGTATGAACGTCGTATTAGCGAACTGACCGG
TAAAATTACCAAAAGCGCGAAAGAAAAAGTTCAGCGCAGTCTGAAACATGAGGA
TATTAATCTGCAAGAGATTATTAGCGCAGCCGGTAAAGAACTGTCAGAAGCATT
TAAACAGAAAACCAGCGAAATTCTGTCACATGCACATGCAGCACTGGATCAGCC
GCTGCCGACCACCCTGAAAAAACAAGAAGAAAAAGAAATCCTGAAAAGCCAGC
TGGATAGCCTGCTGGGTCTGTATCATCTGCTGGACTGGTTTGCAGTTGATGAAA
GCAATGAAGTTGATCCGGAATTTAGCGCACGTCTGACCGGCATTAAACTGGAAA
TGGAACCGAGCCTGAGCTTTTATAACAAAGCCCGTAATTATGCCACCAAAAAAC
CGTAT

FIG 40 (cont'd)

AGCGTCGAAAAATTCAAACTGAACTTTCAGATGCCGACCCTGGCAAGCGGTTG
GGATGTTAATAAAGAAAAAAACAACGGTGCCATCCTGTTCGTGAAAAATGGCCT
GTATTATCTGGGTATTATGCCGAAACAGAAAGGTCGTTATAAAGCGCTGAGCTT
TGAACCGACGGAAAAAACCAGTGAAGGTTTTGATAAAATGTACTACGACTATTTT
CCGGATGCAGCCAAAATGATTCCGAAATGTAGCACCCAGCTGAAAGCAGTTAC
CGCACATTTTCAGACCCATACCACCCCGATTCTGCTGAGCAATAACTTTATTGAA
CCGCTGGAAATCACCAAAGAGATCTACGATCTGAATAACCCGGAAAAAGAGCC
GAAAAAATTCCAGACCGCATATGCAAAAAAAACCGGTGATCAGAAAGGTTATCG
TGAAGCGCTGTGTAAATGGATTGATTTCACCCGTGATTTTCTGAGCAAATACAC
CAAAACCACCAGTATCGATCTGAGCAGCCTGCGTCCGAGCAGCCAGTATAAAG
ATCTGGGCGAATATTATGCAGAACTGAATCCGCTGCTGTATCATATTAGCTTTCA
GCGTATTGCCGAGAAAGAAATCATGGACGCAGTTGAAACCGGTAAACTGTACCT
GTTCCAGATCTACAATAAAGATTTTGCCAAAGGCCATCATGGCAAACCGAATCT
GCATACCCTGTATTGGACCGGTCTGTTTAGCCCTGAAAATCTGGCAAAAACCTC
GATTAAACTGAATGGTCAGGCGGAACTGTTTTATCGTCCGAAAAGCCGTATGAA
ACGTATGGCACATCGTCTGGGTGAAAAAATGCTGAACAAAAAACTGAAAGACCA
GAAAACCCCGATCCCGGATACACTGTATCAAGAACTGTATGATTATGTGAACCA
TCGTCTGAGCCATGATCTGAGTGATGAAGCACGTGCCCTGCTGCCGAATGTTAT
TACCAAAGAAGTTAGCCACGAGATCATTAAAGATCGTCGTTTTACCAGCGACAA
ATTCTTTTTTCATGTGCCGATTACCCTGAATTATCAGGCAGCAAATAGCCCGAG
CAAATTTAACCAGCGTGTTAATGCATATCTGAAAGAACATCCAGAAACGCCGAT
TATTGGTATTGATCGTGGTGAACGTAACCTGATTTATATCACCGTTATTGATAGC
ACCGGCAAAATCCTGGAACAGCGTAGCCTGAATACCATTCAGCAGTTTGATTAC
CAGAAAAAACTGGATAATCGCGAGAAAGAACGTGTTGCAGCACGTCAGGCATG
GTCAGTTGTTGGTACAATTAAAGACCTGAAACAGGGTTATCTGAGCCAGGTTAT
TCATGAAATTGTGGATCTGATGATTCACTATCAGGCCGTTGTTGTGCTGGAAAA
CCTGAATTTTGGCTTTAAAAGCAAACGTACCGGCATTGCAGAAAAAGCAGTTTA
TCAGCAGTTCGAGAAAATGCTGATTGACAAACTGAATTGCCTGGTGCTGAAAGA
TTATCCGGCTGAAAAAGTTGGTGGTGTTCTGAATCCGTATCAGCTGACCGATCA
GTTTACCAGCTTTGCAAAAATGGGCACCCAGAGCGGATTTCTGTTTTATGTTCC
GGCACCGTATACGAGCAAAATTGATCCGCTGACCGGTTTTGTTGATCCGTTTGT
TTGGAAAACCATCAAAAACCATGAAAGCCGCAAACATTTTCTGGAAGGTTTCGA
TTTTCTGCATTACGACGTTAAAACGGGTGATTTCATCCTGCACTTTAAAATGAAT
CGCAATCTGAGTTTTCAGCGTGGCCTGCCTGGTTTTATGCCTGCATGGGATATT
GTGTTTGAGAAAAACGAAACACAGTTCGATGCAAAAGGCACCCCGTTTATTGCA
GGTAAACGTATTGTTCCGGTGATTGAAAATCATCGTTTCACCGGTCGTTATCGC
GATCTGTATCCGGCAAATGAACTGATCGCACTGCTGGAAGAGAAAGGTATTGTT
TTTCGTGATGGCTCAAACATTCTGCCGAAACTGCTGGAAAATGATGATAGCCAT
GCAATTGATACCATGGTTGCACTGATTCGTAGCGTTCTGCAGATGCGTAATAGC
AATGCAGCAACCGGTGAAGATTACATTAATAGTCCGGTTCGTGATCTGAATGGT
GTTTGTTTTGATAGCCGTTTTCAGAATCCGGAATGGCCGATGGATGCAGATGCA
AATGGTGCATATCATATTGCACTGAAAGGACAGCTGCTGCTGAACCACCTGAAA
GAAAGCAAAGATCTGAAACTGCAAAACGGCATTAGCAATCAGGATTGGCTGGC
ATATATCCAAGAACTGCGTAAC**CCTAAAAAAAAAACGCAAAGTG**AAGCTTGCGG
CCGCACTCGAGCACCACCACCACCACCAC

FIG 41A

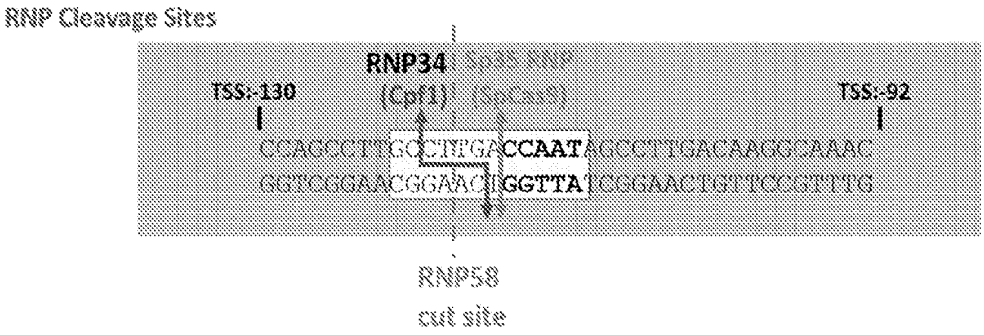


FIG 41B

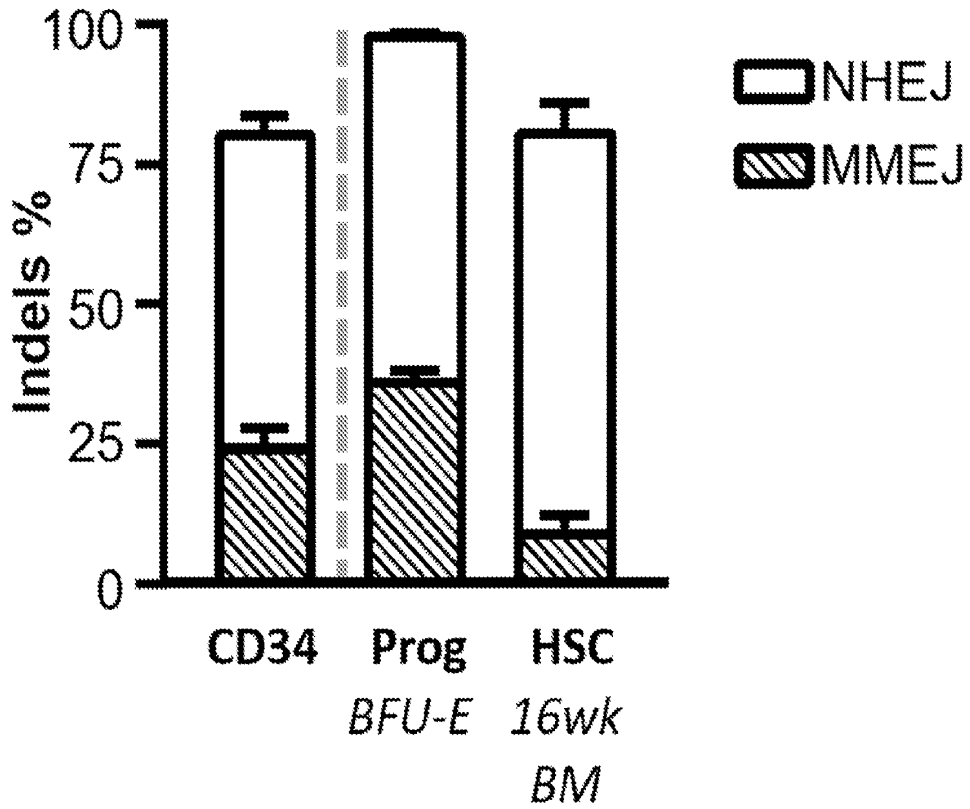


FIG 41C

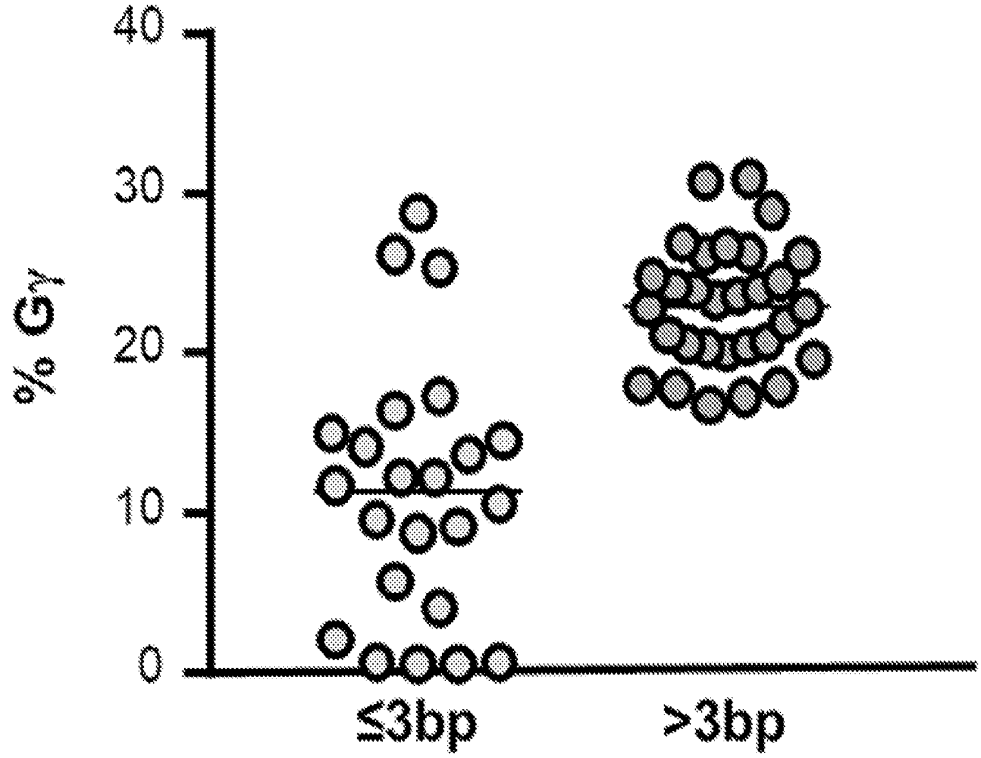


FIG 41D

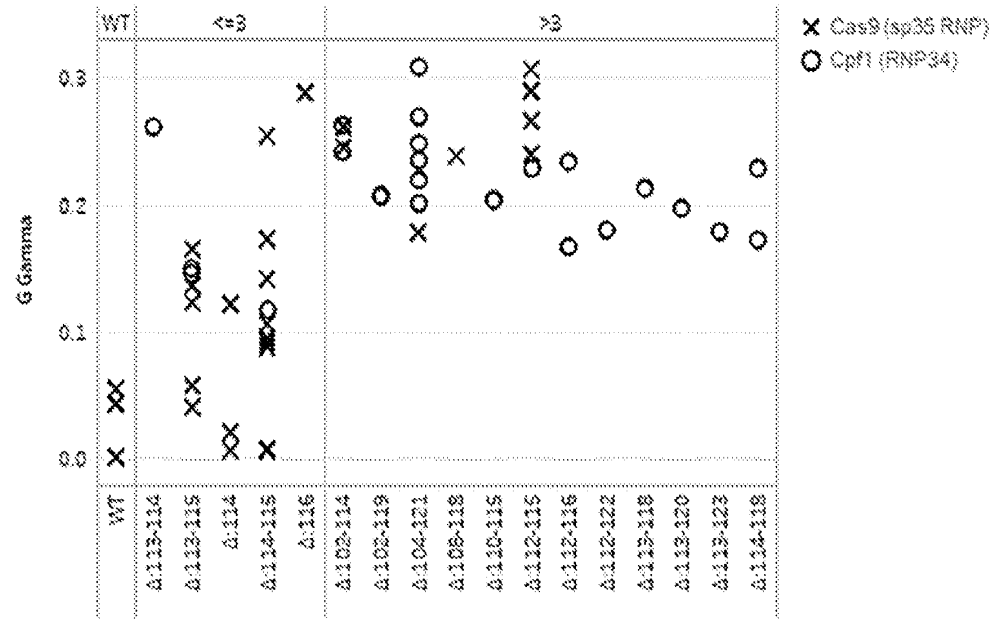

WT
<=3
>3
Cas9 (sp35 RNP)
Cpf1 (RNP34)
G Gamma
0.0
0.1
0.2
0.3
WT
Δ:113-114
Δ:113-115
Δ:114
Δ:114-115
Δ:116
Δ:102-114
Δ:102-119
Δ:104-121
Δ:108-118
Δ:110-115
Δ:112-115
Δ:112-116
Δ:112-122
Δ:113-118
Δ:113-120
Δ:113-123
Δ:114-118

FIG 41E

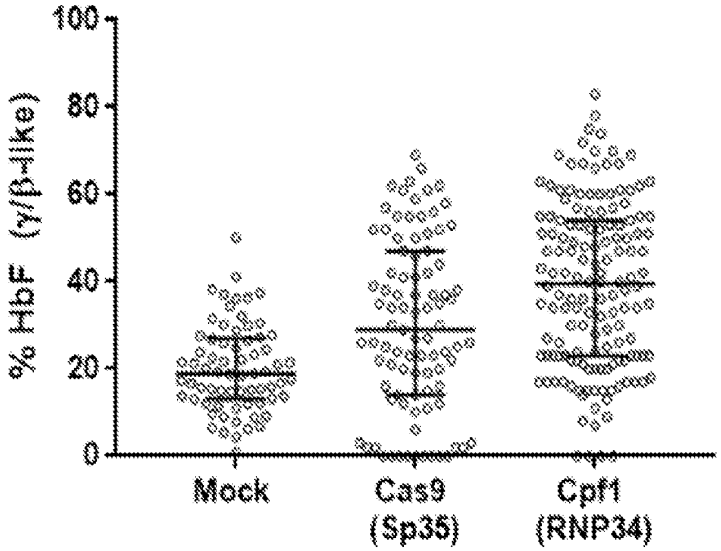


FIG 42A

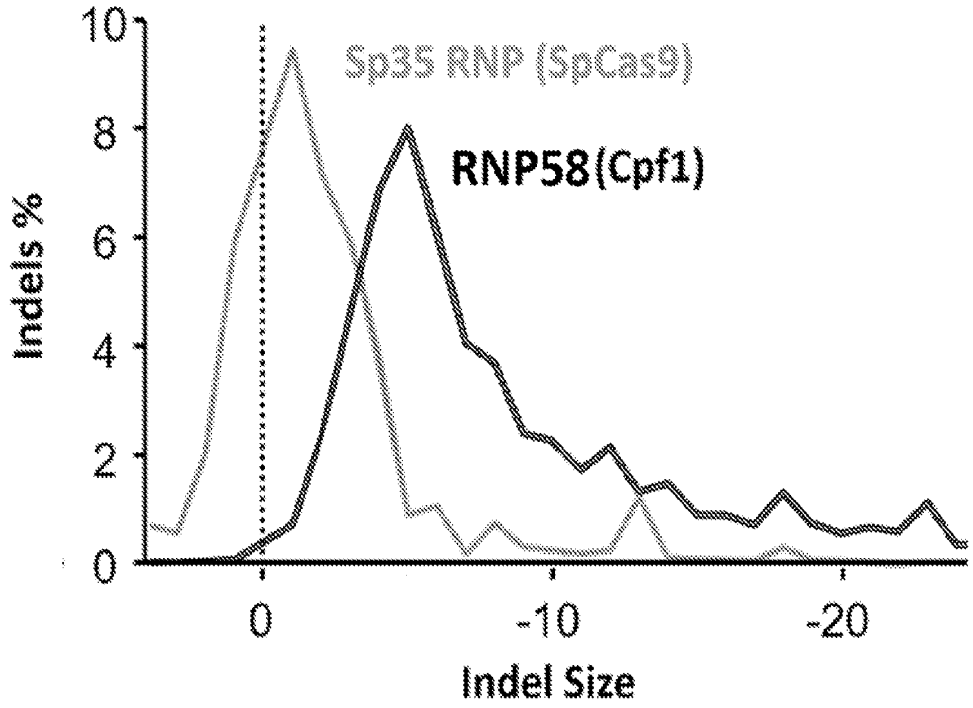


FIG 42B

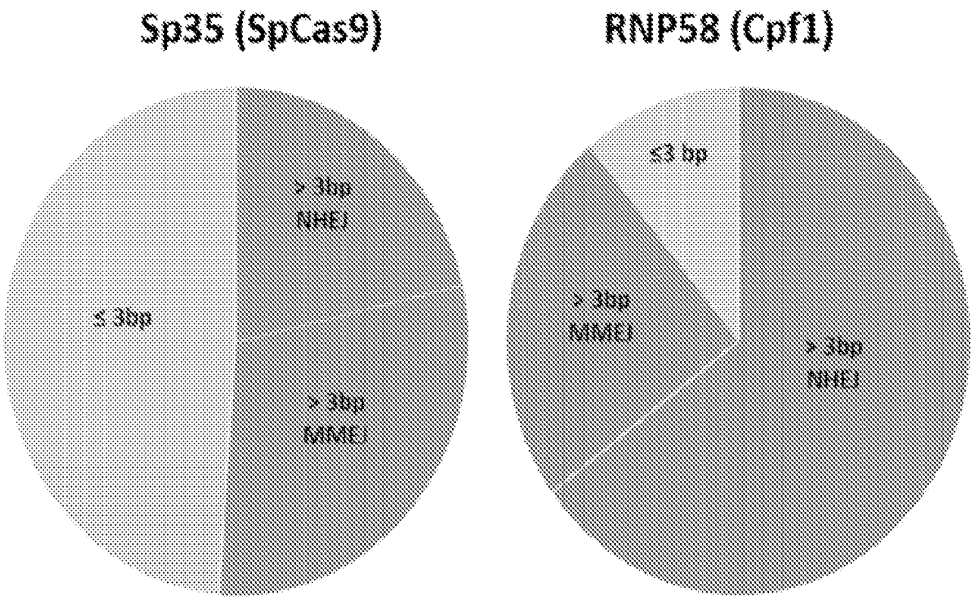

Sp35 (SpCas9)
≤ 3bp
> 3bp NHEJ
> 3bp MMEJ
RNP58 (Cpf1)
≤3 bp
> 3bp MMEJ
> 3bp NHEJ

FIG 42C

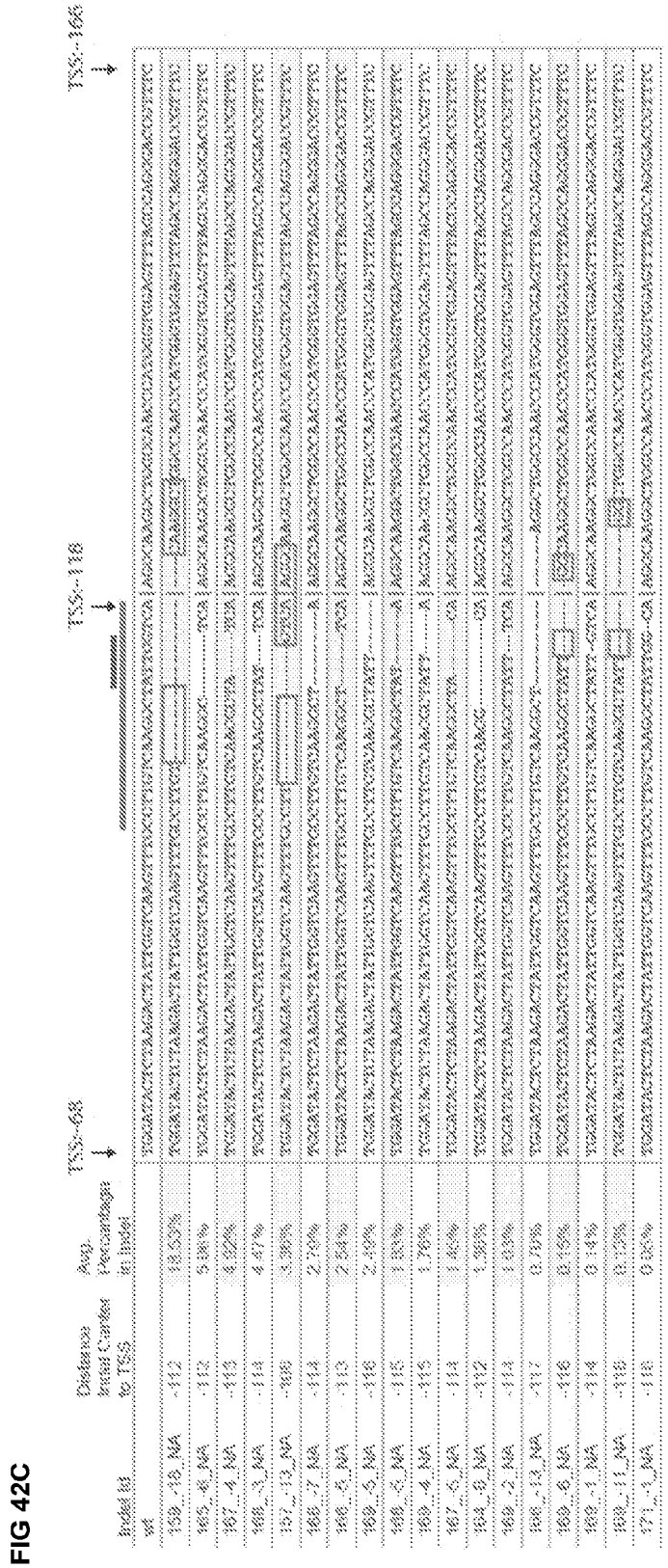

Indel Id
Distance from Center to TSS
Avg. Percentage in Indel
TSS--68
TSS--116
TSS--166
wt
159_-10_NA -110 18.50%
165_-6_NA -112 5.38%
167_-4_NA -113 4.82%
168_-3_NA -114 4.47%
162_-13_NA -108 3.38%
168_-7_NA -114 2.79%
166_-5_NA -113 2.34%
169_-5_NA -116 2.40%
169_-5_NA -115 1.82%
169_-4_NA -115 1.76%
167_-5_NA -114 1.42%
164_-9_NA -112 1.36%
169_-2_NA -114 1.03%
166_-13_NA -117 0.70%
169_-6_NA -116 0.15%
169_-1_NA -114 0.14%
169_-11_NA -116 0.13%
171_-1_NA -116 0.06%

FIG 42D

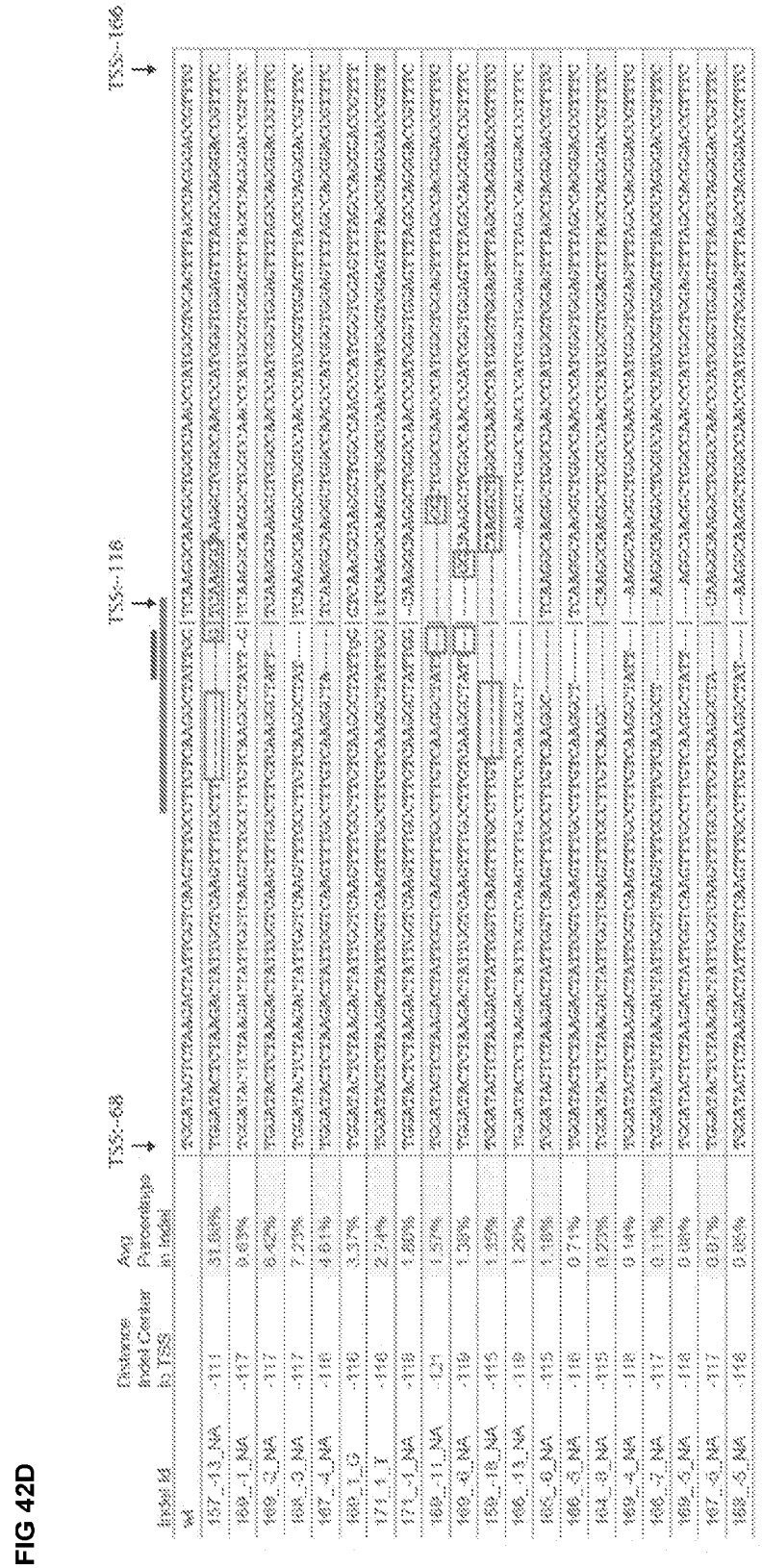

TSS--68
TSS--116
TSS--166
Indel Id
Distance Indel Center to TSS
Avg Percentage in Indel
wt
167_-13_NA -111 31.80%
169_-1_NA -117 0.69%
169_-3_NA -117 0.42%
168_-3_NA -117 7.73%
167_-4_NA -116 4.61%
168_1_C -116 3.37%
171_1_T -116 2.74%
171_-1_NA -118 1.86%
169_-11_NA -121 1.57%
169_-8_NA -119 1.38%
169_-18_NA -115 1.35%
166_-13_NA -119 1.20%
165_-6_NA -115 1.18%
166_-5_NA -116 0.71%
164_-9_NA -115 0.70%
169_-4_NA -118 0.14%
168_-7_NA -117 0.11%
169_-5_NA -118 0.08%
167_-3_NA -117 0.07%
169_-5_NA -116 0.06%

FIG 42E

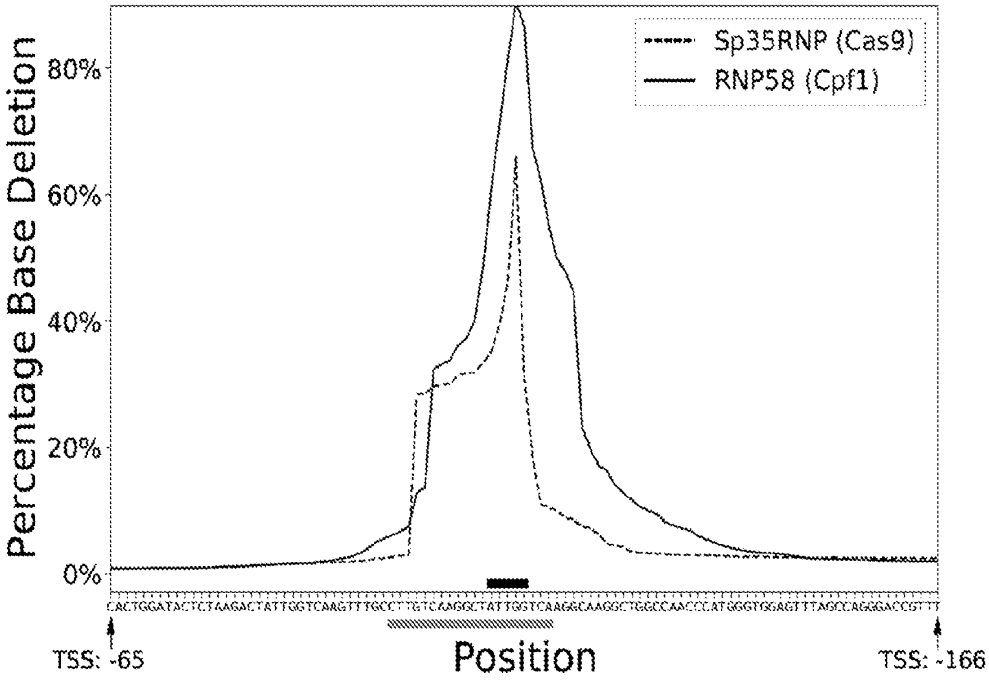

Sp35RNP (Cas9)
RNP58 (Cpf1)
Percentage Base Deletion
80%
60%
40%
20%
0%
TSS: -65
Position
TSS: -166

FIG 42F

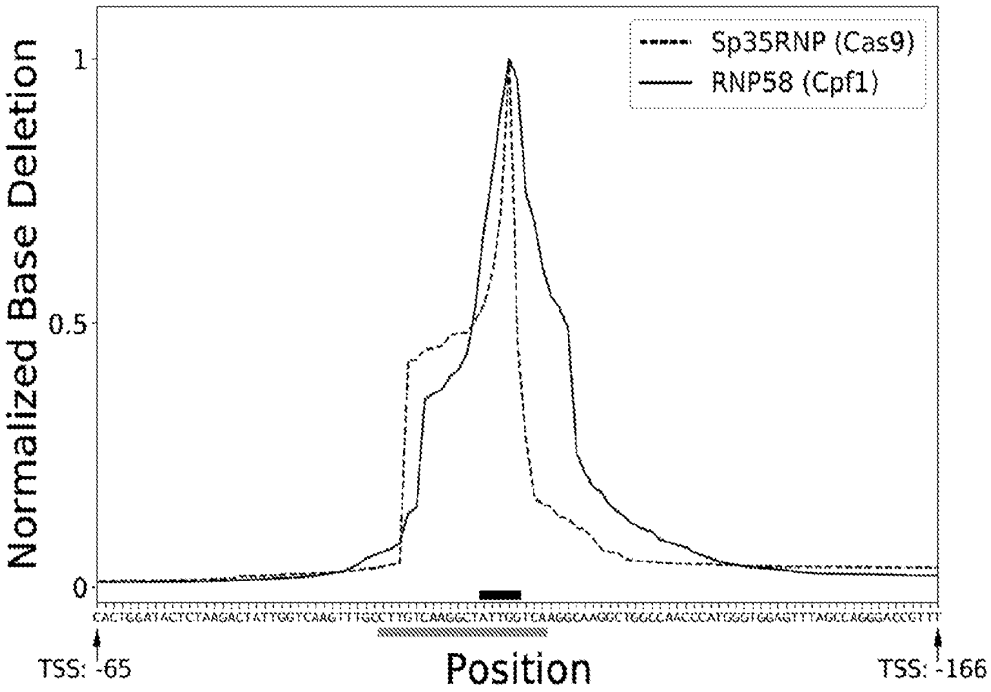

Sp35RNP (Cas9)
RNP58 (Cpf1)
Normalized Base Deletion
1
0.5
0
TSS: -65
Position
TSS: -166

FIG 42G

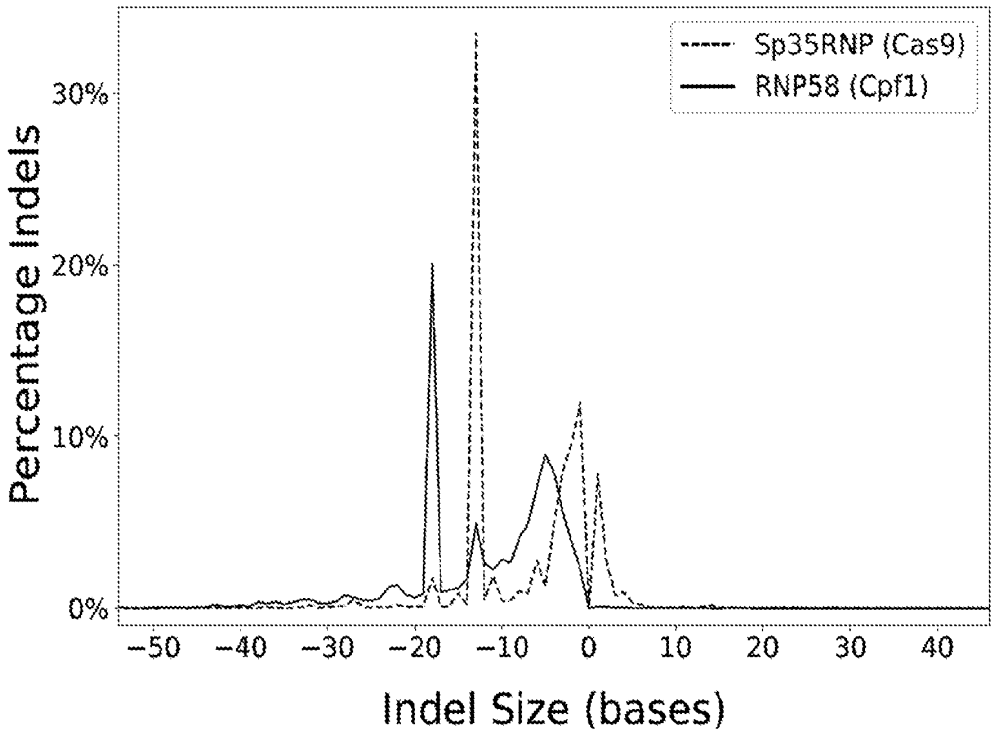

Sp35RNP (Cas9)
RNP58 (Cpf1)
Percentage Indels
30%
20%
10%
0%
−50
−40
−30
−20
−10
0
10
20
30
40
Indel Size (bases)

FIG 42H

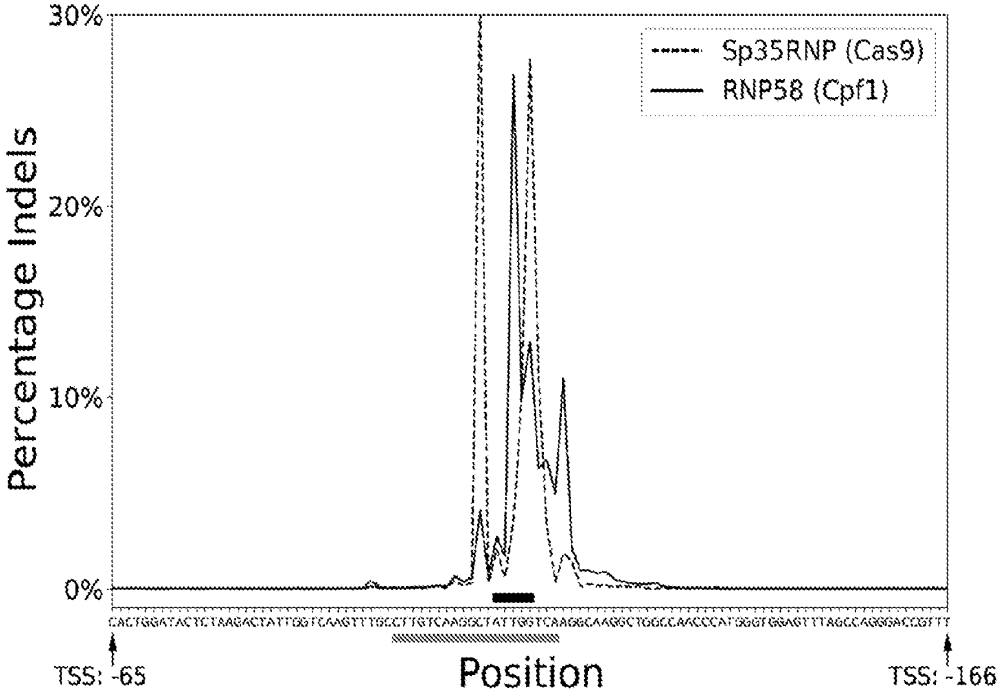

30%
20%
10%
0%
Percentage Indels
Sp35RNP (Cas9)
RNP58 (Cpf1)
TSS: -65
Position
TSS: -166

FIG 42I

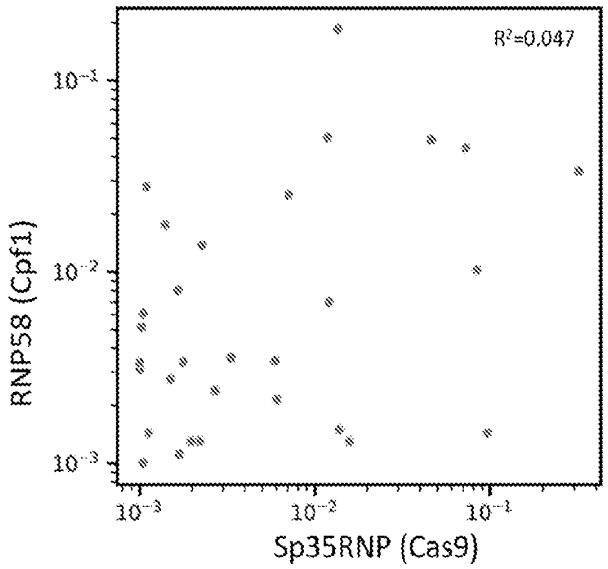


FIG 42J

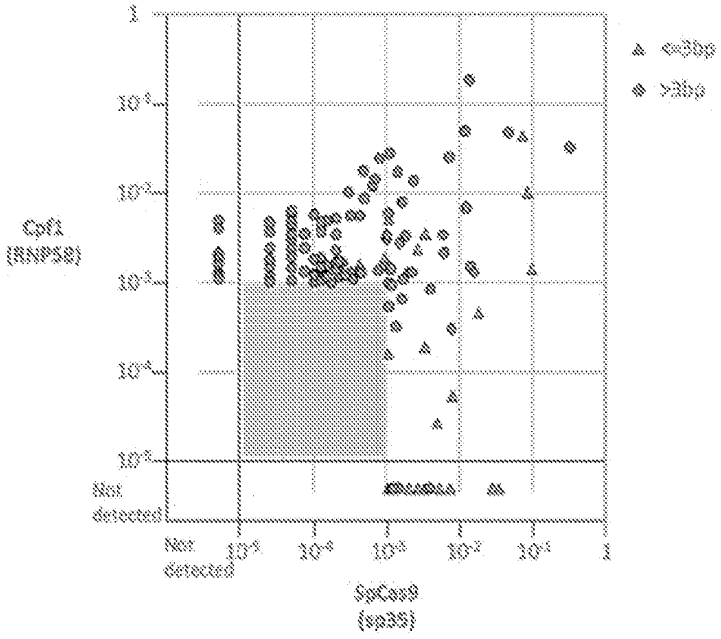


FIG 43A

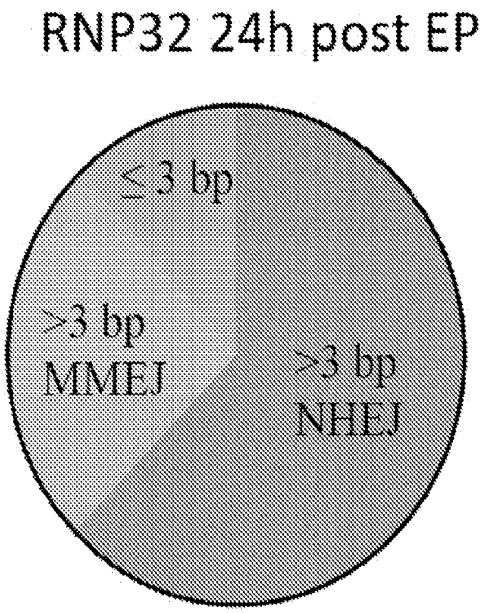


FIG 43B

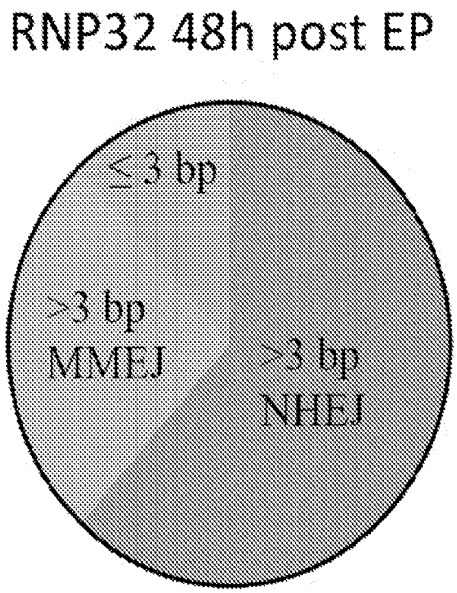


FIG 43C

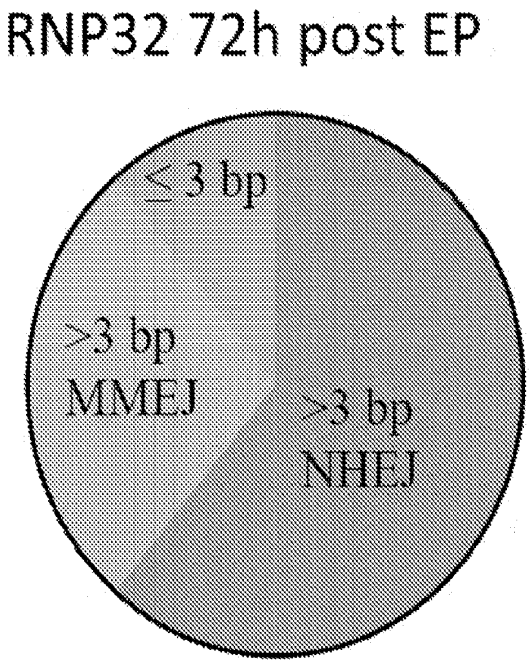


FIG 43D

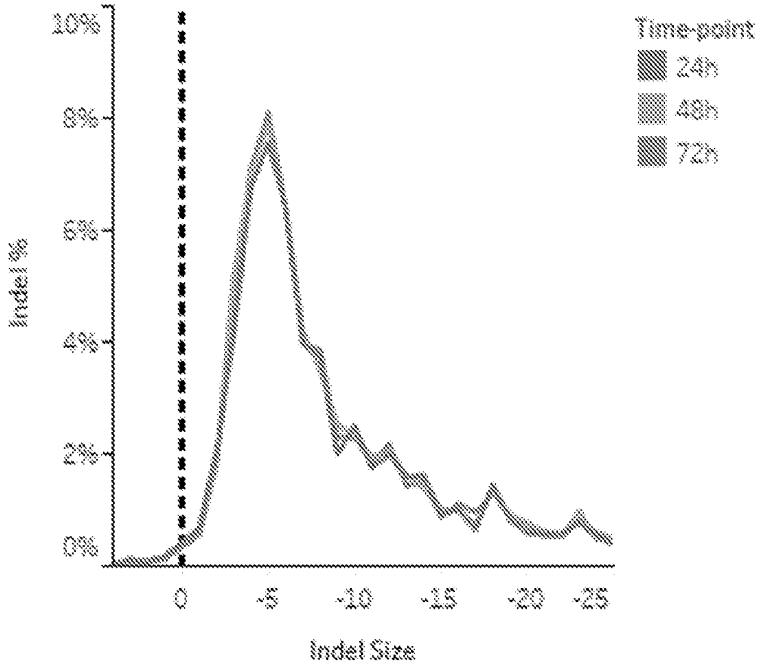


FIG 43

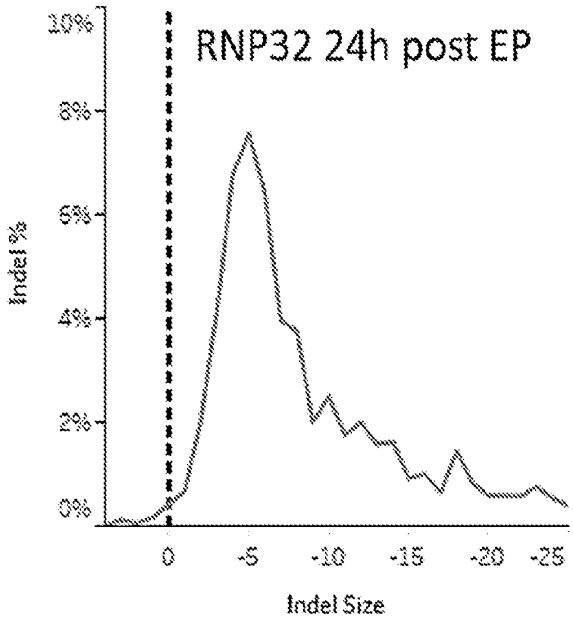


FIG 43F

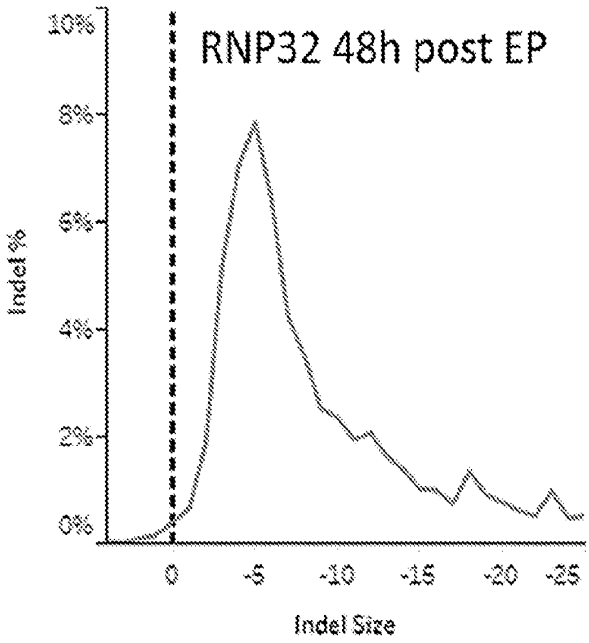


FIG 43G

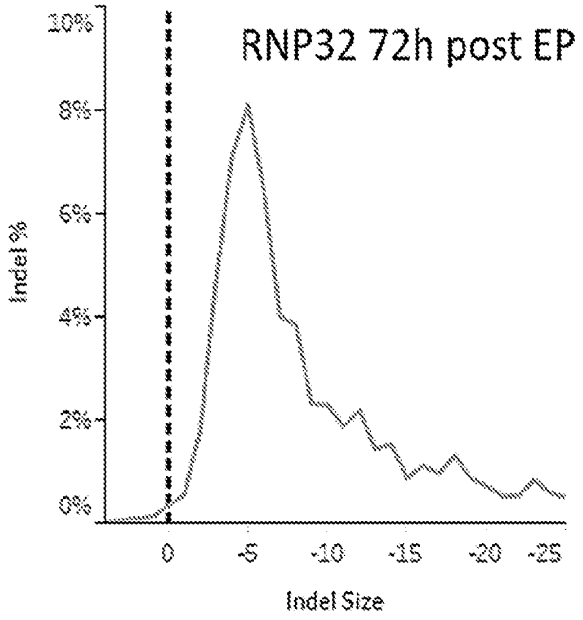


FIG 43H

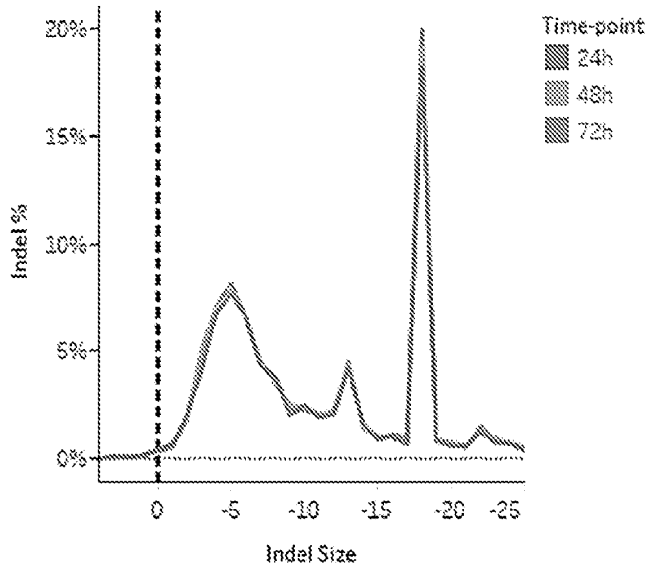


FIG 43I

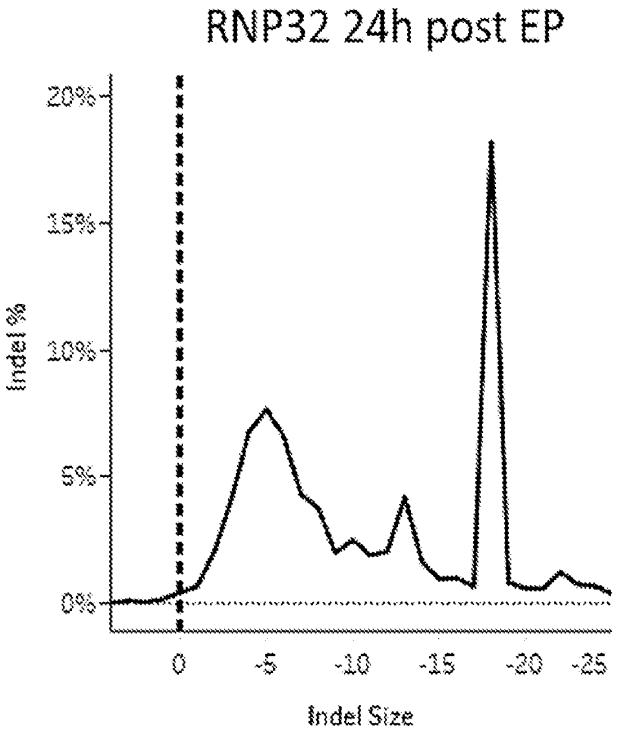


FIG 43J

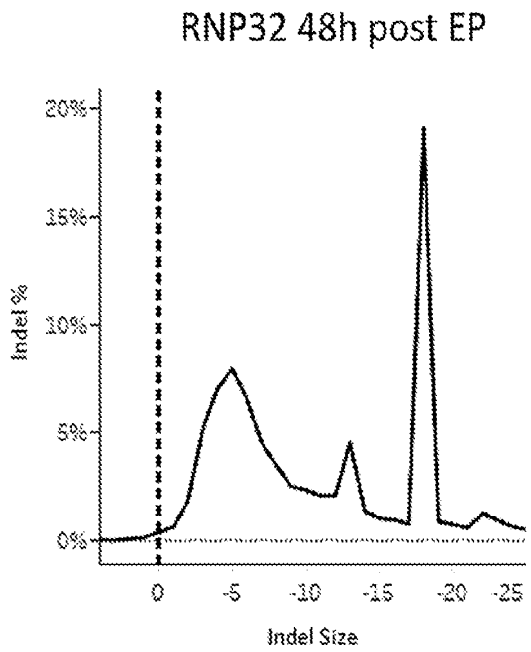


FIG 43K

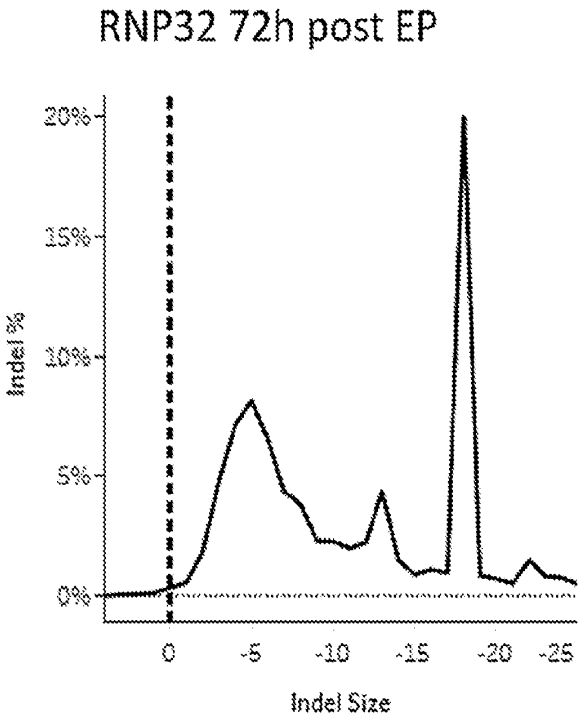


FIG 43L

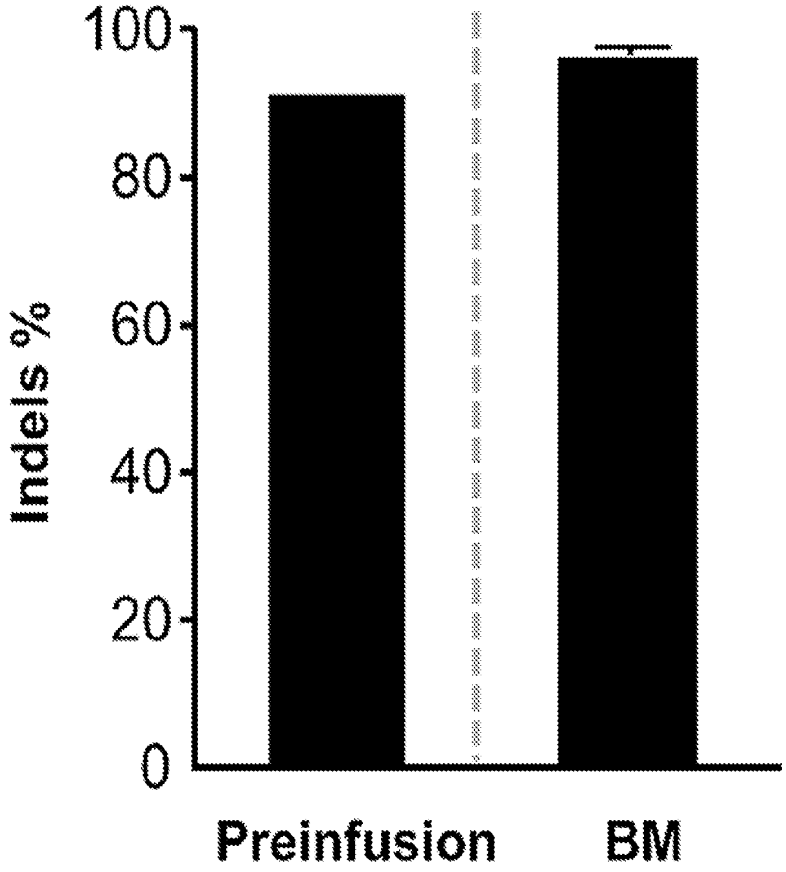


FIG 43M

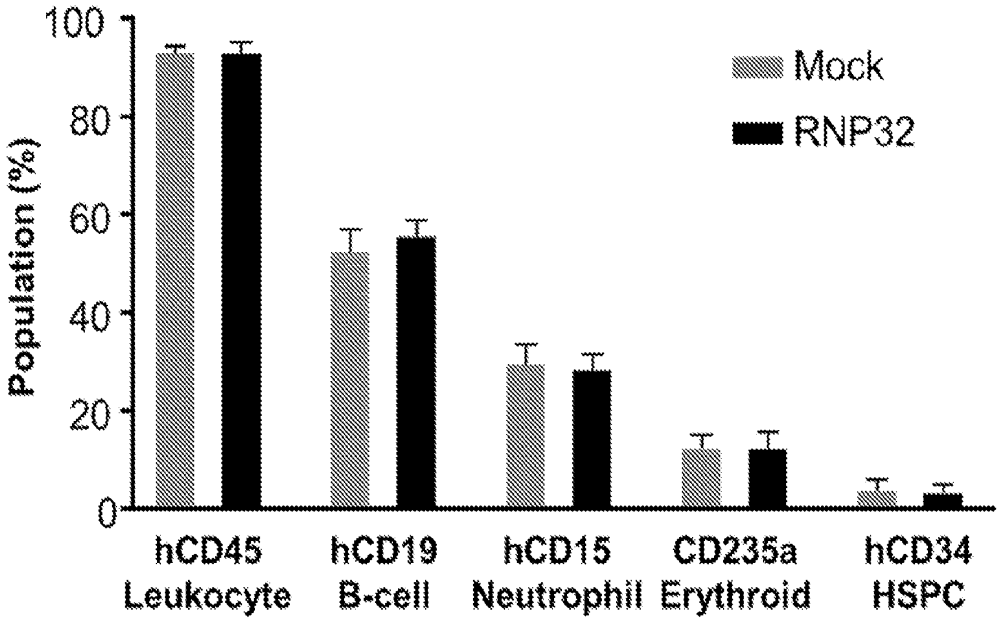

100
80
60
40
20
0
Population (%)
Mock
RNP32
hCD45
Leukocyte
hCD19
B-cell
hCD15
Neutrophil
CD235a
Erythroid
hCD34
HSPC

FIG 43N

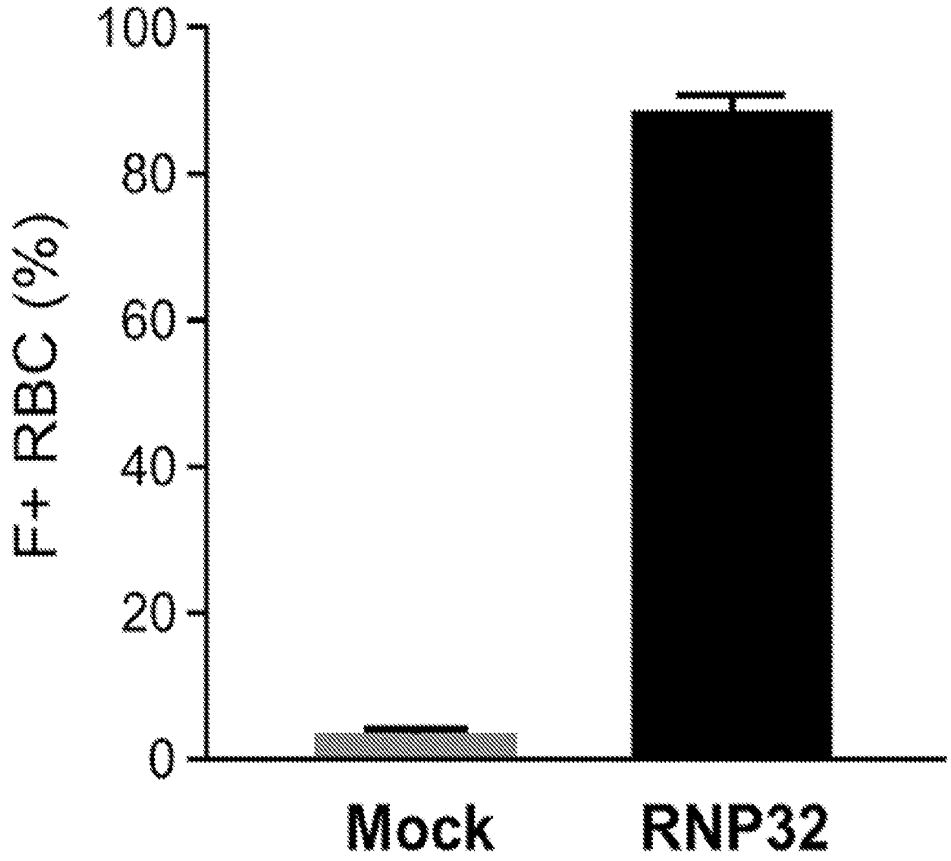


FIG 43O

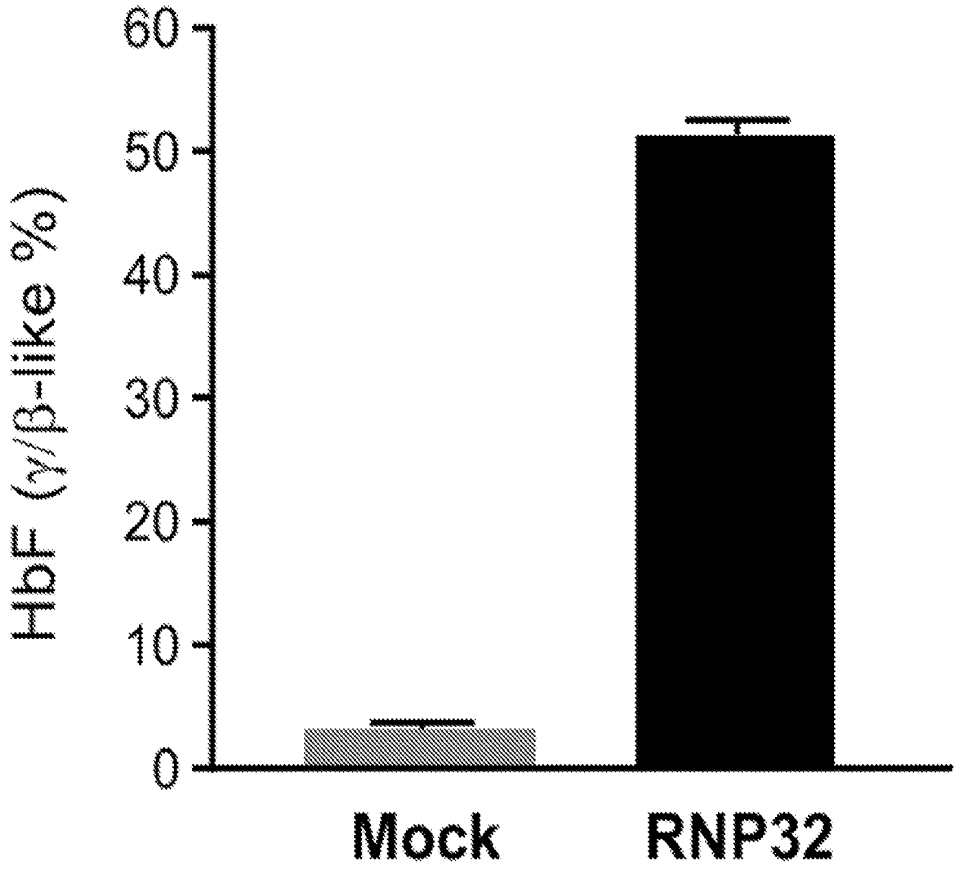


FIG 44A

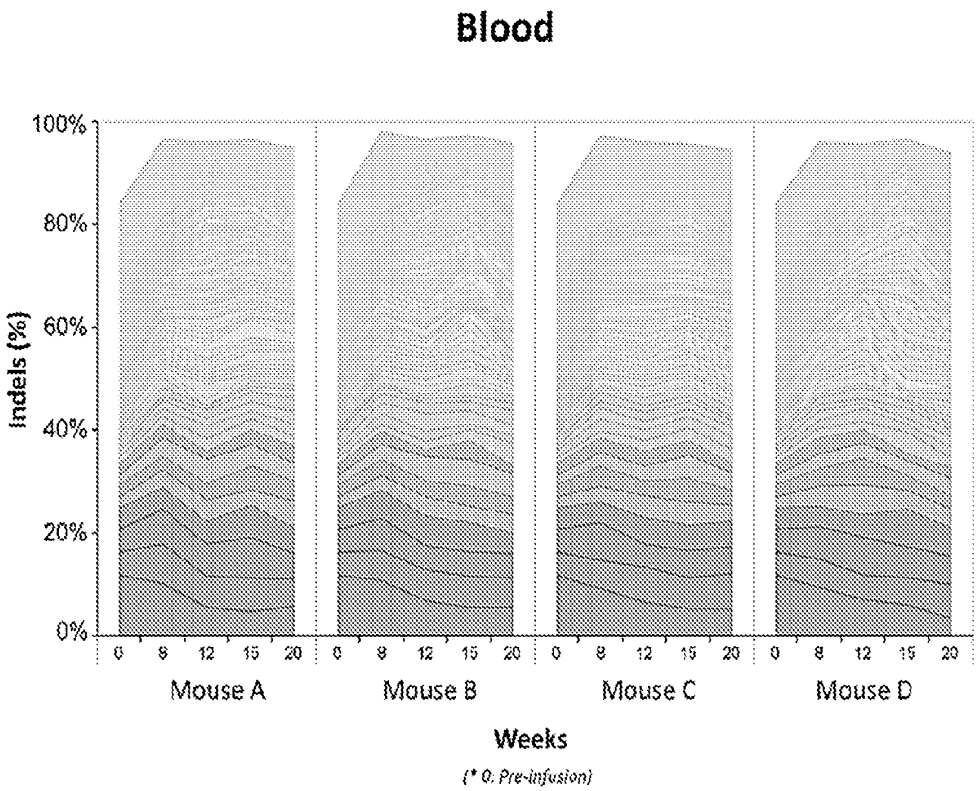

Blood
100%
80%
60%
40%
20%
0%
Indels (%)
0
8
12
15
20
Mouse A
Mouse B
Mouse C
Mouse D
Weeks
(* 0: Pre-infusion)

FIG 44B

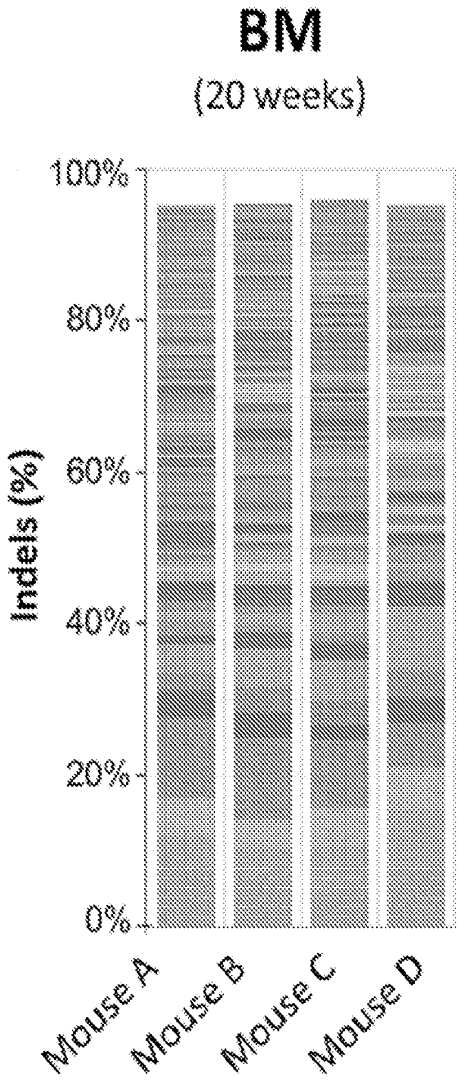

BM
(20 weeks)
Indels (%)
100%
80%
60%
40%
20%
0%
Mouse A
Mouse B
Mouse C
Mouse D

FIG 45

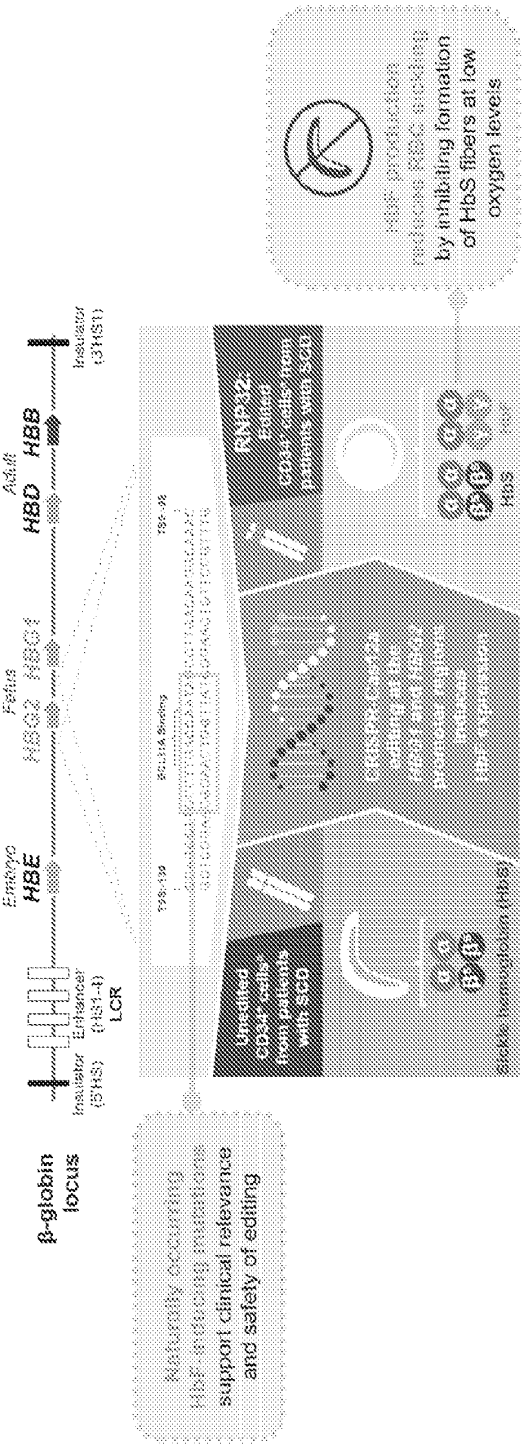

β-globin locus
Insulator (5'HS)
Enhancer (HS1-4)
LCR
Embryo
HBE
Fetus
HBG2
HBG1
Adult
HBD
HBB
Insulator (3'HS)
Naturally occurring HbF-enhancing mutations support clinical relevance and safety of editing
HbF induction restores HbS binding by inhibiting formation of HbS fibers at low oxygen levels
BCL11A Binding
BNP42
Fetal globin induction
Sickle transcription (HbS)
HbS

FIG 46A

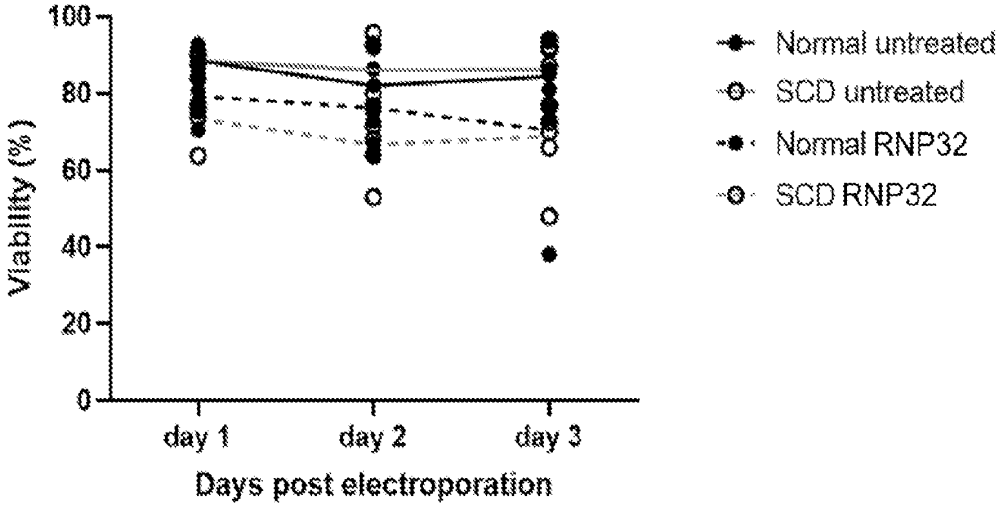

100
80
60
40
20
0
Viability (%)
day 1
day 2
day 3
Days post electroporation
Normal untreated
SCD untreated
Normal RNP32
SCD RNP32

FIG 46B

Experiment ID	Donor Lot	Donor Type	Treatment	Cell Number Per OC100 [a]	Day 1 Viability (%)	Day 2 Viability (%)	Day 3 Viability (%)
SCD011	CEL211-001	SCD	untreated		90.3	95.8	92.3
			RNP32	0.90×10^6	76.5	76.9	77.0
	CEL238-001	Normal	untreated		91.8	92.8	94.0
			RNP32	0.70×10^6	88.4	86.4	86.3
SCD014	CEL240-001	SCD	untreated		89.8	80.0	94.0
			RNP2	0.75×10^6	75.6	71.7	70.0
	CEL241-001	SCD	untreated		87.6	76.0	87.4
			RNP32	1.00×10^6	77.1	68.1	72.6
	CEL239-001	SCD	untreated		90.8	92.7	91.5
			RNP32	1.00×10^6	73.8	64.0	76.8
	CEL211-001	SCD	untreated		84.1	[b]	66.0
			RNP32	0.50×10^6	63.8	53.0	48.0
	CEL238-001	normal	untreated		85.6	76.8	85.2
			RNP32	1.00×10^6	77.5	82.0	81.1
	CEL021-014	normal	untreated		83.7	67.0	73.4
			RNP32	0.40×10^6	70.6	64.0	38.1
	CEL042-001	normal	untreated		92.9	91.9	85.1
			RNP32	1.00×10^6	80.7	72.9	76.2

FIG 47A

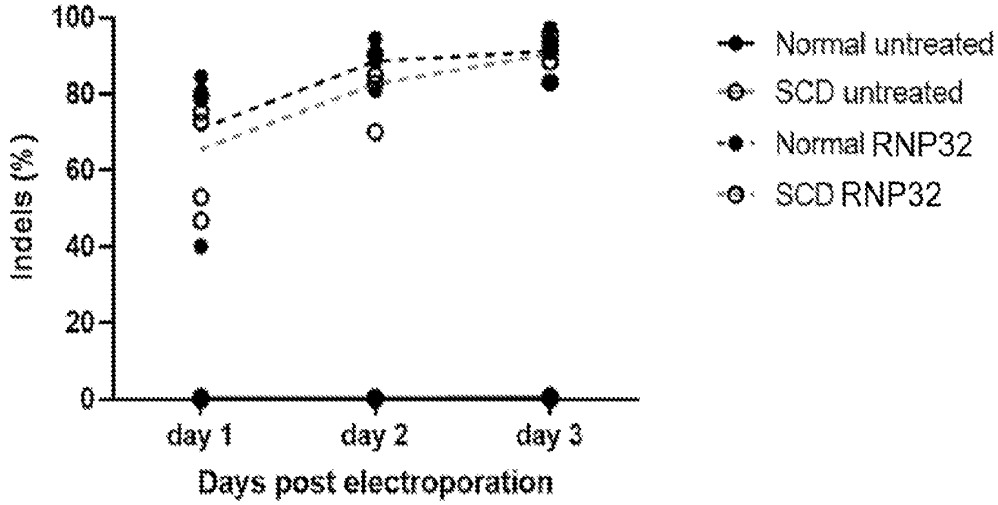

100
80
60
40
20
0
Indels (%)
day 1
day 2
day 3
Days post electroporation
Normal untreated
SCD untreated
Normal RNP32
SCD RNP32

FIG 47B

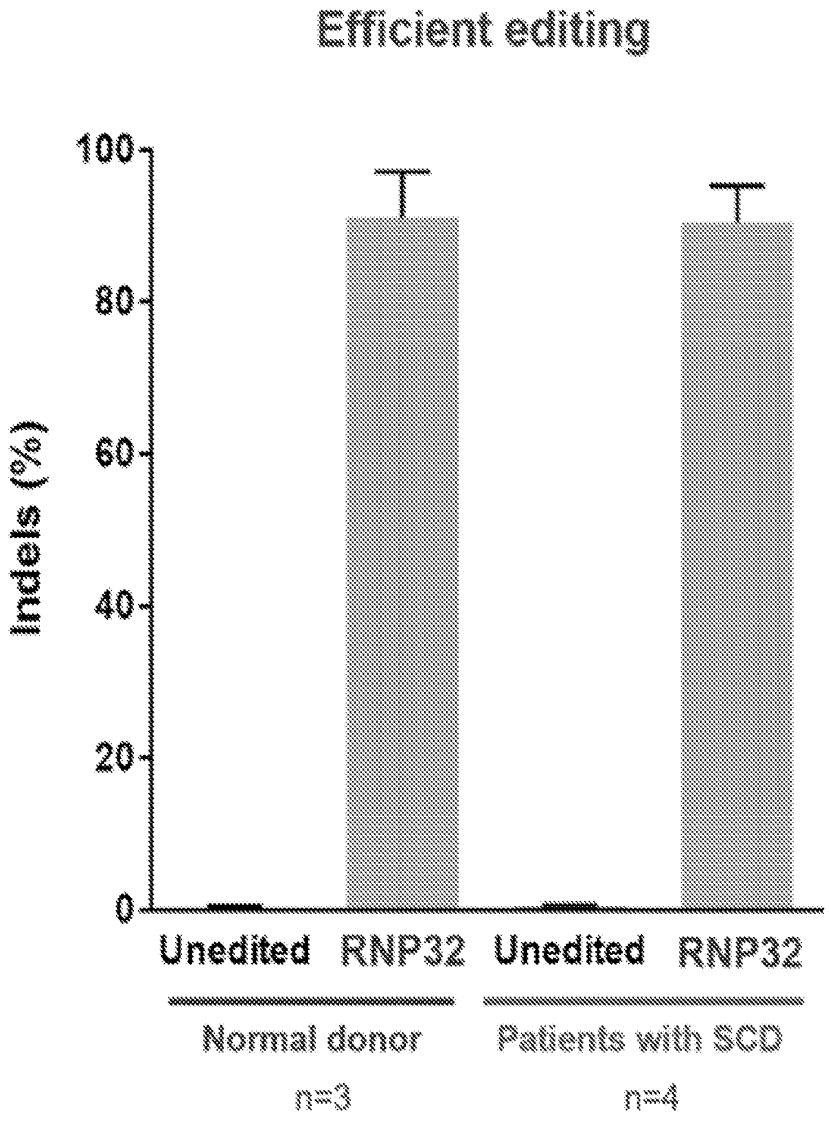

Efficient editing
Indels (%)
100
80
60
40
20
0
Unedited
RNP32
Unedited
RNP32
Normal donor
n=3
Patients with SCD
n=4

FIG 47C

Experiment ID	Donor Lot	Donor Type	Treatment	Day 1 Indels (%)	Day 2 Indels (%)	Day 3 Indels (%)
SCD011	CEL211-001	SCD	untreated	0.08	0.07	0.08
			RNP32	79.50	90.50	95.78
	CEL238-001	Normal	untreated	0.08	0.06	0.08
			RNP32	84.39	94.42	97.17
SCD014	CEL240-001	SCD	untreated	0.45	0.29	0.76
			RNP32	75.38	85.05	91.26
	CEL241-001	SCD	untreated	0.33	0.35	0.38
			RNP32	72.63	84.80	93.47
	CEL239-001	SCD	untreated	0.50	0.34	0.44
			RNP32	52.92	82.78	88.58
	CEL211-001	SCD	untreated	ND	ND	0.49
			RNP32	46.78	70.06	83.04
	CEL238-001	normal	untreated	0.72	0.38	0.32
			RNP32	80.70	90.58	93.61
	CEL021-014	normal	untreated	ND	ND	0.26
			RNP32	40.07	80.95	83.13
	CEL042-001	normal	untreated	0.43	0.27	0.67
			RNP32	78.37	88.50	90.98

FIG 48A

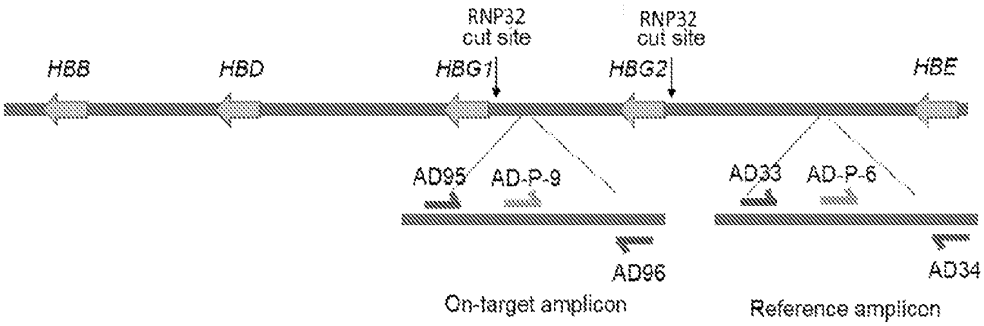


FIG 48B

Assay	Primer/Probe	Description	Sequence
ddPCR Assay6	AD-P-6	ddPCR_GG_DownRef_ Probe1	TCCCTGATGAACATAGATGCC AACA (SEQ ID NO:1271)
	AD33	ddPCR_GG_DR1_F	ACTGGGAAAGAACATAACCAA (SEQ ID NO:1272)
	AD34	ddPCR_GG_DR1_R	GATATGCTGTCGGATGCTATT A (SEQ ID NO:1273)
ddPCR Assay9	AD-P-9	DA2_Probe2	AGAATGTTCTTTGGCAGGTAC TGTGG (SEQ ID NO:1274)
	AD95	Test DA2_2f	CGTAGCTCTTCTATGCTCAG (SEQ ID NO:1275)
	AD96	Test_DA2_2R	CCATGTACACGCACATCT (SEQ ID NO:1276)

FIG 48C

Master Mix 1 ddPCR Assay 6		Master Mix 2 ddPCR Assay 9	
Amount (µL)	Material	Amount (µL)	Material
18	AD33	18	AD95
18	AD34	18	AD96
5	AD-P-6	5	AD-P-9
59	DNAse/RNAse free water	59	DNAse/RNAse free water
100	**TOTAL**	**100**	**TOTAL**

FIG 48D

Master Mix 3 (per sample)	
Amount (µL)	**Material**
11.5	ddPCR SuperMix (no dUTP) 2x
1.15	Master Mix 1
1.15	Master Mix 2
9.2	Diluted DNA
23	**TOTAL Per Sample**

FIG 49A

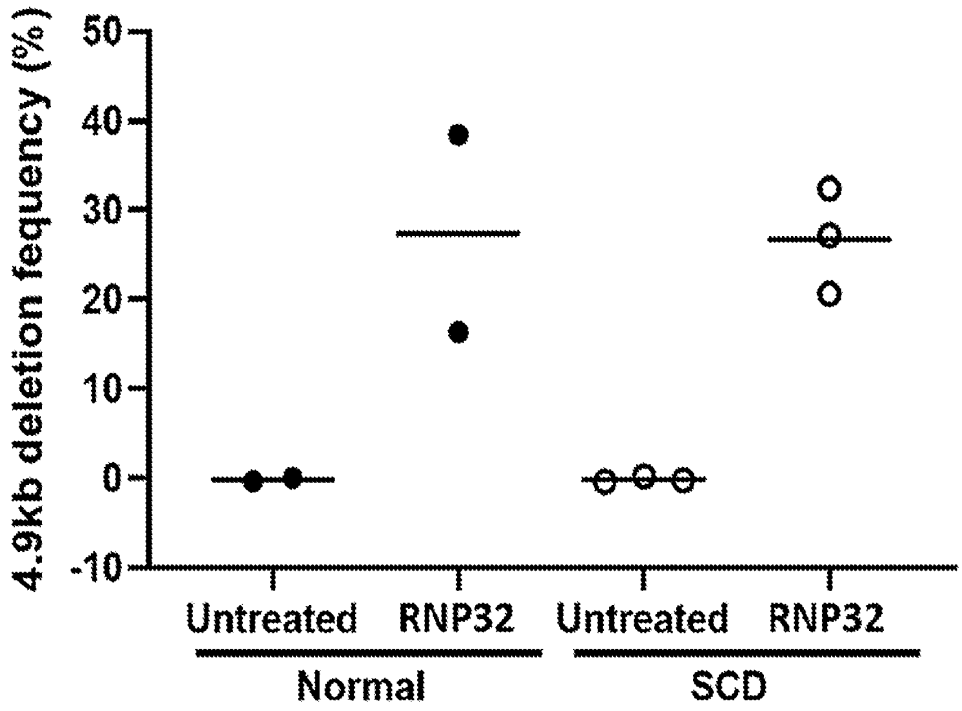


FIG 49B

Experiment ID	Donor Lot	Donor Type	Treatment	Day 1 4.9kb Deletion (%)
SCD014	CEL240-001	SCD	untreated	-0.4
			RNP32	27.2
	CEL241-001	SCD	untreated	-0.2
			RNP32	32.4
	CEL239-001	SCD	untreated	0.2
			RNP32	20.7
	CEL211-001	SCD	untreated	NA
			RNP32	NA
	CEL238-001	normal	untreated	-0.3
			RNP32	38.5
	CEL021-014	normal	untreated	NA
			RNP32	NA
	CEL042-001	normal	untreated	0.0
			RNP32	16.4

FIG 50A

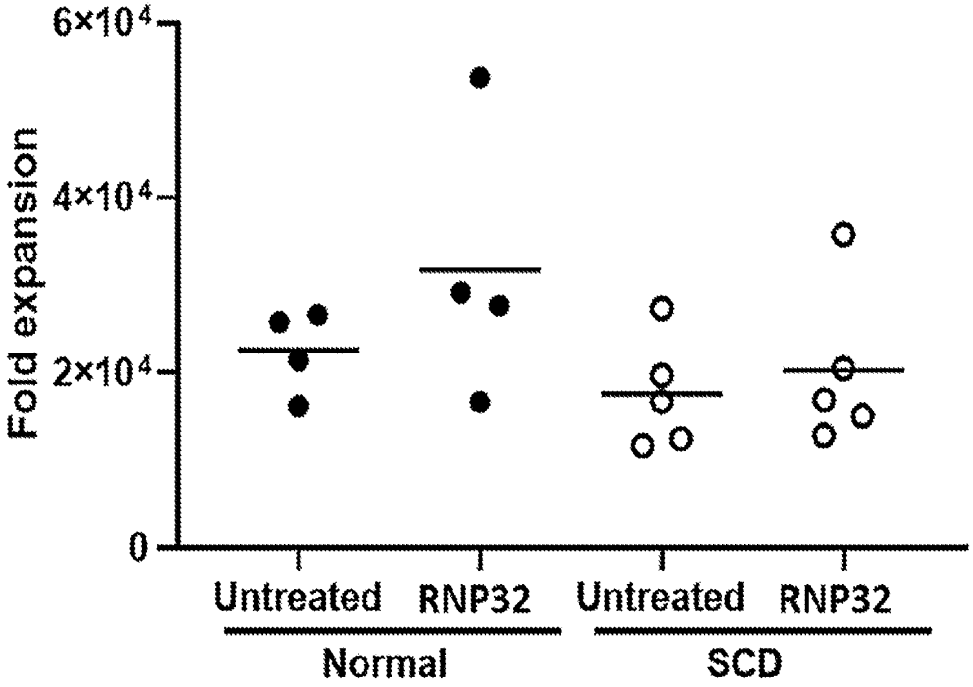


FIG 50B

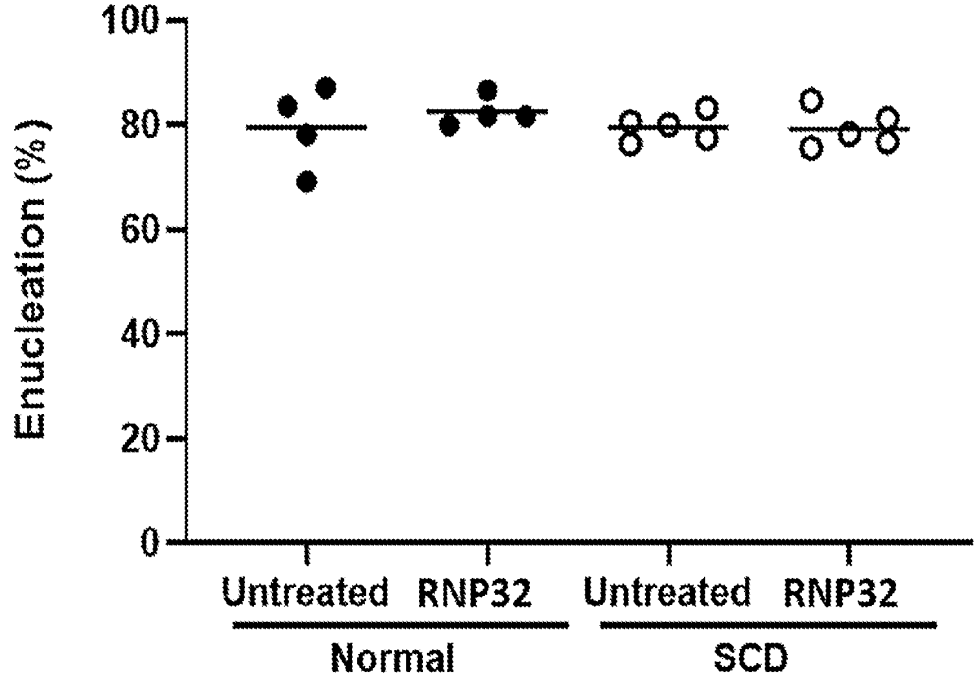


FIG 51A

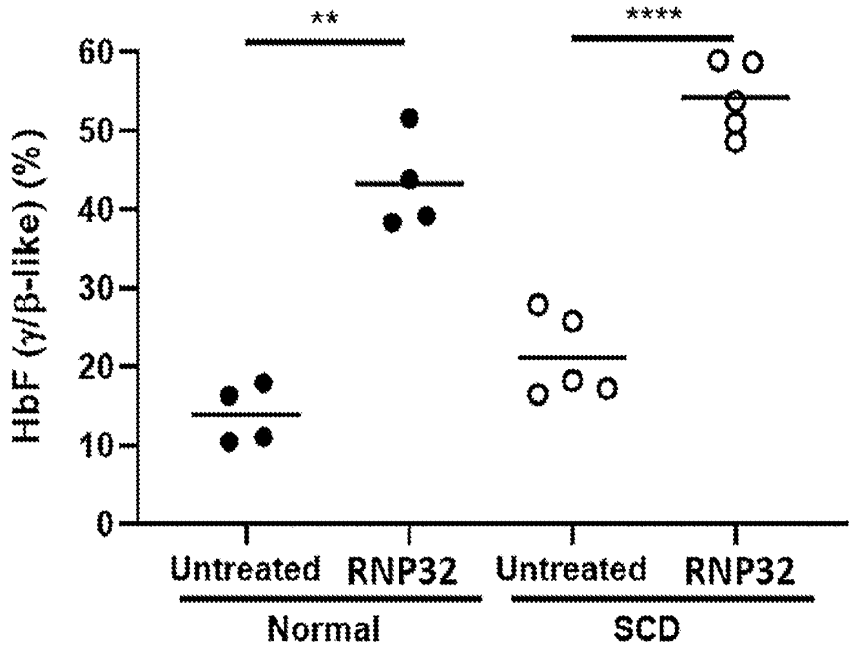


FIG 51B

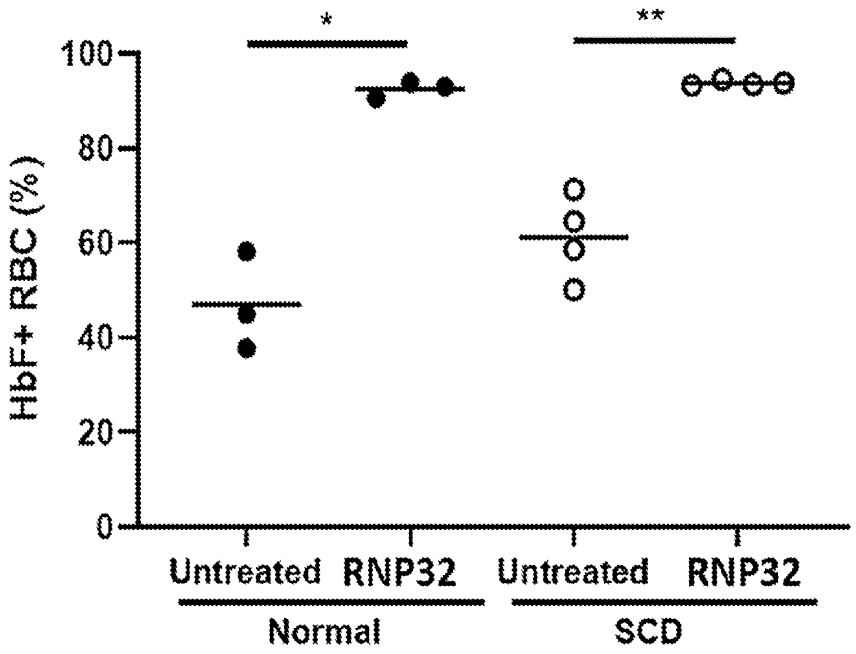


FIG 51C

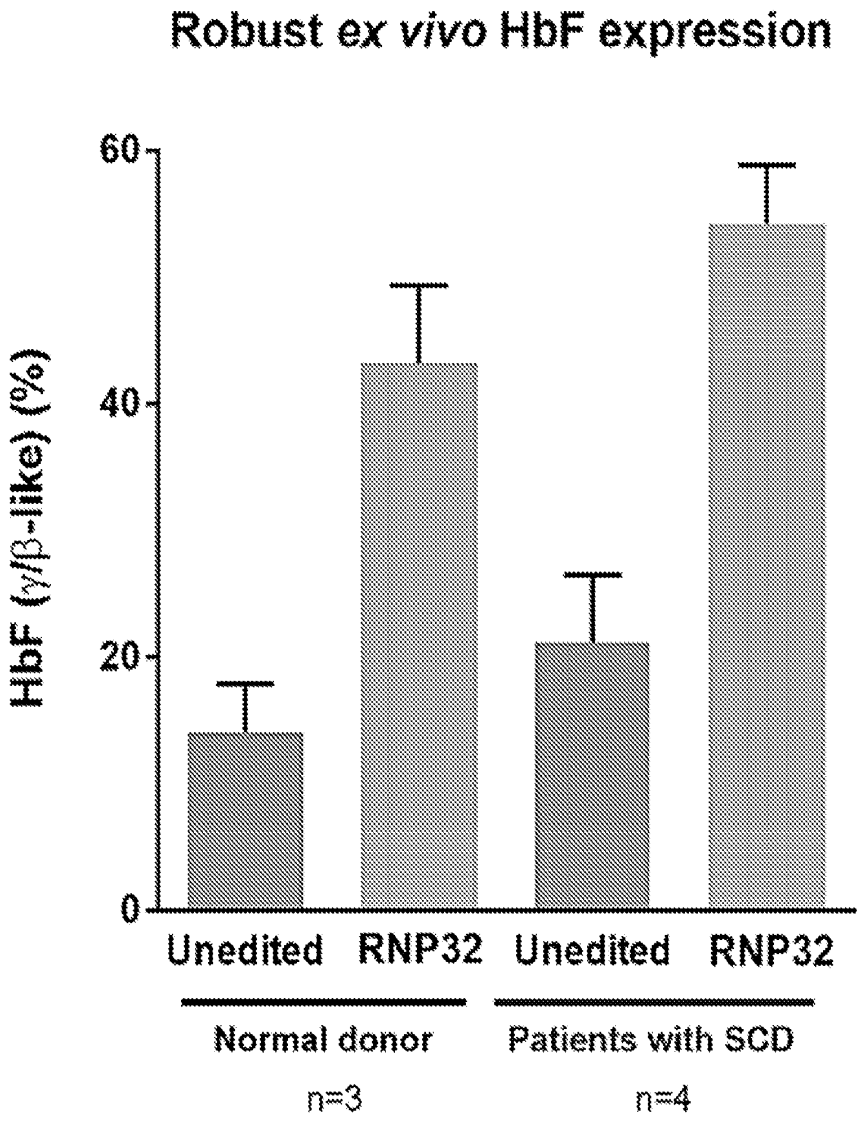
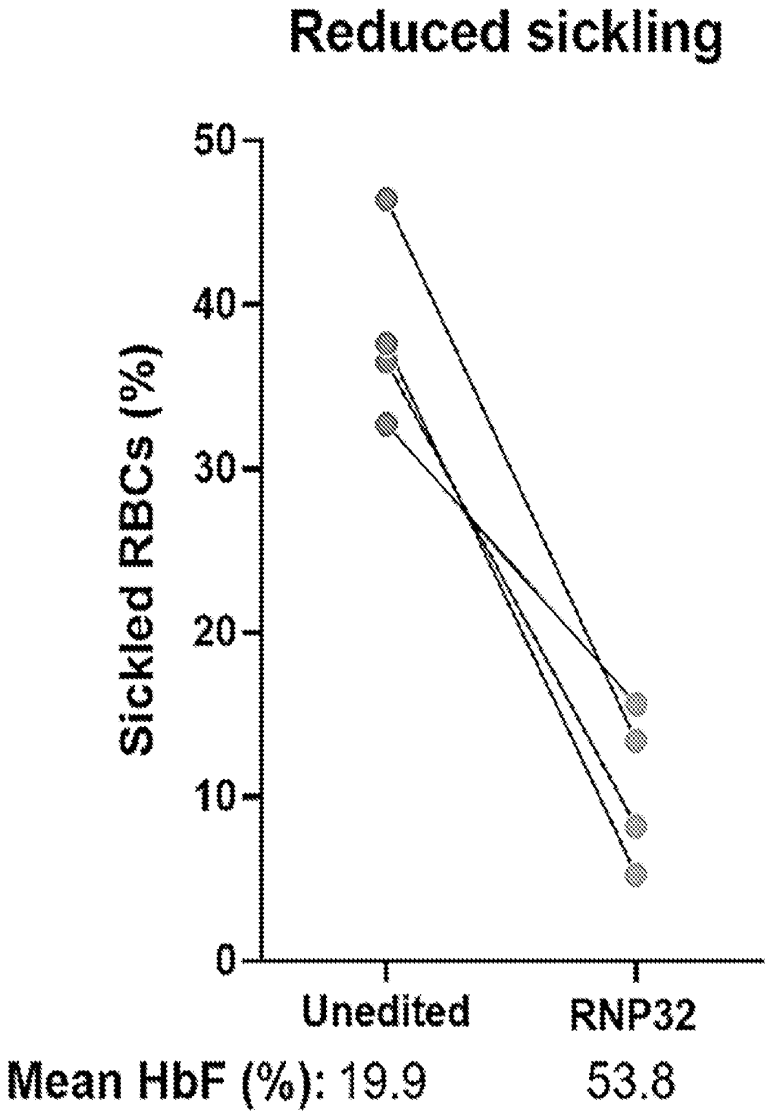

Robust ex vivo HbF expression
HbF (γ/β-like) (%)
60
40
20
0
Unedited
RNP32
Unedited
RNP32
Normal donor
n=3
Patients with SCD
n=4

FIG 52


Reduced sickling
Sickled RBCs (%)
50
40
30
20
10
0
Unedited
RNP32
Mean HbF (%): 19.9
53.8

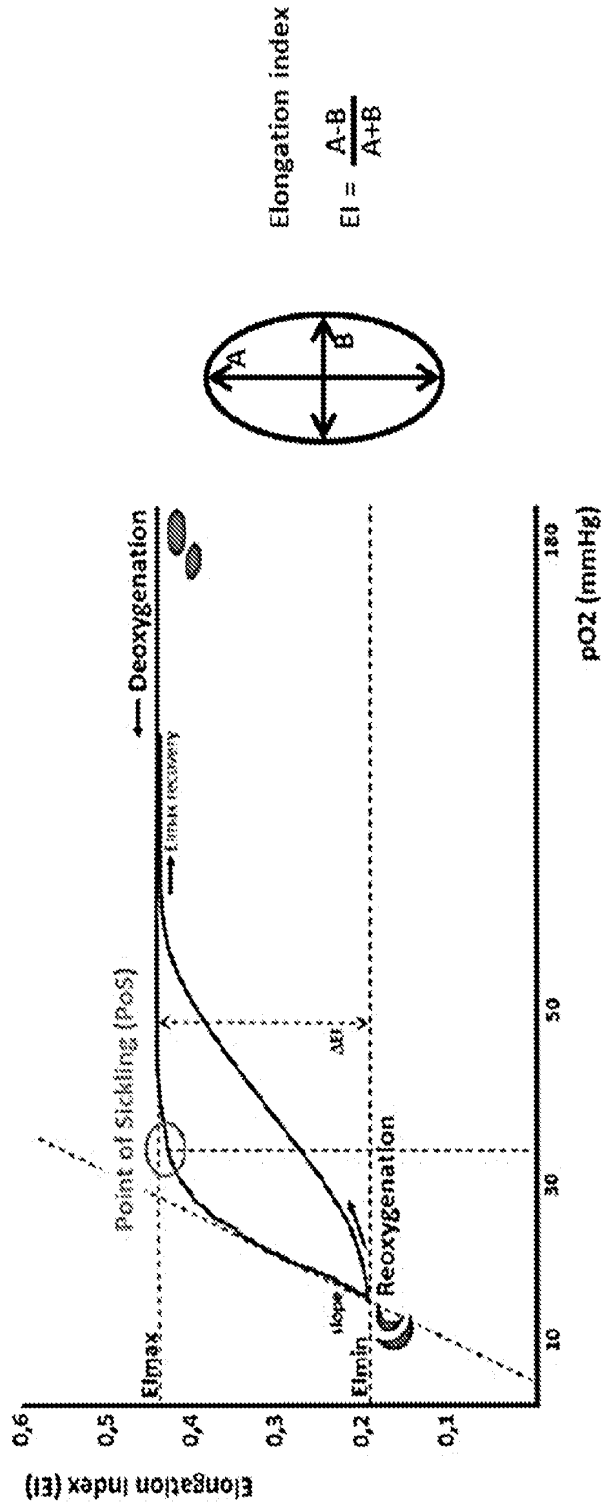

Elongation index
EI = A-B / A+B
A
B
Deoxygenation
Emax recovery
Point of Sickling (PoS)
ΔEI
Reoxygenation
Slope
Emax
Emin
Elongation Index (EI)
0,6
0,5
0,4
0,3
0,2
0,1
10
30
50
180
pO2 (mmHg)

FIG 53A

FIG 53B

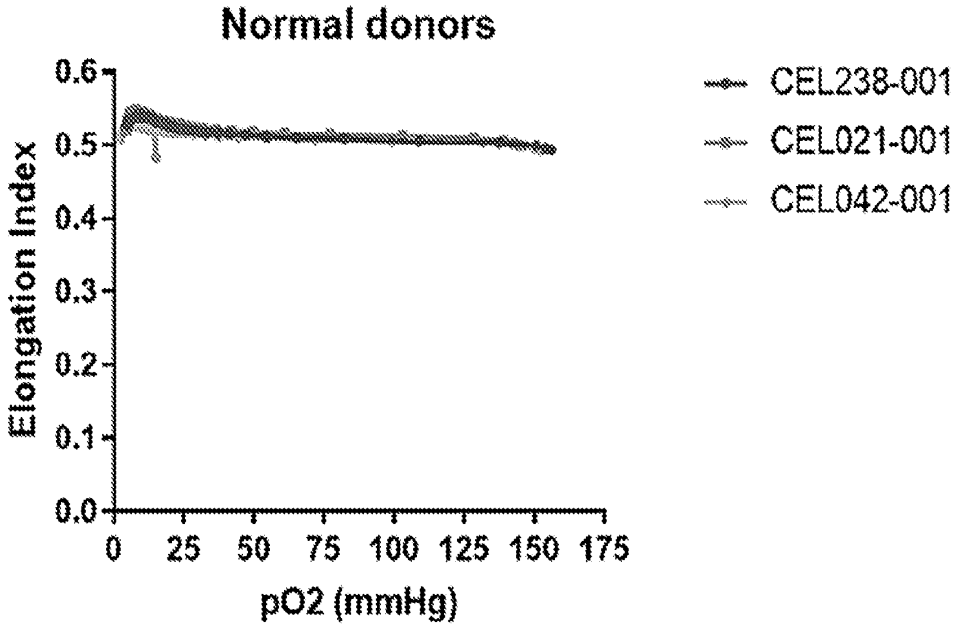

Normal donors
Elongation Index
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
25
50
75
100
125
150
175
pO2 (mmHg)
CEL238-001
CEL021-001
CEL042-001

FIG 53C

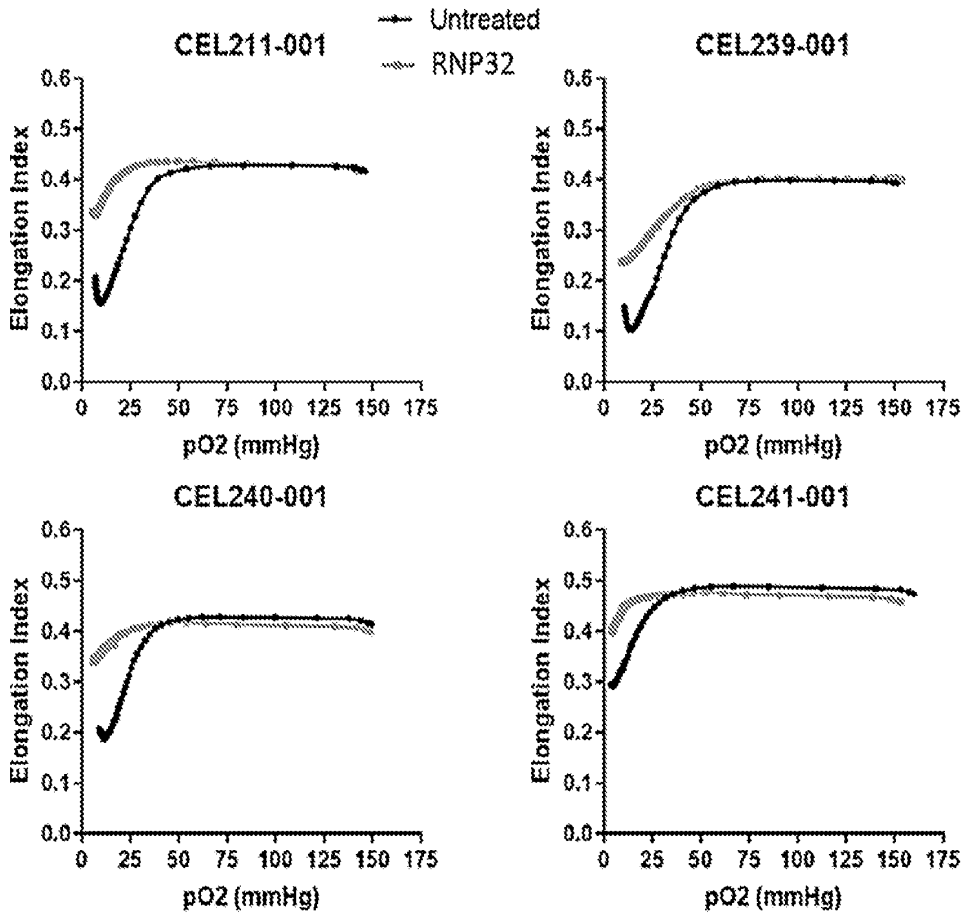

Untreated
RNP32
CEL211-001
CEL239-001
CEL240-001
CEL241-001
Elongation Index
pO2 (mmHg)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0
25
50
75
100
125
150
175

FIG 53D

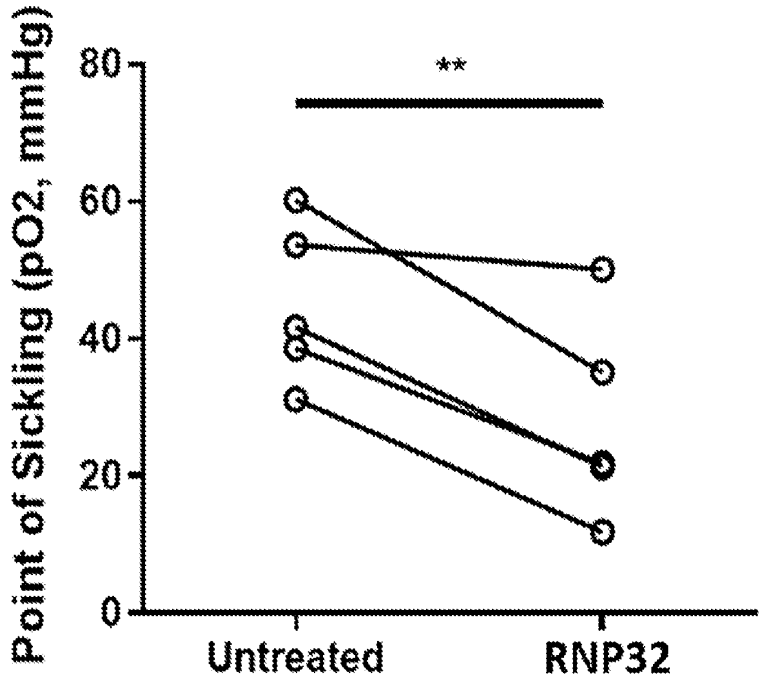


FIG 53E

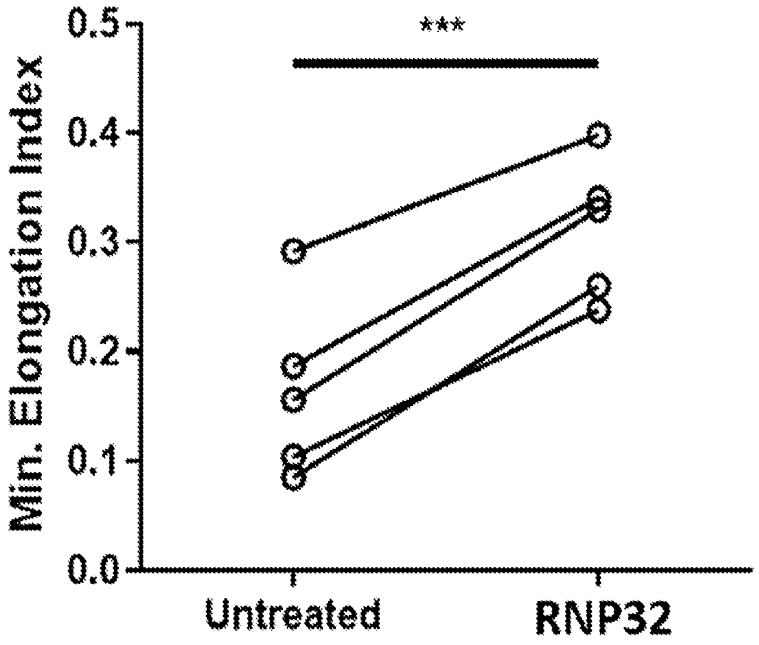


FIG 54A

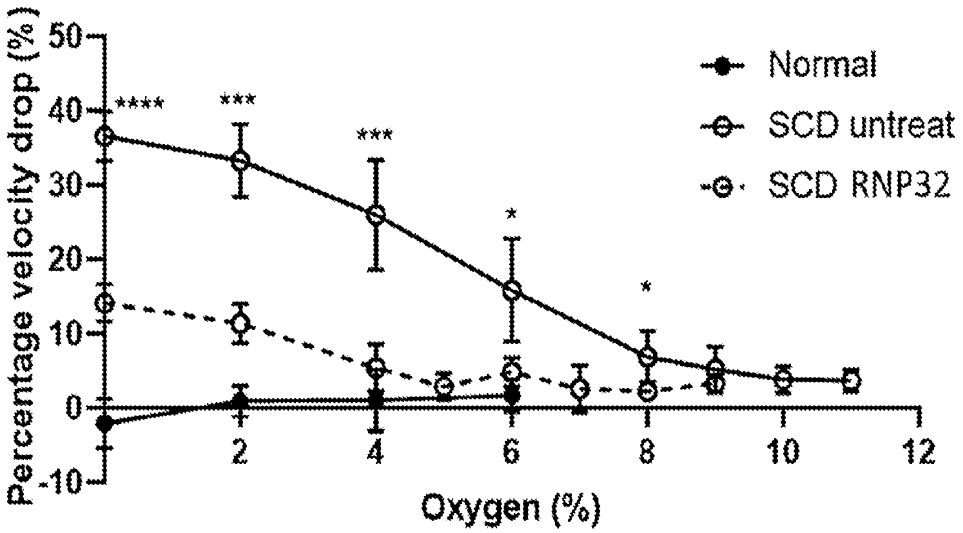

Percentage velocity drop (%)
50
40
30
20
10
0
-10
****
***
***
*
*
2
4
6
8
10
12
Oxygen (%)
Normal
SCD untreat
SCD RNP32

FIG 54B

Exp ID	Donor Lot	Donor Type	Sample ID[a]	Treatment	Oxygen Concentration										
					Median Fractional Velocity Drop (%)										
					0%	2%	4%	5%	6%	7%	8%	9%	10%	11%	
SCD 011	CEL211-011	SCD	SCD donor 1	untreated	38.54	39.05	31.64		23.16		10.76	9.23	5.13	4.71	
			Edited donor 1	RNP32	16.67	12.55	6.65	3.59	7.5	6.65	3.05	3.30			
			SCD donor 2[b]	untreated	41.89	40.14	36.59		26.68		15.17		7.85		
			Edited donor 2[b]	RNP32	21.52	16.90	12.50								
	CEL238-001	normal	Healthy 1	untreated	0.51	-1.23	0.50		0.16						
			NA	RNP32											
SCD 014	CEL240-001	SCD	SCD donor 3	untreated	34.12	29.80	19.12		14.61		4.45	2.18	1.81		
			Edited donor 3	RNP32	10.76	10.53	1.69	0.81							
	CEL241-001	SCD	SCD donor 4	untreated	38.48	32.36	28.66		17.02		8.85	5.52	4.59	2.67	
			Edited donor 4	RNP32	13.72	12.25	4.55	4.2	4.62	-0.52					
	CEL239-001	SCD	SCD donor 5	untreated	39.67	37.42	33.38		19.45		8.10	3.75			
			Edited donor 5	RNP32	16.47	14.28	10.01		4.28	3.16	1.39				
	CEL211-001	SCD	SCD donor 6	untreated	32.07	27.62	17.12		4.79		2.25				
			Edited donor 6	RNP32	12.93	7.26	3.92		2.92	0.99					
	CEL238-001	normal	Healthy 2	untreated	-2.44	3.50	1.41		3.20						
			NA	RNP32											
	CEL021-014	normal	Healthy 3	untreated	-6.63	-0.13	6.16								
			NA	RNP32											
	CEL042-001	normal	Healthy 4	untreated	0.22	1.50	-3.88								
			NA	RNP32											

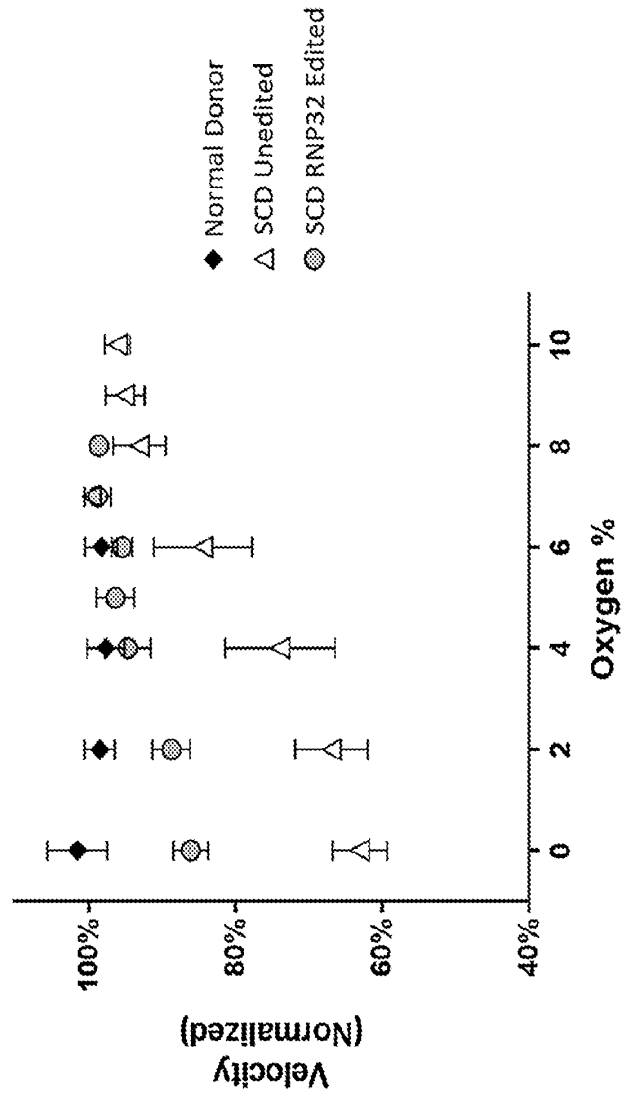

Normal Donor
SCD Unedited
SCD RNP32 Edited
Velocity (Normalized)
100%
80%
60%
40%
0
2
4
6
8
10
Oxygen %

FIG 54C

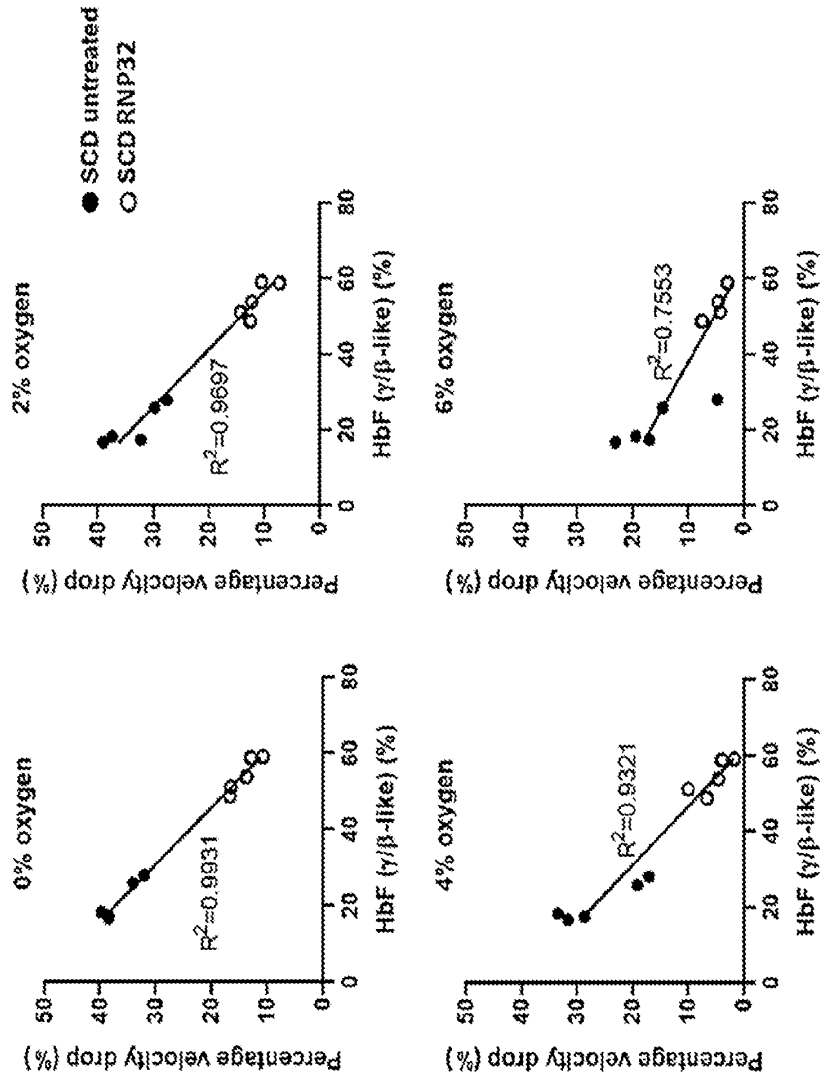

SCD untreated
SCD RNP32
2% oxygen
R^2=0.9697
6% oxygen
R^2=0.7553
0% oxygen
R^2=0.9931
4% oxygen
R^2=0.9321
Percentage velocity drop (%)
HbF (γ/β-like) (%)
0
10
20
30
40
50
60
80

FIG 54D

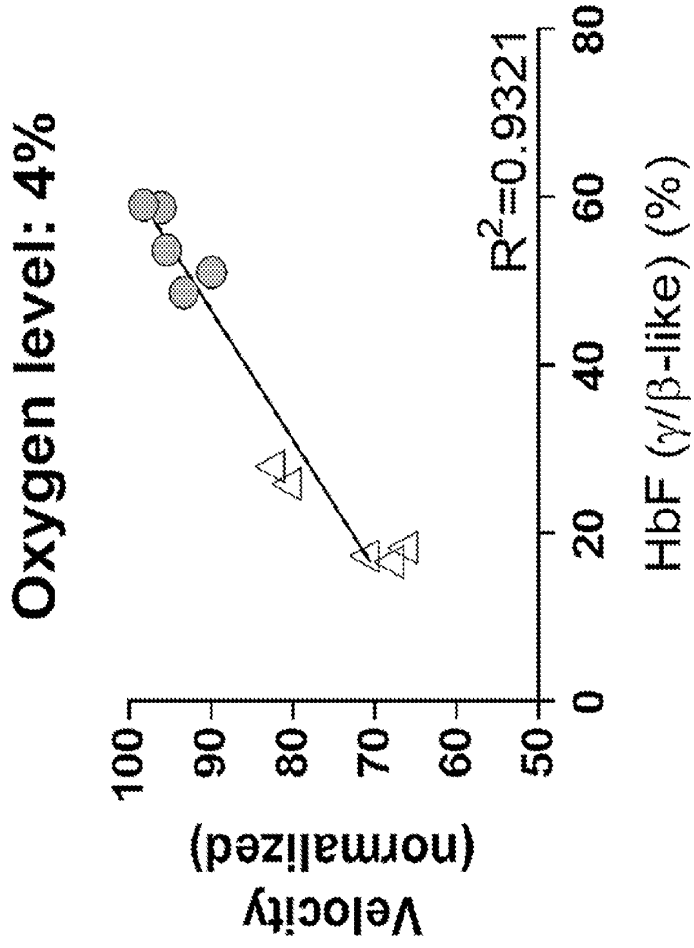

Oxygen level: 4%
Velocity (normalized)
100
90
80
70
60
50
0
20
40
60
80
HbF (γ/β-like) (%)
R2=0.9321

FIG 54E

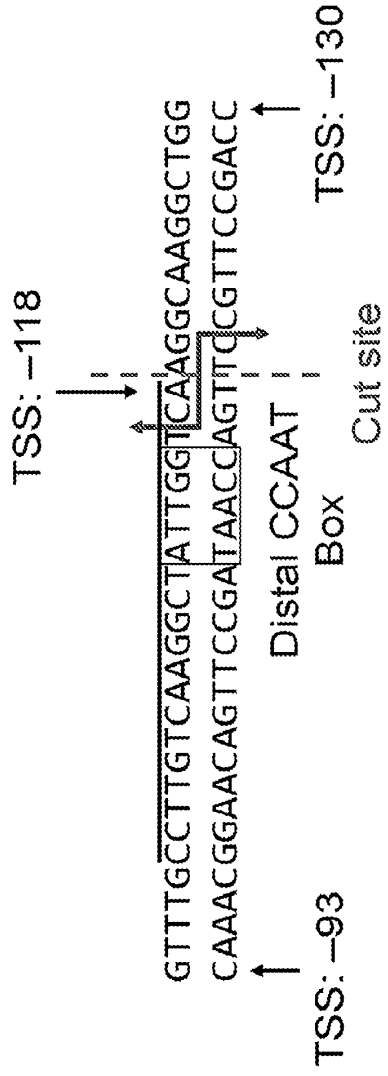

TSS: −118
GTTTGCCTTGTCAAGGCTATTGGTCAAGGCAAGGCTGG
CAAACGGAACAGTTCCGATAACCAGTTCCGTTCCGACC
TSS: −93
Distal CCAAT
Box
Cut site
TSS: −130

FIG 55A

FIG 55B

Oligonucleotide (OLI) used for sequencing			
Type	**Primer ID**	**Sequence**	**Genomic hg38 coordinates**
Forward primer	OLI13753	ACACTCTTTCCCTACACGACGC TCTTCCGATCTCATGGCGTCTG GACTAGGAG (SEQ ID NO:1268)	chr11: 5249801-5249821 (+)
Reverse primer	OLI13752	GTGACTGGAGTTCAGACGTGT GCTCTTCCGATCTAAACACATT TCACAATCCCTGAAC (SEQ ID NO:1269)	chr11: 5250276-5250300 (–)
*underlined sequence is gene specific and corresponds to the coordinates.			

FIG 55C

Amplicon (AMP) used for analysis		
Amplicon ID	**Sequence**	**Genomic hg38 coordinates**
AMP3322	CATGGCGTCTGGACTAGGAGCTTATTGATAA CCTCAGACGTTCCAGAAGCGAGTGTGTGGA ACTGCTGAAGGGTGCTTCCTTTTATTCTTCA TCCCTAGCCAGCCGCCGGCCCCTGGCCTCAC TGGATACTCTAAGACTATTGGTCAAGTTTGC CTTGTCAAGGCTATTGGTCAAGGCAAGGCT GGCCAACCCATGGGTGGAGTTTAGCCAGGG ACCGTTTCAGACAGATATTTGCATTGAGATA GTGTGGGGAAGGGGCCCCCAAGAGGATACT GCTAATTTTTTTTATAGCCTTTGCCTTGTTCC GATTCAGTCATTCCAGTTTTTCTCTAATTTAT TCTTCCCTTTAGCTAGTTTCCTTCTCCCATCA TAGAGGATACCAGGACTTCTTTTGTCAGCCG TTTTTTACCTTCTTGTCTCTAGCTCCAGTGAG GCCTGTAGTTTAAAGCTAAAGCATGTACCA ATTTTTGAAAAGTTCAGGGATTGTGAAATGT GTTT (SEQ ID NO:1270)	chr11: 5249801-5250300 (+) → HBG1 chr11: 5254725-5255228 (+)→ HBG2

FIG 55D

Example of indel_id with explanation	
Indel_id‡	**Explanation**
159_-18_NA	A deletion of 18 bases starting after position 159 inside the amplicon
170_3_AGT	An insertion of AGT starting after position 170 inside the amplicon

FIG 56

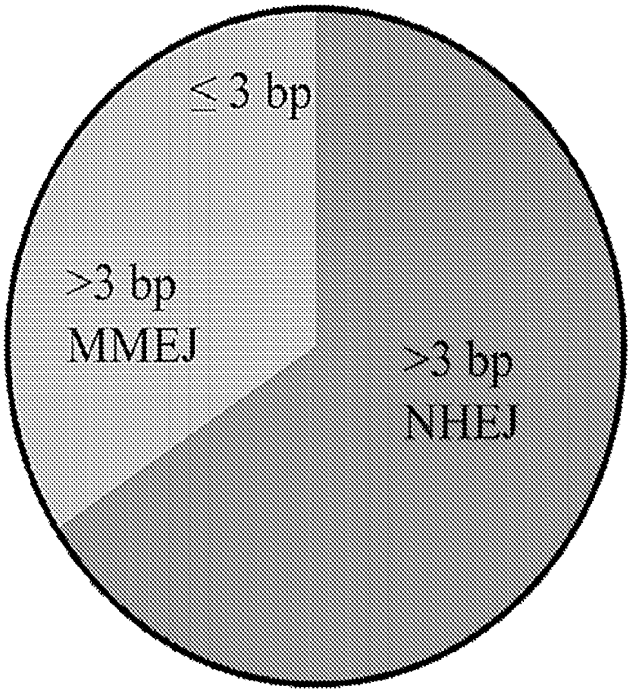
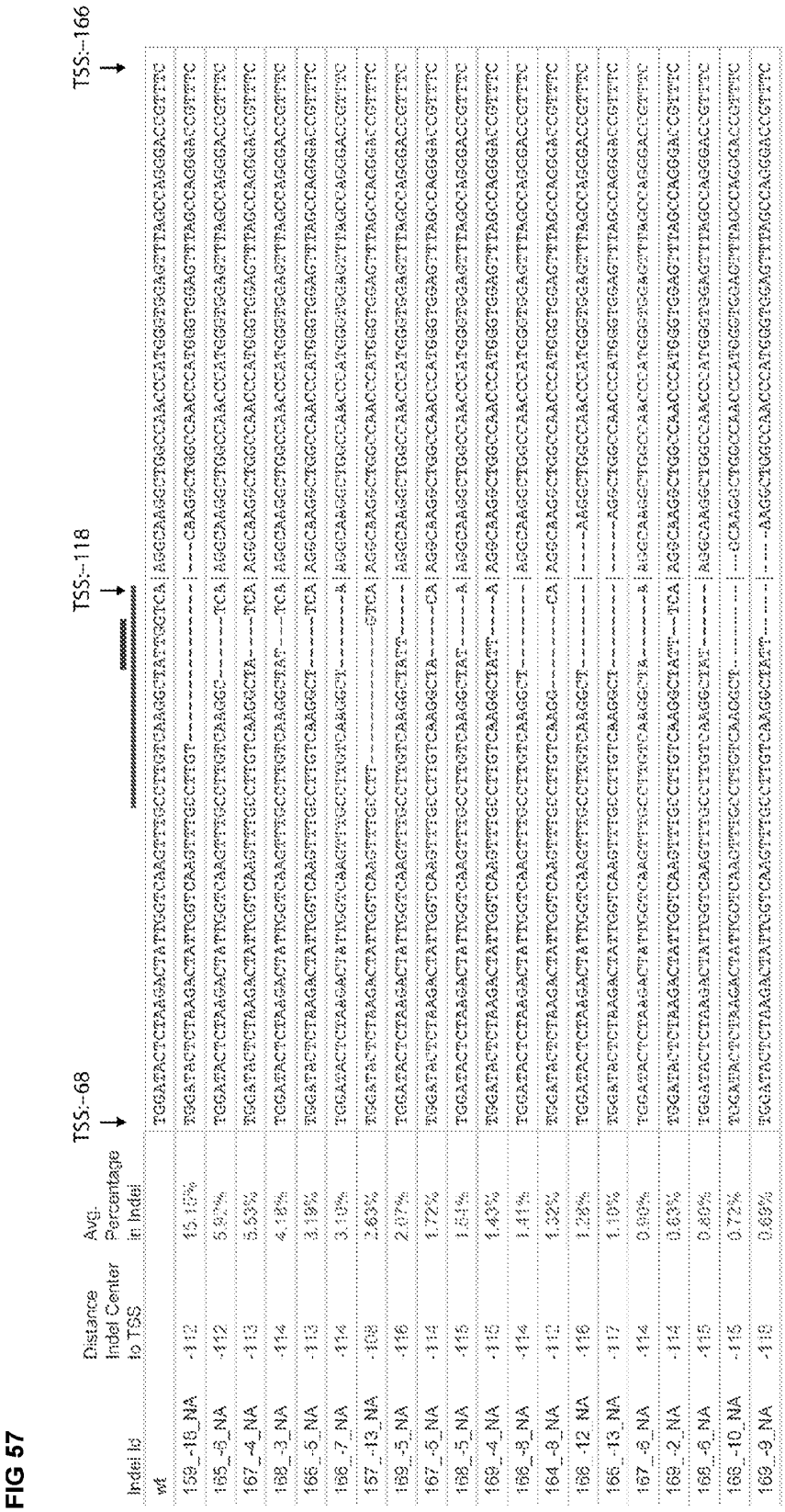


FIG 57


TSS:-166
TSS:-118
TSS:-68
Indel Id
Distance Indel Center to TSS
Avg. Percentage in Indel
wt
159_-18_NA
165_-8_NA
167_-4_NA
168_-3_NA
166_-5_NA
166_-7_NA
157_-13_NA
169_-5_NA
167_-6_NA
168_-5_NA
169_-4_NA
166_-8_NA
164_-8_NA
166_-12_NA
166_-13_NA
167_-6_NA
169_-2_NA
168_-6_NA
166_-10_NA
169_-9_NA

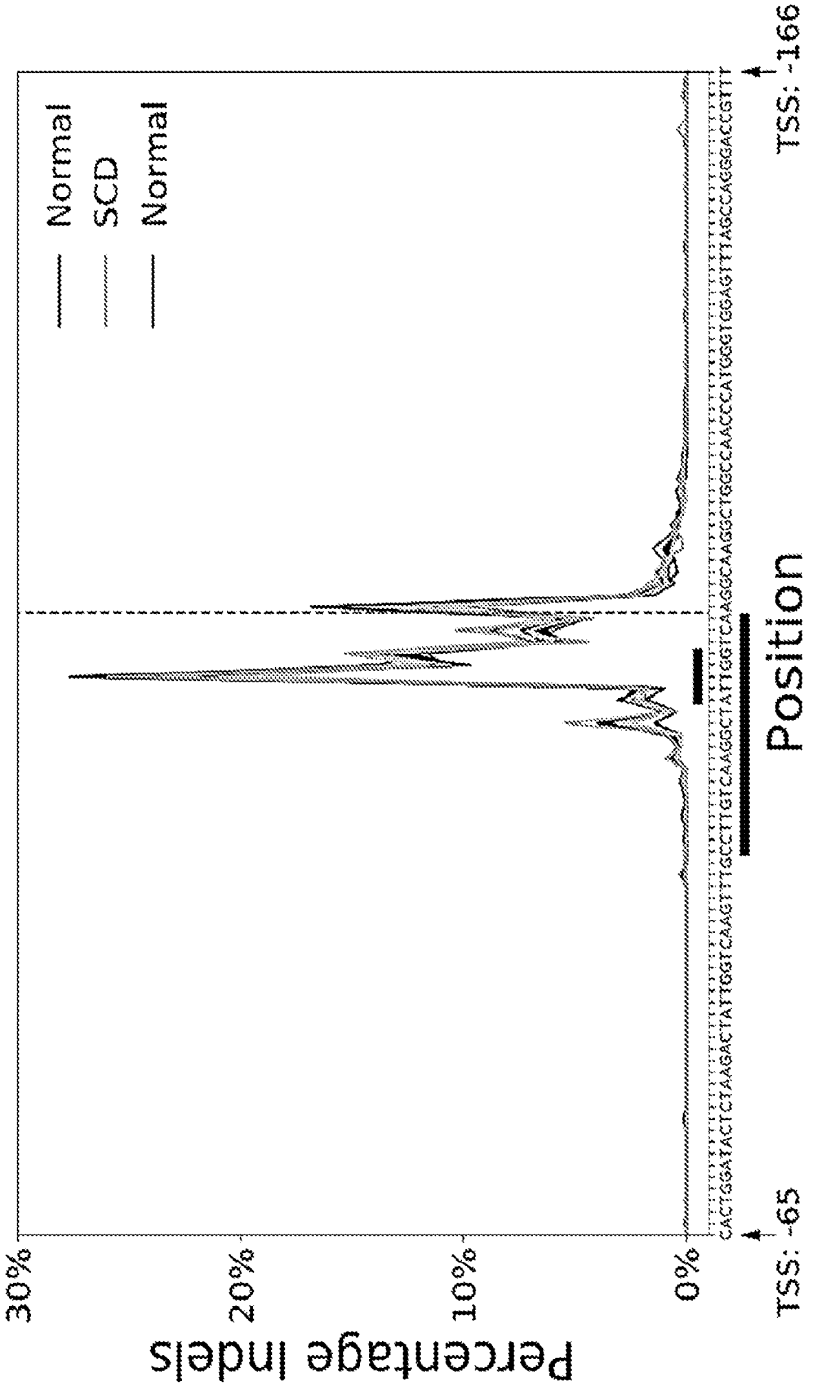

Normal
SCD
Normal
30%
20%
10%
0%
Percentage Indels
Position
TSS: -65
TSS: -166
CACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATTGGTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGGGACCGTTT

FIG 58

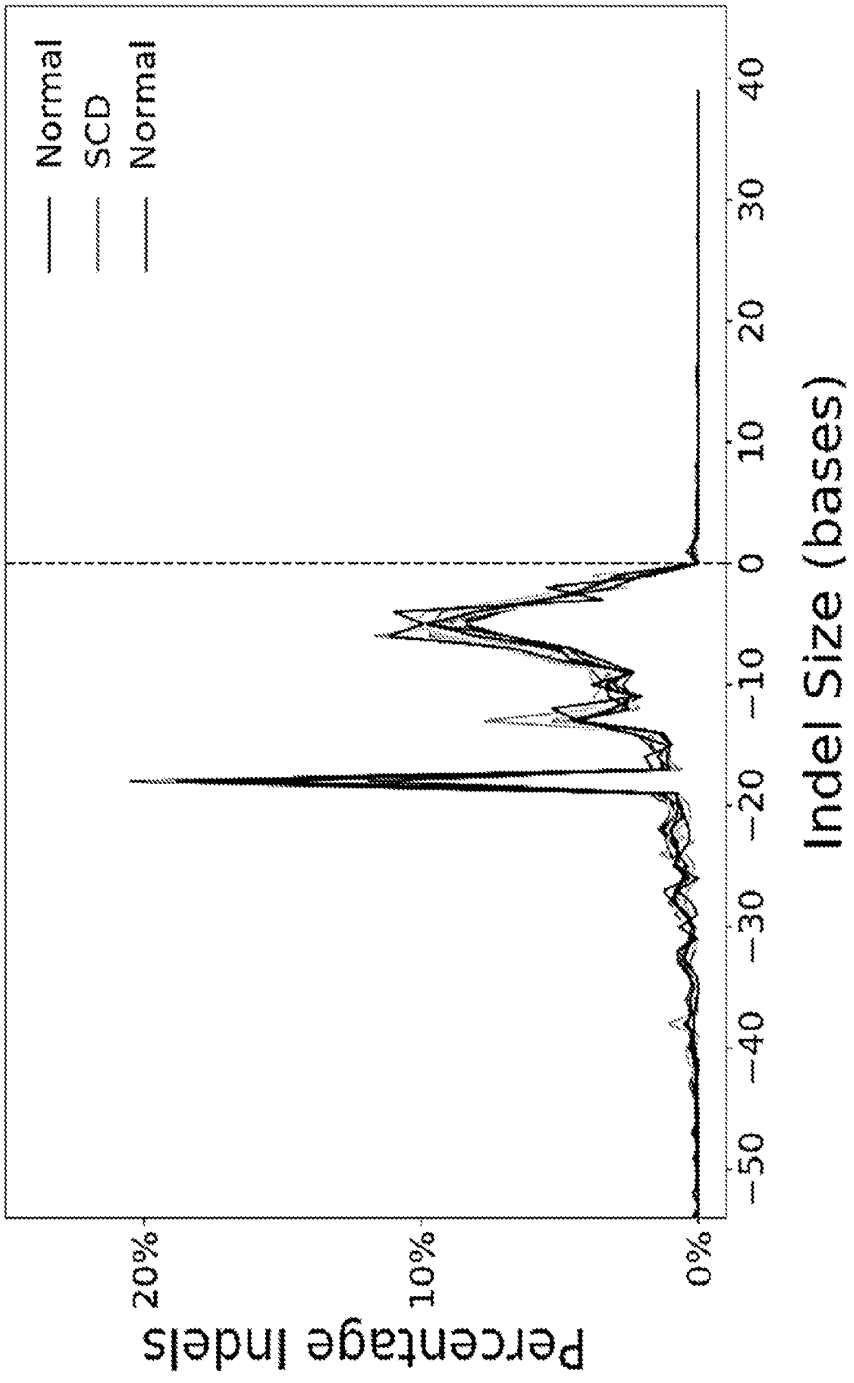

Normal
SCD
Normal
Percentage Indels
20%
10%
0%
−50
−40
−30
−20
−10
0
10
20
30
40
Indel Size (bases)

FIG 59

FIG 60

Indel Length	Percentage
50	0.00%
49	0.00%
48	0.00%
47	0.00%
46	0.00%
45	0.00%
44	0.00%
43	0.00%
42	0.00%
41	0.00%
40	0.00%
39	0.00%
38	0.00%
37	0.00%
36	0.00%
35	0.00%
34	0.00%
33	0.00%
32	0.00%
31	0.00%
30	0.00%
29	0.00%
28	0.00%
27	0.00%
26	0.00%
25	0.00%
24	0.00%
23	0.00%
22	0.00%
21	0.00%
20	0.00%
19	0.00%
18	0.00%
17	0.00%
16	0.01%
15	0.01%

FIG 60 (cont'd)

14	0.01%
13	0.00%
12	0.01%
11	0.01%
10	0.00%
9	0.01%
8	0.01%
7	0.01%
6	0.01%
5	0.03%
4	0.02%
3	0.06%
2	0.10%
1	0.20%
-1	2.48%
-2	3.70%
-3	5.50%
-4	8.31%
-5	9.30%
-6	8.39%
-7	5.50%
-8	4.46%
-9	2.88%
-10	3.29%
-11	2.64%
-12	3.09%
-13	4.77%
-14	1.85%
-15	1.24%
-16	1.32%
-17	1.07%
-18	16.46%
-19	1.13%
-20	0.75%
-21	0.69%
-22	1.07%
-23	0.79%
-24	0.78%

FIG 60 (cont'd)

-25	0.57%
-26	0.36%
-27	0.68%
-28	0.72%
-29	0.47%
-30	0.27%
-31	0.21%
-32	0.39%
-33	0.56%
-34	0.22%
-35	0.13%
-36	0.30%
-37	0.19%
-38	0.38%
-39	0.13%
-40	0.19%
-41	0.09%
-42	0.10%
-43	0.12%
-44	0.07%
-45	0.06%
-46	0.03%
-47	0.06%
-48	0.03%
-49	0.04%
-50	0.02%

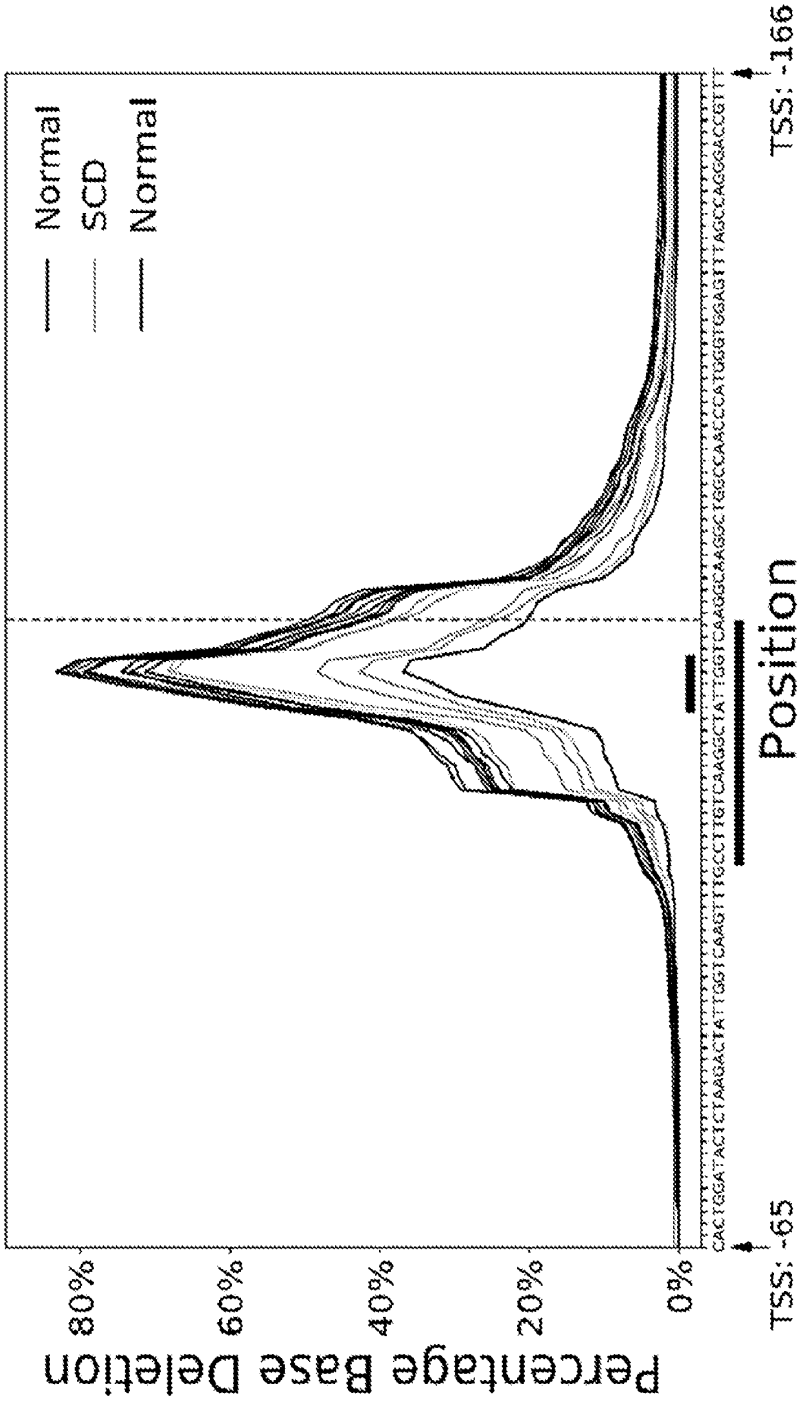

Percentage Base Deletion
80%
60%
40%
20%
0%
Normal
SCD
Normal
Position
TSS: -65
TSS: -166

FIG 61

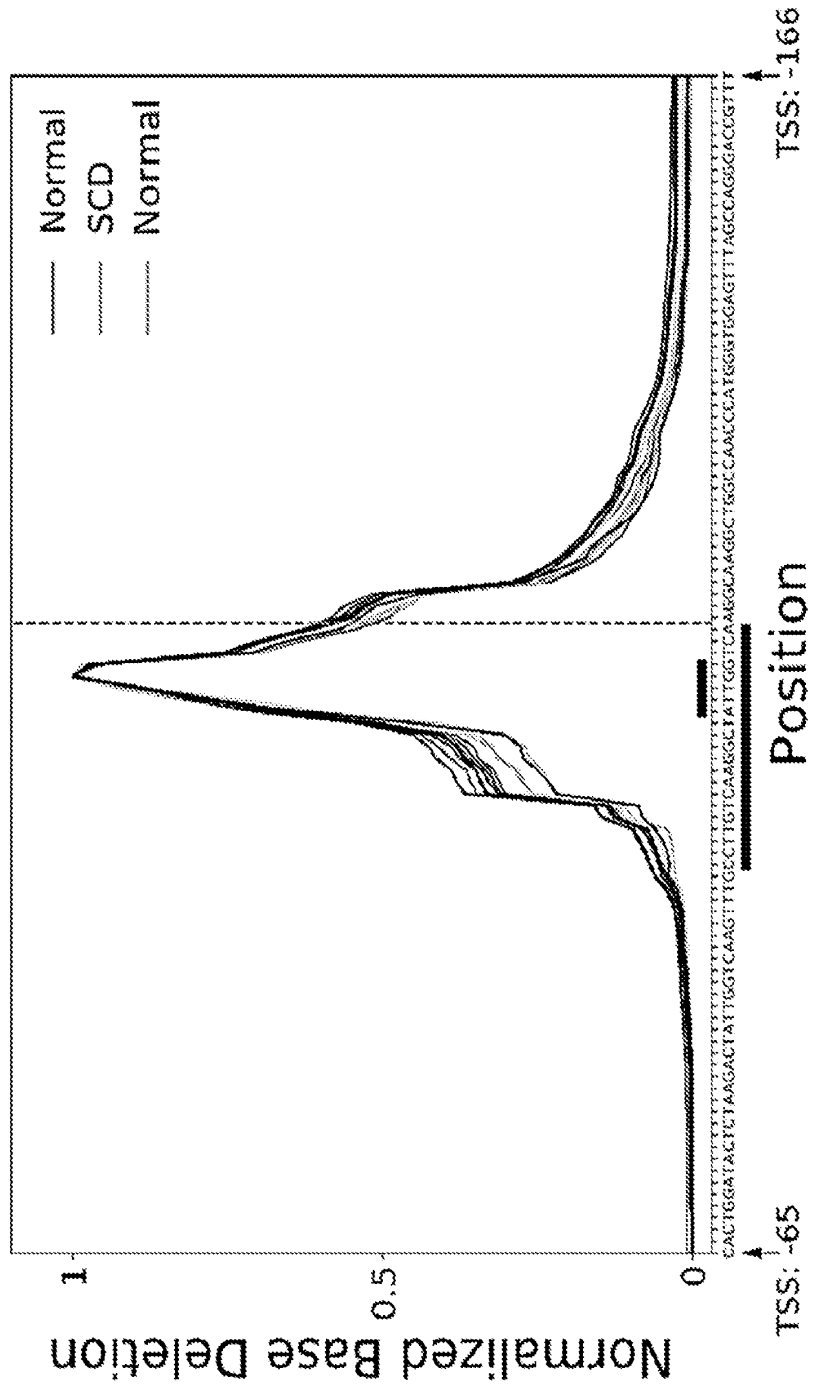

Normalized Base Deletion
1
0.5
0
Normal
SCD
Normal
TSS: -65
TSS: -166
Position

FIG 62

FIG 63

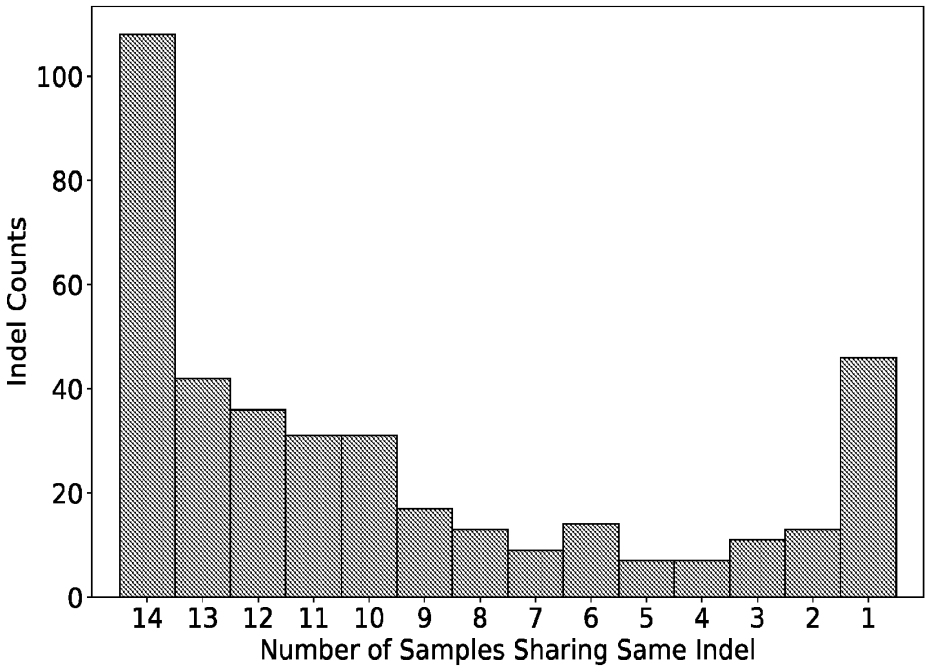

Indel Counts
0
20
40
60
80
100
14 13 12 11 10 9 8 7 6 5 4 3 2 1
Number of Samples Sharing Same Indel

FIG 64

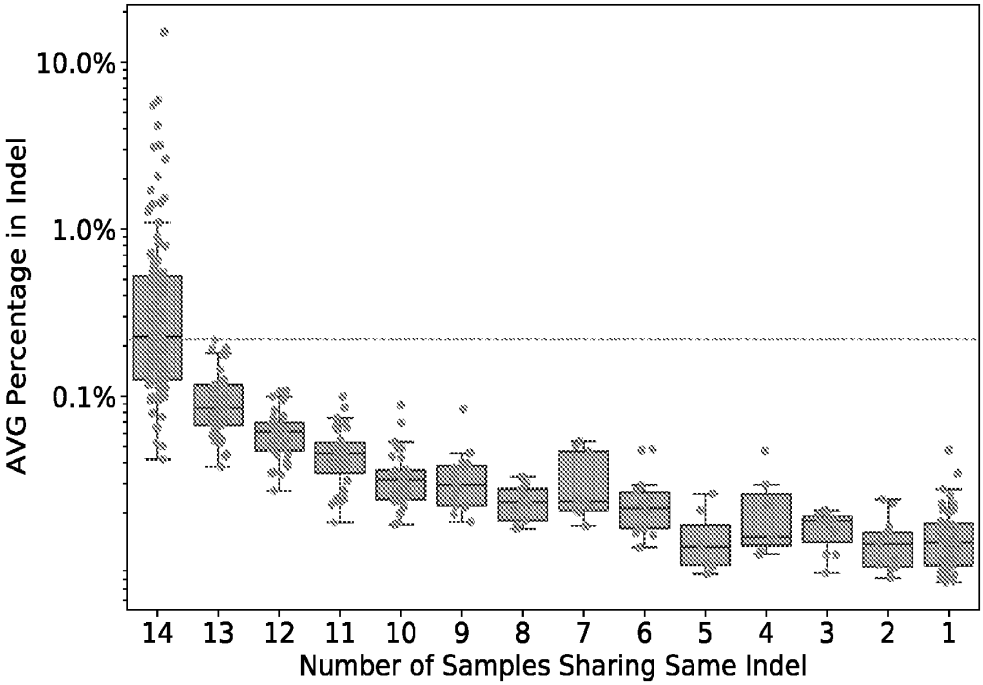

10.0%
1.0%
0.1%
AVG Percentage in Indel
14 13 12 11 10 9 8 7 6 5 4 3 2 1
Number of Samples Sharing Same Indel

FIG 65A

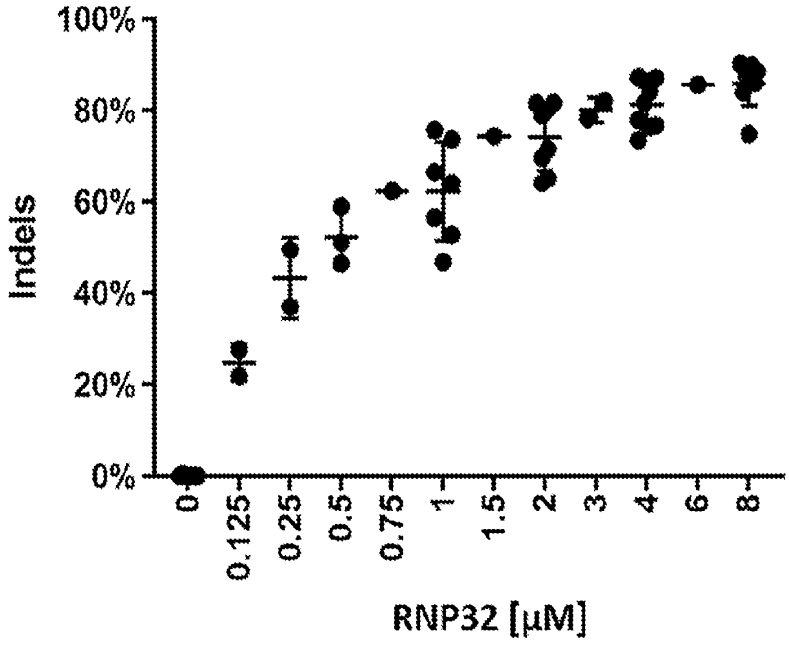


FIG 65B

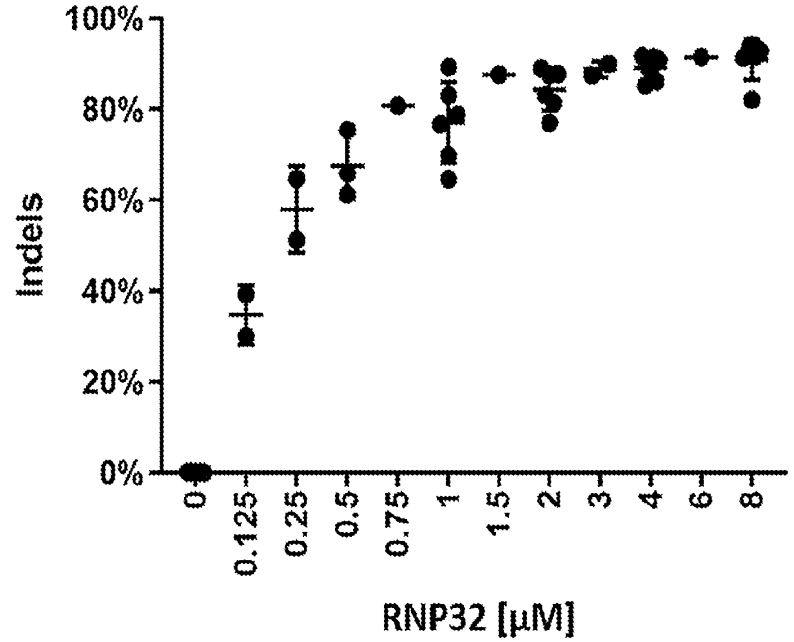


FIG 66

Experiment ID	MAX064	MAX065		MAX066		MAX072	MAX074	MAX076	MAX078	MAX080
Donor Lot	CEL045-002	CEL045-002		CEL045-002		CEL046-004	CEL046-004	CEL046-004	CEL021-021	CEL047-002
RNP Conc. (μM)	Viability (%)									
0	92.3	85.2		92.3	-	92.0	84.1	94.4	89.4	86.9
0.125	-	-	-	-	-	91.8	-	-	89.3	-
0.25	-	-	-	-	-	92.0	-	-	89.4	-
0.5	-	-	-	-	-	89.6	-	92.7	90.4	-
0.75	-	-	-	-	-	-	-	-	88.1	-
1	-	81.6	82.0*	91.1	89.5	90.7	-	92.5	90.4	85.9
1.5	-	-	-	-	-	-	-	-	90.6	-
2	90.3	80.3	84.5*	88.4	90.0	90.0	-	92.8	90.8	82.9
3	-	-	-	-	-	89.5	-	-	92.5	-
4	90.5	84.7	88*	91.9	89.4	89.9	88.4	93.5	94.0	82.4
6	-	-	-	-	-	-	-	-	93.7	-
8	92.0	83.8	89.5*	92.1	91.1	90.0	-	93.1	93.4	88.2

FIG 67

Experiment ID	MAX064	MAX065		MAX066		MAX072	MAX074	MAX076	MAX078	MAX080
Donor Lot	CEL045-002	CEL045-002		CEL045-002		CEL046-004	CEL046-004	CEL046-004	CEL021-021	CEL047-002
RNP Conc. (μM)	Indel levels (%) in total CD34+ cells									
0	0.08	0.08		0.11		0.13	0.11	0.09	0.21	0.4
0.125	-	-	-	-	-	21.95	-	-	27.59	-
0.25	-	-	-	-	-	37.05	-	-	49.56	-
0.5	-	-	-	-	-	51.14	-	46.52	58.97	-
0.75	-	-	-	-	-	-	-	-	62.37	-
1	-	52.76	5.64*	73.72	66.47	64.02	-	46.75	75.71	56.5
1.5	-	-	-	-	-	-	-	-	74.4	-
2	80.26	64.13	12.84*	81.62	81.57	69.61	-	65.11	78.91	71.5
3	-	-	-	-	-	78.20	-	-	82.00	-
4	87.02	73.43	21.82*	87.20	86.56	76.46	76.70	77.93	84.17	81.6
6	-	-	-	-	-	-	-	-	85.50	-
8	89.82	74.83	35.96*	88.45	90.19	86.15	-	83.96	87.43	86

FIG 68

Experiment ID	MAX064	MAX065		MAX066		MAX072	MAX074	MAX076	MAX078	MAX080
Donor Lot	CEL045-002	CEL045-002		CEL045-002		CEL046-004	CEL046-004	CEL046-004	CEL021-021	CEL047-002
RNP Conc. (μM)	Indel levels (%) in total CD34+ cells									
0	-	0.08		0.12		0.09	0.09	0.03	0.27	-
0.125	-	-	-	-	-	30.11	-	-	39.4	-
0.25	-	-	-	-	-	51.24	-	-	64.67	-
0.5	-	-	-	-	-	65.90	-	61.26	75.49	-
0.75	-	-	-	-	-	-	-	-	80.79	-
1	-	69.86	9.79*	83.08	76.69	78.81	-	64.66	89.26	-
1.5	-	-	-	-	-	-	-	-	87.59	-
2	-	77.03	20.61*	87.75	87.64	83.11	-	81.40	88.99	-
3	-	-	-	-	-	87.48	-	-	90.04	-
4	-	85.22	36.30*	90.84	91.38	88.37	89.41	86.21	91.69	-
6	-	-	-	-	-	-	-	-	91.48	-
8	-	82.16	52.19*	91.37	92.91	93.62	-	91.51	94.19	-

FIG 69A

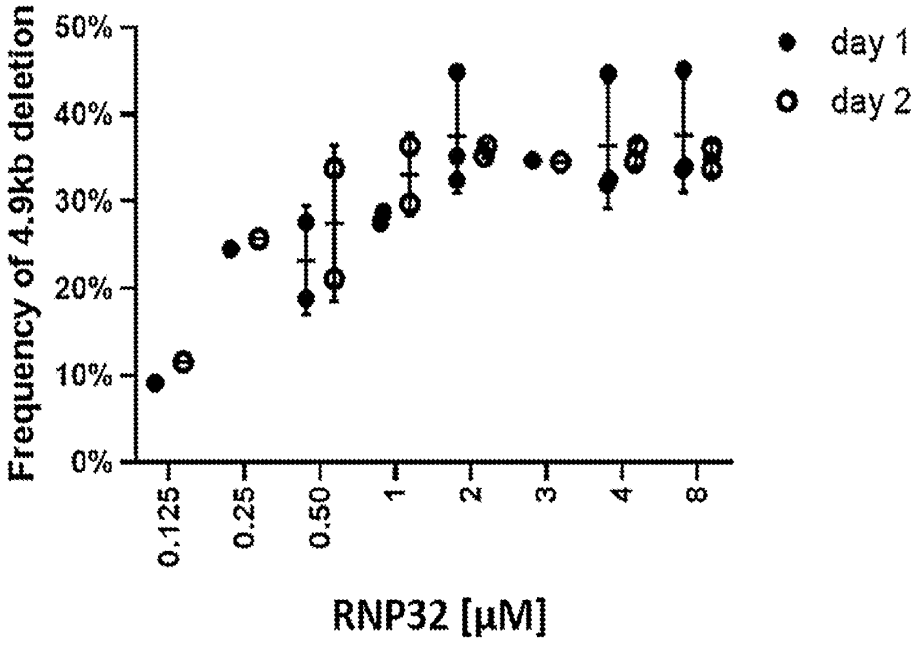


FIG 69B

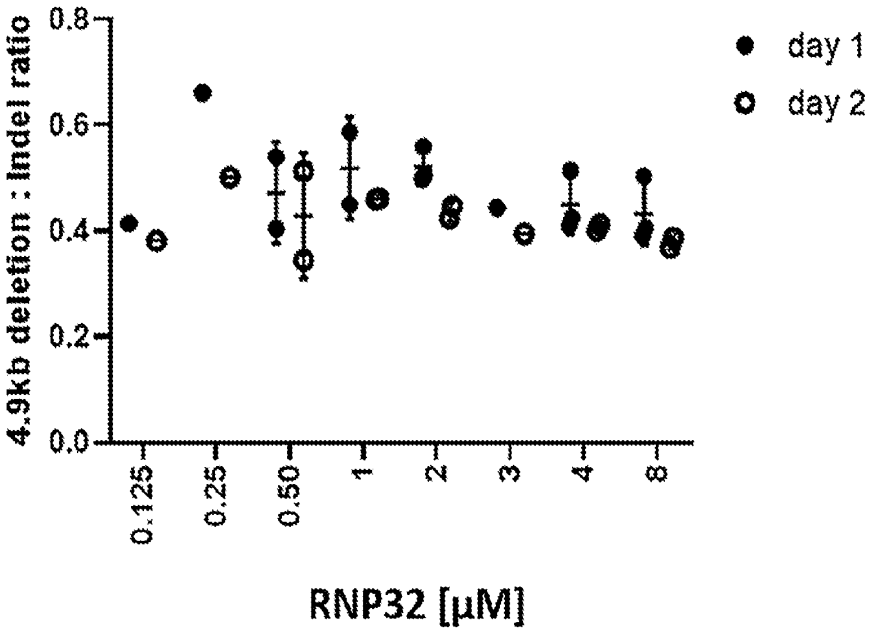


FIG 70

Experiment ID	MAX064		MAX072		MAX076	
Day post electroporation	Day 1	Day 2	Day 1	Day 2	Day 1	Day 2
RNP Concentration (μM)	4.9kb fragment deletion frequency (%) in total CD34+ cells					
0.125	-	-	9.1	11.5	-	-
0.25	-	-	24.5	25.7	-	-
0.5	-	-	27.6	33.8	18.8	21.1
1	-	-	28.8	36.4	27.45	29.7
2	44.9	-	35.2	35.2	32.45	36.4
3	-	-	34.7	34.5	-	-
4	44.7	-	32.5	36.4	31.95	34.5
8	45.2	-	33.6	36.2	34.0	33.7

FIG 71A

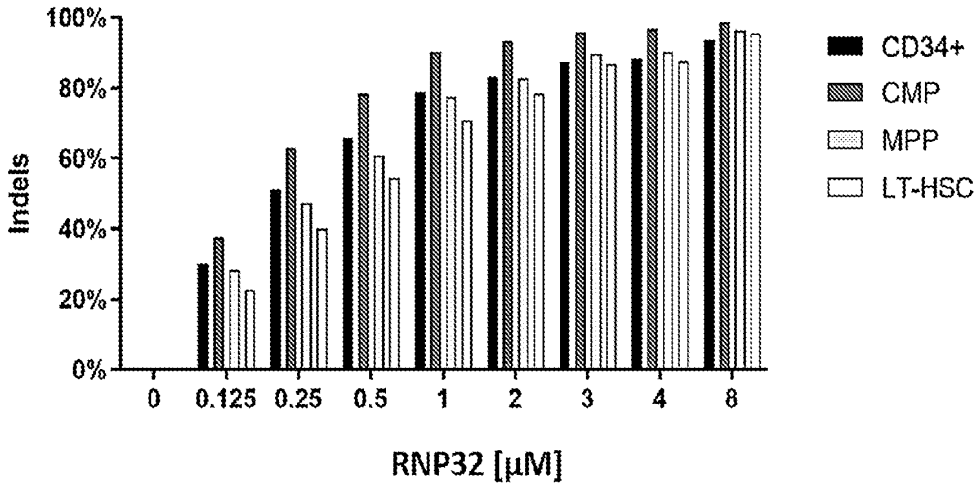


FIG 71B

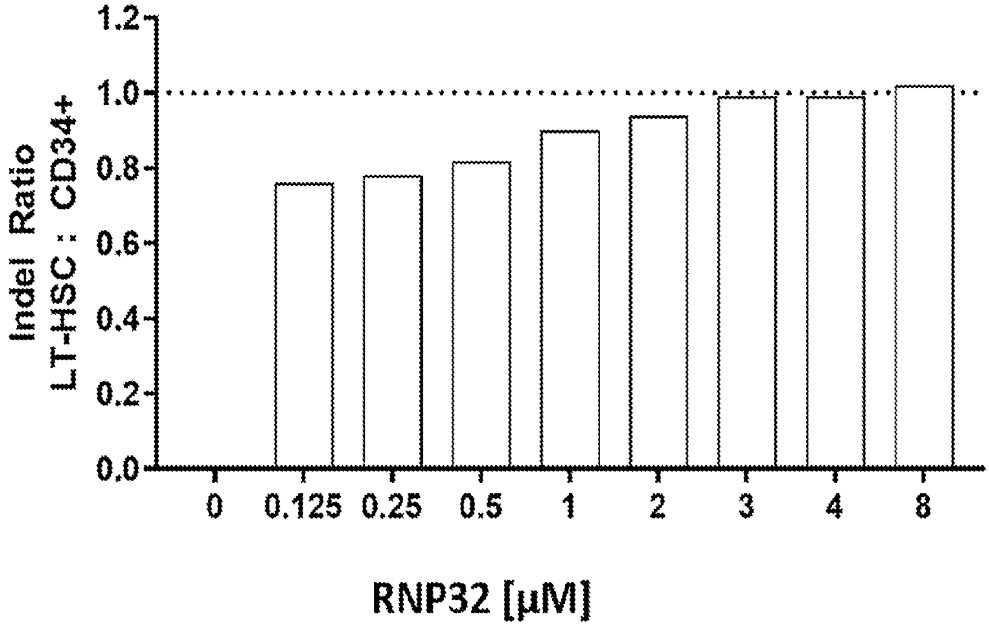


FIG 72

RNP Concentration (μM)	Indel levels (%)			
	Total CD34+ cells	CMP	MPP	LT-HSC
0	0.09			
0.125	30.11	37.58	28.36	22.77
0.25	51.24	63.04	47.37	40.00
0.5	65.90	78.35	60.96	54.32
1	78.81	90.40	77.57	70.77
2	83.11	93.57	82.83	78.46
3	87.48	95.85	89.63	86.71
4	88.37	96.80	90.27	87.74
8	93.62	98.81	96.33	95.64

FIG 73A

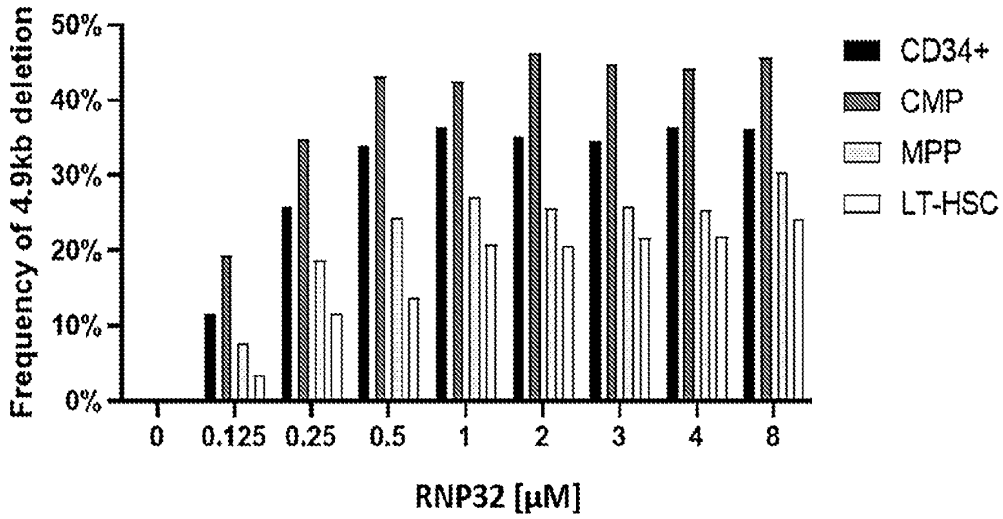


FIG 73B

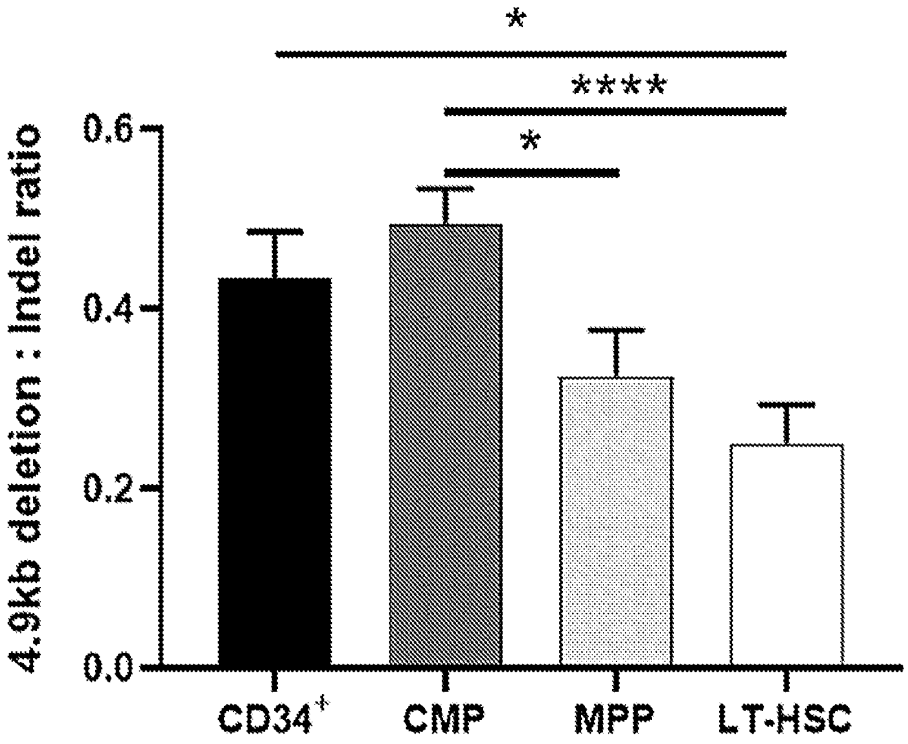


FIG 74

RNP Concentration (μM)	4.9kb fragment deletion frequency (%)			
	Total CD34+ cells	CMP	MPP	LT-HSC
0	0.1			
0.125	11.5	19.3	7.6	3.5
0.25	25.7	34.7	18.7	11.7
0.5	33.8	43.1	24.4	13.6
1	36.4	42.5	27.0	20.8
2	35.2	46.3	25.5	20.5
3	34.5	44.7	25.7	21.6
4	36.4	44.2	25.3	21.8
8	36.2	45.6	30.3	24.0

SUBSTITUTE SHEET (RULE 26)