

(a)

gag atgggtgagagcgtcgtatlaagcgggggagaattagataaaagggaaaaaattcggtaagccagggggaaagaacaaataaactaaaacata
syngag atgggctgcccgccagcgtgctgtcgggcccgcagctggaccgctgggagaagatccgcttgcgccccggcgcaaaaagaagtacaagctgaagcaca

gag tagtatgggcaagcagggagctagaacgattcgcagtttaattcctggccttttagagacatcagaaggctgtagacaaatactgggacagctacaacatc
syngag tctgttgggcccagccgcaactggagcgtcttcgcccgtgaaacccgggctcctggagaccagcaggggtgcccagagatcctcggcaactgcagccagc

gag ccttcagacaggatcagaagaacttagatcatlataatacaaataggcagtcctctattgtgtgcatcaaaggatagatgtaaaagacaccaaggaagcc
syngag cctgcaaacccggcagcagggagctgcgacgctgtacaacaccgtggccacgctgtactcgtccaccagcgcacgaaatcaaggatadgaagagggcc

gag tttagataagatagaggaagagcaaaaacaaagttaagaaaaggcaccagcaagcagcagctgacacagaaacaacagccaggtcagccaaaattacccta
syngag ctggataaaaatcgaagaggaacagaataagagcaaaaagaaggcccaacagggcccgggacacgggacacgaaacaggtcagccagaactacccca

gag tagtgcagaacctccaggggcaaatggtacatcagccatatacctagaaactttaaatacattgggtaaaagttagtagaagagaaggctttcagccaga
syngag tctgtgcagaacctccaggggagatgggtgacacagccatctcccggcagcgtgaaacgctgggtgaaagtggtggaagagaaggcttttagcccgga

gag agtaatacccattgtttcagcattatcagaaggagccaccacaagatttaaataccatgctaaaacagtggggggacatcaagcagccatgcaaatg
syngag ggtgatacccattgtttcagcctctgtcagaggagccaccccaagatctgaaacacatgctcaacacagtggggggacacagggcccatgcaatgagct

gag ttaaaagagaccatcaatgaggaagctgcagaatgggatagatgcatccagtgcatgagggccatattgaccagccagatgagagaaccaaggggaa
syngag ctgaaaggagaccatcaatgaggaagctgcccgaatgggatcgtgtgcatccgggtgacagcagggcccatcgcacggggccagatgctgtagcaccgggct

gag gtgacatagcaggaactactagtacccttcaggaacaaataggatggatgacacataatccacctatcccagtaggagaaatctataaaagatggataat
syngag cagacatcgccggaaacgactagtacccttcaggaacagatcggctggatgacaaacaacccaccatcccgggtgggagaaatctacaaacgctggatcat

gag cctgggatlaaataaaatagtaagaatgtatagccctaccagcatctctggacataagacaaggaccaaggaacccttttagagactatgtagaccgatcc
syngag cctgggctgaaacaagatcgtgcatgtatagccctaccagcatcctggacatccgcaaggcccgaaggaaccctttcgcgactacgtggaccggttc

gag tataaaaactctaagagccgagcaagcttcacaagaggtaaaaaatggatgacagaaacctgttgggtccaaatgcaaccagatgttaagactattt
syngag tacaaaaagctccgcccagcagggctagccaggggtgaaagaactggatgaccgaaacctgttgggtccagaacgcaaccggactgcaagacgatcc

gag taaaagcattgggaccagagcagactagaagaatgatgacagcatgtcaggagtggggggaccggccataaaagcaagagttttggctgaagcaat
syngag tgaagggcctgggcccagcggctaccctagagaaatgatgaccgctgtcaggagtgggcggaccggccaacagggcagcgtcctggctgagggcat

gag gagccaagtaacaaatcagctaccataatgatacagaaaggcaattttagaaccaagaagactgttaagtgttcaattgtggcaagaaggccac
syngag gagccaggtgaccaactcagctaccatcatgatgcagcgggcaactttcgaaccaacgcaagatcgtcaagtgttcaactgtggcaagaaggccac

gag atagccaaaaattgcagggcccctaggaaaaaggctgttggaaatgtgaaaggaaggacaccaaatgaaagattgtactgagagacaggtcaactcc
syngag acagcccgcactcagggcccctaggaaaaaggctgttggaaatgtgaaaggaaggacaccaaatgaaagattgtactgagagacaggtcaactttt

gag tggggaaagatctggccttcccacaagggaaggccagggaaatttcttcagagcagaccagaccacaacagcccaccagaagagagcttcaggtttgggga
syngag tagggaaagatctggccttcccacaagggaaggccagggaaatttcttcagagcagaccagaccacaacagcccaccagaagagagcttcaggtttgggga

gag agagacaacaactccctctcagaagcaggagccgatagacaaggaactgtatcctttagcttcctcagatcactctttggcagcagaccctcgtcacia
syngag agagacaacaactccctctcagaagcaggagccgatagacaaggaactgtatcctttagcttcctcagatcactctttggcagcagaccctcgtcacia

Liu et al Supplemental Fig. 1a
Nucleotide sequence alignment of the HIV-1 NL4-3 *gag* gene and a previously described codon-optimized *syngag*

(b)

syngag atgggcycccgcagcgtgctgtgggcygagctggaaccgtgggagaagatccgctgcycccggcgcaaaaagaagtacaagctgaagcaca
NSG atgggtcggagagcgtcgtactttcaggcgggagcttgacaagtgggagaagataagacttcgacctggagggaaaaagcagtacaagtttaagcaca

syngag tctgtgtgggcyagccgcgaactggagcgttcgcygtaaccgcggctcctggagaaccagcaggggtgcccagatcctcggccaactgcagccag
NSG ttgtttgggttcaagagaacttgagaggttcgctgtaaacccaggattacttgaacttctgaggatgcccagagatttagggcaattacagccttc

syngag cctgcaaacggcagcagggagctgcygagcctgtacaacacgtggccagcgtgtactgctccaccagcgcacgaaatcaaggatcagaaagggcc
NSG attacaactgggtcggaggagttagcatccctttacaacacatattgctgttttatactgcgtacaccagagaattgacgtcaaggacactaaagaggct

syngag ctggataaaatcgaagaggaaacagaataagcgaaaaagaaggcccaacaggcgcgcygacccggacacagcaaccaggtcagccagaactaccca
NSG cttgacaaaattgaagaggaaacagaataagctcaaaaagaagctcaacaggctgcygagacagggcaataattcacaggtcgcagaactatccaa

syngag tctgtgcagaacaaccagggcgagatggtgcaccaggcattctcccgcgacgctgaacgcctgggtgaaggtggtggaagagaaggcttttagcgcgga
NSG ttgtacaanaattgcaaggccagatggtcaccaaagcattctgcacgcgacattgaatgcatgggttaaggtcgttaggaaaaagcgttctcgcctga

syngag ggtgataccatgttctcagccctgtcagaggagaccaccccagaatctgaacacatgctcaaacagtggtgggacaccaggcgcctatgcagatg
NSG ggttatcccagatgttctcggctcttagcagggggctacacctcaggatttaaatcaaatcgaatcaggtaggaggggcaccaggctgcaatgcagatg

syngag ctgaaggagaccatcaatgaggaggtgccgaatgggactgctgcatccgtgacgcagcagggcccaatcgcaaccggccagatgctgagccacggggct
NSG cttaaagaaaacgataaacgaaagggcgtgagtggaagcagctgcaccctgctgcatgcccagaccgatagccctgggcaaatgaggagcccgagaggga

syngag cagacatcgcgggaaccgactagtacccttcaggaaacagatcggctggatgaccaaacaaccaccatcccgggtgggagaaatctacaacgctggatcat
NSG gcgatatttggggcacaacgtaaccgttgcaagagcagattgggtggatgacgcacaaccggcaattcctgttggggagataacaagaggtggatgat

syngag cctgggctgaacaagatcgtgctgcatgataagcctacagcattcctggacatccgccaaggccggaaggaacccttctcgcactacgtggaacgggttc
NSG attggggcttaacaagattgttaggatgtactcgcagcgtcgatactagacataggcagggccctaaagagcggctcagggactacgtgacagatc

syngag tacaanaacgctccgcygagcaggttagccaggaggtgaagaactggatgaccgaaadccctgctgggtccagaaccggaaccggactgcaagacgatcc
NSG tacaagacactgagggcggaaacaggcgtcgcaggaagttgaagaactggatgacggagagcgttattagtgcaagaccgaaatcctgactgcaaaaacat

syngag tgaaggcctgggcccagcggctaccctagaggaaatgatgacgcctgtcagggagtgggcggaccggccacaaggcacgctcctggctgaggccat
NSG tgaaggcgttagggccgggggcaacgctggaggagatgatgacgcttgcgaaggcgttaggaggccagggcacaaggcgaggggttagcggaggcgt

syngag gagccaggtgaccaactccgctaccatcatgatgagcgcggcaacttctggaaaccaacgcaagatcgtcaagtgttcaactgtggcaagaaggccac
NSG gtcgcaggttacgaaccggcaacgattatgatcgaagggaaatctcagaatcagaggaaaacggtaaatgcttcaactgaggaaaggaggacac

syngag acagcccgaactgcagggccctaggaaaaagggctgttggaaaatgtggaaggaaggabaccaaatgaaagattgtactgagagacaggctaatTTTT
NSG atcgcaaaagaactgcagagcaccgctaaagaaaggatgctggaagtcggcaaaagaggccaccagatgaaggactgcacagaacgtcaagcaaatTTT

syngag tagggaagatctggccttcccacaaggggaaggccagggaatttcttcagagcagaccagagccaaacagccccaccagaagagagcttcaggtttgggga
NSG tggggaagatctggccttcccacaaggggaaggccagggaatttcttcagagcagaccagagccaaacagccccaccagaagagagcttcaggtttgggga

syngag agagacaacaactccctctcagaagcaggagccgatagacaaggaactgtatcctttagcttccctcagatcactctttggcagcgaccctcgtcaca
NSG agagacaacaactccctctcagaagcaggagccgatagacaaggaactgtatcctttagcttccctcagatcactctttggcagcgaccctcgtcaca

Liu et al Supplemental Fig. 1b

Nucleotide sequence alignment of two codon optimized sequences *syngag* and *NSG* that encode HIV-1 Gag

(c)

gag atgggtgcgagagcgtcggat taa g c g g g g a g a a t t a g a t a a a t g g g a a a a a t t c g g t t a a g g c c a g g g g a a a g a a c a a t a t a a a c t a a a c a t a
NSG atgggtgcgagagcgtcggat c t t c a g g c g g g a g c t g a c a a g t g g g a g a a g a t a a g a c t c g a c c t g g a g g g a a a a g c a g t a c a a g t t a a g c a c a

gag t a g t a t g g g c a a g c a g g g a g c t a g a a c g a t t c g c a g t t a a t c c t g g c c t t t t a g a g a c a t c a g a a g g c t g t a g a c a a a t a c t g g g a c a g c t a c a a c c a t c
NSG t t g t t t g g g c t c a a g a g a a c t t g a g a g t t c g c t g t a a a c c a g g a t t a c t t g a a a c t t c t g a g g g a t g c c g a c a g a t t t a g g g c a a t t a c a g c c t t c

gag c c t t c a g a d a g g a t c a g a a g a a c t t a g a t c a t t a t a t a a t a c a a t a g c a g t c c t c t a t t g t g t g c a t c a a g g a t a g a t g t a a a g a c a c c a a g g a a g c c
NSG a t t a c a a a c t g g g t c g g a g g a g t t a c g a t c c c t t t a c a a c a c t a t t g c t g t t t a t a c t g c g t a c a c c a g a g a a t t g a c g t c a a g g a c a c t a a a g a g g c t

gag t t a g a t a a g a t a g a g g a a g a g c a a a a c a a a g t a a g a a a a g g c a c a g c a a g c a g c t g a c a c a g g a a a c a a c a g c c a g g t c a g c c a a a t t a c c c t a
NSG c t t g a c a a a a t t g a a g a g g a c a g a a t a a g t c t a a a a g a a a g c t c a a c a g g c t g c g g c a g a c a c g g g c a a t a a t t c a a a g t g c g a g a a c t a t c c a a

gag t a g t g c a g a a c t c c a g g g c a a a t g g t a c a t c a g g c c a t a t c a c t a g a a c t t t a a t g c a t g g g t a a a g t a g t a g a a g a g a a g g c t t c a g c c c a g a
NSG t t g t a c a a a a t t g c a a g g c c a g a t g g t t c a c c a a g c g a t t t c g c c a c g g a c a t t g a a t g c a t g g g t a a a g g t c g t t g a g g a a a a g g c t t c t c g c c t g a

gag a g t a a t a c c c a t g t t t t c a g c a t t a t c a g a a g g a g c a c c c c a c a a g a t t t a a a t a c c a t g c t a a a c a c a g t g g g g g a c a t c a a g a g c c a t g c a a a t g
NSG g g t t a t c c c g a t g t t c t c g g c t t a g c a g g g g g c t a c a c c t c a g g a t t a a a t a c a a t g c t g a a t a c g g t a g g a g g g c a c c a g g c t g c a a t g c a g a t g

gag t t a a a a g a g a c c a t c a a t g a g a a g c t g c a a t g g g a t a g a t t g c a t c c a g t g c a t g c a g g g c c t a t t g c a c d a g g c c a g a t g a g a a c c a a g g g g a a
NSG c t t a a g g a a a c g a t a a a c g a a g a g g c c g t g a g t g g g a c a g g t g c a c c c t g t g c a t g c g g a c c g a t a g c c c t g g g c a a a t g a g g g a g c c g a g a g g a

gag g t g a c a t a g c a g g a a c t a c t a g t a c c c t t c a g g a a c a a a t a g g a t g g a t g a c a c a t a a t c c a c c t a t c c a g t a g g a g a a t c t a t a a a g a t g g a t a a t
NSG g c g a t a t t g c g g c a c a a d g t c a a c g t t g c a a a g a g a g a t t g g t g g a t g a c g c a c a a c c g c c a a t t c c t g t t g g g a g a t a t a c a a g a g g t g g a t t a t

gag c c t g g g a t t a a a t a a a t a g t a a g a a t g t a t a g c c c t a c c a g c a t t c t t g g a c a t a a g a c a a g g a c c a a a g g a a c c c t t t a g a g a c t a t g t a g a c c g a t t c
NSG a t t g g g g c t t a a c a a g a t t g t t a g g a t g t a c t c g c g a g c t g a t a c t a g a c a t t a g g c a g g g c c t a a a g a g c c g t t c a g g g a c t a c g t t g a c a g a t t c

gag t a t a a a a c t c t a a g a g c c g a g c a a g c t t c a c a a g a g g t a a a a a t t g g a t g a c a g a a a c c t t g t t g g t c c a a a t g c g a a c c c a g a t t g t a a g a c t a t t t
NSG t a c a a g a c a c t g a g g c g g a a c a g g c t g c a g g a a g t t a a g a a c t g g a t g a c g g a a c g t t a t a g t g c a g a a c g a a a t c c t g a c t g c a a a c a a t a t

gag t a a a a g c a t t g g g a c c a g g a g c g a c a c t a g a a g a a t g a t g a c a g a t g t c a g g a g t g g g g g a c c c g g c c a t a a a g c a a g a g t t t g g c t g a a g c a a t
NSG t g a a g g c g t t a g g g c c g g g g a a c g c t g g a g g a g a t g a t g a c g g c t t g c c a a g g c g t a g g a g g c c a a g g c a c a a g g c a g g g t g t t a g e g g a g g c g a t

gag g a g c c a a g t a a c a a t c c a g c t a c c a t a a t g a t a c a g a a a g g c a a t t t t a g g a a c c a a g a a a g a c t g t t a a g t g t t c a a t t g t g g c a a a g a a g g g c a c
NSG g t c g a g g t t a c g a a c c c g g c a a c g a t t a t g a t c c a a a g g g a a a t t t c a g a a a t c a g a g g a a a a c g g t a a a a t g c t t c a a c t g c g g a a a g g a g g a c a c

gag a t a g c c a a a a a t t g c a g g g c c c t a g a a a a a g g g c t g t t g g a a a t g t g g a a g g a a g g a c a c c a a a t g a a a g a t t g t a c t g a g a g a c a g g c t a a c t c c
NSG a t c g c a a a g a a c t g c a g a g c a c c g c g t a a g a a a g g a t g c t g g a a g t g c g g c a a a g a g g g c c a c c a g a t g a a g g a c t c a c a g a a c g t c a a g c a a a t t t c c

gag t g g g a a g a t c t g g c c t t c c c a a g g g a a g g c c a g g a a t t t c t t c a g a g c a g a c c a g a g c c a a c a g c c c c a c c a g a a g a g a g c t t c a g g t t t g g g a
NSG t g g g a a g a t c t g g c c t t c c c a a g g g a a g g c c a g g g a a t t t c t t c a g a g c a g a c c a g a g c c a a c a g c c c c a c c a g a a g a g a g c t t c a g g t t t g g g a

gag a g a g a c a a c a a c t c c c t c t c a g a a g c a g a g c c g a t a g a c a a g g a a c t g t a t c c t t a g c t t c c c t c a g a t c a c t c t t t g g c a g o g a c c c c t g t c a a a
NSG a g a g a c a a c a a c t c c c t c t c a g a a g c a g a g c c g a t a g a c a a g g a a c t g t a t c c t t a g c t t c c c t c a g a t c a c t c t t t g g c a g o g a c c c c t g t c a a a

Liu et al Supplemental Fig. 1c
Nucleotide sequence alignment of the HIV-1 NL4-3 *gag* gene and codon-optimized *NSG*

(d)

```
Gag      MGARASVLSGGELDKWEKIRLRPGGKKQYKCLKHIVWASRELERFAVNPGLLETSEGCRQI
NSG      MGARASVLSGGELDKWEKIRLRPGGKKQYKCLKHIVWASRELERFAVNPGLLETSEGCRQI
synGag   MGARASVLSGGELDRWEKIRLRPGGKKKYKCLKHIVWASRELERFAVNPGLLETSEGCRQI
*****

Gag      LGQLQPSLQGTGSEELRSLYNTIAVLYCVHQRIDVKDTKEALDKIEEEQNKSKKKAQQAAA
NSG      LGQLQPSLQGTGSEELRSLYNTIAVLYCVHQRIDVKDTKEALDKIEEEQNKSKKKAQQAAA
synGag   LGQLQPSLQGTGSEELRSLYNTVATLYCVHQRIEIKDTKEALDKIEEEQNKSKKKAQQAAA
*****

Gag      DTGNNSQVVSQNYPIVQNLQGMVHQAI SPRTLNAWVKVVEEKAFSPEVIMFSAALSEGAT
NSG      DTGNNSQVVSQNYPIVQNLQGMVHQAI SPRTLNAWVKVVEEKAFSPEVIMFSAALSEGAT
synGag   DTGHSNQVVSQNYPIVQNIQGMVHQAI SPRTLNAWVKVVEEKAFSPEVIMFSAALSEGAT
***

Gag      PQDLNNTMLNTVGGHQAAMQMLKETINEEAAEWDRLHPVHAGPIAPGQMREPRGSDIAGTT
NSG      PQDLNNTMLNTVGGHQAAMQMLKETINEEAAEWDRLHPVHAGPIAPGQMREPRGSDIAGTT
synGag   PQDLNNTMLNTVGGHQAAMQMLKETINEEAAEWDRVHPVHAGPIAPGQMREPRGSDIAGTT
*****

Gag      STLQEQIGWMTHNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
NSG      STLQEQIGWMTHNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
synGag   STLQEQIGWMTNPNPPIPVGEIYKRWIILGLNKIVRMYSPTSILDIRQGPKEPFRDYVDRF
*****

Gag      YKTLRAEQASQEVKNWMTETLLVQANANPDCKTILKALGPGATLEEMMTACQGVGGPGHKA
NSG      YKTLRAEQASQEVKNWMTETLLVQANANPDCKTILKALGPGATLEEMMTACQGVGGPGHKA
synGag   YKTLRAEQASQEVKNWMTETLLVQANANPDCKTILKALGPAATLEEMMTACQGVGGPGHKA
*****

Gag      RVLAEAMSQVTNPATIMI QKGNFRNQRKTVKCFNCGKEGHI AKNCRAPRKKGCWKCGKEG
NSG      RVLAEAMSQVTNPATIMI QKGNFRNQRKTVKCFNCGKEGHI AKNCRAPRKKGCWKCGKEG
synGag   RVLAEAMSQVTNSATIMMQRGNFRNQRKIVKCFNCGKEGHTARNCRAPRKKGCWKCGKEG
*****

Gag      HQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPEPTAPPEESFRFGEETTTSPSQQEPID
NSG      HQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPEPTAPPEESFRFGEETTTSPSQQEPID
synGag   HQMKDCTERQANFLGKIWPSHKGRPGNFLQSRPEPTAPPEESFRFGEETTTSPSQQEPID
*****

Gag      KELYPLASLRSLFGSDPSSQ
NSG      KELYPLASLRSLFGSDPSSQ
synGag   KELYPLASLRSLFGSDPSSQ
*****
```

Liu et al Supplemental Fig. 1d

Amino acid sequence alignment of the Gag proteins encoded by NL4-3 *gag*, *NSG* and *syn**gag*. Asterisks indicate amino acid residues that are identical in these three proteins.