

1 GGACCTGCAAGGCCCCACC 20
21 CACGCCGGGGCCGAGTAACCCCGCCGCTCTGGCACGCGGGCTCAGGCTGTCGCG 80
81 GGAGCTCCTTGAAGCCCGTGGGAAACCTGGGCGCTTGGCGGAGACGCTTTGTGGCT 140
141 GCTCTCTGGAGCATTTTTTCCCTTAGAGCTGTTGCTGTTTTCTGTGACGCCGCTTT 200
201 TCTGAAAGACGGTGTGCGCTTTGGCTTATCTGGACGCTCCCAAAGTCCCGGAA 260
261 AGGACACGGCCGGACACCTGGGGCTTGGTTCCGGCCGGAAGTCCGGGAGGTGTAGG 320
321 GCTAGAGTTCTGCGCTGGCGGGCGGTTCTGCTGCTGCTGCGCGGCCGCTGGGCT 380
381 ATGGGGCGCGGCTCGGGGCTGGCGGCTCCAGCAACAGCAGCGCCGGAGGACGCGGAG 440
1 M G R R S R G R R L Q Q Q R P E D A E 20
441 GATGGCCGCGAGGGTGTGAAAGCGCGGCGAGCGGGCTGGGAAGGAGGCTACCCGGAG 500
21 D G A E G G G K R G E A G W E G G Y P E 40
501 ATCGTCAAGGAGAACAAAGCTGTTCGAGCACTACTACAGGAGCTCAAGATCGTGCCGGAG 560
41 I V K E N K L F E H Y Y Q E L K I V P E 60
lep56
561 GCGAGTGGGGCAGTTCATGGACGCTCTCAGGGAGCGCTCCGGCCACTTTAAGAATT 620
61 G E W G Q F M D A L R E P L P A T L R I 80
621 ACTGTTACAAAAGCCACGAAAAGAGATTCTCCATTGCTTAAAGAACAAATATTTTAAAG 680
81 T G Y K S H A K E I L H C L K N K Y F K 100
681 GAATTTGGAGACTGGAGGTGACGGTCAAGAAAGTTGAAAGTTCACAGCCACTGAGTTGG 740
101 E L E D L E V D G Q K V E V P Q P L S W 120
lep77-1
741 ATCTCTGAAGAACTTGCCTGGCACACAAATTAAGTCAAAAATCTTGAGAAAATCGCCA 800
121 Y P E E L A W H T N L S R K I L R K S P 140
801 CACTTGGAAAAGTTTCACTAGTTCAGTTAGTGAACAGAATCTGGAATATAGTCGT 860
141 H L E K F H Q F L V S E T E S G N I S R 160
861 CAAGAAGCTGTAGCATGATCCACCACTGCTCCTCAAGCTGCGGCTCATCATAAGATC 920
161 Q E A V S M I P P L L L N V R P H H K I 180
921 TTAGATATGTGTCAGCACCTGGCTCAAAGACCACACAGTTAATGAAATGCTACATGCC 980
181 L D M C A A P G S K T T Q L I E M L H A 200
lep77-2
981 GACATGAATGTCCCTTTCAGAGGGATTGTTATTGCGAATGATGTGGAACAAGCGC 1040
201 D M N V P F P E G F V I A N D V D N K R 220
1041 TGCTACTGCTGTCCATCAAGCAAGAGGCTGAGCAGCCCTGCATCATGGTGGTCAAC 1100
221 C Y L L V H Q A K R L S S P C I M V V N 240
1101 CATGATGCTCAGCATACCCAGGCTCCAGATAGATGTGGACGGCAGGAAAGATCCTC 1160
241 H D A S S I P R L Q I D V D G R K E I L 260
1161 TTCTATGATCAATTTTATGTGATGTCCTTGCAGTGGAGACGGCACTATGAGAAAAAC 1220
261 F Y D R I L C D V P C S G D G T M R K N 280
1221 ATTGATGTTGAAAAAGTGGACACCTTAAATAGCTTGCAGCTACATGCTTACAGCTG 1280
281 I D V W K K W T T L N S L Q L H G L Q L 300
1281 CGGATTGCAACACGCGGGGCTGAACAGCTGGCTGAAGGTGGAAGGATGGTATTCCAGG 1340
301 R I A T R G A E Q L A E G G R M V Y S T 320
1341 TGTTCACTAAACCTTATTGAGGATGAAGCAGTCATAGCATCTTACTGGAAAAAGTGAA 1400
321 C S L N P I E D E A V I A S L L E K S E 340
1401 GGTGCTTTGGAGCTGTGATGTGCTAATGAATGCCAGGCTGAAGTGGATGCCTGGA 1460
341 G A L E L A D V S N E L P G L K W M P G 360
1461 ATCACACAGTGAAGGTAATGACGAAGATGGCGAGTGTACAGACTGGGACGCTGTT 1520
361 I T Q W K V M T K D G Q W F T D W D A V 380
1521 CCTCACAGCAGACACCCAGATCCGACTACCATGTTCCCTCCGAAGGACCCAGAAAAAG 1580
381 P H S R H T Q I R P T M F P P K D P E K 400
1581 CTGACGGCCATGCACCTGGAGCGATGCCTTAGGATATTACCCATCATCAGAATCTGGA 1640
401 L Q A M H L E R C L R I L P H H Q N T G 420
1641 GGGTTTTTTTGGCAGTATTTGGTAAAAATCTTCAATGCCGTGGAATAAACGTGAGCCA 1700
421 G F F V A V L V K K S S M P W N K R Q P 440
1701 AAGCTTCAGGTAATCTGCAGAGACAGAGAAAGCACAGCTGAGCCCTGCAGATCTC 1760
441 K L Q G K S A E T R E S T Q L S P A D L 460
1761 ACAGAAGGGAAACCACAGATCCCTCTAAGCTGGAAGTCCGTCATTACAGAACTGGT 1820
461 T E G K P T D P S K L E S P S F T G T G 480
lep60
1821 GACACAGAAATAGCTCATGCAACTGAGGATTTAGAGAATAATGGCAGTAAGAAGATGGC 1880
481 D T E I A H A T E D L E N N G S K K D G 500
1881 GTGTGTGCTCCTCCATCAAGAAAAATGAAGTTATTTGGATTTAAGAAGATCCATTT 1940
501 V C G P P P S K K M K L F G F F K E D P F 520
1941 GTATTTATCTGAAGATGACCCATTATTTCCACTATTGAGAATTTATGCTTTGGAT 2000
521 V F I P E D D P L F P P I E K F Y A L D 540
2001 CCTTATTCCCAAGGATGAATTTGTTAACTCGGACTACAGAAGGGAAGAAAGGAGCTC 2060
541 P S F P R M N L L T R T T E G K K R Q L 560
2061 TACATGGTTTCTAAGGAGTGGCGAATGTGCTGCTGAATAACAGTGAGAAGATGAAGTT 2120
561 Y M V S K E L R N V L L N N S E K M K V 580
2121 ATTAACACGGGGATCAAGTCTGGTGTAGAATAACACGGGTGAAGGTTGACTGTGCT 2180
581 I N T G I K V W C R N N S G E E F D C A 600
2181 TTCCGGCTGGCAGGGGGAATATATACATTGTATCCATTTATTAAGTCAAGAAATTT 2240
601 F R L A R E G I Y T L Y P F I N S R I I 620
2241 ACTGTATCAATGGAAGATGTTAAGTACTGTTGACCCAGGAAATCCCTTTTATGAAAA 2300
621 T V S M E D V K I L L T Q E N P F F R K 640
2301 CTCAGCAGTGAAGCTACAGTCAAGCAAGGACCTGGCAAGGGAAGCATCGCTGAAG 2360
641 L S S E T Y S Q A K D L A K G S I V L K 660
2361 TATGAACAGATTCTGGAATCCAGACGCTCTGCAAGTCCCATCGCTTATCGGATGG 2420
661 Y E P D S A N P D A L Q C P I V L C G W 680
2421 CGGGAAAGGCTCCTTCAAGTCTTTGTCGCAAGAAATGAAGGCTTATTATCTCAGG 2480
681 R G K A S I R T F V P K N E R L H Y L R 700
2481 ATGATGGGGCTGGAGTATTTGGGAGAAAAGAAGGAGGGGTTATCCTCACAATGAG 2540
701 M M G L E V L G E K K K E G V I L T N E 720
2541 AGTGACGCAGCCGGACAGCCAGCAATGACGTGACTGAGGGACAGAGCAGGAGAG 2600
721 S A A S T G Q P D N D V T E G Q R A G E 740
2601 CCCAACGCCAGATGCAGAAGGCGCAACAGTCCAGCAGTGCAGCAGGCTGTGACCCG 2660
741 P N S P D A E E A N S P D V T A G C D P 760
2661 GCGGGGCTCCTCCACCCCGGTGAGCAGGCCAAGGAGCGGGGGCCACACCCCTCACA 2720
761 A G V H P P R * 767
2721 CGCAAACTGGCTCTTCTGCTCACTGGTGTGAAACAAATCCAGAGCAGCTGTGGC 2780
2781 CTGTAAGCATATATTTCTAATGACTGCAGACTGGTGGATCATTAGGAGCTTCTGAATG 2840
2841 ACCAGGACTGCTTCTTTGGAGCTGATGAAATGACTCTTTTACGCTGTTAGAATCAC 2900
2901 TTGTTTTATTGTTCTTTGGCAGCTGGTCTAGTGTCTTTTGTGCGGAATAGAC 2960
2961 TTTCAAAGTTGTACTTCTATCAAGAAACAAACTGCCCTTGCAAGAAATTCAGGCTTT 3020
3021 TGTTAAGCCTGTATTGGCTTAAGGTGCAGTATTTTAAATTTATTTATAGAAAAGAA 3080
3081 TCTATAAATCTTGGGAAAGTGTATAAGCTTTAAATTAATTCATTGAGCTGCACCTCA 3140
3141 GTGGTGTGCTTAACATGCAAGTGGGGTTAATCTGAGGCTCAGATGACTTTGTGCTC 3200
3201 TTTGGAATAAAGGTAATAAATAACTCCAGAGTAAGAGCTGTATCGTGAATGTCATA 3260
3261 CTAATTTATTGAGGGGACTTATGTGCTTTTAAATGAAGTGCCTTACAATTTTATT 3320
3321 TTAATGGGGTGGGACTCTTGAATATTTCAATAAAATGATAAAAATATAAAAAAAA 3380

Supplementary Figure S1 Full-length cDNA sequences of NSUN2 and their predicted amino acid sequences. Peptide sequences indicated in red (lep56, lep77-1, lep77-2, and lep60) were determined by Edman analysis. The DNA sequences in blue indicate PCR primer sequences for 5'-RACE.