



1 GGATCTGCATCGCTCCGGTGCCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCC
201 GTGAACGTTCTTTTTCGAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCGAGGGCTCGCATCTCTCTTCACGCGCCCGCCCTACCTGAGGCC
301 GCCATCCACGCGGTTGAGTCGCGTTTGCGCCCTCCCGCTGTGGTGCCTCTGAAGTGCCTCCGCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCGGGCTCTCCACGCTTTGCCTGACCTGCTTGTCTAACTCTACGCTTTTGTTCGTTT

AgeI (552)

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGGGCGCTACCTGAGATCACCGGTCCACATGGCAAGGCTGGGAAACTGCAGCCTGACTTGGGCTGC
1 ▶ M A R L G N C S L T W A A
601 CCTGATCATCCTGCTGCTCCCGGAAGTCTGGAGGAGTGGGGCACATCAGTGTCTCAGCCCCATCGTCCACTGGGGGATCCCATCACAGCCTCCTGC
13▶ L I I L L L P G S L E E C G H I S V S A P I V H L G D P I T A S C
701 ATCATCAAGCAGAAGTGCAGCCATCTGGACCCGGAGCCACAGATTCTGTGGAGACTGGGAGCAGAGCTTCAGCCCGGGGAGGCAGCAGCGTGTGTCTG
47▶ I I K Q N C S H L D P E P Q I L W R L G A E L Q P G G R Q Q R L S
801 ATGGGACCAGGAATCTATCATCACCCTGCCCCACCTCAACCACACTCAGGCCTTTCTCTCTGCTGCTGAACTGGGGAACAGCCTGCAGATCCTGGA
80▶ D G T Q E S I I T L P H L N H T Q A F L S C C L N W G N S L Q I L D
901 CCAGTTGAGCTGCGCGCAGGCTACCCAGCCATACCCACAACCTCTCCTCATGSAACCTCACAACCAGCAGCCTCATCTGCCAGTGGGAGCCA
113▶ Q V E L R A G Y P P A I P H N L S C L M N L T T S S L I C Q W E P
1001 GGACCTGAGACCCACCTACCCACCAGCTTCACTCTGAAGAGTTTCAAGAGCCGGGCAACTGTGACACCAAGGGGACTCCATCCTGGACTGCGTGCCCA
147▶ G P E T H L P T S F T L K S F K S R G N C Q T Q G D S I L D C V P
1101 AGGACGGGAGAGCCACTGCTGCATCCCACGAAACACCTGCTGTTGTACCAGAATATGGGCATCTGGGTGCAGGCAGAGAATGCGCTGGGGACAGCAT
180▶ K D G Q S H C C I P R K H L L L Y Q N M G I W V Q A E N A L G T S M
1201 GTCCCCAAGTGTGCTTGTATCCCATGGATGTTGTGAAACTGGAGCCCCCATGCTGCGGACCATGGACCCAGCCCTGAAGCGGCCCTCCCGAGGCA
213▶ S P Q L C L D P M D V V K L E P P M L R T M D P S P E A A P P Q A
1301 GGCTGCCTACAGCTGTGCTGGAGCCATGGCAGCCAGGCTGCACATAAATCAGAAGTGTGAGCTGCGCCACAAGCCGAGCGTGGAGAAGCCAGCTGGG
247▶ G C L Q L C W E P W Q P G L H I N Q K C E L R H K P Q R G E A S W
1401 CACTGGTGGGCCCCCTCCCTTGAGGGCCTTCACTGATGAGCTCTGCGGGCTCTCCAGCCACGGCTACACCCTGCAGATACGCTGCATCCGCTGGCC
280▶ A L V G P L P L E A L Q Y E L C G L L P A T A Y T L Q I R C I R W P
1501 CCTGCCTGGCCACTGGAGCGACTGGAGCCCAGCCTGGAGTGAGAATACCGAACGGGCCCCACTGTCAGACTGGACACATGGTGGCGGAGAGGAGCAG
313▶ L P G H W S D W S P S L E L R T T E R A P T V R L D T W W R Q R Q
1601 CTGGACCCAGGACAGTGCAGCTGTTCTGGAAGCCAGTGCCTTGGAGGAAGCAGCGGACGGATCCAAGGTTATGTGGTTTCTTGGAGACCCCTCAGGC
347▶ L D P R T V Q L F W K P V P L E E D S G R I Q G Y V V S W R P S G
1701 AGGCTGGGGCCATCCTGCCCTTGCAACACCACAGAGCTCAGCTGCACCTTCCACCTGCCTCAGAAGCCAGGAGGTGGCCCTTGTGGCCTATAACTC
380▶ Q A G A I L P L C N T T E L S C T F H L P S E A Q E V A L V A Y N S
1801 AGCCGGGACCTCTCGTCCACTCCGGTGGTCTTCTCAGAAAGCAGAGGCCAGCTCTGACCAGACTCCATGCCATGGCCCGAGACCCCTCACAGCCTCTGG
413▶ A G T S R P T P V V F S E S R G P A L T R L H A M A R D P H S L W
1901 GTAGGCTGGGAGCCCCCAATCCATGGCCTCAGGGCTATGTATTGAGTGGGGCTGGGCCCCCCAGCGCGAGCAATAGCAACAAGACTGGAGGATGG
447▶ V G W E P P N P W P Q G Y V I E W G L G P P S A S N S N K T W R M
2001 AACAGAATGGGAGAGCCACGGGTTTCTGCTGAAGGAGAATCAGGCCCTTTCAGCTCTATGAGATCATCGTACTCCCTTGTACCAGGACACCATGGG
480▶ E Q N G R A T G F L K E N I R P F Q L Y E I I V T P L Y Q D T M G
2101 ACCCTCCAGCATGTCTATGCTACTCTCAAGAAATGGCTCCCTCCATGCCAGAGCTGCATCTAAAGCAATTGGCAAGACTGGGCAAGCAGCTGGAG
513▶ P S Q H V Y A Y S Q E M A P S H A P E L H L K H I G K T W A Q L E
2201 TGGGTGCCTGAGCCCCCTGAGCTGGGGAAGAGCCCCCTTACCCTACACCATCTTCTGGACCAACGCTCAGAACCAGTCTTCTCCGCCATCCTGAATG
547▶ W V P E P P E L G K S P L T H Y T I F W T N A Q N Q S F S A I L N
2301 CCTCTCCCGTGGCTTTGCTCCTCATGGCCTGGAGCCCGCAGTCTGTATCACATCCACTCATGGCTGCCAGCCAGGCTGGGGCCACCAACAGTACAGT
580▶ A S S R G F V L H G L E P A S L Y H I H L M A A S Q A G A T N S T V
2401 CCTCACCTGATGACCTTGACCCAGAGGGTGGAGCTACACATCCTGGGCCTGTCGGCCTCTGCTGTTGCTCACCTGCCTCTGTGGAAGTGC
613▶ L T L M T L T P E G S E L H I I L G L F G L L L L L L T C L C G T A
2501 TGGCTCTGTTGACGCCCCAAGGAAAGTCCCTCTGGCCAAGTGTCCAGACCCAGCTCACAGCAGCCTGGGCTCCTGGGTGCCACAATCATGGAGG
647▶ W L C C S P N R K N P L W P S V P D P A H S S L G S W V P T I M E
2601 AGGATGCCTTCCAGCTGCCCGGCTTGGCACGCCACCCATACCAAGCTCACAGTGTGGAGGAGGATGAAAAGAAGCCGGTGCCTGGGAGTCCCATAA
680▶ E D A F Q L P G L G T P P I T K L T V L E E D E K K P V P W E S H N
2701 CAGCTCAGAGACTGTGGCCTCCCACTCTGGTCCAGACCTATGTGCTCCAGGGGACCCAAAGAGCAGTTTCCACCCAGCCCCAATCCAGTCTGGACC
713▶ S S E T C G L P T L V Q T Y V L Q G D P R A V S T Q P Q S Q S G T
2801 AGCGATCAGGTCCTTTATGGGAGCTGCTGGCAGCCCAAGGCCAGGGCCAGGCACTATCTCCGCTGTGACTCCACTCAGCCCCCTTGGCGGGCC
747▶ S D Q V L Y G Q L L G S P T S P G P G H Y L R C D S T Q P L L A G
2901 TCACCCAGCCCAAGTCTATGAGAACTCTGGTCCAGGCCAGCCCTTGGGACCTGGTAACCCAGCCCCAAGCCAGGAGGAGCAGTGTGTCTT
780▶ L T P S P K S Y E N L W F Q A S P L G T L V T P A P S Q E D D C V F

NheI (3088)

3001 TGGCCACTGCTCAACTTCCCCCTCCTGCAGGGGATCCGGGTCCATGGGATGGAGGCGCTGGGGAGCTTCTAGGGCTTCTGGGGTTCCTAGCTGGCCA
813▶ G P L L N F P L L Q G I R V H G M E A L G S F •
3101 GACATGATAAGATACATTGATGAGTTTGGACAAACCACAAC TAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTG

3201 TAACCATTATAAGCTGCAATAAAACAAGTTAAACAACAACAATTGCATTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTA
3301 AAACCTCTACAAATGTGGTATGGAATTCTAAAATACAGCATAGCAAACTTTAACCTCCAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATA
3401 AGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTTGCAGCCTCACCTTCTTTCATGGAGTTAAGATATAGTGATTTCCTCAAGGTTTGA
3501 ACTAGCTCTTCATTTCTTTATGTTTTAAATGCACTGACCTCCACATTCCTTTTTAGTAAAATATTCAGAAAATAATTTAAATACATCATTGCAATGAAA
3601 ATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTCATAATATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTTAATAGA
3701 AATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCTGGTGTACTTGAGGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTCATCTC
3801 AATGAGCACAAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTGCACATGCCACAGGGGCTGACCACCCTGATGGATCTGTCACCTCATCAGAGTAG
118 I L V F C D P A Y D S I L E R C M G C P S V V R I S R D V E D S Y
3901 GGGTGCCTGACAGCCACAATGGTGTCAAAGTCTTCTGCCCGTTGCTCACAGCAGACCCAATGGCAATGGCTTCAGCACAGACAGTACCCTGCCAATGT
84 P H R V A V I T D F D K Q G N S V A S G I A I A E A C V T V R G I Y
4001 AGGCCTCAATGTGGACAGCAGAGATGATCTCCCGAGTCTGGTCTGATGGCCGCCCGACATGGTGTCTGTTGTCCTCATAGAGCATGGTATCTTCTC
51 A E I H V A S I I E G T K T R I A A G V H H K N D E Y L M T I K E
4101 AGTGGCGACCTCCACCAGCTCCAGATCTGCTGAGAGATGTTGAAGTCTTCATGGTGGCCCTCTATAGTGAGTCGTATTATACTATGCCGATATACTA
18 T A V E V L E L D Q Q S I N F T K M
4201 TGCCGATGATTAATTGTCAAACAGCGTGGATGGCGTCTCCAGCTTATCTGACGGTCTACTAACGAGCTCTGCTTATATAGACCTCCACCCTACACGC
4301 CTACCGCCCATTTGCGTCAATGGGGCGGAGTTGTTACGACATTTTGGAAAGTCCCCTGATTACTAGTCAAACAAACTCCCATTGACGTCAATGGGGT
4401 GGAGACTTGAAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATGACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGAT
4501 GTACTGCCAAGTAGGAAAGTCCCATAGGTCATGACTGGGCATAATGCCAGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATA
4601 TGATACACTTGATGACTGCCAAGTGGGCGATTTACCGTAAATACTCCACCCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACG
4701 TCATTATTGACGTCAATGGGCGGGGTCGTTGGGCGGTGAGCCAGGCGGGCCATTTACCGTAAAGTTATGTAACGCTGCAGGTTAATTAAGAACATGTGA
4801 GCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGGTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAAATCGACGCT
4901 CAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTTCCCTGGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCTGCCGCTTAC
5001 CGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTGTTGCTCCAAGCTG
5101 GGCTGTGTGCACGAACCCCGTTCAGCCCGACCGTGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGG
5201 CAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGT
5301 ATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTT
5401 GTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTACGGGGTCTGACGCTCAGTGAACGAAACTCACGTT
5501 AAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCCAATAAAATATCTTTATTTTATTACATCTGTGTGTTGGTTTTTTGTGTGAA
5601 TCGTAACTAACATACGCTCTCCATCAAACAAACGAAACAAACAAACTAGCAAATAGGCTGTCCCAGTGCAAGTGCAGGTGCCAGAACATTTCTCT
5701 ATCGAA