



1 GGATCTGCATCGCTCCGGTGCCGTCAGTGGGAGAGCGCACATGCCACAGTCCCGAGAAGTTGGGGGAGGGTGGCAATTGAACGGGTGCCTA  
101 GAGAAGGTGGCGCGGGTAAACTGGGAAAGTATGTCGTGACTGGCTCCGCCTTTTTCCCGAGGGTGGGGGAGAACCCTATATAAGTGCAGTAGTCGCC  
201 GTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGCTGAAGCTTCAGAGGGCTCGCATCTCTCTTACGCGCCCCGCCCTACCTGAGGCC  
301 GCCATCCACGCCGGTTGAGTCGCGTTTCTGCCGCTCCCGCCTGTGGTGCTCTGAACTGCGTCCGCCGTCTAGGTAAGTTTAAAGCTCAGGTCGAGACC  
401 GGGCCTTTGTCCGGCGCTCCCTTGAGCCTACCTAGACTCAGCCGGCTCTCCACGCTTTGCTGACCTGCTTGTCTCAACTCTACGCTTTTGTTCGTTT

**AgeI (552)**

501 TCTGTTCTGCGCCGTTACAGATCCAAGCTGTGACCGCGCCTACCTGAGATCACCGGTCATCATGATTGTGGACAAGCTGCTGGACGACAGCCGCGGGCGG  
1 M I V D K L L D D S R G G  
601 AGAGGGGCTGCGGGACGCGGGCGGGCGGCTGCGGCTCATGACCAGCCCGCTCAACCTGAGTACTTCTACGGCGCGTCCGCGCCCGCCGCGCCCGGGC  
13 E G L R D A A G G C G L M T S P L N L S Y F Y G A S P P A A A P G  
701 GCCTGCGACGCCAGCTGCTCGGTCTTGGGCCCTCGGCGCCGGCTCGCCCGCTCCGACTCTCGACTTCTCTCTGCCTCGTGGTGTCTCTCTGCG  
47 A C D A S C S V L G P S A P G S P G S D S S D F S S A S S V S S C  
801 GCGCGTGGAGTCCCGGTCGAGAGGGCGGCCCGCGCCGAGCGCCAGCAGTTGAGCCCATATGGGGTGGCAGGCAGCAGAGAGGCCCTTTCAAGG  
80 G A V E S R S R G G A R A E R Q P V E P H M G V G R Q Q R G P F Q G  
901 TGTTCCGGTAAAGAACTCAGTGAAGAACTCTGTTGCACATCCGAAGTCATAAACAAGGCTTCTGGCCAAAGTGTGGATGATTTAAGACACAAGT  
113 V R V K N S V K E L L L H I R S H K Q K A S G Q A V D D F K T Q G  
1001 GTGAACATAGAACAGTTCAGAGAATTGAAGAACACAGTATCATACAGTGGGAAAAGGAAAGGCCCGATTCTGTCTGATGGACCTGCTTGCAAAAGGC  
147 V N I E Q F R E L K N T V S Y S G K R K G P D S L S D G P A C K R  
1101 CAGCTCTGTTGCATTCCCAATTTTTGACACCACCTCAACACCAACGCCCGGGGAGAGCATGGAAGATGTTTCAATGAACCCAAACAGGAGAGCAG  
180 P A L L H S Q F L T P P Q T P T P G E S M E D V H L N E P K Q E S S  
1201 TGCTGATCTGCTTCAAGACATTATCAACATTAAGAATGAATGCAGCCCGTTCCCTGAACACAGTTCAGTTAGCTGGTGAACCCCGTGGTGGTCCCT  
213 A D L L Q N I I N I K N E C S P V S L N T V Q V S W L N P V V V P  
1301 CAGAGCTCCCCGAGAGCAGTGTGAGACTTCCATGGAGGGCAGGTCTTTTCCACCTCAGAAATGCCAACATTCAAGTCAGGGGCTCCCAACAAA  
247 Q S S P A E Q C Q D F H G G Q V F S P P Q K C Q P F Q V R G S Q Q  
1401 TGATAGACCAGGCTTCCCTGTACCAGTATTCTCCACAGAACCAGCATGTAGAGCAGCAGCCACACTACACCCACAAACCAACTCTGGAATACAGTCTTT  
280 M I D Q A S L Y Q Y S P Q N Q H V E Q Q P H Y T H K P T L E Y S P F  
1501 TCCATACCTCCCAGTCCCGGCTTATGAACCAAACCTCTTTGATGGTCCAGAATCAGATTTTGCACAAACCAAAGCTTAGTTTCCCTTCTTGGTGT  
313 P I P P Q S P A Y E P N L F D G P E S Q F C P N Q S L V S L L G D  
1601 CAAAGGGAATCTGAGAATATTGTAATCCCATGCAGACTTCTCCAGTGTTCAGCAGCAAAATGATGCTCACTTGCACAGTTCAGCATGATGCCAGCA  
347 Q R E S E N I A N P M Q T S S S V Q Q N D A H L H S F S M M P S  
1701 GCGCTGTGAGGCTGTTGGGCGACGAGATGGCCTCTGACTTTCAAACACTTCACTGCCATTCTCAAACATGGGAAATCCAATGAACACCACAGTTC  
380 S A C E A M V G H E M A S D S S N T S L P F S N M G N P M N T T Q L  
1801 AGGAAATCACTTTTTAGTGGCAGGTGGAGCAGGAAGAAAGCAAATTTGGCAAATATTTCCCAAGACCAGTTTCTTTCAAAGGATGCAGATGGTGACAG  
413 G K S L F Q W Q V E Q E E S K L A N I S Q D Q F L S K D A D G D T  
1901 TTCCTTATATTGCTGTTGCCAAGGGAGAAGGCACTTTCCTATGTTCTTGAAGAAAGATGAATGCACTTCACTGCTGGATATTAAGAGCACAAATG  
447 F L H I A V A Q G G R R A L S Y V L A R K M N A L H M L D I K E H N  
2001 GACAGATGCCTTTCAGTGGCAGTGGCTGCCAATCAGCATCTCATTGTGCAGGATCTGGTGAACATCGGGGCACAGTGAACACCACAGATGCTGGGG  
480 G Q S A F Q V A V A A N Q H L I V Q D L V N I G A Q V N T D C W G  
2101 AAGAACACCTCTGCATGTGTGCTGAGAAGGGCACTCCAGGTGCTTCAAGCGATTGAGAAGGGAGCAGTGGGAAGTAATCAGTTTGTGGATCTTGA  
513 R T P L H V C A E K G H S Q V L Q A I Q K G A V G S N Q F V D L E  
2201 GCAACTAACTATGATGGCCTGACTCCCCTTCACTGTGCAGTATAGCCACAATGCTGTGGTCCATGAACCCAGAGAAATCAACAGCCTCATTACCTG  
547 A T N Y D G L T P L H C A V I A H N A V V H E L Q R N Q Q P H S P  
2301 AAGTTCAGGAGCTTTACTGAAGAATAAGAGTCTGTTGATACCATTAAGTGCCATAATCAAATGGGAGCAGCGTGGAAAGCGAAGGATCGCAAAAGTGG  
580 E V Q E L L L K N K S L V D T I K C L I Q M G A A V E A K D R K S G  
2401 CCGCACAGCCCTGATTTGGCAGTGAAGAAGCAAATCTGAACTCATTCCGCTCTTTTGGAGTGGCCAGTTGCCTGTCTTTGTGAATGCAAAGCT  
613 R T A L H L A A E E A N L E L I R L F L E L P S C L S F V N A K A  
2501 TACAATGGCAACTGCCTCCATGTTGCTGCCAGTTCAGTATCGTTGACACAATTAGATGCTGTCCGCTGTTGATGAGGAAGGGAGCAGACCCAA  
647 Y N G N T A L H V A A S L Q Y R L T Q L D A V R L L M R K G A D P  
2601 GTACTCGAACTTGGAGAACGAACAGCCAGTGCATTTGGTCCCGATGGCCCTGTGGGAGAACAGATCCGACGTATCCTGAAGGAAAGTCCATTGAGCA  
680 S T R N L E N E Q P V H L V P D G P V G E Q I R R I L K G K S I Q Q

**NheI (2738)**

2701 GAGAGCTCCACCGTATTAGTCCATTAGCTTGGAGCCTCTAGCTGGCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCA  
713 R A P P Y •  
2801 GTGAAAAAATGCTTTATTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAACAAGTTAACAACAACAATTGCATTCA  
2901 TTTATGTTTCAGGTTTCAGGGGAGGTGTGGGAGGTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTATGGAATTTCAAATACAGCATAGCAAACT  
3001 TTAACCTCAAATCAAGCCTCTACTTGAATCCTTTTCTGAGGGATGAATAAGGCATAGGCATCAGGGGCTGTTGCCAATGTGCATTAGCTGTTGACGCC  
3101 TCACCTTCTTTCATGGAGTTTAAAGATATAGTGTATTTCCCAAGTGTGAACTAGCTCTTCATTTCTTTATGTTTTAAATGCAGTACCTCCACATTC

3201 CTTTTAGTAAAATATTCAGAAATAATTTAAATACATCATTGCAATGAAAATAAATGTTTTTTATTAGGCAGAATCCAGATGCTCAAGGCCCTTCATAAT  
3301 ATCCCCAGTTTAGTAGTTGGACTTAGGGAACAAAGGAACCTTAAATAGAAATTGGACAGCAAGAAAGCGAGCTTCTAGCTTTAGTTCCTGGTGTACTTG  
3401 AGGGGATGAGTTCCTCAATGGTGGTTTTGACCAGCTTGCCATTCATCTCAATGAGCACAAAGCAGTCAGGAGCATAGTCAGAGATGAGCTCTCTGCACA  
134 L P I L E E I T T K V L K G N M E I L V F C D P A Y D S I L E R C M  
3501 TGCCACAGGGGCTGACCACCCTGATGGATCTGTCCACCTCATCAGAGTAGGGGTGCCTGACAGCCACAATGGTGTCAAAGCTTCTGCCCGTTGCTCAC  
101 G C P S V V R I S R D V E D S Y P H R V A V I T D F D K Q G N S V  
3601 AGCAGACCAATGGCAATGGCTTCAGCACAGACAGTGACCCTGCCAATGTAGGCCTCAATGTGGACAGCAGAGATGATCTCCCAGTCTTGGTCTGTAG  
68 A S G I A I A E A C V T V R G I Y A E I H V A S I I E G T K T R I  
3701 GCCGCCCCGACATGGTGTCTGTGTCCTCATAGAGCATGGTGTCTTCTCAGTGGCGACCTCCACCAGCTCCAGATCTGCTGAGAGATGTTGAAGTCT  
34 A A G V H H K N D E Y L M T I K E T A V E V L E L D Q Q S I N F T K  
3801 TCATGGTGGCCCTCTATAGTGAGTCGTATTATACTATGCCGATATACTATGCCGATGATTAATTGTCAA AACAGCGTGGATGGCGTCTCCAGCTTATCT  
1 M  
3901 GACGGTCACTAAACGAGCTCTGCTTATATAGACCTCCACCGTACACGCCTACCGCCATTTGCGTCAATGGGGCGGAGTTGTACGACATTTTGAA  
4001 GTCCCGTTGATTTACTAGTCAAACAAACTCCATTGACGTCAATGGGGTGGAGACTTGAAATCCCCGTGAGTCAAACCGCTATCCACGCCATTGATG  
4101 TACTGCCAAAACCGCATCATCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCATAAGGTCATGTACTGGGCATAATGCC  
4201 AGGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCGATTTACCGTAAATACTCCAC  
4301 CCATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTGACGTCAATGGGCGGGGTCTTGGGCGGTGAGCCAGGCGGG  
4401 CCATTTACCGTAAGTTATGTAACGCTGCAGGTTAATTAAGAACATGTGAGCAAAGGCCAGCAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGC  
4501 GTTTTTCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTT  
4601 CCCCTGGAAGCTCCCTCGTGCCTCTCTGTCCGACCCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGAAGCGTGGCGCTTTCTCATA  
4701 GCTCAGCTGTAGGTATCTCAGTTCGGTGTAGGTGCTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCGTGCCTTATCCGG  
4801 TAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCT  
4901 ACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGAAAAAGAGTTG  
5001 GTAGTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCC  
5101 TTTGATCTTTTACGGGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTTGGTCATGGCTAGTTAATTAACATTTAAATCAGCGGCCGCA  
5201 ATAAAATATCTTTATTTTATTACATCTGTGTGTTGGTTTTTTGTGTGAATCGTAACTAACATACGCTCTCCATCAAACAAACGAAACAAACAAACT  
5301 AGCAAATAGGCTGTCCCAGTGAAGTGCAGGTGCCAGAACATTTCTATCGAA