

Additional data file 2

File format: pdf

Title and description of the data:

Nucleotide sequence alignment between human *NAT8* gene and its duplicated pseudogene *NAT8B*.

The alignment includes the entire gene along with 541 bp of the 5' upstream and 700 bp of 3' downstream regions. The untranslated mRNA and the protein-coding regions are marked with thin and bold lines, respectively. Nucleotides typed in **bold** indicate SNPs detected by resequencing and their position relative to ATG. Divergent positions between *NAT8* and *NAT8B* overlapping with SNPs listed for the human *NAT8* gene in NCBI dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) are boxed.

← Untranslated mRNA

← Protein-coding regions

□ SNPs in dbSNP database that may represent PSVs

-1255 SNP positions detected by resequencing

PCR primer "NAT8 2kb F"

NAT8_29_10_06 CAGAGGAACTGCCCAGCCAAGCTCAGG**C**AAAA**T**TGCTGACT**T**CAGACTCATGAACTAAAT 60
NAT8B_20_10_06 CAGAGGAACTGCCCAGCCAAGCTCAGG**C**AAAA**T**TGCTGACT**T**CAGACTCATGAGCTAAAT 60

NAT8_29_10_06 AAATATTTGTTTCATTTGGTGGTTGTTCTTATGCAACATTATGCTGGCACTAGAAAAATG 120
NAT8B_20_10_06 AAATATTTGATTTTCATTTTATGATTGTTCTTATGCAACATTATCCTGGCACTAGAAAAATG 120

859

NAT8_29_10_06 AAACAGATGGGAATCAA**A**AAGACTGAATGAAGCAGAAACCCAGAAGCTACCCACACAAGCA 180
NAT8B_20_10_06 AAACAGATGGGAATCA**A**AAAAAGACTGAATGAAGCAGAAACCCAGAAGCTACCCACAGGCA 180

NAT8_29_10_06 ATCACATGGGACTCAGATG**T**GAATGGACAATAAGCTTTTATTGCA**TT**SAAAGGTCATTGC 240
NAT8 B20_10_06 ATCGCATGGGACTCAGATG**T**TAATGGACAATAAGCTTTTATTGCA**CA**SAAAGGTCATTGC 240

NAT8_29_10_06 AG**T**G**A**AGGTTGGGGATTGCTTGCTGCTACAG**C**TGAATGGATTCTATTCTGACCAATACA 300
NAT8B 20_10_06 AG**T**C**A**GAGGTTGGGGATTGCTTGCTGCTACAG**C**CGAATGGATTCTATTCTGACCAATACA 300

NAT8_29_10_06 **C**ACAGAAAG**A**GAT**C**A**C**AGACT**T**CC**T**A**C**CT**T**A**A**AGAAGGGAGGT**G**TAGATGAAATGAAC 360
NAT8B 20_10_06 **G**AAGGAAAG**A**AAT**C**ATAGAG**C**CC**T**G**C**CT**G**AC**C**AGAAGGGAGGT**G**ATAGATGAAATGAAC 360

NAT8_29_10_06 TGTATGAAG**G**CCACT**T**AGCCTGGCCACAC**C**AGAAGAAGGACTGGCCGCTTCTTGAA 420
NAT8B 20_10_06 TGTATGAAG**A**TCAC**C**AGCCTGGCCACAC**T**GAAGAAGGACTGGCCGCTTCTTGAA 420

NAT8_29_10_06 GCCCA**T**GCTCTGGTAGAG**G**CCATAGCAGAGAGCTGGAT**G**T**G**CCGGTGTCCAGGATAAC 480
NAT8B 20_10_06 GCCCA**A**GCTCTGGTAGAG**C**CCATGGCAGAGAGCTGGAT**G**T**T**G**C**CGGTGTCCAGGACAAC 480

NAT8_29_10_06 TTCACTGTAGCCCTGGTCCC**G**GGCAA**A**CTGGAGGACAGTCCTGACCAGGGCTTTTGCTAT 540
NAT8B 20_10_06 TTCACTGTAGCCCTAGTCCC**G**GGCAA**A**CTGGAGGACAGTCCTGACCAGGGCTTTTGCTAT 540

428

NAT8_29_10_06 CCCCTGACGACGGTGCTCACTGTCCACA**A**AGAGATGAAACAGCTGCAACCGCTTCTCCCT 600
NAT8B_20_10_06 CCCCTGACCACGGTGCTCATTGTCCACAGAGAGATGAAACAGCTGCAACCGCTTCTCCCT 600

390 **361**

NAT8_29_10_06 CAAGGTGGGATCATCAACAGGCAGAGCTCCTACC**A**TGCCACCACCTTCTCTTCAGACTC 660
 NAT8B_20_10_06 CAAGGTGGGATCATCAACGGGCAGAGCTCCTACTGTGCCACCACCTTCTCTTCAGATT 660

NAT8_29_10_06 AGCCACCCAGAAGCAGGAGCCACGCTCACTCAGGTAGGATTTGGTAATGTCAGACATGTC 720
 NAT8B_20_10_06 ACCCACCCAGAAGCAGGAGCCACACTCACTCAGGTAGGATTTGGTGATGTCAGACATGTC 720
 * *****

218

NAT8_29_10_06 TGTGCACAATGTCATGTCCACATACTCCGTCAGGGTFTTTTGGCAAGGAACCACAGGGC 780
 NAT8B_20_10_06 TGTGCGCAATGTCATGTCTACATAACCGTCAGGGTFTTTTGGCAAGGAACCACAGGGC 780

NAT8_29_10_06 AGGGAAGAGGCTGATGCTGAACACGAGGGCTAGAAGCCAGGATCCAGAGACCAGGAGTAG 840
 NAT8B_20_10_06 AGGAAGGAGGCTGAGGCTGAACACGAGGGCCAGAATCCAGGAGCCAGAGACCAGGAGTAG 840
 *** * *****

NAT8_29_10_06 GGCGAGGGGCCCCCAAGTAAGAGTATGAGGGTTCGAGGCAGCTTCAGCAATTCGCCGAA 900
 NAT8B_20_10_06 GGCAAGGGGCCCCCAAGTAAGAGTATGAGGGTTCGAGGCAGCTTCAGTAATTCGCCGAA 900
 *** **** *****

41

NAT8_29_10_06 GGTGGCTGGGGCATGCTCGGCCATCCCCGGGAGAGCAAGCCACAACCCACTGGCGGT 960
 NAT8B_20_10_06 GGTGGCTGGGGCGTGTTCGCCATCCCCGGGAGAGCAAGCCACGACCTACTTGCGGTC 960

NAT8_29_10_06 GCTCTCCTGGTATTTGCGGATGTGACAAGGAGCCATGGACAGACTTCTGTGTCTGAAATC 1020
 NAT8B_20_10_06 GCTCTCCTGGTATTTGCGGATGTGATAAGGAGCCATGGACAGACTTCTGTGTCTGAAATC 1020

←

NAT8_29_10_06 TCTAAGTCCAGGAGACAGAGCTAGGCCTCCCTGCACGCCTTTACGTCAGAC-TGGGGAAG 1079
 NAT8B_20_10_06 ACTGAGTCCAGGAAACAGAGCTAGGCCTCCCTGCACGCCTTTACGCCAGGCCTGGGGAAG 1080
 ** *****

NAT8_29_10_06 AGAGAGAAAATCATGTGAGTGCCTGCATGGCTCCCAGCCTTGCAGGAACAGGGAGACCTG 1139
 NAT8B_20_10_06 AGAGAGGAAACCACGTGAGTGCCTGCATGGCTCCCAGCCTTGCAGGAACAGGGAGACCTG 1140
 ***** ** *

NAT8_29_10_06 CTCAAGGGTGTGTCTTTCCGCGTTTGCCCATGAGGAAGCAGGCCTCTGCCACTGGGAGAA 1199
 NAT8B_20_10_06 CTTAAGGGTGTGTCTTTCCGCGTTTGCCCATGAGGAAGCAGGCCTCTGCCACTGGGAGAA 1200
 ** *****

-222

NAT8_29_10_06 GGTAGGGTCAGAAAGACCTGCATTGGAATCTGGTTCATCCCTTTCTAGCTGTGTGACCT 1259
 NAT8B_20_10_06 GGTAGCATGAGAAAGACCTGCATTGGAATCTGGTTCATCCCTTTCTAGCTGTGTGACCT 1260
 ***** * *****

NAT8_29_10_06 TAGGCATGTCACTTAACCTCTCTGAGCCTCCATTTCTTCAACTATAGAACAGGGGAGAAG 1319
 NAT8B_20_10_06 TGGGCATGTCACTTAACCTCTCTGAGCCTCCACTTCTTCACTATAGAACAGGGGAGAAG 1320
 * *****

-365

NAT8_29_10_06 AAAACCTGCCTTCTAGAGTTGTTTTGAAGATTGAATAAGATGCCTTCTATAAAGCCCTGG 1379
 NAT8B_20_10_06 AAAACCTGCCTTCTAGAGTTGTTTTGAAGATTGAATAAAAATACCTTCTATAAAGCCCTGA 1380

NAT8_29_10_06 TAGATGAACTACCCACCGCTCTTCCCTTTTTCTGTTTTCTTTTTCTGGCTCTCTCAGCTTC 1439
 NAT8B_20_10_06 TAGATGAACTACCCACCACTCTCC---CTCTGTTTTCTTTTTCTGGCTCTCTCAGCTTC 1437

NAT8_29_10_06 TCCGTGCATTGAGGGGGCACCTACCTCAGGCAGGCCACACGTGCAACCCTTTTGTCTTC 1499
 NAT8B_20_10_06 TTGGTGCATTGAGAGAGTACCTACCTCAGGCAGGCCACATGCCCAACCCTTTGGTGTTC 1497
 * * * * *

NAT8_29_10_06 TCCAGGGACCATAACATCGCAAAAATTCTCTGTCCAA-CACAAGTTTCAGGAAAGGCTCCC 1558
 NAT8B_20_10_06 TCCAGGGACCATAACATCGCAGAAATCCTGTGACCAAGCACAAAGTTTCAGGAAAGGCTCCC 1557
 * * * * *

NAT8_29_10_06 TCCAGCTTTATTGTAAGTTTTTAGGTTGCTGAAGTGGCATAAGAGCCAAGCATAAAAAAC 1618
 NAT8B_20_10_06 TCCAGCTTTATTGTAAGTTTTTAGGTTGCTGAAGTGGCATAAGTGCCAAGCATAAAAAAC 1617
 * * * * *

NAT8_29_10_06 AGCCCAGCCTCCAGATCCCAAGGAAGGTGCATGCTCAGAGCACTGTCTTCCACCCAGTC 1678
 NAT8B_20_10_06 AGCCCAGCCTCCAGATCCCAAGGAAGGTGCATGCTCAGAGCACTGTCTTCCACCCAGTC 1677
 * * * * *

NAT8_29_10_06 CCCTCACCTGTCACCACAAGAGCCTTGAGTGTCCAGGGAGCTTCAGCTCTCAGCCGCTCA 1738
 NAT8B_20_10_06 CCCTCACCTGTCACCACAAGAGCCTTGAGTGTCCAGGGAGCTTCAGCTCTCAGCCGCTCA 1737
 * * * * *

-763

NAT8_29_10_06 CTGGCATCCAGGAGATACCGTCTGGGGCTTGCTCACCCCTGGAAAGTAGACTGCCTGCCT 1798
 NAT8B_20_10_06 CTGGCATCCAGGAGATACCGTCTGGGGCTTGCTCACCCCTGGAAAGCAGGCTGCCTGCCT 1797
 * * * * *

-803

NAT8_29_10_06 TATTGCTTGGCTCCTGAATGTTTGGTAATCATCAACCCTAGGGTCAAAGAGCTGGGCTGT 1858
 NAT8B_20_10_06 CATTGCTTGGCTCCTGAATGTTTGGTAATCATCAACCCTAGGGTCAAAGAGCTGGGCTGT 1857
 * * * * *

-911

NAT8_29_10_06 AACACAGCAGGGAGCCCAGTGTGCATTCCTCTTTGTGGTCAAAGGGAA**CG**ACCTGGAGGG 1918
 NAT8B_20_10_06 AACACAGCAGGGAGCCCAGTGTGCATTCCTCTTTATAGTCAAAGGGAATGGCCTGGAGGG 1917
 * * * * *

NAT8_29_10_06 GGAGACTCAGGGTCTGGAACCGCCTGGGTTCCCTTGCCCTTCTGGAAGGCCTGCCACCTCT 1978
 NAT8B_20_10_06 GGAAACTCAGGGTCTGGAACCGCCTGGGTTCCCTTGCCCTTCTGGAAGGCCTGCCACCTCT 1977
 * * * * *

-1041

NAT8_29_10_06 CTTGCCTGTT**C**AGGCCCGTGCATCTCTTGCCACCTGGCACACTTAGTACTCTCGGT**CC** 2038
 NAT8B_20_10_06 CTTGCCTGTT**C**AGGCCCTGGGCATCTTTCTGCCACCTGGCACACTTGATACTCTC----- 2032
 * * * * *

NAT8_29_10_06 CTGTCCACATATACGATCCACCCTGGTCCCTTGTACACATATACGACGAGCCTGAGAACTG 2098
 NAT8B_20_10_06 -----TGAGATCCACCTTGGTCCCTGTACACATATACCACGAGCCTGAGAACTG 2081
 * * * * *

-1152/-1153

NAT8_29_10_06 TACCATT**C**ACCC**TG**CAATCAGTCACCATTGGCCAGAAGCCGAGTCCCC**TG**CCTGGCAAC 2158
 NAT8B_20_10_06 TGCCATT**C**ACCC**TG**CACTCAGTCACCATTGGCCAGAAGCCGAGTCCCC**C**ACTTGGCAAC 2141
 * * * * *

NAT8_29_10_06 AAGGCCCTTCC**C**AGCTAGGCTTCTCCTCGCTGAGGTTTCCCTGGTCCACGAAGCATCTGT 2218
 NAT8B_20_10_06 AAGGCCCTTCC**C**AGCTAGGCTTCTCCATGCTGAGGTTTCCCTGGTCCACAAAGCATCGAT 2201
 * * * * *

-1255

NAT8_29_10_06 TGTTGGCAGTTCTTGCACCTCACCAACCCCCATCTCCACCCATTCATTGATTCAG 2278
 NAT8B_20_10_06 TGTTGGCAGTTCTTGCACCTCACCAACCCCCACCTCCACCCATTCATTGATTCAG 2261

NAT8_29_10_06 TCAACAGTTGTGTGTCATGCACTTCCTGTGTCTGGGGTCATTTTCAGCCATGAGCAAGGT 2338
 NAT8B_20_10_06 TCAACAGGTGTGTGCGATGCACTTCCCATGTCTGGGGTCATTTTCAGCCATGAGCAAGGT 2321

PCR primer "NAT8 2kb R"

NAT8_29_10_06 TGACATTCGGCTTGACAGGCATGGCTGTGTGTGGCTTGTGTGTGTTCTGGCTGTTGGTCA 2398
 NAT8B_20_10_06 TGACATTCGGCT-GACAGGCATGGCTGTGTGTGGCTTGTGTGTGTTCTGGCTGTTGGTCA 2380

NAT8_29_10_06 GAGGCTGTGCCACACCCGTTGTAAATATTTTGCACCTGCAGATGTTGGGGAGGCCCAAAT 2458
 NAT8B_20_10_06 GAGGCTGTGCCACACCCGTTGTAAATATTTTGCACCTGCAGATGTTGGAGAGGCCCAAAT 2440

NAT8_29_10_06 TCATGCTCTT
 NAT8B_20_10_06 TCAAGCTCTT
 *** *****