POU2F1
SEQUENCING OF A CANDIDATE GENE FOR HEREDITARY HEARING IMPAIRMENT

By Stud. Med. Garth Tylden
Supervised by Toril Fagerheim

A 5th year research thesis
Faculty of Medicine, University of Tromsø
Written at the Department of Medical Genetics, University Hospital of North Norway/
Institute of Clinical Medicine, University of Tromsø
June - September 2003

Acknowledgements

This research thesis was executed at the Department of Medical Genetics at the University Hospital of Northern Norway (UNN) during the spring and summer of 2003.

Firstly, I would like to thank my supervisor Toril Fagerheim for her informal, enthusiastic guidance and her ability to not appear condescending in the face of stupid questions.

A special thanks to Bo Dreyer for his invaluable contributions, to Øivind Nilssen, Vigdis Brox and Helle Klenow for their help and insights and to Mona Nystad, Bente Guntvedt and Christoffer Jonsrud for giving me a glimpse of other aspects of every day life at DMG UNN.

Finally, I would like to thank all the members of the department for generating the positive atmosphere that made this undertaking a pleasurable learning experience.

Glossary

Caloric testing	A test for vestibular function, made by irrigating the external auditory meatus with either hot or cold water; this normally causes stimulation of the vestibular apparatus, resulting in nystagmus: in vestibular disease, the response may be reduced or absent.
cDNA (complementary DNA)	DNA that is synthesized *in vitro* from a messenger RNA template that corresponds to expressed sequences of genomic DNA. Viral reverse transcriptase is used to synthesise cDNA.
cM (centimorgan)	A unit of measure of recombination frequency. One centimorgan is equal to a 1% chance that a marker at one genetic locus will be separated from a marker at a second locus due to crossing over in a single generation. In human beings, 1 centimorgan is equivalent, on average, to 1million base pairs.
DFNA7	Autosomal dominant non-syndromic deafness locus number 7. The letter A indicates dominant pattern of inheritance. Recessive loci are denoted *DFNB*.
Exon	A coding sequence of DNA within a gene, which survives the processing of RNA in the cell nucleus to become part of a spliced messenger RNA in the cytoplasm. In the primary RNA transcript neighbouring exons are separated by introns.
Intron	A noncoding sequence of DNA within a gene, that is transcribed into the primary RNA transcript but is then cut out by RNA splicing, leaving a mature mRNA that is then translated in the cytoplasm. Introns are poorly conserved and of variable length, but the regions at the ends are self complementary, allowing a hairpin structure to form naturally in the hnRNA, this is the cue for removal by RNA splicing. Introns are thought to play an important role in allowing rapid evolution of proteins by exon shuffling.
Locus	The site in a linkage map or on a chromosome where the gene/marker is located. Any one of the alleles of a gene may be present at this site.
Lod score	The total relative probability, expressed on a logarithmic scale, that a linkage relationship exists among selected loci. Lod is an acronym for "logarithmic odds."
Microsatellite markers	Small repetitive DNA sequences dispersed in genomes. The lengths of these elements are hyper-variable and thus are highly polymorphic, making them ideal in genomic fingerprinting applications
PCR (polymerase chain reaction)	The first practical system for *in vitro* amplification of DNA and as such one of the most important recent developments in molecular biology. Two synthetic oligonucleotide primers, which are complementary to two regions of the target DNA (one for each strand) to be amplified, are added to the target DNA (that need not be pure), in the presence of excess deoxynucleotides and Taq polymerase, a heat stable DNA polymerase. In a series (typically 30) of temperature cycles, the target DNA is repeatedly denatured (around 90°C), annealed to the primers (typically at 50-60°C) and a daughter strand extended from the primers (72°C). As the daughter strands themselves act as templates for subsequent cycles, DNA fragments matching both primers are amplified exponentially, rather than linearly. The original DNA need thus be neither pure nor abundant and the polymerase chain reaction has accordingly become widely used not only in research, but in clinical diagnostics and forensic science.
Presbyacusis	Loss of ability to perceive or discriminate sounds as a part of the aging process.
Primer	Short pre-existing polynucleotide chain to which new deoxyribonucleotides can be added by DNA polymerase.
SSCP analysis	Single stranded conformational polymorphism analysis Technique for detecting point mutations in genes by amplifying a region of genomic DNA (using asymmetric PCR) and running the resulting product on a high quality gel. Single base substitutions can alter the secondary structure of the fragment in the gel, producing a visible shift in its mobility.
Transcript	Term used to refer to the various segments of messenger RNA (mRNA) that result from transcription of a gene.
Transcription factor	Protein required by RNA polymerase for activation of transcription of eukaryotic genes. Binds to promoter region upstream from the activated gene.

Summary

Hearing impairment (HI) is the most common sensory debility. In developed countries prelingual HI affects about 1 in 1000 children. Genetics plays an important role in congenital HI and doubtlessly also in presbyacusis. The HI locus *DFNA7* was mapped by Fagerheim et al in a large Norwegian family where affected individuals exhibit non-syndromic autosomal dominant post lingual progressive high-tone hearing loss. The locus spans a 22 cM region between the microsatellite markers D1S104 and D1S466 on chromosome 1 (1q21-q23). Within this region, only 0.8-cM from the D1S196 marker which gave the highest lod-score, lies the *POU2F1* gene. *POU2F1* encodes a ubiquitously expressed transcription factor and is one of several candidate genes in the *DFNA7* locus. *POU2F1* belongs to the POU domain family of transcription factors, 2 of which have earlier been shown to cause nonsyndromic HI. In this study, the genomic structure of *POU2F1* was deduced using different genomic databases. The gene consisted of 21 exons which, according to sequence information listed at ENSEMBLE, gave rise to 5 transcripts varying from 2-17 exons in length. The purpose of the study was to identify potential mutations, so although some exons were less well documented than others, all 21 exons were sequenced. The template DNA was the genomic DNA of 2 affected and 2 unaffected members of the Norwegian family mentioned above. Additionally *POU2F1* cDNA from a healthy donor was sequenced to control the consensus sequence data.

There were no apparent differences between affected and unaffected subjects. However, one difference between the subject DNA and the consensus sequence was found, namely a single G>C substitution at position 143 in exon 2. The substitution was shared by all 4 subjects and thus probably represents a polymorphism. Unfortunately, no cDNA sequence data was obtained covering exon 2 for comparison. No information regarding single nucleotide polymorphisms in this exon was available at the NCBI database and as such, we have no inkling of the frequency of this polymorphism.

cDNA sequences were resolved covering exons 1 and 6-21, evidence was also found suggesting the existence of transcripts containing exons 2 and 3. All of the cDNA sequences that were resolved matched the consensus and subject sequences.

The candidacy of *POU2F1* as the antecedent mutation of *DFNA7* is weakened. However, rejection at this stage is precluded by methodical shortcomings and uncertainty about the existence of additional exons.

Introduction

Incidence of hearing loss

Hearing impairment (HI) is a common affliction. In developed countries prelingual HI affects about 1 in 1000 children[1,6]. An additional 1 in 1000 children develop a post lingual hearing deficit. Incidence of HI increases with increasing age. A handicapping loss of hearing is manifested in 3 in 1000 individuals between the ages of 30 and 50 years, rising to 23 in 1000 of those between 60 and 70 years. An estimated 50 % of octogenarians have a handicapping loss of hearing[1,5].

Genetics of hearing impairment

Genetics plays an important role in congenital HI and doubtlessly also in presbyacusis. About half of all cases of congenital HI are caused by genetic defects[2]. The majority (70%) of these mutations affect only the auditory system (nonsyndromic) making early diagnosis difficult. In the remaining 30% HI is one of a number of clinical features present in affected individuals (syndromic). Nonsyndromic HI is further divided into autosomal dominant (20%) and recessive (80%) types (ADD and ARD), X-linked HI (~1%) and HI linked to loci in mitochondrial DNA (< 1 %).

Syndromic HI is conductive in most cases[4]. ARD is almost always sensorineural and severe. ADD is usually post-lingual and progressive consistent with hearing loss due to accumulation of pathology.

A large number of genes can give rise to non-syndromic HI. To date more than 70 loci have been mapped and 25 nuclear genes identified. Of the identified genes, 16 give rise to ADD[3].

In addition to being a common aetiology of congenital HI and early onset progressive hearing loss, mutations in genes will no doubt be found to play an important role in presbyacusis.

Non-syndromic autosomal dominant deafness locus number 7 (*DFNA7*)

DFNA7 was discovered in a large Norwegian family where affected individuals exhibit non-syndromic autosomal dominant post lingual progressive high-tone hearing loss. Debut age and progression vary but hearing loss does not set in until after age 4-5 years. In the majority of affected family members hearing loss is greater than 45 dB by age 15 years. However the highest frequencies are affected first and individuals retain acceptable social hearing and normal lingual skills. *DFNA7* is thus a mild form of HI.

DFNA7 was mapped to 1q21-q23 on chromosome 1 by linkage analysis. It is a large locus. Bordered by the microsatellite markers D1S104 at the centromeric end and D1S466 at the telomeric end, it stretches 22 cM in length and contains over 100 genes. No other reports of families with linkage to this region have been published since its discovery and there has been no progress in refining the genetic region. Because of the large size of the region studies have focused on identification and testing of candidate genes, of which there are several[8,9]. So far myelin protein zero (MPZ or P0), Cx40 and Cx50 have been tested by SSCP analysis and/or sequencing.

POU domain class 2 transcription factor 1; *POU2F1*

POU2F1 (alternatively *OTF-1* and *OCT-1*) is one of several candidate genes located in 1q21-q23. The gene spans ~ 200 kb of the genome and probably contains more than 21 exons[14,15] though currently confirmed transcripts contain 16 exons or fewer. It encodes a helix-loop-helix transcription factor which is a member of the POU-domain family of proteins. The POU-domain is a structural motif containing two DNA-binding domains, a 75 amino acid POU-specific domain and a 60 amino acid POU-homeodomain separated by a short 20 amino acid spacer. The POU-homeodomain is related to the homeodomains of many other transcription factors. Members of the POU-domain family are important during development of the nervous system[12].

POU2F1 is an attractive candidate gene for *DFNA7* for a few reasons:

1. Similar genes involved in earlier characterised deafness loci with similar clinical manifestations.
2. Close proximity of *POU2F1* to the micro satellite marker D1S196 which gave the highest lod-score.
3. *POU2F1* is expressed in rat cochlea during embryogenesis[12].

Mutant forms of two genes related to *POU2F1*, namely *POU3F4* and *POU4F3* have been shown to cause progressive sensorineural hearing loss. The *POU3F4* gene is involved in X-linked progressive sensorineural HI combined with stapes fixation[10] and *POU4F3* is involved in autosomal dominant progressive sensorineural hearing loss[11]. The *POU2F1* gene is located on chromosome 1 only 0.8-cM from the D1S196 marker which gave the highest lod-score with *DFNA7*[8]. The *POU2F1* gene has been reported to be expressed in the rat cochlea during embryogenesis, consistent with its contribution to inner ear development[7].

The protein product of *POU2F1*, OCT-1 is ubiquitously expressed. The fact that OCT-1 is expressed in all eukaryotic cells may seem to undermine its candidacy as a genetic determinant of disease in such specialised tissues as those found in the inner ear. However there are potentially many ways in which a mutation in OCT-1 might lead to disease. OCT-1 may participate in tissue specific transcription via direct protein-protein interactions and protein-protein interactions mediated by the various OCT-1 DNA binding sites. Also, splice variants and tissue specific isoforms of OCT-1 (and mouse homologues) have been characterised[13,14,15]. The occurrence of a mutation affecting only interactions that take place in the inner ear is a plausible explanation as is mutation within a tissue-specific exon of an alternatively spliced isoform of Oct-1. The inner ear is an immensely complex organ and even slight aberrations of expression at key points during embryogenesis could lead to disease.

Material and methods

Clinical material

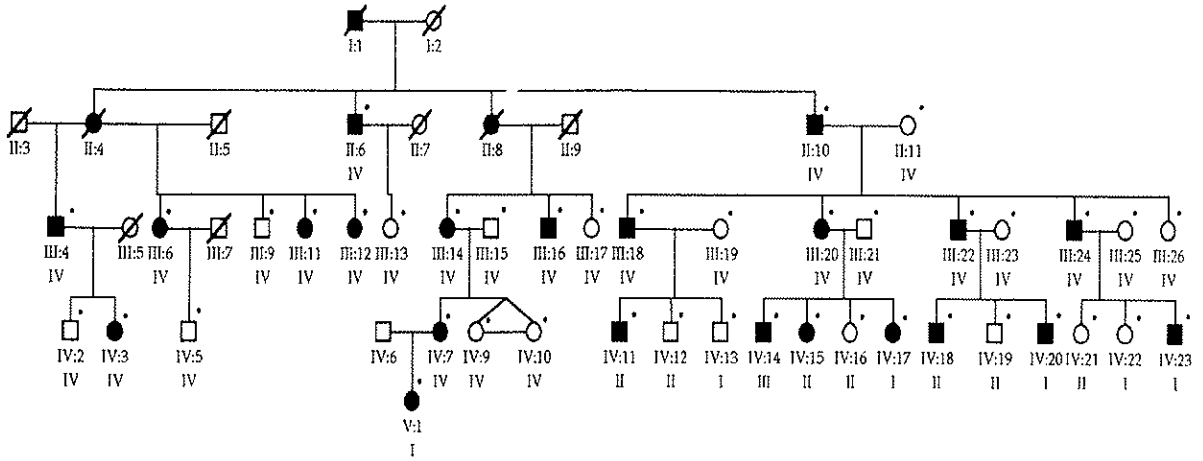


Figure 1: The pedigree for the Norwegian deafness family. The 42 family members included in the study carried out by Fagerheim et al[8] are indicated with a dot and have their corresponding liability class number below their pedigree number. Liability class number I (10-15 years), II (16-20 years), III (21-25 years) and IV (>25 years) have the penetrance values of 0.5, 0.75, 0.90 and 0.99 respectively.

Subjects III:17, III:21, III:18 and III:24 were selected and assigned numbers 1 – 4 respectively. As indicated in figure 1, subjects 1 and 2 exhibited symptoms, whilst subjects 3 and 4 did not.

Acquisition of the consensus sequence and primer design

The consensus sequence, the exon boundaries and the flanking intron sequences of *POU2F1* were obtained by searching the nucleotide databases GenBank at NCBI (http://www.ncbi.nlm.nih.gov) and ENSEMBL (http://www.ensembl.org/) with the search terms *POU2F1* and OCT-1. "Blast 2 sequences" (http://www.ncbi.nlm.nih.gov/gorf/bl2.html) was used to compare exons of the new transcripts with those of the original transcript (see results).

All primers were designed using the web-based primer design program Primer 3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). The same primers were used for both PCR and sequencing reactions.

Exonal and intronal primers.

Two sequencing goals demanded two types of primers. Intronal primers (see table 1 for primer sequences) were required to obtain full length sequences of the individual exons from the genomic DNA of the patients. To ensure optimal resolution of the target exon and flanking intron bases, primers were designed to amplify the target exon and a "buffer region" of at least 40 base pairs either side of the exon.

Exonal primers (table 2) on the other hand were used to sequence *POU2F1* cDNA from a healthy subject in order to corroborate the sequence information available at the databases. After attempts at amplifying entire cDNAs failed, we chose to sequence the cDNA in a series of overlapping fragments. Because of uncertainties concerning the initial exons the first primer set had 5 forward primers. The primer sets are colour coded in table 3.

Before commencing sequencing of the subject DNA, the first batch of intronal primers was tested on genomic DNA isolated from an arbitrary whole blood sample from the blood bank at the University Hospital of Tromsø. Testing of the second batch of primers (for exons 1-3 and 7) and the exonal primers was deemed unnecessary because of the good results obtained with the first batch.

Mixing of reagents and thermal cycling parameters were identical to those used in PCR (se tables 4 and 5).

Table 1: Primers used for PCR and sequencing of genomic DNA from subjects. To provide optimal templates for sequencing, primers were designed to amplify the target exon and at least 40 intronal base pairs either side of the exon.

Name	Sequence (5'-3')	Size of PCR prod.
POU2-EX1-F1	AGAGGAGGGAGGAGGCAAG	277
POU2-EX1-R1	CATCTCCCCGCCGTTAAT	
POU2-EX2-F1	GGGCATGACTCAAAGAGGTC	399
POU2-EX2-R1	GCATACCAAATGGGCTTCAT	
POU2-EX3-F1	CCCCCACTGTGTGTAATTGTT	250
POU2-EX3-R1	CACCTGGATTTATCAGATGCTAC	
POU2-EX4-F1	TGTGTGATGGGTTTGTTGGA	364
POU2-EX4-R1	ACATCCAGCCTGGTCATGTT	
POU2-EX5-F1	TTGCTTTTCCCCTTAATCCA	263
POU2-EX5-R1	GCATAGTCAACCCGACACAC	
POU2-EX6-F1	CAGTTTTGCCTCCTCAATCC	207
POU2-EX6-R1	GTTACATGGAAACTTTCTTCTTTGTG	
POU2-EX7-F1	GCAGATGCAATCGGGTTTAT	490
POU2-EX7-R1	TTGACCAGAAGCATTGCTGA	
POU2-EX8-F1	GGTGCCTGATGAATGTTGGT	254
POU2-EX8-R1	TCAGGCCCCAGTACTTTACG	
POU2-EX9-F1	GGGTAGGGTGGGTTGAGAAG	446
POU2-EX9-R1	TTCCTGTCCATTTTTGGTTCA	
POU2-EX10-F1	TGATTCCTAAAAGATGGGGTTTT	260
POU2-EX10-R1	CCAATGGCAGTTCTGACACA	
POU2-EX11-F1	AGCAGTTGGCTTGCATTAGG	357
POU2-EX11-R1	AGGCCCCACACTAACAATCA	
POU2-EX12-F1	CAGAAGTCATCAGCTGGAAGC	307
POU2-EX12-R1	AATAGTTTCAGGGCCAAAATGA	
POU2-EX13-F1	TTTCTTTTGCCATGTGTTCG	218
POU2-EX13-R1	CCAAAAGTCTCCCCAAACAA	
POU2-EX14-F1	TCCATAAATGTGGCTCTTTCC	353
POU2-EX14-R1	TGTGGAATCCATGAATACAGAGAG	
POU2-EX15-F1	AGGCTTTAAGCACTGGTGAGA	280
POU2-EX15-R1	CCAATTTATATAACAATCCCCATGT	
POU2-EX16-F1	CTTGGATGTAGCTTATTTTGTCAGTT	297
POU2-EX16-R1	TTAAGTACTACACAGACAATCCCATGT	
POU2-EX17-F1	CCTGACGTGATTATGCCAGTAG	308
POU2-EX17-R1	AAATTGCAAAAATTAGCCCTTG	
POU2-EX18-F1	CGAATAGCAGCTTTCCAACA	336
POU2-EX18-R1	TTAGTTAGGGCCTTAATTCACAATTT	
POU2-EX19-F1	AAAGCCACCATTCTTCCAAA	455
POU2-EX19-R1	GGATGACCTCCACCTCAGAC	
POU2-EX20-F1	AAAATGAGACCTGCGTCTGC	230
POU2-EX20-R1	TCTGTCTCAAGCACACACACA	
POU2-EX21-F1	CCTGGTGGGTTGTAGGAAAA	479
POU2-EX21-R1	GCTGGCAGTCCAATCACAC	

Table 2: Primers for amplification and sequencing of *POU2F1* cDNA.

Name	Sequence (5'-3')
IE-EX1-F1	GAGGAGCAGCGAGTCAAGAT
IE-EX2-F1	CAGTGATGCCAGCAAAATGT
IE-EX3-F1	AGAAGAGCTTTCCTGCCTTTTT
IE-EX3-R1	GACTGCATTCTTTCTTTGAACTGA
IE-EX4-F1	GGTTGGAAAGGATATTGGTGTACT
IE-EX5-F1	CCACCCCAAACTGCTACCT
IE-EX6-F1	CCGTCAGAAACCAGTAAACCA
IE-EX6-R1	TTGCCATCTCCACTCTCCAT
IE-EX11-F1	ACTACTCCAGCAGGCACAGG
IE-EX11-R1	GCGATCTGTATGGGCTGAGA
IE-EX16-F1	TCTGATTCGTCCCTCTCCAG
IE-EX16-R1	CTCTAAGGCCACACGGATGT
IE-EX19-F1	CACCACCTCCAACAACACAG
IE-EX19-R1	CAGGCTTGGGTTTAGTCCTG
IE-EX21-R1	CGAGAGGAAGCCAATCACAT

Table 3: Location and orientation of the exonal primers for cDNA sequencing with expected product sizes and resulting fragmnets. Primer sets and fragments are colour coded. Sequences of primers are presented in table 2.

Exon	E-primers	Prod. size*	Resulting fragments
1	↓↓**	140/112	
2	↓	247	***
3	↑↓	150	*** A
4	↓	154	***
5	↓	156	***
6	↓↑	509	
7			
8			B
9			
10			
11	↑↓	823	
12			
13			C
14			
15			
16	↓↑	735	
17			D
18			
19	↑↓	836	
20			E
21	↑		

*Product sizes are affixed to the adjacent forward primer and denote the length in base pairs between the forward and reverse primers. Absence of intervening exons was assumed during calculation for the first 5 forward primers. These figures represent therefore minimum product size.

**The two arrows here represent one forward primer which is compatible with two reverse primers.

***Intervening exons did not appear in cDNA fragment A.

Template acquisition

Isolation of genomic DNA

DNA was extracted from EDTA blood using the standard protocol for the Applied Biosystems GENEPURE 341 Nucleic Acid Purification System.

Synthesis of cDNA

RNA from a healthy donor was isolated from whole blood using the PAXgene Blood RNA System from PreAnalytiX and cDNA was synthesised using the SuperScript™ First-Strand Synthesis System for RT-PCR from Invitrogen. The manufacturer's protocols were followed without deviation. Briefly outlined, stabilised mRNA was reverse transcribed using the retrovirus protein reverse transcriptase. Oligo dT primers targeting the poly A tails of the mRNA served as starting points for reverse transcription. The RNA template was subsequently digested with RNase leaving single stranded cDNA ready for PCR.

Automatic sequencing

PCR

Prior to sequencing the target DNA was amplified by polymerase chain reaction (PCR). A touchdown-PCR protocol (touchdown refers to diminishing annealing temperature) was used in order to circumvent non-specific priming, thereby increasing the efficiency of PCR. Reagents were mixed in polypropylene PCR tubes as indicated in table 4 and subsequently thermally cycled in a MBS 0.2G/S thermocycler from Hybaid. The cycling parameters are shown in table 5.

Table 4: Mixing of reagents for PCR.

Reagent	Concentration		Pr. reaction	
Primer F	20	Pmol/µl	0.25	µl
Primer R	20	Pmol/µl	0.25	µl
RedTaq* PCR mix.	2.5	U/µl	7.5	µl
Template	5	ng/µl	2	µl
Water			6	µl
Total			15	µl

* From Sigma

Table 5: Cycling parameters for PCR. Extension times were lengthened to 50 seconds during amplification of cDNA fragments.

Phase	Cycles	Temp. (°C)	Time
Initial denature	1	94	8 min
Denature		95	20 sec
Anneal	2	63	20 sec
Extend		72	20 sec
Denature		95	20 sec
Anneal	2	61	20 sec
Extend		72	20 sec
Denature		95	20 sec
Anneal	2	59	20 sec
Extend		72	20 sec
Denature		95	20 sec
Anneal	30	57	20 sec
Extend		72	20 sec
Cool	1	4	∞

Electrophoresis

To ensure that PCR produced the desired product, samples were run on agarose gels with a DNA size marker. The band pattern produced was correlated with the expected product sizes (see tables 1 and 3 and figures 4-6).

A suitably sized plastic casting mould and well comb were selected and levelled. Two hundred ml 1 X TBE buffer was added to 2.0 g agarose in a flask. The mixture was brought to boil in a microwave oven and then cooled to ~ 55°C. The gel mixture was then poured into the casting mould so that the depth of the liquid was ~ 5-6 mm. The gel was left to polymerise at room temperature for 20 minutes and the remaining gel mixture was stored at 55°C for later use. The stiffened gel was placed in an electrophoresis chamber containing 1 x TBE buffer with the wells positioned adjacent to the negative electrode (DNA migrates towards the positive electrode). The level of the electrophoresis buffer was adjusted so that the gel was covered. DNA samples (already containing RedTaq loading buffer) were applied to the wells along with the DNA size marker (1 kb plus, figure 2). The electrophoresis was run at 90 V for ~45 minutes. Following electrophoresis the gel was soaked in ethidium bromide solution (10 mg/ml). Bands were visualised and photographed using a Dual Intensity Transilluminator (ULTRAVIOLET PRODUCTS) UVP Imagestore 5000 (SONY). (UV light at 302 nm).

1X TBE buffer
0.8 M Tris base
0.1 M NaOAc
20 mM EDTA
pH adjusted to 8.0 with acetic acid

Figure 2: Ready-Load™ "1 Kb Plus" DNA ladder.

Cleaning the PCR product

In order to remove primers and inactivate (dephosphorylate) residual free nucleotides, which would otherwise disturb the sequencing reaction, the PCR product was treated with the two enzymes exonuclease I and shrimp alkaline phosphatase (EXO-SAP). To the PCR product, of which ~10 µl remained, 1 µl of a pool containing 1 U SAP and 5 U EXO/µl was added. The reaction was then incubated at 37°C for 1 hour and subsequently deactivated at 85°C for 15 minutes.

Sequencing

The reagents were mixed as shown in table 6 and the tubes were subjected to temperature cycling as shown in table 7.

Table 6: Mixing of reagents for sequencing PCR

Component	Concentration	Volume (µl)
DNA template (PCR product)	~50 ng/µl	3
Big Dye™ terminator RR mix*	10 ×	2
Big Dye™ sequencing buffer*	6.67 ×	3
Primer (either 3' or 5')	3 pmol/µl	1
MilliQ water		11
Total		20

*PE Applied Biosystems

Table 7: Cycling parameters for sequencing PCR

Phase	No. cycles	Temperature (°C)	Time
Denature		95	30 sec
Anneal	30	50	5 sec
Extend		60	4 min
Cool	1	4	∞

Precipitation of the sequencing product

The automatic sequencing team at UiTø requires that the sequencing product be precipitated and dried before delivery.

The sequencing PCR product (20 µl) was added to a mixture of 2 µl 3M sodium acetate (pH 4.6) and 50 µl 96 % ethanol in a micro-centrifuge tube. The mixture was incubated at room temperature for 20 minutes and then centrifuged for 15 minutes at 13000 rpm in a Biofuge 13 (HERAEUS, SEPATECH). The supernatant was removed by suction and the pellet washed with 250 µl 70% ethanol, vortexing briefly. The purified DNA was then centrifuged at 13000 rpm for 5 minutes, again removing the supernatant by suction. Lastly the pellet was air dried for 20 minutes and stored at -20°C prior to automatic sequencing.

Sequencing electrophoresis

The sequencing reaction products were run on an ABI 377 Automatic DNA sequencer.

Sequence analysis

Sequencing 21 exons from 4 patients in both directions generated a large amount of data. The sequence data was analysed using the programs Sequence Analysis 3.7 and Seqscape. A consensus sequence consisting of 21 exons was created with which the sequence information from the cDNA donor and the 4 patients was aligned. The alignments were studied both automatically and manually, comparing with the chromatogram alignments when necessary.

Results and discussion

Acquisition of the consensus sequence and primer design

Searches of the database at ENSEMBL at two different stages (2 months apart) gave differing results. Initially a gene with a single transcript was obtained. Upon searching a second time the gene had 4 transcripts which contained 4 additional exons while 2 were omitted. We chose to include all of the exons found in both searches in the study. The 21 exons with ENSEMBL IDs and sizes are presented in table 8.

The reason for the differing search results lies in the nature of the ENSEMBL system and unfinished human genome project. ENSEMBL is an automated system for tracking, assembling and analysing sequenced fragments of the human genome (as well as other species). The database is constantly evolving due to the ever increasing flood of new sequence fragments and activity was intense in the months leading up to April 2003 at which point 99 % of the human genome project was expected to be finished.

Table 8: Comparison of the 5 ENSEMBL transcripts of pou2f1. Exons are numbered 1-21 in order of occurrence in genomic DNA. Initial 9 characters of the ENSEMBL exon IDs (ENSE00000) are omitted to save space. Sets of exons with matching DNA sequences are aligned along the same row (note that one exon may have more than one ID). Exons with no matching counterparts are highlighted in blue.

No.	Orig. ENST-00000308928	ENST00000-271411	ENST00000-308928	ENST00000-308943	ENST00000-318156	Exon size (bp)
1		1218788	1200261	1218961	1200261	51/61/64
2			1218885		1218794	205/204
3				1200259		101
4	1200252					98
5	1200238					137
6	0958560	1218925	0958560		1218925	66/67
7			1218948			290
8	0958561	0958561	0958561		0958561	101
9	0958562	0958562	0958562		0958562	54
10	0958563	0958563	0958563		0958563	120
11	0958564	0958564	0958564		0958564	189
12	0958565	0958565	0958565		0958565	127
13	0958566	0958566	0958566		0958566	95
14	1200208	0958567	0958567		0958567	174
15	1200205	0958568	0958568		0958568	142
16	1200204	0958569	0958569		0958569	140
17	1200201	0958570	0958570		0958570	180
18	1200248	1157574	1157574		1157574	106
19	1157566	1157566	1157566		1157566	346
20	0958573	0958573	0958573		0958573	89
21	1218941	0958574*	1218941		1218941	387/604

*0958574 is the same as 1218941 but includes a 3'untranslated region.

Testing of primers

The first batch of intronal primers (for exons 4-6 and 8-21) gave good results immediately without need for optimisation (see figure 3).

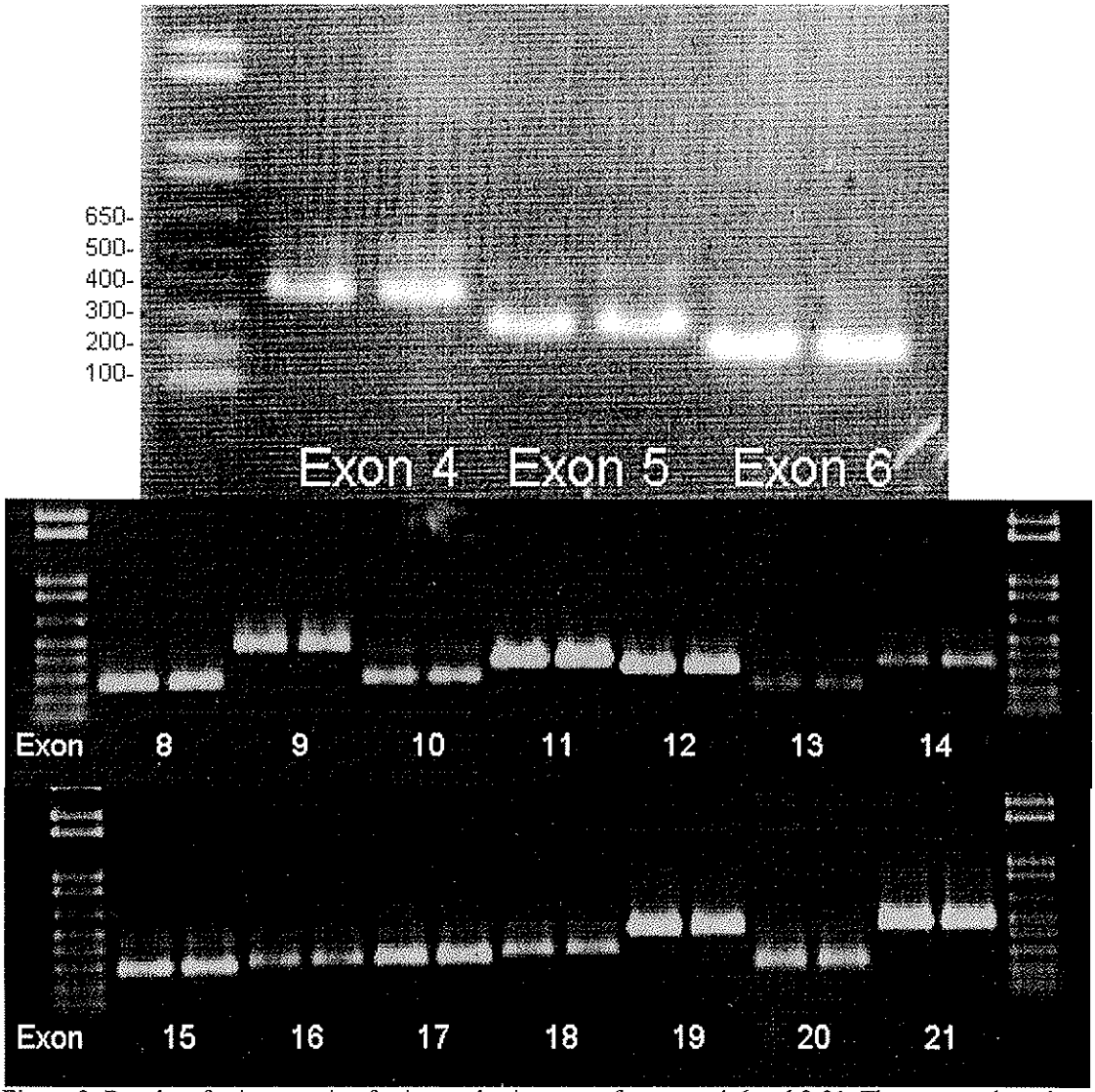


Figure 3: Results of primer testing for intronal primer sets for exons 4-6 and 8-21. The expected product sizes are presented in table 1. Template DNA is genomic DNA isolated from an arbitrary whole blood sample from blood bank at the University Hospital of Northern Norway. Mixing of reagents and thermal cycling parameters were identical to those used in pre-sequencing PCR (see tables 4 and 5). Primer sets for exons 1-3 and 7 were not tested prior to use.

Automatic sequencing

PCR

Subject DNA

All of the intronal primer sets gave visible bands of varying intensity in the vicinity of the expected product sizes though some gave extra bands (see figures 4 and 5).

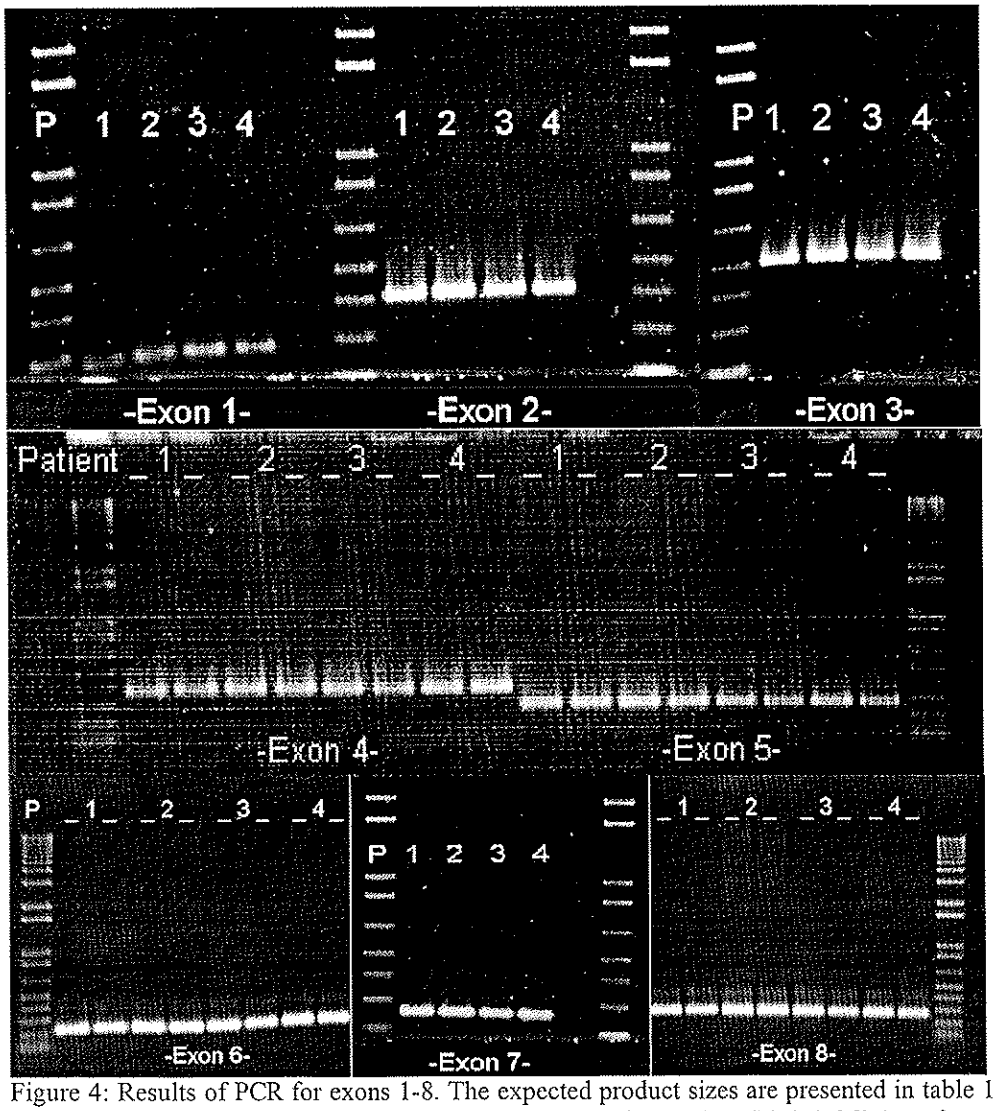


Figure 4: Results of PCR for exons 1-8. The expected product sizes are presented in table 1. Template DNA is genomic DNA isolated from subjects 1-4, here denoted as patient (P) 1-4. Mixing of reagents and thermal cycling parameters are presented in tables 4 and 5.

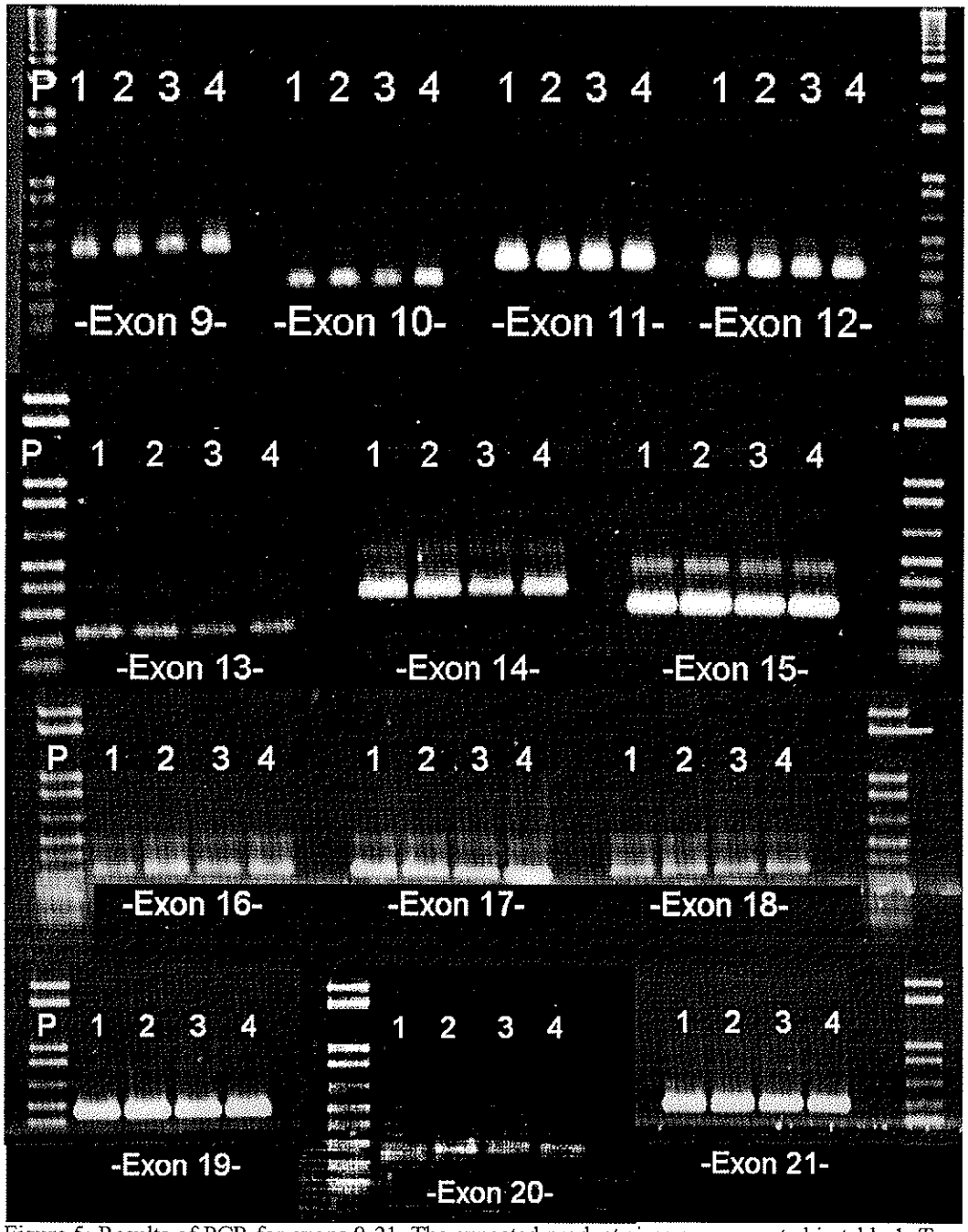


Figure 5: Results of PCR for exons 9-21. The expected product sizes are presented in table 1. Template DNA is genomic DNA isolated from subjects 1-4, here denoted as patient (P) 1-4. Mixing of reagents and thermal cycling parameters are presented in tables 4 and 5.

cDNA

Because the cDNA was in short supply the 4 forward primers IE-EX2-F1, IE-EX3-F1, IE-EX4-F1, IE-EX5-F1 and the reverse primer IE-EX3-R1 (see table 2 and 3) were not used in PCR. Instead the first forward primer in the set (IE-EX1-F1) was used in the hope that it would amplify all the intervening exons. In tune with a well known set of natural laws, the polymerase chain reaction favoured the amplification of the shortest possible sequence. However, a smear of ghost bands trailed the main product (see figures 6; A and 7) spiking the curiosity and providing an opportunity to use the 4 afore mentioned forward primers. Thus, 1:10 and 1:100 dilutions of fragment A were used as template in nested PCR in an attempt to establish the identity of the extra exons in the ghost bands. The results are illustrated in figure 8.

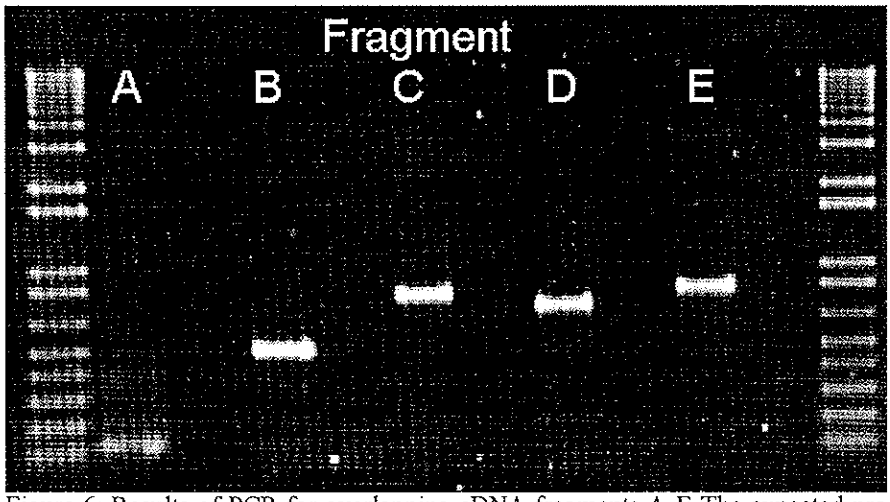


Figure 6: Results of PCR for overlapping cDNA fragments A-E The expected product sizes are presented in table 3. Primer sets were as follows A = IE-EX1-F1 + IE-EX6-R1, B = IE-EX6-F1 + IE-EX11-R1 C = IE-EX11-F1 + IE-EX16-R1 D = IE-EX16-F1 + IE-EX19-R1, E = IE-EX19-F1 + IE-EX21-R1. Template DNA was cDNA isolated from a normal subject. Mixing of reagents and thermal cycling parameters are presented in tables 4 and 5 (note that extension time was 50 seconds).

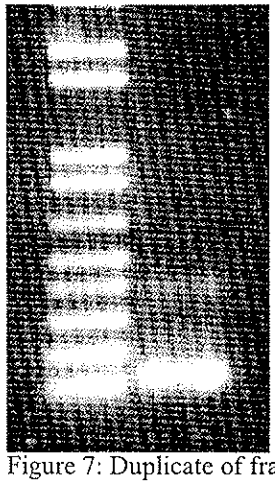


Figure 7: Duplicate of fragment A in figure r4 adjusted to bring out the extra bands.

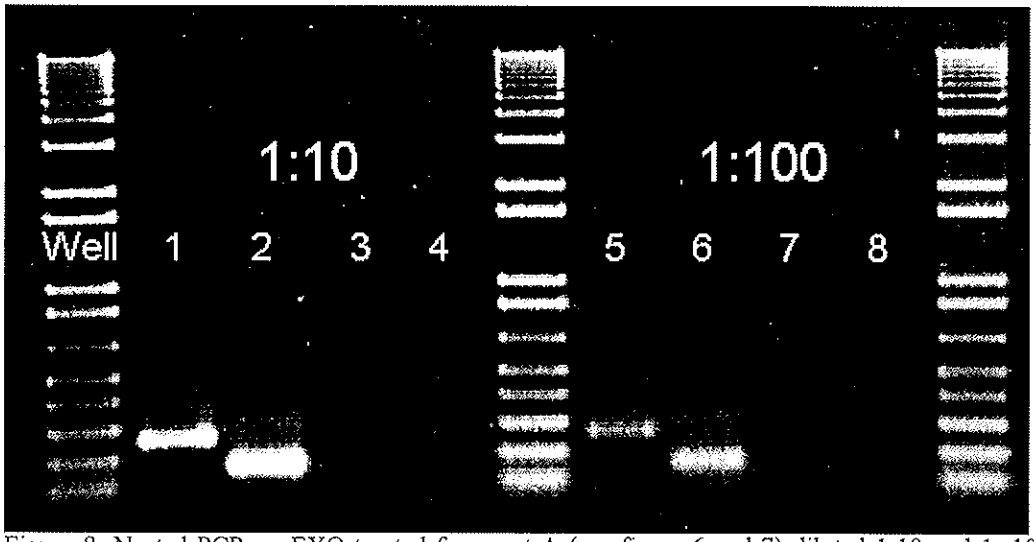


Figure 8: Nested PCR on EXO-treated fragment A (see figure 6 and 7) diluted 1:10 and 1: 100. All reactions contained the reverse primer IE-EX6-R1.Reactions 1 and 5 contained the forward primer IE-EX2-F1, while reactions 2 and 6 contained IE-EX3-F1, reactions 3 and 7 contained IE-EX4-F1 and finally reactions 4 and 8 contained IE-EX5-F1. Mixing of reagents and thermal cycling parameters were identical to those used in pre-sequencing PCR.

The presence of a band in well 1 (see figure8) indicates that at least part of the new ENST0000308928 transcript (see table 8) is present in blood. Presence of a band in well 2 is more surprising and suggests that a novel transcript including exons 3 (ENSE00001200259) and 6 (ENSE00001218925/0958560) is also present in blood. This may however simply be a reflection of the tendency towards long non-specific transcripts in blood. There was No evidence supporting co-occurrence of exon1 with exons 4 and 5 in transcripts.

Sequencing

Subject DNA

Sequencing from the genomic DNA of the 4 subjects went more or less smoothly and full length sequences for all the exons were attained (see alignments in appendix). Although there were many aberrations in the form of non-discriminate (N) bases, gaps, inserts and substitutions, all were cancelled by the complementary strand or, failing that, by a new copy of the same strand. There were no obvious differences between the 4 subjects. There were, however uncancelled differences between the subject sequences and the consensus sequence at two sites. The first was a G>C substitution at position 143 in exon 2 (ENSE00001218885). All 4 subject sequences carried the substitution, implying a polymorphism. Unfortunately there is no cDNA sequence of exon 2 to substantiate the implication. The second, almost unworthy of mention, was a G>A/N substitution in the 8th base of the 3' flanking intron sequence of exon 20 (mutations in bases immediately adjacent to exons can give rise to problems during mRNA splicing). Only forward sequences appeared to be disturbed (see figure accompanying exon 20 in appendix), which suggests that it was simply an artefact of sequencing. The substitution is mentioned here because it remained uncancelled in subjects 2 and 3. All attempts at reverse sequencing exon 20 of subjects 2 and 3 resulted in unreadable double sequences. The sequences were displaced 4 positions over the entire length indicating mispriming perhaps as a result of a mutation the primer binding site in the intron but more likely because of a sub-optimal primer sequence (see figure accompanying exon 20 in appendix).

cDNA

Sequencing Pou2f1 cDNA was less successful and exons 2-5 and 7 were not covered. The 5' end of exon 1 was not resolved and there were several uncancelled Ns as well as a G>C substitution near the 3' end of exon 21. However the substitution and all but two of the Ns could be attributed to artefacts of sequencing (see figures accompanying exons 16 and 21 in appendix).

Conclusion

No mutations were found in the 21 exons studied. This makes *POU2F1* a less likely candidate for *DFNA7*. However, *POU2F1*s candidacy cannot be rejected unequivocally at this point for several reasons.

Firstly, apart from the ubiquitously expressed OCT-1 protein, two tissue specific isoforms of human OCT-1 have been discovered in lymfocytes[13,14] and several murine isoforms have been reported[15]. Given the high level of evolutionary conservation between these two species, there may well be several human isoforms of OCT-1 yet to be discovered. Perhaps a cochlea specific isoform with 1 or more unique exons exists. The fact that the sequence information regarding the original *POU2F1* gene transcript was removed and 4 new transcripts were added to the ENSEMBL database during the course of this study illustrates the incomplete understanding of the array of expression products.

Secondly, *DFNA7* exhibits autosomal dominant inheritance, meaning that only one chromosome is affected. We cannot be certain that the sequences produced in this study represent both chromosomes. Theoretically the mutation itself could cause sequencing of the affected chromosome to fail so that only the normal sequence is resolved. For example, a large insertion could increase the length of the mutated exon so that the shorter, non-mutated exon is preferentially amplified during PCR. This is particularly relevant for the newly discovered exons which have not yet been tested for large rearangements by southern blot or SSCP.

Thirdly, mutations affecting the *POU2F1* gene need not lie within the gene itself. A mutation in an enhancer or promoter (control) region could cause reduced expression. Such mutations are more challenging to associate to disease and in the case of *POU2F1* no control regions are as yet defined.

In an incidental finding, a novel transcript including exons 3 (ENSE00001200259) and 6 (ENSE00001218925/0958560) (compare with table 8) appeared to be present in blood. This finding may simply be a reflection of the tendency towards long non-specific transcripts in blood or it may indicate the existence of a novel isoform of OCT-1.

In order to rule out mutation in *POU2F1* as the antecedent of *DFNA7*, future studies could include mutational analysis (by SSCP and sequencing) of all new exons and control regions as the consensus data is presented at ENSEMBL. Screening for expression of OCT-1 and its possible isoforms in cDNA libraries from human cochlea at different stages of development, may also be helpful.

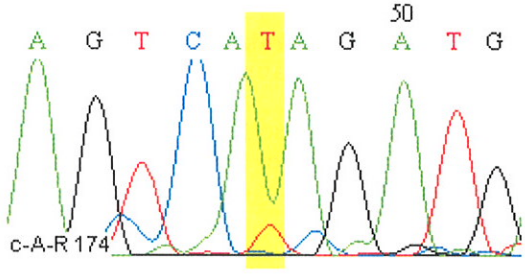
References

1. Morton NE. Genetic epidemiology of hearing impairment. Ann N Y Acad Sci. 1991;630:16-31.
2. Marazita ML, Ploughman LM, Rawlings B, Remington E, Arnos KS, Nance WE. Genetic epidemiological studies of early-onset deafness in the U.S. school-age population. Am J Med Genet. 1993 Jun 15;46(5):486-91.
3. Petersen MB. Non-syndromic autosomal-dominant deafness. Clin Genet. 2002 Jul;62(1):1-13.
4. Petit C. Genes responsible for human hereditary deafness: symphony of a thousand. Nat Genet. 1996 Dec;14(4):385-91.
5. Nadol JB Jr. Hearing loss. N Engl J Med. 1993 Oct 7;329(15):1092-102.
6. Anderssen SH, Andresen J, Andersen R, Sponheim L. [Universal neonatal hearing screening of infants with otoacoustic emissions]. Tidsskr Nor Laegeforen. 2002 Sep 20;122(22):2187-9.
7. Ryan AF, Crenshaw EB III, Simmons DM. Gene expression in normal and abnormal inner ears. Ann N Y Acad Sci. 1991;630:129-32.
8. Fagerheim T, Nilssen O, Raeymaekers P, Brox V, Moum T, Elverland HH, Teig E, Omland HH, Fostad GK, Tranebjaerg L. Identification of a new locus for autosomal dominant non-syndromic hearing impairment (DFNA7) in a large Norwegian family. Hum Mol Genet. 1996 Aug;5(8):1187-91.
9. Tranebjaerg L, Elverland HH, Fagerheim T. DFNA7. Adv Otorhinolaryngol. 2000;56:97-100.
10. de Kok YJ, van der Maarel SM, Bitner-Glindzicz M, Huber I, Monaco AP, Malcolm S, Pembrey ME, Ropers HH, Cremers FP. Association between X-linked mixed deafness and mutations in the POU domain gene POU3F4. Science. 1995 Feb 3;267(5198):685-8
11. Vahava O, Morell R, Lynch ED, Weiss S, Kagan ME, Ahituv N, Morrow JE, Lee MK, Skvorak AB, Morton CC, Blumenfeld A, Frydman M, Friedman TB, King MC, Avraham KB. Mutation in transcription factor POU4F3 associated with inherited progressive hearing loss in humans. Science. 1998 Mar 20;279(5358):1950-4.
12. He X, Treacy MN, Simmons DM, Ingraham HA, Swanson LW, Rosenfeld MG. Expression of a large family of POU-domain regulatory genes in mammalian brain development. Nature. 1989 Jul 6;340(6228):35-41. Erratum in: Nature 1989 Aug 24;340(6235):662.
13. Pankratova EV, Deyev IE, Zhenilo SV, Polanovsky OL. Tissue-specific isoforms of the ubiquitous transcription factor Oct-1. Mol Genet Genomics. 2001 Oct;266(2):239-45.
14. Luchina NN, Krivega IV, Pankratova EV. Human Oct-1L isoform has tissue-specific expression pattern similar to Oct-2. Immunol Lett. 2003 Feb 3;85(3):237-41.
15. Riss J, Laskov R. Expression of novel alternatively spliced isoforms of the oct-1 transcription factor. Biochim Biophys Acta. 1999 Feb 16;1444(2):295-8.

Appendix

Note: "P1, P2, P3 and P4" refer to subjects 1-4. "c-A, c-B, c-C, c-D and c-E" denote cDNA fragments A-E (see table 3). –F = forward sequence and –R = reverse sequence.

```
Exon 1 (ense 00001200261)

P1-F 141  ---------- ---------- ---------- -------TAT TTGGGCGCNA GGGGGGAGGG GAGCCNGAGC GAGGGAGGGT
P1-R 142  -GNNNNNNNN NNNNNNNNTT TAGAGGAGGG AGGAGGCAAG CAGGGCGGGA GGGGGGAGGG GAGCCAGAGC GAGGGAGGGT
P2-F 143  ---------- ---------- ---------- ---------- ----GCGGGA GGGGGGAGGG GAGCNAGAGC GAGGGAGGGT
P2-R 144  GNNNNNNNNN NNNNNNNNTT TAGAGGAGGG AGGAGGCAAG CAGGGCGGGA GGGGGGAGGG GAGCCAGAGC GAGGGAGGGT
P3-F 145  ---------- ---------- ---------- ---------- ----GCGGGA GGGGGGAGGG GAGCNAGAGC GAGGGAGGGT
P3-R 146  ---------- ------NNTT TAGAGGAGGG AGGAGGCAAG CAGGGCGGGA GGGGGGAGGG GAGCCAGAGC GAGGGAGGGT
P4-F 147  ---------- ---------- ---------- ---------- ----GCGGNA GGGGGGAGGG GAGCNNGAGC GAGGGAGGGT
P4-R 148  -GNNNNNNNN NNNNNNNNTT TAGAGGAGGG AGGAGGCAAG CAGGGCGGGA GGGGGGAGGG GAGCCAGAGC GAGGGAGGGT
                   10         20         30         40         50         60         70         80

REF                                    taaaatat tcAAAATGGC GGACGGAGGA GCAGCGAGTC A-AGATGAGA GTTCAGCCGC
c-A-F 173                                       -------- ---------- ---------- -------GAGN GTTCAGCCGC
c-A-R 174                                              NNCNNCCC CCNTTGAGGA GCAGCGAGTC ATAGATGAGA GTTCAGCCGC
P1-F 141  TTATCGACCG GGCGATTTTG GTTAAAATAT TCAAAATGGC GGACGGAGGA GCAGCGAGTC A-NGATGAGA GTTCAGCCGC
P1-R 142  TTATCGACCG GGCGATTTTG GTTAAAATAT TCAAAATGGC GGACGGAGGA GCAGCGAGTC A-AGATGAGA GTTCAGCCGC
P2-F 143  TTATCGACCG GGCGATTTTG GTTAAAATAT TCAAAATGGC GGACGGAGGA GCAGCGAGTC A-AGATGAGA GTTCAGCCGC
P2-R 144  TTATCGACCG GGCGATTTTG GTTAAAATAT TCAAAATGGC GGACGGAGGA GCAGCGAGTC A-AGATGAGA GTTCAGCCGC
P3-F 145  TTATCGACCG GGCGATTTTG GTTAAAATAT TCAAAATGGC GGACGGAGGA GCAGCGAGTC A-AGATGAGA GTTCAGCCGC
P3-R 146  TTATCGACCG GGCGATTTTG GTTAAAATAT TCAAAATGGC GGACGGAGGA GCAGCGAGTC A-AGATGAGA GTTCAGCCGC
P4-F 147  TTATCGACCG GGCGATTTTG GTTAAAATAT TCAAAATGGC GGACGGAGGA GCAGCGAGTC A-AGATGAGA GTTCAGCCGC
P4-R 148  TTATCGACCG GGCGATTTTG GTTAAAATAT TCAAAATGGC GGACGGAGGA GCAGCGAGTC A-AGATGAGA GTTCAGCCGC
                   90        100        110        120        130        140        150        160

REF       GGCGGCAGCA GCAGCAGgta atcatta
c-A-F 173 GGCGGCAGCA GCAGCAG
c-A-R 174 GGCGGCAGCA GCAGCAG
P1-F 141  GGCGGCAGCA GCAGCAGGTA ATCATTACAG CATTTTACAT ATTCATATTC ATACTCAACC CCGGCTCCCG CTGCCCCCCC
P1-R 142  GGCGGCAGCA GCAGCAGGTA ATCATTACAG CATTTTACAT ATTCATATTC ATACTCAACC CCGGCTCCCG CTGCCCCCCC
P2-F 143  GGCGGCAGCA GCAGCAGGTA ATCATTACAG CATTTTACAT ATTCATATTC ATACTCAACC CCGGCTCCCG CTGCCCCCCC
P2-R 144  GGCGGCAGCA GCAGCAGGTA ATCATTACAG CATTTTACAT ATTCATATTC ATACTCAACC CCGGCTCCCG CTGCCCCCCC
P3-F 145  GGCGGCAGCA GCAGCAGGTA ATCATTACAG CATTTTACAT ATTCATATTC ATACTCAACC CCGGCTCCCG CTGCCCCCCC
P3-R 146  GGCGGCAGCA GCAGCAGGTA ATCATTACAG CATTTTACAT ATTCATATTC ATACTCAACC CCGGCTCCCG CTGCCCCCCC
P4-F 147  GGCGGCAGCA GCAGCAGGTA ATCATTACAG CATTTTACAT ATTCATATTC ATACTCAACC CCGGCTCCCG CTGCCCCCCC
P4-R 148  GGCGGCAGCA GCAGCAGGTA ATCATTACAG CATTTTACAT ATTCATATTC ATACTCAACC CCGGCTCCCG CTGCCCCCCC
                   170        180        190        200        210        220        230        240

P1-F 141  CCGCGACTTA GCATAATTTA TTAGTACTCA GGATTATTAT AATTAACGGC GGGGAGANGN NAANNNNNNT TNNNNNNNNN
P1-R 142  CCGCGACTTA GCATAATTTA TTAGTACTCA NNATGNGNNC AAA------- ---------- ---------- ----------
P2-F 143  CCGCGACTTA GCATAATTTA TTAGTACTCA GGATTATTAT AATTAACGGC GGGGAGANGN AANNNNNNNN NNNNNNNNNN
P2-R 144  CCGCGACTTA GCATAATTTA TTAGTACTCA NNANGNNNNC ANNA------ ---------- ---------- ----------
P3-F 145  CCGCGACTTA GCATAATTTA TTAGTACTCA GGATTATTAT AATTAACGGC GGGGAGANGN NANNNNNNNN NNNNNNNNNN
P3-R 146  CCGCGACTTA GCATAATTTA TTAGTATCAN ATGNGNCA-- ---------- ---------- ---------- ----------
P4-F 147  CCGCGACTTA GCATAATTTA TTAGTACTCA GGATTATTAT AATTAACGGC GGGGAGANGG NANNNNNNNN NNNNNNNNNN
P4-R 148  CCGCGACTTA GCATAATTTA TTAGTACNC- ---------- ---------- ---------- ---------- ----------
                   250        260        270        280        290        300        310        320

P1-F 141   C 283
P1-R 142   - 281
P2-F 143   C 276
P2-R 144   - 283
P3-F 145   N 276
P3-R 146   - 261
P4-F 147   C 276
P4-R 148   - 267
```



Exon 2 (ense 00001218885)

```
P1-F 149  ---------- ---------- ---------- ------TNAT GNAGAATCTG TTTTTTAGTG CCGGGTGAAA TAACTAAATA
P1-R 150  --NNNNNNNN NNNNNTTTTG CGNCC-GCTC AAAGAGGTCT GTAGAATCTG TTTTTTAGTG CCGGGTGAAA TAACTAAATA
P2-F 151  ---------- ---------- ---------- ---------- -GNGAATCTG TTTTTTAGTG CCGGGTGAAA TAACTAAATA
P2-R 152  --NNNNNNNN NNNNNTNTTG GNGCC-GCTC AAAGAGGTCT GTAGAATCTG TTTTTTAGTG CCGGGTGAAA TAACTAAATA
P3-F 153  ---------- ---------- ---------- ---------- TGAGAATCTG TTTTTTAGTG CCGGGTGAAA TAACTAAATA
P3-R 154  NNNNNNNNNN NNNNNTTTTG CNGC--GCTC AAAGAGGTCT GTAGAATCTG TTTTTTAGTG CCGGGTGAAA TAACTAAATA
P4-F 155  ---------- ---------- ---------- ---------- TGAGAATCTG TTTTTTAGTG CCGGGTGAAA TAACTAAATA
P4-R 156  -NNNNNNNNN NNNNNTTTTG NGNCCANCTC AAAGAGGTCT GTAGAATCTG TTTTTTAGTG CCGGGTGAAA TAACTAAATA
                 10         20         30         40         50         60         70         80

                                                    tat gattctaGCT ACTACTGGGC TGTAAACAGT GATGCCAGCA
P1-F 149  TTTCGTTTTA AAAGGAAGGA ATTAATCAGG AACCTTTTAT GATTCTAGCT ACTACTGGGC TGTAAACAGT GATGCCAGCA
P1-R 150  TTTCGTTTTA AAAGGAAGGA ATTAATCAGG AACCTTTTAT GATTCTAGCT ACTACTGGGC TGTAAACAGT GATGCCAGCA
P2-F 151  TTTCGTTTTA AAAGGAAGGA ATTAATCAGG AACCTTTTAT GATTCTAGCT ACTACTGGGC TGTAAACAGT GATGCCAGCA
P2-R 152  TTTCGTTTTA AAAGGAAGGA ATTAATCAGG AACCTTTTAT GATTCTAGCT ACTACTGGGC TGTAAACAGT GATGCCAGCA
P3-F 153  TTTCGTTTTA AAAGGAAGGA ATTAATCAGG AACCTTTTAT GATTCTAGCT ACTACTGGGC TGTAAACAGT GATGCCAGCA
P3-R 154  TTTCGTTTTA AAAGGAAGGA ATTAATCAGG AACCTTTTAT GATTCTAGCT ACTACTGGGC TGTAAACAGT GATGCCAGCA
P4-F 155  TTTCGTTTTA AAAGGAAGGA ATTAATCAGG AACCTTTTAT GATTCTAGCT ACTACTGGGC TGTAAACAGT GATGCCAGCA
P4-R 156  TTTCGTTTTA AAAGGAAGGA ATTAATCAGG AACCTTTTAT GATTCTAGCT ACTACTGGGC TGTAAACAGT GATGCCAGCA
                 90        100        110        120        130        140        150        160

          AAATGTTACT TCAGCTGATG AAGTGATGCT GTTTCGAGAA TTTGAAAGCA ATTTTTCAGT GGATAAAGAA GTTGACAGCA
P1-F 149  AAATGTTACT TCAGCTGATG AAGTGATGCT GTTTCGAGAA TTTGAAAGCA ATTTTTCAGT GGATAAAGAA GTTGACAGCA
P1-R 150  AAATGTTACT TCAGCTGATG AAGTGATGCT GTTTCGAGAA TTTGAAAGCA ATTTTTCAGT GGATAAAGAA GTTGACAGCA
P2-F 151  AAATGTTACT TCAGCTGATG AAGTGATGCT GTTTCGAGAA TTTGAAAGCA ATTTTTCAGT GGATAAAGAA GTTGACAGCA
P2-R 152  AAATGTTACT TCAGCTGATG AAGTGATGCT GTTTCGAGAA TTTGAAAGCA ATTTTTCAGT GGATAAAGAA GTTGACAGCA
P3-F 153  AAATGTTACT TCAGCTGATG AAGTGATGCT GTTTCGAGAA TTTGAAAGCA ATTTTTCAGT GGATAAAGAA GTTGACAGCA
P3-R 154  AAATGTTACT TCAGCTGATG AAGTGATGCT GTTTCGAGAA TTTGAAAGCA ATTTTTCAGT GGATAAAGAA GTTGACAGCA
P4-F 155  AAATGTTACT TCAGCTGATG AAGTGATGCT GTTTCGAGAA TTTGAAAGCA ATTTTTCAGT GGATAAAGAA GTTGACAGCA
P4-R 156  AAATGTTACT TCAGCTGATG AAGTGATGCT GTTTCGAGAA TTTGAAAGCA ATTTTTCAGT GGATAAAGAA GTTGACAGCA
                 170        180        190        200        210        220        230        240

          CGATTTGTTG GATGTGATGA AGGATTAATG AGCATACACC TTCACTTGTA TTAGCTTAAG ATGGAATGGT TCTGGGCAAT
P1-F 149  CGATTTGTTG GATGTGATGA AGGATTAATC AGCATACACC TTCACTTGTA TTAGCTTAAG ATGGAATGGT TCTGGGCAAT
P1-R 150  CGATTTGTTG GATGTGATGA AGGATTAATC AGCATACACC TTCACTTGTA TTAGCTTAAG ATGGAATGGT TCTGGGCAAT
P2-F 151  CGATTTGTTG GATGTGATGA AGGATTAATC AGCATACACC TTCACTTGTA TTAGCTTAAG ATGGAATGGT TCTGGGCAAT
P2-R 152  CGATTTGTTG GATGTGATGA AGGATTAATC AGCATACACC TTCACTTGTA TTAGCTTAAG ATGGAATGGT TCTGGGCAAT
P3-F 153  CGATTTGTTG GATGTGATGA AGGATTAATC AGCATACACC TTCACTTGTA TTAGCTTAAG ATGGAATGGT TCTGGGCAAT
P3-R 154  CGATTTGTTG GATGTGATGA AGGATTAATC AGCATACACC TTCACTTGTA TTAGCTTAAG ATGGAATGGT TCTGGGCAAT
P4-F 155  CGATTTGTTG GATGTGATGA AGGATTAATC AGCATACACC TTCACTTGTA TTAGCTTAAG ATGGAATGGT TCTGGGCAAT
P4-R 156  CGATTTGTTG GATGTGATGA AGGATTAATC AGCATACACC TTCACTTGTA TTAGCTTAAG NTGGAATGGT TCTGGGCAAT
                 250        260        270        280        290        300        310        320

          ATAAAATAAC AGgtttttcca gt
P1-F 149  ATAAAATAAC AGGTTTTTCCA GTTTATTTTT ATTAC-TGTA CTTGCTTGTT TATAATATTA GGCATATTAA TTAACACCTA
P1-R 150  ATAAAATAAC AGGTTTTTCCA GTTTATTTTT ATTACGTGTA CTTGCTTGTT TATAATATTA GGCATATTAA TTNCGNNNNC
P2-F 151  ATAAAATAAC AGGTTTTTCCA GTTTATTTTT ATTAC-TGTA CTTGCTTGTT TATAATATTA GGCATATTAA TTAACACCTA
P2-R 152  ATAAAATAAC AGGTTTTTCCA GTTTATTTTT ATTAC-TGTA CTTGCTTGTT TATAATATTA GGCATATTAA NAACNNNNNN
P3-F 153  ATAAAATAAC AGGTTTTTCCA GTTTATTTTT ATTAC-TGTA CTTGCTTGTT TATAATATTA GGCATATTAA TTAACACCTA
P3-R 154  ATAAAATAAC AGGTTTTTCCA GTTTATTTTT ATTAC-TGTA CTTGCTTGTT TATAATATTA GGCAT-TTAA NTNC------
P4-F 155  ATAAAATAAC AGGTTTTTCCA GTTTATTTTT ATTAC-TGTA CTTGCTTGTT TATAATATTA GGCATATTAA TTAACACCTA
P4-R 156  ATAAAATAAC AGGTTTTTCCA GTTTATTTTT ATTAC-TGTA CTTGCTTGTT TATAATATTA GGCATATTAA NAAC------
                 330        340        350        360        370        380        390        400

P1-F 149  TGAAGCCCAT TTGANTGCCA AANNNNNNNN NNNNNNN-- 400
P1-R 150  AA-------- ---------- ---------- --------- 399
P2-F 151  TGAAGCCCAT TTGAATGGCA AANNNNNNNN NNNNNNN-- 395
P2-R 152  AANNAT---- ---------- ---------- --------- 402
P3-F 153  TGAAGCCCAT TTGAGTNGCA AANNNNNNNN NNNNNNNN- 397
P3-R 154  ---------- ---------- ---------- --------- 390
P4-F 155  TGAAGCCCAT TTGTGGGCCA CANNNNNNNN NNNNNNNNN 398
P4-R 156  ---------- ---------- ---------- --------- 392
                 410        420        430
```

Exon 3 (ense 00001200259)

```
P1-F 165  TNTNAANGGA NGGGGNNTTN GGNNTTGGTA NNNAGGGATG GGAGGTTNTT AGNNTANGGG CGTNGCAAAT AAAGGCAAAG
P1-R 166  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
P2-F 167  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
P2-R 168  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
P3-F 169  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
P3-R 170  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
P4-F 171  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
P4-R 172  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
                 10         20         30         40         50         60         70         80

P1-F 165  GGNGGANGAA GGATTGGTAG ANNNNGGGAA ACATTTTTNG ANAANNAAAA GAGNAGANNA GTCAGAAAAG GANAGNGNAT
P1-R 166  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
P2-F 167  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
P2-R 168  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
P3-F 169  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
P3-R 170  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
P4-F 171  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
P4-R 172  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ----------
                 90        100        110        120        130        140        150        160

P1-F 165  TTNAAAAANG TTANANNAAG GGTGGGTGCG TGTAGGGNCT GGCANCCCCC ACTGTGTGTA ATTGTTCTCT CTTTAAGTGT
P1-R 166  ---------- ---------- ---------- ---------- ---------- ---------- ---------- -CNNNNAGTG
P2-F 167  ---------- ---------- ---------- -------NNN NNNCNCCCCC ACTGTGTGTA ATTGTTCTCT CTTTAAGTGT
P2-R 168  ---------- ---------- ---------- ---------- ---------- ---------- -----TNNNN NCNNNNNNNG
P3-F 169  ---------- ---------- ------NNNN NNNNNNNNNN NNNCNCCCCC ACTGTGTGTA ATTGTTCTCT CTTTAAGTGT
P3-R 170  ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------G
P4-F 171  ---------- ---------- ---------- -------NNN NNNCTCCCCC ACTGTGTGTA ATTGTTCTCT CTTTAAGTGT
P4-R 172  ---------- ---------- ---------- ---------- ---------- ---------- ---------- -------TGT
                 170        180        190        200        210        220        230        240

                                                                      gtgt tcatagACTG GAAAAGTAAG AAGAGCTTTC
P1-F 165  TAGGTTATTT TATACAGCTT ACTCAGTTTT GCCTTTTTCT CTTTTTGTGT TCATAGACTG GAAAAGTAAG AAGAGCTTTC
P1-R 166  TAGGTTATTT TATACAGCTT ACTCAGTTTT GCCTTTTTCT CTTTTTGTGT TCATAGACTG GAAAAGTAAG AAGAGCTTTC
P2-F 167  TAGGTTATTT TATACAGCTT ACTCAGTTTT GCCTTTTTCT CTTTTTGTGT TCATAGACTG GAAAAGTAAG AAGAGCTTTC
P2-R 168  TANGTTATTT TATACAGCTT ACTCAGTTTT GCCTTTTTCT CTTTTTGTGT TCATAGACTG GAAAAGTAAG AAGAGCTTTC
P3-F 169  TAGGTTATTT TATACAGCTT ACTCAGTTTT GCCTTTTTCT CTTTTTGTGT TCATAGACTG GAAAAGTAAG AAGAGCTTTC
P3-R 170  TAGGTTATTT TATACAGCTT ACTCAGTTTT GCCTTTTTCT CTTTTTGTGT TCATAGACTG GAAAAGTAAG AAGAGCTTTC
P4-F 171  TAGGTTATTT TATACAGCTT ACTCAGTTTT GCCTTTTTCT CTTTTTGTGT TCATAGACTG GAAAAGTAAG AAGAGCTTTC
P4-R 172  TAGGTTATTT TATACAGCTT ACTCAGTTTT GCCTTTTTCT CTTTTTGTGT TCATAGACTG GAAAAGTAAG AAGAGCTTTC
                 250        260        270        280        290        300        310        320

         CTGCCTTTTT AATTACCAAA CTACTCTCAG TTTTCAATGA ATCAGTTCAA AGAAAGAATG CAGTCTTTCT ATACCTGgta
P1-F 165 CTGCCTTTTT AATTNCCAAA CTACTCTCAG TTTTCAATGA ATCAGTTCAA AGAAAGAATG CAGTCTTTCT ATACCTGGTA
P1-R 166 CTGCCTTTTT AATTACCAAA CTACTCTCAG TTTTCAATGA ATCAGTTCAA AGAAAGAATG CAGTCTTTCT ATACCTGGTA
P2-F 167 CTGCCTTTTT AATTACCAAA CTACTCTCAG TTTTCAATGA ATCAGTTCAA AGAAAGAATG CAGTCTTTCT ATACCTGGTA
P2-R 168 CTGCCTTTTT AATTACCAAA CTACTCTCAG TTTTCAATGA ATCAGTTCAA AGAAAGAATG CAGTCTTTCT ATACCTGGTA
P3-F 169 CTGCCTTTTT AATTACCAAA CTACTCTCAG TTTTCAATGA ATCAGTTCAA AGAAAGAATG CAGTCTTTCT ATACCNG-TA
P3-R 170 CNGCCTTTTT AATTACCAAA CTACTCTCAG TTTTCAATGA ATCAGTTCAA AGAAAGAATG CAGTCTTTCT ATACCTGGTA
P4-F 171 CTGCCTTTTT AATTACCAAA CTACTCTCAG TTTTCAATGA ATCAGTTCAA AGAAAGAATG CAGTCTTTCT ATACCTGGTA
P4-R 172 CTGCCTTTTT AATTACCAAA CTACTCTCAG TTTTCAATGA ATCAGTTCAA AGAAAGAATG CAGTCTTTCT ATACCTGGTA
                 330        340        350        360        370        380        390        400

         aatattt
P1-F 165 AATATTTTT-- GGNAAN-AAA GCNNGNACGG NNGAANCG-- ---------- ---------- ---------- ------ 435
P1-R 166 AATATTTTCT GCTAAATAAA GCTATTAATT TAGTAGCATC TGATAAATCC AGGTGAGNNN NNNNNNNNNN NNNNNN 245
P2-F 167 AATATTTTCT GCTAAA-AAA G--------- ---------- ---------- ---------- ---------- ------ 223
P2-R 168 AATATTTTCT GCTAAATAAA GCTATTAATT TAGTAGCATC TGATAAATCC AGGTGAGNNN NNNNNNNNNN NNNNNN 251
P3-F 169 AAN-TTTTCT GCTAAA-AAA GCTNTNNNNN NNNNA----- ---------- ---------- ---------- ------ 246
P3-R 170 AATATTTTCT GCTAAATAAA GCTATTAATT TAGTAGCATC TGATAAATCC AGGTGAGNNN NNNNNNNNNN NNNNNN 237
P4-F 171 AATATTTTN- GCTAAA-AAA G--------- ---------- ---------- ---------- ---------- ------ 222
P4-R 172 AATATTTTCT GCTAAATAAA GCTATTAATT TAGTAGCATC TGATAAATCC AGGTGAGNNN NNNNNNNNNN NNNNNN 239
                 410        420        430        440        450        460        470
```

Exon 4 (ense 00001200252)

```
P1-F 5     ---------- ---------- ---------- ---------- TTNNTNTCTT TTG-AAATAA TTATTGATAA TGCTTGAGGT
P1-R 6     NNNNNNNNNN NNNNNNNNNN NNNTTGTGNG ATGGGTTTGT TGGACTTTTT TTGTAAATAA TTATTGATAA TGCTTGAGGT
P2-F 7     ---------- ---------- ---------- ---------- -TNNNCTTTT TTG-AAATAA TTATTGATAA TGCTTGAGGT
P2-R 8     NNNNNNNNNN NNNNNNNNNN NNNTTGTGTG ATGGGTTTGT TGGACTTTTT TTGTAAATAA TTATTGATAA TGCTTGAGGT
P3-F 9     ---------- ---------- ---------- ---------- TTNNNCTTTT TTG-AAATAA TTATTGATAA TGCTTGAGGT
P3-R 10    NNNNNNNNNN NNNNNNNNNN NNNTTGTGNG ATGGGTTTGT TGGACTTTTT TTGTAAATAA TTATTGATAA TGCTTGAGGT
P4-F 11    ---------- ---------- ---------- ---------- -----CTTTT TTG-AAATAA TNATTGATAA TGCTTGAGGT
P4-R 12    NNNNNNNNNN NNNNNNNNNN NNNTTGTGTG ATGGGTTTGT TGGACTTTTT TTGTAAATAA TTATTGATAA TGCTTGAGGT
                  10         20         30         40         50         60         70         80

P1-F 5     ACTAAGTATT TTTATTACAG AGAATTGAAA TATCATTAAG TGTTTTCTGT AAGTAAGTTA ATAGGACACT AAATTATTGG
P1-R 6     ACTAAGTATT TTTATTACAG AGAATTGAAA TATCATTAAG TGTTTTCTGT AAGTAAGTTA ATAGGACACT AAATTATTGG
P2-F 7     ACTAAGTATT TTTATTACAG AGAATTGAAA TATCATTAAG TGTTTTCTGT AAGTAAGTTA ATAGGACACT AAATTATTGG
P2-R 8     ACTAAGTATT TTTATTACAG AGAATTGAAA TATCATTAAG TGTTTTCTGT AAGTAAGTTA ATAGGACACT AAATTATTGG
P3-F 9     ACTAAGTATT TTTATTACAG AGAATTGAAA TATCATTAAG TGTTTTCTGT AAGTAAGTTA ATAGGACACT AAATTATTGG
P3-R 10    ACTAAGTATT TTTATTACAG AGAATTGAAA TATCATTAAG TGTTTTCTGT AAGTAAGTTA ATAGGACACT AAATTATTGG
P4-F 11    ACTAAGTATT TTTATTACAG AGAATTGAAA TATCATTAAG TGTTTTCTGT AAGTAAGTTA ATAGGACACT AAATTATTGG
P4-R 12    ACTAAGTATT TTTATTACAG AGAATTGAAA TATCATTAAG TGTTTTCTGT AAGTAAGTTA ATAGGACACT AAATTATTGG
                  90        100        110        120        130        140        150        160

                tggaa aagagATGAA GGTTGGAAAG GATATTGGTG TACTGTTGAG ATTGGCTGGG GTGAAACCAG TGGTGCTAAA
P1-F 5     TAGATTGGAA AAGAGATGAA GGTTGGAAAG GATATTGGTG TACTGTTGAG ATTGGCTGGG GTGAAACCAG TGGTGCTAAA
P1-R 6     TAGATTGGAA AAGAGATGAA GGTTGGAAAG GATATTGGTG TACTGTTGAG ATTGGCTGGG GTGAAACCAG TGGTGCTAAA
P2-F 7     TAGATTGGAA AAGAGATGAA GGTTGGAAAG GATATTGGTG TACTGTTGAG ATTGGCTGGG GTGAAACCAG TGGTGCTAAA
P2-R 8     TAGATTGGAA AAGAGATGAA GGTTGGAAAG GATATTGGTG TACTGTTGAG ATTGGCTGGG GTGAAACCAG TGGTGCTAAA
P3-F 9     TAGATTGGAA AAGAGATGAA GGTTGGAAAG GATATTGGTG TACTGTTGAG ATTGGCTGGG GTGAAACCAG TGGTGCTAAA
P3-R 10    TAGATTGGAA AAGAGATGAA GGTTGGAAAG GATATTGGTG TACTGTTGAG ATTGGCTGGG GTGAAACCAG TGGTGCTAAA
P4-F 11    TAGATTGGAA AAGAGATGAA GGTTGGAAAG GATATTGGTG TACTGTTGAG ATTGGCTGGG GTGAAACCAG TGGTGCTAAA
P4-R 12    TAGATTGGAA AAGAGATGAA GGTTGGAAAG GATATTGGTG TACTGTTGAG ATTGGCTGGG GTGAAACCAG TGGTGCTAAA
                 170        180        190        200        210        220        230        240

           AGATGCAGAC ACCCCCCATG ATATAGAAAG GAGgtaacag gca
P1-F 5     AGATGCAGAC ACCCCCCATG ATATAGAAAG GAGGTAACAG GCAATAATTG TTGAGTCAGA CATTTCTGCT TGTTCAACTA
P1-R 6     AGATGCAGAC ACCCCCCATG ATATAGAAAG GAGGTAACAG GCAATAATTG TTGAGTCAGA CATTTCTGCT TGTTCAACTA
P2-F 7     AGATGCAGAC ACCCCCCATG ATATAGAAAG GAGGTAACAG GCAATAATTG TTGAGTCAGA CATTTCTGCT TGTTCAACTA
P2-R 8     AGATGCAGAC ACCCCCCATG ATATAGAAAG GAGGTAACAG GCAATAATTG TTGAGTCAGA CATTTCTGCT TGTTCAACTA
P3-F 9     AGATGCAGAC ACCCCCCATG ATATAGAAAG GAGGTAACAG GCAATAATTG TTGAGTCAGA CATTTCTGCT TGTTCAACTA
P3-R 10    AGATGCAGAC ACCCCCCATG ATATAGAAAG GAGGTAACAG GCAATAATTG TTGAGTCAGA CATTTCTGCT TGTTCAACTA
P4-F 11    AGATGCAGAC ACCCCCCATG ATATAGAAAG GAGGTAACAG GCAATAATTG TTGAGTCAGA CATTTCTGCT TGTTCAACTA
P4-R 12    AGATGCAGAC ACCCCCCATG ATATAGAAAG GAGGTAACAG GCAATAATTG TTGAGTCAGA CATTTCTGCT TGTTCAACTA
                 250        260        270        280        290        300        310        320

P1-F 5     GTTATATGGT CAGTACAGAA TAGAGGGTAC AAATGGCTAT TCATATAGAA CATGACCAGG CTGGATGTAA NNNNCNNNN-
P1-R 6     GTTATATGGT CAGTACAGAA TAGAGGGTAC AAATGGCTAT TCAGNGNNNA A--------- ---------- ----------
P2-F 7     GTTATATGGT CAGTACAGAA TAGAGGGTAC AAATGGCTAT TCATATAGAA CATGACCAGG CTGGATGTAA ANNNCNNNN-
P2-R 8     GTTATATNGT CAGTACAGAA TAGAGGGTAC AAA-GGCT-T TCAGNNNNNA A--------- ---------- ----------
P3-F 9     GTTATATGGT CAGTACAGAA TAGAGGGTAC AAATGGCTAT TCATATAGAA CATGACCAGG CTGGATGTAA NNNNNNNNNN
P3-R 10    GTTATATNGT CAGTACAGAA TAGAGGGTAC AAA-GGCTAT TCATTNNNGN AANNAA---- ---------- ----------
P4-F 11    GTTATATGGT CAGTACAGAA TAGAGGGTAC AAATGGCTAT TCATATAGAA CATGACCAGG CTGGATGTAN ANNNCNNNN-
P4-R 12    GTTATATNGT CAGTACAGAA TAGAGGGTAC AAATGGCTAT TCAGNNNNNA A--------- ---------- ----------
                 330        340        350        360        370        380        390        400

P1-F 5     358
P1-R 6     371
P2-F 7     357
P2-R 8     369
P3-F 9     359
P3-R 10    375
P4-F 11    353
P4-R 12    371
```

Exon 5 (ense 00001200238)

```
                                                                            ttcgct taagAACATA
P1-F 13    ---------- ---------- ---------- -----TTNNT TNNNNNNNAC GCAACCCCCC TCTTTTCGCT TAAGAACATA
P1-R 14    NNNNNNNNNN NNNNGNTTTG CTTTTCCCCT TAATCCACTT TCCACCCTAC GCAACCCCCC TCTTTTCGCT TAAGAACATA
P2-F 15    ---------- ---------- ---------- -----TTNNN TNCCNNNCAC GCAACCCCCC TCTTTTCGCT TAAGAACATA
P2-R 16    -NNNNNNNNN NNNNGNTTTG CTTTTCCCCT TAATCCACTT TCCACCCTAC GCAACCCCCC TCTTTTCGCT TAAGAACATA
P3-F 17    ---------- ---------- ---------- -----TTNNN TTCCNNCCAC GCAACCCCCC TCTTTTCGCT TAAGAACATA
P3-R 18    NNNNNNNNNN NNNNNNTTTG CTTTTCCCCT TAATCCACTT TCCACCCTAC GCAACCCCCC TCTTTTCGCT TAAGAACATA
P4-F 19    ---------- ---------- ---------- ------TTNN TTCCTNNCTC GCAACCCCCC TCTTTTCGCT TAAGAACATA
P4-R 20    -NNNNNNNNN NNNGGTTTTG CTTTTCCCCT TAATCCACTT TCCACCCTAC GCAACCCCCC TCTTTTCGCT TAAGAACATA
                   10         20         30         40         50         60         70         80

           C-TGTAGATT TGTTAGAAAT AGTAGTACCT TCTTTCCCCA CCCCAAACTG CTACCTGTTT CTTCCTTGTT TGGACTCTCT
P1-F 13    C-TGTAGATT TGTTAGAAAT AGTAGTACCT TCTTTCCCCA CCCCAAACTG CTACCTGTTT CTTCCTTGTT TGGACTCTCT
P1-R 14    C-TGTAGATT TGTTAGAAAT AGTAGTACCT TCTTTCCCCA CCCCAAACTG CTACCTGTTT CTTCCTTGTT TGGACTCTCT
P2-F 15    NCTGTAGATT TGTTAGAAAT AGTAGTACCT TCTTTCCCCA CCCCAAACTG CTACCTGTTT CTTCCTTGTT TGGACTCTCT
P2-R 16    C-TGTAGATT TGTTAGAAAT AGTAGTACCT TCTTTCCCCA CCCCAAACTG CTACCTGTTT CTTCCTTGTT TGGACTCTCT
P3-F 17    C-TGTAGATT TGTTAGAAAT AGTAGTACCT TCTTTCCCCA CCCCAAACTG CTACCTGTTT CTTCCTTGTT TGGACTCTCT
P3-R 18    C-TGTAGATT TGTTAGAAAT AGTAGTACCT TCTTTCCCCA CCCCAAACTG CTACCTGTTT CTTCCTTGTT TGGACTCTCT
P4-F 19    C-TGTAGATT TGTTAGAAAT AGTAGTACCT TCTTTCCCCA CCCCAAACTG CTACCTGTTT CTTCCTTGTT TGGACTCTCT
P4-R 20    C-TGTAGATT TGTTAGAAAT AGTAGTACCT TCTTTCCCCA CCCCAAACTG CTACCTGTTT CTTCCTTGTT TGGACTCTCT
                   90        100        110        120        130        140        150        160

           GCTCCTCTAG AAGATTTCAC AGCAATGCTG GACTGCAGTG ACTATGTTCT AGgtgggtac ca
P1-F 13    GCTCCTCTAG AAGATTTCAC AGCAATGCTG GACTGCAGTG ACTATGTTCT AGGTGGGTAC CAGCACAGTT CTTTTCTTAG
P1-R 14    GCTCCTCTAG AAGATTTCAC AGCAATGCTG GACTGCAGTG ACTATGTTCT AGGTGGGTAC CAGCACAGTT CTTTTCTTAG
P2-F 15    GCTCCTCTAG AAGATTTCAC AGCAATGCTG GACTGCAGTG ACTATGTTCT AGGTGGGTAC CAGCACAGTT CTTTTCTTAG
P2-R 16    GCTCCTCTAG AAGATTTCAC AGCAATGCTG GACTGCAGTG ACTATGTTCT AGGTGGGTAC CAGCACAGTT CTTTTCTTAG
P3-F 17    GCTCCTCTAG AAGATTTCAC AGCAATGCTG GACTGCAGTG ACTATGTTCT AGGTGGGTAC CAGCACAGTT CTTTTCTTAG
P3-R 18    GCTCCTCTAG AAGATTTCAC AGCAATGCTG GACTGCAGTG ACTATGTTCT AGGTGGGTAC CAGCACAGTT CTTTTCTTAG
P4-F 19    GCTCCTCTAG AAGATTTCAC AGCAATGCTG GACTGCAGTG ACTATGTTCT AGGTGGGTAC CAGCACAGTT CTTTTCTTAG
P4-R 20    GCTCCTCTAG AAGATTTCAC AGCAATGCTG GACTGCAGTG ACTATGTTCT AGGTGGGTAC CAGCACAGTT CTTTTCTTAG
                  170        180        190        200        210        220        230        240

P1-F 13    AAGCTTTAAT ATGGAAATAG AGTGTGTCGG GTTGACTATG CAANANNNNN NNNNNNNNNN - 264
P1-R 14    AAGCTTTAAT ANGAAANNAN NAANNNA--- ---------- ---------- ---------- - 266
P2-F 15    AAGCTTTAAT ATGGAAATAG AGTGTGTCGG GTTGACTATG CANANNNNNN NNNNNNNNNN N 266
P2-R 16    AAGCTTTAAT ANGAAANNAN NAANNNA--- ---------- ---------- ---------- - 265
P3-F 17    AAGCTTTAAT ATGGAAATAG AGTGTGTCGG GTTGACTATG CAAAANNNNN NNNNNNNNNN N 265
P3-R 18    AAGCTTTAAT ANGAAANNAG NAANNNA--- ---------- ---------- ---------- - 266
P4-F 19    AAGCTTTAAT ATGGAAATAG AGTGTGTCGG GTTGACTATG CAAANNNNNN NNNNNNNNNN N 264
P4-R 20    AAGCTTTAAT AGGAAGTGAG GAATCAA--- ---------- ---------- ---------- - 265
                  250        260        270        280        290        300
```

Exon 6 (ense 00000958560)

```
P1-F 21    ---------- ---------- ---------- ---------- ---------- -TCTCCCCGT TTTTTAAAAA CCTTATTCTC
P1-R 22    NNNNCNNGNN NCNGACNNNN NGCTCAGTTT TGCCTCCTCA ATCCCATCTC CATCCCCAGT TTTTTAAAAA CCTTATTCTC
P2-F 23    ---------- ---------- ---------- ---------- ----TNNNTN CNCTCCCCGT TTTTTAAAAA CCTTATTCTC
P2-R 24    NNNACNGGNN NCNGACNGNN NGCTCAGTTT TGCCTCCTCA ATCCCATCTC CATCCCCAGT TTTTTAAAAA CCTTATTCTC
P3-F 25    ---------- ---------- ---------- ---------- ----TTNNTT CCCTCCCCGT TTTTTAAAAA CCTTATTCTC
P3-R 26    NNNACNNGNN NCNGACNGNN NGCTCAGTTT TGCCTCCTCA ATCCCATCTC CATCCCCAGT TTTTTAAAAA CCTTATTCTC
P4-F 27    ---------- ---------- ---------- ---------- ----TNNNTN NNNTNCCCGT TTTTTAAAAA CCTTATTCTC
P4-R 28    -NNNACGGNN NCNNACNGNN NGCCCAGTTT TGCCTCCTCA ATCCCAACTC CATCCCCAGT TTTTTAAAAA CCTTATTCTC
                  10         20         30         40         50         60         70         80

                ttt attgcagA-C TCAAGAATGA ACAATCCG-- TCAGAAACCA GTAAACCATC TATGGAGAGT GGAGATGGCA
C-A-F 173                 A-C TCAAGAATGA ACAATCCG-- TCAGAAACCA GTAAACCATC TATGGAGAGT GGAGATGGCA
C-A-R 174                 A-C TCAAGAATGA ACAATCCG-- TCAGAAACCN NNNNNNNGNN NAAA----------------
C-B-F 175                 --- ---------- ---------- ---------- ---------- ---GN-GAGT GGAGATGGCA
C-B-R 176                 NNN NNNNNNNNNN NNNCCCCC-- CCC-AAACCA GTAAACCATC TATGGAGAGT GGAGATGGCA
P1-F 21    C-TCTGATTT ATTGCAGA-C TCAAGAATGA ACAATCCG-- TCAGAAACCA GTAAACCATC TATGGAGAGT GGAGATGGCA
P1-R 22    C-TCTGATTT ATTGCAGA-C TCAAGAATGA ACAATCCG-- TCAGAAACCA GTAAACCATC TATGGAGAGT GGAGATGGCA
P2-F 23    C-TCTGATTT ATTGCAGA-C TCAAGAATGA ACAATCCG-- TCAGAAACCA GTAAACCATC TATGGAGAGT GGAGATGGCA
P2-R 24    C-TCTGATTT ATTGCAGA-C TCAAGAATGA ACAATCCG-- TCAGAAACCA GTAAACCATC TATGGAGAGT GGAGATGGCA
P3-F 25    C-TCTGATTT ATTGCAGA-C TCAAGAATGA ACAATCCG-- TCAGAAACCA GTAAACCATC TATGGAGAGT GGAGATGGCA
P3-R 26    C-TCTGATTT ATTGCAGA-C TCAAGAATGA ACAATCCG-- TCAGAAACCA GTAAACCATC TATGGAGAGT GGAGATGGCA
P4-F 27    C-TCTGATTT ATTGCAGA-C TCAAGAATGA ACAATCCG-- TCAGAAACCA GTAAACCATC TATGGAGAGT GGAGATGGCA
P4-R 28    CCTCTGATTT ATTGCAGAAC TCAAGAATGA ACAATCCCGT CCAGAAACCA GTAAACCNNC TANGGAGAGT GGAGATGGCA
                  90        100        110        120        130        140        150        160

           ACACAGgtaa gagttt
C-A-F 173  AAGANN
C-A-R 174  ------
C-B-F 175  ACACAG
C-B-R 176  ACACAG
P1-F 21    ACACAGGTAA GAGTTTTCTG ATCTAGCTTT TTAATTAACT CTAGTAGAGC ACAAAGAAGA AAGTTTCCAT GTAACACNNN
P1-R 22    ACACAGGTAA GAGTTTTCTG ATCTAGCTTT TTAATTAACT CTAGTAGAGN GNNAAGAANN AA--------- ----------
P2-F 23    ACACAGGTAA GAGTTTTCTG ATCTAGCTTT TTAATTAACT CTAGTAGAGC ACAAAGAAGA AAGTTTCCAT GTAACACNNN
P2-R 24    ACACAGGTAA GAGTTTTCTG ATCTAGCTTT TTAATTAACT CTAGTAGAGN GNNNNNAA-- ---------- ----------
P3-F 25    ACACAGGTAA GAGTTTTCTG ATCTAGCTTT TTAATTAACT CTAGTAGAGC ACAAAGAAGA AAGTTTCCAT GTAACACNNN
P3-R 26    ACACAGGTAA GAGTTTTCTG ATCTAGCTTT TTAATTAACT CTAGTAGAGN GNNNNNAANN AA--------- ----------
P4-F 27    ACACAGGTAA GAGTTTTCTG ATCTAGCTTT TTAATTAACT CTAGTAGAGC ACAAAGAAGA AAGTTTCCAT GTAACACANN
P4-R 28    ACACAGGTAA GAGTTTTNTG NN-TAGTTTT T--ATTAA-T NTAGNGGNGN NNNNA----- ---------- ----------
                  170        180        190        200        210        220        230        240

P1-F 21    NNNNNNNN-- ---------- ---- 193
P1-R 22    ---------- ---------- ---- 218
P2-F 23    NNNNNNNNNN NNNNNNNNNN NNNN 216
P2-R 24    ---------- ---------- ---- 214
P3-F 25    NNNNNNNNNN NNNNNNNNNN NNNN 216
P3-R 26    ---------- ---------- ---- 218
P4-F 27    NNNNNNN--- ---------- ---- 199
P4-R 28    ---------- ---------- ---- 210
                  250        260
```

Exon 7 (ense 00001218948)

```
P1-F 157  ---------- ---------- ---------- ---------- ---------- ---------- -------CCT ATTAAAGACT
P1-R 158  --NNNNNNNN NNNNNNTNCT NNTCNCAG-- ---ATCGGGT TTATGATCAG GACAGCCCTA TCTATATCCT ATTAAAGACT
P2-F 159  ---------- ---------- ---------- ---------- ---------- -GAAGCCCTA TCTATATCCT ATTAAAGACT
P2-R 160  -----NNNNN NNNNNNTNNN NNNTGCAN-- ---ATCGGGT TTATGATCAG GACAGCCCTA TCTATATCCT ATTAAAGACT
P3-F 161  ---------- ---------- ---------- ---------- ---------C NGAAGCCCTA TCTATATCCT ATNAAAGACT
P3-R 162  --NNNNNNNN NNNNNNTNNT NNTCNCGC-- ---ATCGGGT TTATGATCAG GACAGCCCTA TCTATATCCT ATTAAAGACT
P4-F 163  ---------- ---------- ---------- ---------- ------GNNC NGAAGCCCTA TCTATATCCT ATTAAAGACT
P4-R 164  GNNGNNNNNN NNNNNNNNNT TTTGCCAAGG NCAATCGGGT TTATGATCAG GACAGCCCTA TCTATATCCT ATTAAAGACT
                 10        20        30        40        50        60        70        80

                    ttcttttaca GCATGGACCC TTTTATGATA TGGGCACTGA AACTAAAGCA CATGGTGGAA GAAGGATTGG
P1-F 157  GCTTTTTGCC TTCTTTTACA GCATGGACCC TTTTATGATA TGGGCACTGA AACTAAAGCA CATGGTGGAA GAAGGATTGG
P1-R 158  GCTTTTTGCC TTCTTTTACA GCATGGACCC TTTTATGATA TGGGCACTGA AACTAAAGCA CATGGTGGAA GAAGGATTGG
P2-F 159  GCTTTTTGCC TTCTTTTACA GCATGGACCC TTTTATGATA TGGGCACTGA AACTAAAGCA CATGGTGGAA GAAGGATTGG
P2-R 160  GCTTTTTGCC TTCTTTTACA GCATGGACCC TTTTATGATA TGGGCACTGA AACTAAAGCA CATGGTGGAA GAAGGATTGG
P3-F 161  GCTTTTTGCC TTCTTTTACA GCATGGACCC TTTTATGATA TGGGCACTGA AACTAAAGCA CATGGTGGAA GAAGGATTGG
P3-R 162  GCTTTTTGCC TTCTTTTACA GCATGGACCC TTTTATGATA TGGGCACTGA AACTAAAGCA CATGGTGGAA GAAGGATTGG
P4-F 163  GCTTTTTGCC TTCTTTTACA GCATGGACCC TTTTATGATA TGGGCACTGA AACTAAAGCA CATGGTGGAA GAAGGATTGG
P4-R 164  GCTTTTTGCC TTCTTTTACA GCATGGACCC TTTTATGATA TGGGCACTGA AACTAAAGCA CATGGTGGAA GAAGGATTGG
                 90       100       110       120       130       140       150       160

          TAGCATATAG AAACATTTTT AGACAAATGA AAAAGCAAAA AAGTCAGAAA TTACAGTGTA TTTCCATAAA GTTACACCAA
P1-F 157  TAGCATATAG AAACATTTTT AGACAAATGA AAAAGCAAAA AAGTCAGAAA TTACAGTGTA TTTCCATAAA GTTACACCAA
P1-R 158  TAGCATATAG AAACATTTTT AGACAAATGA AAAAGCAAAA AAGTCAGAAA TTACAGTGTA TTTCCATAAA GTTACACCAA
P2-F 159  TAGCATATAG AAACATTTTT AGACAAATGA AAAAGCAAAA AAGTCAGAAA TTACAGTGTA TTTCCATAAA GTTACACCAA
P2-R 160  TAGCATATAG AAACATTTTT AGACAAATGA AAAAGCAAAA AAGTCAGAAA TTACAGTGTA TTTCCATAAA GTTACACCAA
P3-F 161  TAGCATATAG AAACATTTTT AGACAAATGA AAAAGCAAAA AAGTCAGAAA TTACAGTGTA TTTCCATAAA GTTACACCAA
P3-R 162  TAGCATATAG AAACATTTTT AGACAAATGA AAAAGCAAAA AAGTCAGAAA TTACAGTGTA TTTCCATAAA GTTACACCAA
P4-F 163  TAGCATATAG AAACATTTTT AGACAAATGA AAAAGCAAAA AAGTCAGAAA TTACAGTGTA TTTCCATAAA GTTACACCAA
P4-R 164  TAGCATATAG AAACATTTTT AGACAAATGA AAAAGCAAAA AAGTCAGAAA TTACAGTGTA TTTCCATAAA GTTACACCAA
                170       180       190       200       210       220       230       240

          GTGTGCCTGC CTCTCCTGCC TCCCCTTCCA GCTTTTTGTC TTCTGCCATT TCTGAGTCAG CAAGACCCCT CCTGTTCCTC
P1-F 157  GTGTGCCTGC CTCTCCTGCC TCCCCTTCCA GCTTTTTGTC TTCTGCCATT TCTGAGTCAG CAAGACCCCT CCTGTTCCTC
P1-R 158  GTGTGCCTGC CTCTCCTGCC TCCCCTTCCA GCTTTTTGTC TTCTGCCATT TCTGAGTCAG CAAGACCCCT CCTGTTCCTC
P2-F 159  GTGTGCCTGC CTCTCCTGCC TCCCCTTCCA GCTTTTTGTC TTCTGCCATT TCTGAGTCAG CAAGACCCCT CCTGTTCCTC
P2-R 160  GTGTGCCTGC CTCTCCTGCC TCCCCTTCCA GCTTTTTGTC TTCTGCCATT TCTGAGTCAG CAAGACCCCT CCTGTTCCTC
P3-F 161  GTGTGCCTGC CTCTCCTGCC TCCCCTTCCA GCTTTTTGTC TTCTGCCATT TCTGAGTCAG CAAGACCCCT CCTGTTCCTC
P3-R 162  GTGTGCCTGC CTCTCCTGCC TCCCCTTCCA GCTTTTTGTC TTCTGCCATT TCTGAGTCAG CAAGACCCCT CCTGTTCCTC
P4-F 163  GTGTGCCTGC CTCTCCTGCC TCCCCTTCCA GCTTTTTGTC TTCTGCCATT TCTGAGTCAG CAAGACCCCT CCTGTTCCTC
P4-R 164  GTGTGCCTGC CTNTCCTGCC TCCCCTTCCA GCTTTTTGTC TTCTGCCATT TCTGAGTCAG CAAGACCCCT CCTGTTCCTC
                250       260       270       280       290       300       310       320

          CTTCTCAGCC TACTCAGCAT GAAGACAAGG ATGAAGATCT TTGTGATGAT CCACTTCCAC TTAATGAATA gtaaatatat
P1-F 157  CTTCTCAGCC TACTCAGCAT GAAGACAAGG ATGAAGATCT TTGTGATGAT CCACTTCCAC TTAATGAATA GTAAATATAT
P1-R 158  CTTCTCAGCC TACTCAGCAT GAAGACAAGG ATGAAGATCT TTGTGATGAT CCACTTCCAC TTAATGAATA GTAAATATAT
P2-F 159  CTTCTCAGCC TACTCAGCAT GAAGACAAGG ATGAAGATCT TTGTGATGAT CCACTTCCAC TTAATGAATA GTAAATATAT
P2-R 160  CTTCTCAGCC TACTCAGCAT GAAGACAAGG ATGAAGATCT TTGTGATGAT CCACTTCCAC TTAATGAATA GTAAATATAT
P3-F 161  CTTCTCAGCC TACTCAGCAT GAAGACAAGG ATGAAGATCT TTGTGATGAT CCACTTCCAC TTAATGAATA GTAAATATAT
P3-R 162  CTTCTCAGCC TACTCAGCAT GAAGACAAGG ATGAAGATCT TTGTGATGAT CCACTTCCAC TTAATGAATA GTAAATATAT
P4-F 163  CTTCTCAGCC TACTCAGCAT GAAGACAAGG ATGAAGATCT TTGTGATGAT CCACTTCCAC TTAATGAATA GTAAATATAT
P4-R 164  CTTCTCAGCC TACTCAGCAT GAAGACAAGG ATGAAGATCT TTGTGATGAT CCACTTCCAC TTAATGAATA GTAAACNTAT
                330       340       350       360       370       380       390       400

P1-F 157  TTTCTCTCAC TTAGGATTTT CTTTAGCTTA CTTTACTGTA AGAATACAGT ATATAATACA TATACAAATA TGTTTTAGTC
P1-R 158  TTTCTCTCAC TTAGGATTTT CTTTAGCTTA CTTTACTGTA AGAATACAGT ATATAATACA TATACAAATN GTTNNGTNNN
P2-F 159  TTTCTCTCAC TTAGGATTTT CTTTAGCTTA CTTTACTGTA AGAATACAGT ATATAATACA TATACAAATA TGTTTTAGTC
P2-R 160  TTTCTCTCAC TTAGGATTTT CTTTAGCTTA CTTTACTGTA AGAATACAGT ATATAATACA TATACAAATA TGTTTTAGTC
P3-F 161  TTTCTCTCAC TTAGGATTTT CTTTAGCTTA CTTTACTGTA AGAATACAGT ATATAATACA TATACAAATA TGTTTTAGTC
P3-R 162  TTTCTCTCAC TTAGGATTTT CTTTAGCTTA CTTTACTGTA AGAATACAGT ATATAATACA TATACAAATA TGTTTTAGTC
P4-F 163  TTTCTCTCAC TTAGGATTTT CTTTAGCTTA CTTTACTGTA AGAATACAGT ATATAATACA TATACAAATA TGTTTTAGTC
P4-R 164  TTTCTCTCAC TTAGGATTTT CTTTAGCTTA CTTTACTGTA AGAATACAGT ATATAATACA TA-ACAAATN GNNNNGNNNN
                410       420       430       440       450       460       470       480

P1-F 157  AACTGTTTAT GTTATCAGCA ATGCTTGNGA AAAANNNNNN NNNN------  457
P1-R 158  NNNNNGNNNN A--------- ---------- ---------- ----------  484
P2-F 159  AACTGTTTAT GTTATCAGCA ATGCTTGCGA ANAANNNNNN NNN-------  472
P2-R 160  AAC------- ---------- ---------- ---------- ----------  473
P3-F 161  AACTGTTTAT GTTATCAGCA ATGCTNGCNA ANAANNANNN NNN-------  474
P3-R 162  AAC------- ---------- ---------- ---------- ----------  476
P4-F 163  AACTGTTTAT GTTATCAGCA ATGCTCNGTN ANNNANNNNN NNNNNNNNNN  484
P4-R 164  NNNNNGGNNN AA-------- ---------- ---------- ----------  491
                490       500       510       520       530
```

Exon 8 (ense 00000958561)
(P3-F 33 sequence was non-specific only the reverse sequence is included)

```
P1-F 183  ---------- ---------- ---------- ---------- --TTNNACCN NNNNATGCTT -ATTTA-TTT ATTTCTTTTA
P1-R 184  NNNNNNNNNN NNNNNNNNNN TTGGTGCCTG ATGAATGTTG GTAACTGAAA TCTAGTGCTT TATTTCATTT ATTTCTTTTA
P2-F 31   ---------- ---------- ---------- ---------- -----NNNNA NTCTATGCTT TATTTA-TTT ATTTCTTTTA
P2-R 32   NNNNNNNNNN NNNNNNNNNN TTGGTGCCTG ATGAATGTTG GTAACTGAAA TCTAGTGCTT TATTTCATTT ATTTCTTTTA
P3-R 34   NNNNNNNNNN NNNNNNNNNN TTGGTGCCTG ATGAATGTTG GTAACTGAAA TCTAGTGCTT TATTTCATTT ATTTCTTTTA
P4-F 35   ---------- ---------- ---------- ---------- --------AA TCTAGTGCTT TATTTNATTT ATTTCTTTTA
P4-R 36   NNNNNNNNNN NNNNNNNNNN TTGGTGCCTG ATGAATGTTG GTAACTGAAA TCTAGTGCTT TATTTCATTT ATTTCTTTTA
                 10         20         30         40         50         60         70         80

REF                              tgctttctag GCACACAAAC CAATGGTCTG GACTTTCAGA AGCAGCCTGT GCCTGTAGGA
C-B-F 175                                    GCACACAAAC CAATGGTCTG GACTTTCAGA AGCAGCCTGT GCCTGTAGGA
C-B-R 176                                    GCACACAAAC CAATGGTCTG GACTTTCAGA AGCAGCCTGT -CCTGTAGGA
P1-F 183  ATAGTTGAAA TTATTTTGCT TGCTTTCTAG GCACACAAAC CAATGGTCTG GACTTTCAGA AGCAGCCTGT GCCTGTAGGA
P1-R 184  ATAGTTGAAA TTATTTTGCT TGCTTTCTAG GCACACAAAC CAATGGTCTG GACTTTCAGA AGCAGCCTGT GCCTGTAGGA
P2-F 31   ATAGTTGAAA TTATTTTGCT TGCTTTCTAG GCACACAAAC CAATGGTCTG GACTTTCAGA AGCAGCCTGT GCCTGTAGGA
P2-R 32   ATAGTTGAAA TTATTTTGCT TGCTTTCTAG GCACACAAAC CAATGGTCTG GACTTTCAGA AGCAGCCTGT GCCTGTAGGA
P3-R 34   ATAGTTGAAA TTATTTTGCT TGCTTTCTAG GCACACAAAC CAATGGTCTG GACTTTCAGA AGCAGCCTGT GCCTGTAGGA
P4-F 35   ATAGTTGAAA TTATTTTGCT TGCTTTCTAG GCACACAAAC CAATGGTCTG GACTTTCAGA AGCAGCCTGT GCCTGTAGGA
P4-R 36   ATAGTTGAAA TTATTTTGCT TGCTTTCTAG GCACACAAAC CAATGGTCTG GACTTTCAGA AGCAGCCTGT GCCTGTAGGA
                 90        100        110        120        130        140        150        160

          GGAGCAATCT CAACAGCCCA GGCGCAGGCT TTCCTTGGAC ATCTCCATCA Ggtaggaatg t
C-B-F 175 GGAGCAATCT CAACAGCCCN GGCGCAGGCT TTCCTTGGAC ATCTCCATCA G
C-B-R 176 GGAGCAATCT CAACAGCCCA GGCGCAGGCT TTCCTTGGAC ATCTCCATCA G
P1-F 183  GGAGCAATCT CAACAGCCCA GGCGCAGGCT TTCCTTGGAC ATCTCCATCA GGTAGGAATG TTCTGCTCAA CCATCAGTGA
P1-R 184  GGAGCAATCT CAACAGCCCA GGCGCAGGCT TTCCTTGGAC ATCTCCATCA GGTAGGAATG TTCTGCTCAA CCATCAGTGA
P2-F 31   GGAGCAATCT CAACAGCCCA GGCGCAGGCT TTCCTTGGAC ATCTCCATCA GGTAGGAATG TTCTGCTCAA CCATCAGTGA
P2-R 32   GGAGCAATCT CAACAGCCCA GGCGCAGGCT TTCCTTGGAC ATCTCCATCA GGTAGGAATG TTCTGCTCAA CCATCAGTNA
P3-R 34   GGAGCAATCT CAACAGCCCA GGCGCAGGCT TTCCTTGGAC ATCTCCATCA GGTAGGAATG TTCTGCTCAA CCATCAGN-A
P4-F 35   GGAGCAATCT CAACAGCCCA GGCGCAGGCT TTCCTTGGAC ATCTCCATCA GGTAGGAATG TTCTGCTCAA CCATCAGTGA
P4-R 36   GGAGCAATCT CAACAGCCCA GGCGCAGGCT TTCCTTGGAC ATCTCCATCA GGTAGGAATG TTCTGCTCAA CCATCAGTGA
                 170        180        190        200        210        220        230        240

P1-F 183  GAGTGAAAGA TAGAGGCGTA AAGTACTGGG NGCTGAAANN NNNNN----- ---------- -- 241
P1-R 184  GAGNAAAGAN GNGGNCANAA ---------- ---------- ---------- ---------- -- 260
P2-F 31   GAGTGAAAGA TANAGGCGTA AAGTACTGGG GCCCTGAACC ANNNNNNNNN NNNNNNNNNN NN 256
P2-R 32   GAGTGAAAGN AAG------- ---------- ---------- ---------- ---------- -- 253
P3-R 34   GAGNAAAGAN G--------- ---------- ---------- ---------- ---------- -- 250
P4-F 35   GAGTGAAAGA TAGAGGCGTA AAGTACTGGG NNCTGAAANN NNNNNNNNNN NNNNNNNNNN -- 252
P4-R 36   GAGTAAAGA- ---------- ---------- ---------- ---------- ---------- -- 249
                 250        260        270        280        290        300
```

Exon 9 (ense 00000958562)

```
P1-F 37    ---------- ---------- -------TTA ANGGNTGAGC TTGTTCATAA AGTTTTAACT CCGTAAGACA AACTTCTTTG
P1-R 38    -NNNNNNNNN TTTGGGNTNN TGGTTGAGAA GGGACTGAGC TTGTTCATAA AGTTTTAACT CCGTAAGACA AACTTCTTTG
P2-F 39    ---------- ---------- -------TAA ANGGNTGAGC TTGTTCATAA AGTTTTAACT CCGTAAGACA AACTTCTTTG
P2-R 40    -NNNNNNNNT TTGGGGGTNN TGGTTGAGAA GGGACTGAGC TTGTTCATAA AGTTTTAACT CCGTAAGACA AACTTCTTTG
P3-F 41    ---------- ---------- --TAATNAAA NGGNCTGAGC TTGTTCATAA AGTTTTAACT CCGTAAGACA AACTTCTTTG
P3-R 42    NNNNNNNNTT TTGGGGNNNC TGGTTGAGAA GGGACTGAGC TTGTTCATAA AGTTTTAACT CCGTAAGACA AACTTCTTTG
P4-F 43    ---------- ---------- -------TNN NNGGNTGAGC TTGTTCATAA AGTTTTAACT CCGTAAGACA AACTTCTTTG
P4-R 44    -NNNNNNNNT TTGGGGGTNC TGGTTGAGAA GGGACTGAGC TTGTTCATAA AGTTTTAACT CCGTAAGACA AACTTCTTTG
                  10         20         30         40         50         60         70         80

P1-F 37    TCTTCTCAGT TCTAGCTCGG TGACACATAG TAGACTTTAT TTAGTGATGA CAACCCCTTG TGATGAATGT CAACAGTATT
P1-R 38    TCTTCTCAGT TCTAGCTCGG TGACACATAG TAGACTTTAT TTAGTGATGA CAACCCCTTG TGATGAATGT CAACAGTATT
P2-F 39    TCTTCTCAGT TCTAGCTCGG TGACACATAG TAGACTTTAT TTAGTGATGA CAACCCCTTG TGATGAATGT CAACAGTATT
P2-R 40    TCTTCTCAGT TCTAGCTCGG TGACACATAG TAGACTTTAT TTAGTGATGA CAACCCCTTG TGATGAATGT CAACAGTATT
P3-F 41    TCTTCTCAGT TCTAGCTCGG TGACACATAG TAGACTTTAT TTAGTGATGA CAACCCCTTG TGATGAATGT CAACAGTATT
P3-R 42    TCTTCTCAGT TCTAGCTCGG TGACACATAG TAGACTTTAT TTAGTGATGA CAACCCCTTG TGATGAATGT CAACAGTATT
P4-F 43    TCTTCTCAGT TCTAGCTCGG TGACACATAG TAGACTTTAT TTAGTGATGA CAACCCCTTG TGATGAATGT CAACAGTATT
P4-R 44    TCTTCTCAGT TCTAGCTCGG TGACACATAG TAGACTTTAT TTAGTGATGA CAACCCCTTG TGATGAATGT CAACAGTATT
                  90        100        110        120        130        140        150        160

                                t tacttatagG TCCAACTCGC TGGAACAAGT TTACAGGCTG CTGCTCAGTC TTTAAATGTA
C-B-F 175                               G TCCAACTCGC TGGAACAAGT TTACAGGCTG CTGCTCAGTC TTTAAATGTA
C-B-R 176                               G TCCAACTCGC TGGAACAAGT TTACAGGCTG CTGCTCAGTC TTTAAATGTA
P1-F 37    AAACTAGAAC TTCCCCTGAT TACTTATAGG TCCAACTCGC TGGAACAAGT TTACAGGCTG CTGCTCAGTC TTTAAATGTA
P1-R 38    AAACTAGAAC TTCCCCTGAT TACTTATAGG TCCAACTCGC TGGAACAAGT TTACAGGCTG CTGCTCAGTC TTTAAATGTA
P2-F 39    AAACTAGAAC TTCCCCTGAT TACTTATAGG TCCAACTCGC TGGAACAAGT TTACAGGCTG CTGCTCAGTC TTTAAATGTA
P2-R 40    AAACTAGAAC TTCCCCTGAT TACTTATAGG TCCAACTCGC TGGAACAAGT TTACAGGCTG CTGCTCAGTC TTTAAATGTA
P3-F 41    AAACTAGAAC TTCCCCTGAT TACTTATAGG TCCAACTCGC TGGAACAAGT TTACAGGCTG CTGCTCAGTC TTTAAATGTA
P3-R 42    AAACTAGAAC TTCCCCTGAT TACTTATAGG TCCAACTCGC TGGAACAAGT TTACAGGCTG CTGCTCAGTC TTTAAATGTA
P4-F 43    AAACTAGAAC TTCCCCTGAT TACTTATAGG TCCAACTCGC TGGAACAAGT TTACAGGCTG CTGCTCAGTC TTTAAATGTA
P4-R 44    AAACTAGAAC TTCCCCTGAT TACTTATAGG TCCAACTCGC TGGAACAAGT TTACAGGCTG CTGCTCAGTC TTTAAATGTA
                  170        180        190        200        210        220        230        240

           CAGgtaagct ggg
C-B-F 175  CAG
C-B-R 176  CAG
P1-F 37    CAGGTAAGCT GGGACCTGGG ATTATGGGTC AATCTTTTAT TTATTTTTTC TTATATTTTT CTGTAGTGAG CATATATTTT
P1-R 38    CAGGTAAGCT GGGACCTGGG ATTATGGGTC AATCTTTTAT TTATTTTTTC TTATATTTTT CTGTAGTGAG CATATATTTT
P2-F 39    CAGGTAAGCT GGGACCTGGG ATTATGGGTC AATCTTTTAT TTATTTTTTC TTATATTTTT CTGTAGTGAG CATATATTTT
P2-R 40    CAGGTAAGCT GGGACCTGGG ATTATGGGTC AATCTTTTAT TTATTTTTTC TTATATTTTT CTGTAGTGAG CATATATTTT
P3-F 41    CAGGTAAGCT GGGACCTGGG ATTATGGGTC AATCTTTTAT TTATTTTTTC TTATATTTTT CTGTAGTGAG CATATATTTT
P3-R 42    CAGGTAAGCT GGGACCTGGG ATTATGGGTC AATCTTTTAT TTATTTTTTC TTATATTTTT CTGTAGTGAG CATATATTTT
P4-F 43    CAGGTAAGCT GGGACCTGGG ATTATGGGTC AATCTTTTAT TTATTTTTTC TTATATTTTT CTGTAGTGAG CATATATTTT
P4-R 44    CAGGTAAGCT GGGACCTGGG ATTATGGGTC AATCTTTTAT TTATTTTTTC TTATATTTTT CTGTAGTGAG CATATATTTT
                  250        260        270        280        290        300        310        320

P1-F 37    ATAATAGGAA ATGCTTAGAA GTATTACCTT TTCAAAATTA GTGAATCTCG TGAAGATTCA AATAATGATA ATTATTACCA
P1-R 38    ATAATAGGAA ATGCTTAGAA GTATTACCTT TTCAAAATTA GTGAATCTCG TGAAGATTCA AATAATGATA ATTATTACCA
P2-F 39    ATAATAGGAA ATGCTTAGAA GTATTACCTT TTCAAAATTA GTGAATCTCG TGAAGATTCA AATAATGATA ATTATTACCA
P2-R 40    ATAATAGGAA ATGCTTAGAA GTATTACCTT TTCAAAATTA GTGAATCTCG TGAAGATTCA AATAATGATA ATTATTACCA
P3-F 41    ATAATAGGAA ATGCTTAGAA GTATTACCTT TTCAAAATTA GTGAATCTCG TGAAGATTCA AATAATGATA ATTATTACCA
P3-R 42    ATAATAGGAA ATGCTTAGAA GTATTACCTT TTCAAAATTA GTGAATCTCG TGAAGATTCA AATAATGATA ATTATTACCA
P4-F 43    ATAATAGGAA ATGCTTAGAA GTATTACCTT TTCAAAATTA GTGAATCTCG TGAAGATTCA AATAATGATA ATTATTACCA
P4-R 44    ATAATAGGAA ATGCTTAGAA GTATTACCTT TTCAAAATTA GTGAATCTCG TGAAGATTCA AATAATGATA ATTATTACCA
                  330        340        350        360        370        380        390        400

P1-F 37    TTAATGAAAA TCAGTTTTAT ATTATGCTAC TGTGCATGAA CCAAAAACNN GGANANANNN NNN- 436
P1-R 38    TTAATGAAAA TCAGTTTTAT ATT-NGNNNN NNGNNNA--- ---------- ---------- ---- 435
P2-F 39    TTAATGAAAA TCAGTTTTAT ATTNTGCTAC TGTGCATGAA CCAAAAAGGG GGAAAAANNN NNNN 437
P2-R 40    TTAAT-AAAA TCAGTTTTAT ATT-NGCNAC ---------- ---------- ---------- ---- 427
P3-F 41    TTAATGAAAA TCAGTTTTAT ATTNTGCTAC TGTGCATGAA CCAAAAATGN GGGGAAAANN NNNN 442
P3-R 42    TTAATTAAAA TCAGTTTTAT ATTATGCNNC NCNGNNNANA A--------- ---------- ---- 441
P4-F 43    TTAATGAAAA TCAGTTTTAT ATTCTGCTAC TGTGCATGAA CCAAAAACNG GGGAAAANNN NNN- 436
P4-R 44    TTAATTAAAA TCAGTTTTAT ATT-CGNNGG NNGNCAANAA ---------- ---------- ---- 438
                  410        420        430        440        450        460
```

Exon 10 (ense 00000958563)

```
P1-F 45    ---------- ---------- ---------- TNTNTTGANN NCNTACATTT CTTTTAATCA ACNATTTGCA ATCTTTTATT
P1-R 46    NNNNNNNNNN NNNNNNTTGA TTCCTAAAAG ATGGGGTTTT AAGTACATTT CTTTTAATCA ACCATTTGCA ATCTTTTATT
P2-F 47    ---------- ---------- ---------- -TTTTTNCCN AAGTACATTT CTTTTA--AT NACCTTTGCA ATCTTTTATT
P2-R 48    NNANGANNGC NNATNGTTGA TTCCTAAAAG ATGGGGTTTT AAGTACATTT CTTTTAATCA ACCATTTGCA ATCTTTTATT
P3-F 49    ---------- ---------- ---------- ---TNTTNNN NNANGTNATT TTTTTA--TN ACCATTTGCA ATCTTTTATT
P3-R 50    NNANGANNGC NNANNGTTGA TTCCTAAAAG ATGGGGTTTT AAGTACATTT CTTTTAATCA ACCATTTGCA ATCTTTTATT
P4-F 51    ---------- ---------- ---------- TNTTTGNCNN AAGTACATTT CTTTTAATNA ACCATTTGCA ATCTTTTATT
P4-R 52    NNAAGANNGC NANTNGTTGA TTCCTAAAAG ATGGGGTTTT AAGTACATTT CTTTTAATCA ACCATTTGCA ATCTTTTATT
                  10         20         30         40         50         60         70         80

REF                cccacctcag TCTAAATCTA ATGAAGAATC GGGGGATTCG CAGCAGCCAA GCCAGCCTTC CCAGCAGCCT
C-B-F 175                    TCTAAATCTA ATGAAGAATC GGGGGATTCG CAGCAGCCAA GCCAGCCTTC CCAGCAGCCT
C-B-R 176                    TCTAAATCTA ATGAAGAATC GGGGGATTCG CAGCAGCCAA GCCAGCCTTC CCAGCAGCCT
P1-F 45    TCCTACCCAC CCCACCTCAG TCTAAATCTA ATGAAGAATC GGGGGATTCG CAGCAGCCAA GCCAGCCTTC CCAGCAGCCT
P1-R 46    TCCTACCCAC CCCACCTCAG TCTAAATCTA ATGAAGAATC GGGGGATTCG CAGCAGCCAA GCCAGCCTTC CCAGCAGCCT
P2-F 47    TCCTACCCAC CCCACCTCAG TCTAAATCTA ATGAAGAATC GGGGGATTCG CAGCAGCCAA GCCAGCCTTC CCAGCAGCCT
P2-R 48    TCCTACCCAC CCCACCTCAG TCTAAATCTA ATGAAGAATC GGGGGATTCG CAGCAGCCAA GCCAGCCTTC CCAGCAGCCT
P3-F 49    TCCTACCCAC CCCACCTCAG TCTAAATCTA ATGAAGAATC GGGGGATTCG CAGCAGCCAA GCCAGCCTTC CCAGCAGCCT
P3-R 50    TCCTACCCAC CCCACCTCAG TCTAAATCTA ATGAAGAATC GGGGGATTCG CAGCAGCCAA GCCAGCCTTC CCAGCAGCCT
P4-F 51    TCCTACCCAC CCCACCTCAG TCTAAATCTA ATGAAGAATC GGGGGATTCG CAGCAGCCAA GCCAGCCTTC CCAGCAGCCT
P4-R 52    TCCTACCCAC CCCACCTCAG TCTAAATCTA ATGAAGAATC GGGGGATTCG CAGCAGCCAA GCCAGCCTTC CCAGCAGCCT
                  90        100        110        120        130        140        150        160

           TCAGTGCAGG CAGCCATTCC CCAGACCCAG CTTATGCTAG CTGGAGGACA GATAACTGGG gtaagtgttc
C-B-F 175  TCAGTGCAGG CAGCCATTCC CCAGACCCAG CTTATGCTAG CTGGAGGACA GATAACTGGG
C-B-R 176  TCAGTGCAGG CAGCCATTCC CCAGACCCAG CTTATGCTAG CTGGAGGACA GATAACTGGG
P1-F 45    TCAGTGCAGG CAGCCATTCC CCAGACCCAG CTTATGCTAG CTGGAGGACA GATAACTGGG GTAAGTGTTC ACTGAGAGAA
P1-R 46    TCAGTGCAGG CAGCCATTCC CCAGACCCAG CTTATGCTAG CTGGAGGACA GATAACTGGG GTAAGTGTTC ACG-AGAGAA
P2-F 47    TCAGTGCAGG CAGCCATTCC CCAGACCCAG CTTATGCTAG CTGGAGGACA GATAACTGGG GTAAGTGTTC ACTGAGAGAA
P2-R 48    TCAGTGCAGG CAGCCATTCC CCAGACCCAG CTTATGCTAG CTGGAGGACA GATAACTGGG GTAAGTGTTC ACN-AGAGAA
P3-F 49    TCAGTGCAGG CAGCCATTCC CCAGACCCAG CTTATGCTAG CTGGAGGACA GATAACTGGG GTAAGTGTTC ACTGAGAGAA
P3-R 50    TCAGTGCAGG CAGCCATTCC CCAGACCCAG CTTATGCTAG CTGGAGGACA GATAACTGGG GTAAGTGTTC ACG-AGAGA-
P4-F 51    TCAGTGCAGG CAGCCATTCC CCAGACCCAG CTTATGCTAG CTGGAGGACA GATAACTGGG GTAAGTGTTC ACTGAGAGAA
P4-R 52    TCAGTGCAGG CAGCCATTCC CCAGACCCAG CTTATGCTAG CTGGAGGACA GATAACTGGG GTAAGTGTTC ACTGAGAGAA
                  170        180        190        200        210        220        230        240

P1-F 45    TTATAACAAA CTTTTTCTGT GTCAGAACTG CCATTGGANN NNNNNNNNNN NN-- 262
P1-R 46    GGATN-CGGN NG--NCAAAA ---------- ---------- ---------- ---- 256
P2-F 47    TTATAACAAA CTTTTTCTGT GTCAGAACTG CCATTGGAAN NNNNNNNNNN NNN- 260
P2-R 48    TTATAACAAA CNGTTCANAA ---------- ---------- ---------- ---- 259
P3-F 49    TTATAACAAA CTTTTTCTGT GTCAGAACTG CCATTTGGAN NNNNNNNNNN NNNN 259
P3-R 50    ---------- ---------- ---------- ---------- ---------- ---- 238
P4-F 51    TTATAACAAA CTTTTTCTGT GTCAGAACTG CCATTGGAAA ANNNNNNNNN NN-- 262
P4-R 52    TTATAACAAA CT-TTCANAA ---------- ---------- ---------- ---- 259
                  250        260        270        280        290
```

Exon 11 (ense 00000958564)

```
P1-F 53    ---------- ---------- ---------- ---------- ---------- ---------- TTTCC-ATAA TGTGTTCTGG
P1-R 54    ---------- ---NNNNNTT TTAGCAGTG- CTTGCATTAG GAGACTTTCT CCTTCAACAC TTTCCCATAA TGTGTTCTGG
P2-F 55    ---------- ---------- ---------- ---------- -TNNNNANTT TCTCTTNAAC TTTCC-ATAA TGTGTTCTGG
P2-R 56    ---------- --NNNNNTTT TANGCAGTG- CTTGCATTAG GAGACTTTCT CCTTCAACAC TTTCCCATAA TGTGTTCTGG
P3-F 57    ---------- ---------- ---------- ---------- -----NTTTC TCCTTNACAC TTTCC-ATAA TGTGTTCTGG
P3-R 58    NNNNNNNNNN NNNNNNNNNT TTAGCAGNGG CTTGCATTAG GAGACTTTCT CCTTCAACAC TTTCCCATAA TGTGTTCTGG
P4-F 59    ---------- ---------- ---------- ------TNTT NNNNANTTTC TCNTTNACAC TTTCC-ATAA TGTGTTCTGG
P4-R 60    ---------- --NNNNNTNN TANGCAGTG- CTTGCATTAG GAGACTTTCT CCTTCAACAC TTTCCCATAA TGTGTTCTGG
                  10         20         30         40         50         60         70         80

           ttttgtttag CTTACTTTGA CGCCTGCCCA GCAACAGTTA CTACTCCAGC AGGCACAGGC ACAGGCACAG CTGCTGGCTG
C-B-F 175             CTTACTTTGA CGCCTGCCCA GCAACAGTTA CTACTCCAGC AGGCACAGGC ACAGGCACAG CTGCTGGCTG
C-B-R 176             CTTACTTTGA CGCCTGCCCA GCAACAGTTA CTACTCCAGC AGGCACAGGC ACAGGCACAG CTGCTGGCTG
C-C-F 177                        ---- ---------- ---------- ---NNTTGC- CTTCGCACAG CCGCTGGCTG
C-C-R 178                       GNNN NNNNNNNTCC CTTNCNACCG GGGCCNGGNA CANGGCACAG NTGTGGGTTG
P1-F 53    TTTTGTTTAG CTTACTTTGA CGCCTGCCCA GCAACAGTTA CTACTCCAGC AGGCACAGGC ACAGGCACAG CTGCTGGCTG
P1-R 54    TTTTGTTTAG CTTACTTTGA CGCCTGCCCA GCAACAGTTA CTACTCCAGC AGGCACAGGC ACAGGCACAG CTGCTGGCTG
P2-F 55    TTTTGTTTAG CTTACTTTGA CGCCTGCCCA GCAACAGTTA CTACTCCAGC AGGCACAGGC ACAGGCACAG CTGCTGGCTG
P2-R 56    TTTTGTTTAG CTTACTTTGA CGCCTGCCCA GCAACAGTTA NTACTCCAGC AGGCACAGGC ACAGGCACAG CTGCTGGCTG
P3-F 57    TTTTGTTTAG CTTACTTTGA CGCCTGCCCA GCAACAGTTA CTACTCCAGC AGGCACAGGC ACAGGCACAG CTGCTGGCTG
P3-R 58    TTTTGTTTAG CTTACTTTGA CGCCTGCCCA GCAACAGTTA NTACTCCAGC AGGCACAGGC ACAGGCACAG NTGCTGGCTG
P4-F 59    TTTTGTTTAG CTTACTTTGA CGCCTGCCCA GCAACAGTTA CTACTCCAGC AGGCACAGGC ACAGGCACAG CTGCTGGCTG
P4-R 60    TTTTGTTTAG CTTACTTTGA CGCCTGCCCA GCAACAGTTA CTACTCCAGC AGGCACAGGC ACAGGCACAG CTGCTGGCTG
                  90        100        110        120        130        140        150        160

           CTGCAGTGCA GCAGCACTCC GCCAGCCAGC AGCACAGTGC TGCTGGAG-C CACCATCT-C CGCCTCTG-C TGCCACGCCC
C-B-F 175  CTGCAGTGCA GCAGCACTCC GCCAGCCAGC AGCACAGTGC TGCTGGAG-C CACCATCT-C CGCCTCTG-C TGCCACGCCC
C-B-R 176  CTGCAGTGCA GCAGCACTCC GCCAGCCAGC AGCACAGTGC TGCTGGAG-C CACCATCT-C CGCCTNTG-N TGCCACGCCC
C-C-F 177  CTGCAGTGCA GCAGCACTCC GCCAGCCAGC AGCACAGTGC TGCTGGAG-C CACCATCT-C CGCCTCTG-C TGCCACGCNN
C-C-R 178  TT-CAGTGCA ACAGCANTNC GCCAGNCAGC AGCACAGTGG TGGTNGAGGC CCCCATTTTC CGCNTTTGGN TGCCACGGCC
P1-F 53    CTGCAGTGCA GCAGCACTCC GCCAGCCAGC AGCACAGTGC TGCTGGAG-C CACCATCT-C CGCCTCTG-C TGCCACGCCC
P1-R 54    CTGCAGTGCA GCAGCACTCC GCCAGCCAGC AGCACAGTGC TGCTGGAG-C CACCATCT-C CGCCTCTG-C TGCCACGCCC
P2-F 55    CTGCAGTGCA GCAGCACTCC GCCAGCCAGC AGCACAGTGC TGCTGGAG-C CACCATCT-C CGCCTCTG-C TGCCACGCCC
P2-R 56    CTGCAGTGCA GCAGCACTCC GCCAGCCAGC AGCACAGTGC TGCTGGAG-C CACCATCT-C CGCCTNTG-C TGCCACGCCC
P3-F 57    CTGCAGTGCA GCAGCACTCC GCCAGCCAGC AGCACAGTGC TGCTGGAG-C CACCATCT-C CGCCTCTG-C TGCCACGCCC
P3-R 58    CTGCAGTGCA GCAGCACTCC GCCAGCCAGC AGCACAGTGC TGCTGGAG-C CACCATNT-C CGCCTNTG-C TGCCACGCCC
P4-F 59    CTGCAGTGCA GCAGCACTCC GCCAGCCAGC AGCACAGTGC TGCTGGAG-C CACCATCT-C CGCCTCTG-C TGCCACGCCC
P4-R 60    CTGCAGTGCA GCAGCACTCC GCCAGCCAGC AGCACAGTGC TGCTGGAG-C CACCATCT-C CGCCTNTG-C TGCCACGCCC
                  170        180        190        200        210        220        230        240

           ATGACGCAGA TCCCCCTGTC TCAGCCCATA CAGATCGCAC AGgtgagtga gg
C-B-F 175  ATGACGCAGA TCCCCCTGTC TCAGCCCATC GTNGGAGAAA NN
C-B-R 176  ATGACGCAGA TCCCC----- ---------- ---------- --
C-C-F 177  GGGACGCAGA TCCCCCTGTC TCAGCCCATA CAGATCGCAC AG
C-C-R 178  ATGACGCAGA TCCCCCTGTC TCAGCCCATA CAGATCGCAC AG
P1-F 53    ATGACGCAGA TCCCCCTGTC TCAGCCCATA CAGATCGCAC AGGTGAGTGA GGAACTCCAA TAGCTGGGGC AGCAAGTGAG
P1-R 54    ATGACGCAGA TCCCCCTGTC TCAGCCCATA CAGATCGCAC AGGTGAGTGA GGAACTCCAA TAGCTGGGGC AGCAAGTGAG
P2-F 55    ATGACGCAGA TCCCCCTGTC TCAGCCCATA CAGATCGCAC AGGTGAGTGA GGAACTCCAA TAGCTGGGGC AGCAAGTGAG
P2-R 56    ATGACGCAGA TCCCCCTGTC TCAGCCCATA CAGATCGCAC AGGTGAGTGA GGAACTCCAA TAGCTGGGGC AGCAAGTGAG
P3-F 57    ATGACGCAGA TCCCCCTGTC TCAGCCCATA CAGATCGCAC AGGTGAGTGA GGAACTCCAA TAGCTGGGGC AGCAAGTGAG
P3-R 58    ATGACGCAGA TCCCCCTGTC TCAGCCCATA CAGATCGCAC AGGTGAGTGA GGAACTCCAA TAGCTGGGGC AGCAAGTGAG
P4-F 59    ATGACGCAGA TCCCCCTGTC TCAGCCCATA CAGATCGCAC AGGTGAGTGA GGAACTCCAA TAGCTGGGGC AGCAAGTGAG
P4-R 60    ATGACGCAGA TCCCCCTGTC TCAGCCCATA CAGATCGCAC AGGTGAGTGA GGAACTCCAA TAGCTGGGGC AGCAAGTGAG
                  250        260        270        280        290        300        310        320

P1-F 53    GAATAAAGTG GCAGGTTTTA TTTAATGTTA GATTAACTTT TTGATTGTTA GTGTGNGCCC CTAAAANNNN NNNNNNNNNN
P1-R 54    GAATAAAGTG GCAGGTNN-A TNA------- ---------- ---------- ---------- ---------- ----------
P2-F 55    GAATAAAGTG GCAGGTTTTA TTTAATGTTA GATTAACTTT TTGATTGTTA GTGTGNGCCC CTAAAANNNN NNNNNNNNNN
P2-R 56    GAATAAAGTN GCAGGTTTTA TTTAA----- ---------- ---------- ---------- ---------- ----------
P3-F 57    GAATAAAGTG GCAGGTTTTA TTTAATGTTA GATTAACTTT TTGATTGTTA GTG-GNGCCC CTAAAANNNN NNNNNNNNNN
P3-R 58    GAATAAAGTG GCAGGTTTTA TTTAATGTTA GATTAAGTTT NCAA------ ---------- ---------- ----------
P4-F 59    GAATAAAGTG GCAGGTTTTA TTTAATGTTA GATTAACTTT TTGATTGTTA GTGTGNGCCC TAAAANNNNN NNNNNNNNNN
P4-R 60    GAATAAAGTG GCAGGTTTTA TTTAANT--- ---------- ---------- ---------- ---------- ----------
                  330        340        350        360        370        380        390        400

P1-F 53    NNNN 340
P1-R 54    ---- 325
P2-F 55    NNNN 359
P2-R 56    ---- 329
P3-F 57    NNNN 354
P3-R 58    ---- 361
P4-F 59    NN-- 362
P4-R 60    ---- 331
```

Exon 12 (ense 00000958565)

```
P1-F 61    ---------- ---------- ---------- ---------- ------ANCA ACTTTAT--- TTANAATCTC NA-TCNATGT
P1-R 62    NNNNNNNNNN NTTTCCAGAA ATCATCAGCT GGAAGCCTTA TAATTAAGCG AACTTTTATT TCAGAATCTC CAATCCATGT
P2-F 63    ---------- ---------- ---------- ---------- ------AGCG AACTTTTATT TCAGAATCTC CA-TCNATGT
P2-R 64    NNNNNNNNNN NTNTCCAGAA AGCATCAGCT GGAAGCCTTA TAATTAAGCG AACTTTTATT TCAGAATCTC CAATCCATGT
P3-F 65    ---------- ---------- ---------- --------CN NNNATTAGCG ANCTTTTATT TNAGAATCTC CAATCCATGT
P3-R 66    NNNNNNNNNN NTTTCCAGAA ATCATCAGCT GGAAGCCTTA TAATTAAGCG ANCTTTTATT TCAGAATCTC CAATCCATGT
P4-F 67    ---------- ---------- ---------- ---------- -----TANCA ACTTTAT--- TTANAATCTC CA-TCNATGT
P4-R 68    -NNNNNNNNN NNNTCCAGAN AGCATCAGCT GGAAGCCTTA TAATTAAGCG ANCTTTTATT TCAGAATCTC CAATCCATGT
                  10         20         30         40         50         60         70         80

                    a atttttcagG ATCTTCAACA ACTGCAACAG CTTCAACAGC AGAATCTCAA CCTGCAACAG TTTGTGTTGG
C-C-F 177                    G ATCTTCAACA ACTGCAACAG CTTCAACAGC AGAATCTCAA CCTGCAACAG TTTGTGTTGG
C-C-R 178                    G ATCTTCAACA ANTGCAACAG GTTCAACAGC AGAATCTCAA CCTGCAACAG TTTGTGTTGG
P1-F 61    TTTAATTCCA -TTTTTCAGG ATCTTCAACA ACTGCAACAG CTTCAACAGC AGAATCTCAA CCTGCAACAG TTTGTGTTGG
P1-R 62    TTTAATTCCA ATTTTTCAGG ATCTTCAACA ACTGCAACAG CTTCAACAGC AGAATCTCAA CCTGCAACAG TTTGTGTTGG
P2-F 63    TTTAATTCCA ATTTTTCAGG ATCTTCAACA ACTGCAACAG CTTCAACAGC AGAATCTCAA CCTGCAACAG TTTGTGTTGG
P2-R 64    TTTAATTCCA ATTTTTCAGG ATCTTCAACA ACTGCAACAG CTTCAACAGC AGAATCTCAA CCTGCAACAG TTTGTGTTGG
P3-F 65    TTTAATTCCA ATTTTTCAGG ATCTTCAACA ACTGCAACAG CTTCAACAGC AGAATCTCAA CCTGCAACAG TTTGTGTTGG
P3-R 66    TTTAATTCCA ATTTTTCAGG ATCTTCAACA ACTGCAACAG CTTCAACAGC AGAATCTCAA CCTGCAACAG TTTGTGTTGG
P4-F 67    TTTAATTCCA ATTTTTCAGG ATCTTCAACA ACTGCAACAG CTTCAACAGC AGAATCTCAA CCTGCAACAG TTTGTGTTGG
P4-R 68    TTTAATTCCA ATTTTTCAGG ATCTTCAACA ACTGCAACAG CTTCAACAGC AGAATCTCAA CCTGCAACAG TTTGTGTTGG
                  90        100        110        120        130        140        150        160

           TGCATCCAAC CACCAATTTG CAGCCAGCGC AGTTTATCAT CTCACAGACG CCCCAGGGCC AGCAGGgtga gctcct
C-C-F 177  TGCATCCAAC CACCAATTTG CAGCCAGCGC AGTTTATCAT CTCACAGACG CCCCAGGGCC AGCAGG
C-C-R 178  TGCATCCAAC CACCAATTTG CAGCCAGCGC AGTTTATCAT CTCACAGACG CCCCAGGGCC AGCAGG
P1-F 61    TGCATCCAAC CACCAATTTG CAGCCAGCGC AGTTTATCAT CTCACAGACG CCCCAGGGCC AGCAGGGTGA GCTCCTCCTT
P1-R 62    TGCATCCAAC CACCAATTTG CAGCCAGCGC AGTTTATCAT CTCACAGACG CCCCAGGGCC AGCAGGGTGA GCTCCTCCTT
P2-F 63    TGCATCCAAC CACCAATTTG CAGCCAGCGC AGTTTATCAT CTCACAGACG CCCCAGGGCC AGCAGGGTGA GCTCCTCCTT
P2-R 64    TGCATCCAAC CACCAATTTG CAGCCAGCGC AGTTTATCAT CTCACAGACG CCCCAGGGCC AGCAGGGTGA GCTCCTCCTT
P3-F 65    TGCATCCAAC CACCAATTTG CAGCCAGCGC AGTTTATCAT CTCACAGACG CCCCAGGGCC AGCAGGGTGA GCTCCTCCTT
P3-R 66    TGCATCCAAC CACCAATTTG CAGCCAGCGC AGTTTATCAT CTCACAGACG CCCCAGGGCC AGCAGGGTGA GCTCCTCCTT
P4-F 67    TGCATCCAAC CACCAATTTG CAGCCAGCGC AGTTTATCAT CTCACAGACG CCCCAGGGCC AGCAGGGTGA GCTCCTCCTT
P4-R 68    TGCATCCAAC CACCAATTTG CAGCCAGCGC AGTTTATCAT CTCACAGACG CCCCAGGGCC AGCAGGGTGA GCTCCTCCTT
                  170        180        190        200        210        220        230        240

P1-F 61    AGAGCTTATT AGTGGTATAC CAAGGCTGTT CGCTGAATGT TACACATGCA TGAACTACTA TCATTTTGGC CCTGAACCTA
P1-R 62    AGAGCTTATT AGTGGTATAC CAAGGCTGTT CGCTNAATGT TACACATGCA TAANAGANNN AA-------- ----------
P2-F 63    AGAGCTTATT AGTGGTATAC CAAGGCTGTT CGCTGAATGT TACACATGCA TGAACTACTA TCATTTTGGC CCTGAACCTA
P2-R 64    AGAGCTTATT AGTGGTATAC CAAGGCTGTT CGCTNAATGT TACACA-GC- ---------- ---------- ----------
P3-F 65    AGAGCTTATT AGTGGTATAC CAAGGCTGTT CGCTGAATGT TACACATGCA TGAACTACTA TCATTTTGGC CCTGAACTTA
P3-R 66    AGAGCTTATT AGTGGTATAC CAAGGCTGTT CGCTGAATGT TACACATGCA TGAACTAGAN NCAA------ ----------
P4-F 67    AGAGCTTATT AGTGGTATAC CAAGGCTGTT CGCTGAATGT TACACATGCA TGAACTACTA TCATTTTGGC CCTGAAACTA
P4-R 68    AGAGCTTATT AGTGGTATAC CAAGGCTGTT CGNNGAA--- ---------- ---------- ---------- ----------
                  250        260        270        280        290        300        310        320

P1-F 61    TTAAATANNN NNNNNN 285
P1-R 62    ---------- ------ 302
P2-F 63    TTAANANNNN NNNNN- 288
P2-R 64    ---------- ------ 288
P3-F 65    TTAANTANNN NNNNNN 298
P3-R 66    ---------- ------ 304
P4-F 67    TTAANNANNN NNNNNN 287
P4-R 68    ---------- ------ 276
                330
```