Gene expression in thyroid and thyroid cancer

Pauws, E.

Citation for published version (APA):
Pauws, E. (2003). Gene expression in thyroid and thyroid cancer

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 6
**A novel homeobox protein overexpressed in thyroid carcinoma**

Erwin Pauws, Gwendolyn G. Sijmons, Cigdem Yaka, Carrie Ris-Stalpers.

Submitted to Thyroid

---

**Abstract**

Serial Analysis of Gene Expression (SAGE) was applied to compare expression profiles of normal thyroid tissue and papillary thyroid carcinoma (PTC). A SAGE tag corresponding to a partial cDNA called SMAP31 is upregulated about 13-fold in PTC and was selected for further research. BLAST searching the human genome database reveals that the SMAP31 gene is located on chromosome 4q11-12 and contains 6 exons. As a result of alternative splicing 7 transcripts can be identified translating two possible open reading frames (ORF) of 73 and 95 amino acids. Database searching in GenBank's dbEST shows that SMAP31 transcripts are expressed mainly in brain, heart, gingiva and lung tissue. In thyroid tissue three transcripts are present alternatively spliced in the 5'UTR encoding an identical ORF of 73 amino acids. Homology search shows this protein to contain a homeobox domain. Thyroid and/or thyroid carcinoma specific expression of SMAP31 was studied using Northern blot and RT-PCR on a multiple tissue panel. RT-PCR experiments on a cDNA panel containing samples from different normal and tumor tissues shows expression of SMAP31 mRNA in brain, placenta, lung, heart, thyroid and thyroid carcinoma. SMAP31 expression is elevated in 6 PTC tumors compared to 4 normal thyroid controls.
Introduction

With the recent advance of high-throughput genomics the search for novel genes in molecular pathology has been accelerated substantially. Expression profiling techniques such as Serial Analysis of Gene Expression (SAGE) [1, 2] and microarray hybridization together with the progress in the human genome project [3] have provided the tools to search for novel genes. Recently the discovery of novel thyroid-specific genes has been described using these tools combined with the Tissue Preferential Expression (TPE) algorithm [4]. This algorithm calculates the tissue-specificity of selected transcripts generated in a SAGE profile. Using this approach several novel genes involved in thyroid function have been identified [5].

Progress towards the understanding of the molecular basis of thyroid carcinoma has been made with the identification of genetic variation in thyroid tumors [6]. Several genes are differentially expressed in thyroid carcinoma [7]. Some of these genes have been used as a possible diagnostic or prognostic tool, but until now no gene has proven to be comprehensive. To better understand the molecular events associated with thyroid cancer development and to identify novel putative diagnostic markers for differentiated thyroid carcinoma we used SAGE. The comparison and analysis of SAGE profiles from normal thyroid tissue [8] and papillary thyroid carcinoma (Pauws et al. Chapter 5) shows significant up- or downregulation of more than 200 tags. SAGE tags can be linked to known human gene transcripts, but some tags do not correspond to any known human transcript. The latter group (so-called NoMatch tags) corresponds to novel human transcripts and the information of the 10bp SAGE tag can be used to determine the corresponding transcript, gene and protein using public data from the NCBI and CGAP consortia in combination with classical cloning experiments.

We have identified a 10 bp NoMatch SAGE tag (ATGACAGATG) which is upregulated approximately 13 times (2 vs. 27) in a papillary thyroid carcinoma with aggressive metastatic behavior compared to normal thyroid tissue. This particular tag could not be linked to an already known human gene present in the available databases and was therefore investigated as a NoMatch tag. This tag has a TPE value of 75 (range 0-100) compared to a cohort of normal tissues and a TPE value of 76 compared to a cohort of tumor tissues, indicating that its expression is fairly specific for a papillary thyroid carcinoma.

In the current study the genomic organization, the full length coding sequence of the gene as well as tissue specific expression and protein function corresponding to the NoMatch SAGE tag are investigated.

Material & Methods

Tissue and RNA preparation

Tissue was collected from 6 patients who underwent thyroidectomy for suspicion of thyroid cancer. From all tumors, pathologically classified as papillary thyroid carcinoma, samples from tumor and adjacent normal thyroid tissue were collected after microdissection. Normal thyroid gland tissue was obtained from 4 individuals without thyroid pathology after resection at routine autopsy. After homogenization of frozen tissue samples total RNA was extracted using TRIzol (GibcoBRL).

RT-PCR

1st-strand RACE cDNA was synthesized from all thyroid RNA samples using the cDNA synthesis System (Gibco/BRL). Primers were designed on GenBank sequences using the primer design program Primer3 at [www.genome.wi.mit.edu/cgi-bin/primer/primer3.cgi]. Oligonucleotides were synthesized by Sigma Genosys using the following sequences:
Chapter 6

SMAP31; For1_TCAGTGTTTACCTCCTGCC, For2_CTGAGAACGGGTCTACCCT, For3_GAGGGAGGATCGGAATCCT, For4_CCCCCCATGATAGTCTGCTGT, For6_GCATTGACAGCTTCCAAGCTGCT, Rev1_ACAGAGGCCTCGGATTGCTAA; Elongation factor 1α (ELF1α); For_GAACCATCCAGGCCAAATAA, Rev_CCGTTCTTCCACCACTGATT; mitochondrial ATP synthase 6 (ATP6); For_CAGTGATTATAGGCTTTCCCTCTAA, Rev_CAGGGCTATTGGTTGAATGAGTA, [9]. PCR amplification was performed using standard conditions, 2mM MgCl2 and 30" 94°C, 1' 55°C, 1' 72°C for an appropriate number of cycles. For the semi-quantitative RT-PCR the number of cycles for every PCR fragment was individually optimized by performing an amplification using different amounts of cycles (20-30) and choosing a condition where the amplification reaction was in its exponential phase. 10μl of every 25μl PCR reaction was run on a 1.5% agarose gel. Imaging of the gel was performed using the Eagle Eye II System (Stratagene). Band intensity was measured using OneD-scan software from Scanalytics, corrected for background and normalized for PCR fragment size. Ratios of SMAP31 over control transcript ATP6 were calculated.

Northern blotting

RNA gels were prepared using the glyoxal/NaPi method [10] electrophorizing 10 μg of total RNA. Capillary blotting was performed overnight in 20x SSC followed by UV cross linking (1.2J/m2) and baking (80°C). A SMAP31-specific probe was generated by amplification of a 349bp fragment in exon 6 and random prime labelling with 32P. Hybridizations were performed following the Church & Gilbert protocol [11] at 65°C overnight and blots were exposed for 16 hours and analyzed using Phosphor Imager 2.0 (Biosystems).

Results

After the identification of a SAGE-generated NoMatch tag (ATGACAGATG) specifically overexpressed in papillary thyroid carcinoma this tag is used to find the corresponding gene and gene product(s). TAG-to-GENE identification using the SAGEmap program at NCBI [12] generated several corresponding EST sequences. Primers were designed on a partial cDNA sequence called SMAP31 with NCBI accession no.U51712 to amplify the corresponding transcript on a multiple tissue panel (Figure 1).

![Image](image-url)

**Figure 1.** Human multiple tissue panel RT-PCR of SMAP31; top panel shows SMAP31 PCR products (552bp), bottom shows ELF1α PCR products (395bp). Lane abbreviations: Ki=kidney; Li=liver; He=heart; Pl=placenta; Br=brain; Pa=pancreas; Lu=lung; Mu=muscle; Co=Colon; Ov=ovary; Le=leukocytes; Pr=prostate; Si=small intestine; Sp=spleen; Te=testis; Tm=thymus; Thy=thyroid; Tc1,Tc2,Tc3=thyroid carcinoma 1-3; Bc=breast carcinoma; Lc=lung carcinoma; Cc=colon carcinoma; Pc=prostate carcinoma; Oc=ovary carcinoma.
The panel shows expression of SMAP31 in placenta, brain and lung and to a lesser extent in heart. Low SMAP31 expression is also present in the normal thyroid tissue used in this panel while SMAP31 is substantially higher expressed in three papillary thyroid carcinomas, confirming the SMAP31 upregulation present in the comparison of the normal thyroid tissue and papillary thyroid carcinoma SAGE libraries. No other carcinomas present in this panel show expression of SMAP31. The PCR fragment amplified in the RT-PCR was subsequently used as a probe to hybridize a Northern blot containing total RNA from normal thyroid, thyroid carcinoma and placenta (Figure 2). A distinct band can be observed in placenta, thyroid_Normal and thyroid_Tumor. The signal in the thyroid carcinoma lane is more intense than in the thyroid control. The size of the SMAP31 transcript present in all three tissues is estimated at 1.5 kb, an additional smaller transcript (approximately 1.3 kb) is present in the thyroid carcinoma sample.

![Figure 2. Northern blot of SMAP31 mRNA in human thyroid and placenta tissue. Top panel shows SMAP31 mRNA, bottom panel shows ribosomal RNA content for each lane. Lane abbreviations: Pla=placenta; Thy_N=normal thyroid; Thy_T=thyroid tumor (PTC). Arrows indicate position of ribosomal bands. Visible bands at 1.5 and 1.3kb are indicated.](image-url)
To elucidate the genomic organization of the corresponding gene, confirm the size of the transcripts and check for possible additional alternative transcripts, we searched the GenBank database. Using the largest EST corresponding to the original tag in a BLAST search, several cDNA transcripts can be found which are similar to the SMAP31 cDNA. When comparing the cDNA transcripts with a genomic contig from chromosome 4 (NCBI accession no. NT_022853) the gene contains 6 exons spanning about 33kb.

In figure 3A the organization of the SMAP31 gene is shown. Several different gene transcripts are expressed as a result of alternative splicing. Transcripts 1 and 2 contain exons 3, 4 and 6 where exon 3 can be spliced internally by a second donor splice site. Transcript 3 contains exons 1, 4 and 6. Transcripts 1, 2 and 3 are cloned from heart tissue. Transcripts 4 and 5 contain exons 1, 2, 4 and 6 where exon 2 can be spliced internally by a second donor splice site. Transcripts 6 and 7 are similar to transcripts 4 and 5 with the additional presence of exon 5. Transcripts 4, 5, 6 and 7 are cloned from gingival tissue. From these 7 alternatively spliced transcripts two possible ORFs are predicted using the same start codon in exon 4 (Figure 3B). ORF_A uses exons 4 and 6 and is 73 amino acids long (transcripts 1, 2, 3, 4 and 5), ORF_B additionally uses exon 5 and is 95 amino acids long (transcripts 6 and 7). Several cDNA entries in GenBank are identical in their nucleotide sequences: Lung cancer associated Y protein (LACY) and odd homeobox protein 1 (OBI) also contain identical coding regions of 73 amino acids (ORF_A). An alternative transcript of OBI, contains ORF_B containing 95 amino acids. The human gene for SMAP31/LACY/OBI is located on chromosome 4, cytogenetic location 4q11-12. SMAP31 transcripts encoding ORF_A translate to a 73 amino acid protein. BLAST homology search indicates that the protein consists almost entirely of a homeobox domain. Residues 10-62 align 82.5% with the conserved HOX homeodomain. ORF_B translates to a 95 amino acid protein that differs from ORF_A downstream from amino acid residue 48. No homology to any known protein domain is present for ORF_B (Figure 3B). To study which transcripts are expressed in thyroid tissue, different combinations of exons were amplified on normal thyroid cDNA. Forward primers in exons 1 to 6 are combined with a reverse primer in exon 6 (Figure 3C). Figure 3D shows the result this of experiment. An amplification product using a forward primer in exon 1 (A) indicates, according to its size, expression of transcript 3. Two transcripts using exon 2 (transcripts 4 and 5) are represented by a PCR product (B). An aspecific PCR product of approximately 375bp is present in lanes A and B, but does not contain SMAP31 sequence. There is no band visible when amplifying from exon 3, indicating that transcripts 1 and 2 are not present in thyroid tissue (C). The PCR product in (D) shows amplification of transcripts 3, 4 and 5 but not of transcripts 6 and 7 because the fragment amplifying exon 5 (E) is absent indicating no expression of transcripts 6 and 7. Fragment F is used as a positive control for SMAP31 expression, irrespective of the transcript. Figure 3E summarizes the experiment showing that the most prominent transcripts in thyroid are transcripts 4 and 5 containing exons 1, 2a/b, 4 and 6 with sizes of respectively 1505bp and 1549bp. This corresponds to the predominantly observed mRNA size in the Northern blot. The additional smaller mRNA present in Thy_T tissue corresponds probably to transcript 3 with a size of 1382bp.
Figure 3. A - Genomic organization of SMAP31 gene and 7 transcripts arising from alternative splicing. B - SMAP31 predicted protein products. Coding region homologous between ORF_A and ORF_B is in bold. C - SMAP31-specific forward exon primers (ex1-ex6) with SMAP31 reverse primer in most 3' exon 6 generate RT-PCR fragments A-F. D - SMAP31 expression in human thyroid tissue. RT-PCR result of fragments A-F (-) = PCR negative control, (M) = marker. E - Presence (+) or absence (-) of a SMAP31 transcript can be deduced from the RT-PCR result in D. (n.i.) = not informative.
To study the expression level of SMAP31 in thyroid carcinoma in comparison to normal thyroid controls, a panel of 6 thyroid tumors (PTC) with matched controls from surrounding normal tissue and 4 thyroid tissue controls from healthy subjects was used in a semi-quantitative RT-PCR (sqRT-PCR). The relative expression level of SMAP31 was calculated by measuring the ratio of signals between a SMAP31 PCR fragment and a fragment from a housekeeping control transcript, mitochondrial transcript ATP synthase 6. Both amplification reactions were terminated in the exponential phase (SMAP31, 27 cycles; ATP6, 25 cycles). All amplifications are performed in duplo and band intensity was measured. After correcting for background and normalizing for fragment length the ratio between SMAP31 and ATP6 was calculated for every sample. SMAP31 expression is increased in 5 out of 6 thyroid tumor samples (T1,T2,T3,T5,T6) when compared to normal thyroid tissue controls. In 4 out of 6 thyroid tumor samples (T2,T3,T5,T6) SMAP31 expression is increased compared to pathologically normal tissue surrounding the tumor. In 2 thyroid tumor samples (T1,T3) the pathologically normal tissue surrounding the carcinoma shows increased expression of SMAP31 when compared to normal tissue samples from healthy thyroid tissue (Figure 4).

![Figure 4](image-url)

**Figure 4**  sqRT-PCR of SMAP31 expression in thyroid and papillary thyroid carcinoma (PTC). Y-axis depicts relative SMAP31 expression. Tumor samples Thy_T1-6 are divided in PTC (black bars) and surrounding normal tissue (gray bars). Normal thyroid controls N_1-4 are given in white bars.
Discussion

SAGE analysis of normal thyroid tissue and papillary thyroid carcinoma (PTC) has identified several novel candidate genes which may play a role in thyroid physiology and/or tumor biology. One of the most promising new genes of which the expression is upregulated in SAGE libraries from PTC is SMAP31. It was identified because of a thyroid tumor-specific overexpression of the corresponding SAGE tag. This until recently unknown human gene is located on chromosome 4q11-12, consists of 6 exons spanning approximately 33kb. The SMAP31 gene can generate at least 7 different transcripts as a result of alternative splicing. Since most of the splicing occurs in the 5' untranslated region only two putative ORFs can be distinguished. The function of 5'UTR splice variants encoding a similar protein remains speculative but there are reports that small upstream ORFs (uORFs) as small as 2 amino acids in the untranslated region of mRNA transcripts can regulate translation. This can occur through reinitiation, leaky scanning or internal initiation of protein translation [13, 14].

In thyroid tissue as well as in thyroid tumor tissue transcripts containing exons 1, 2, 4 and 6 encoding ORF_A are most abundantly expressed. EST data show that other SMAP31 transcripts are expressed in heart, brain and gingiva tissue. The size of the mRNA expressed in human thyroid tissue is approximately 1.5kb and the ORF_A encodes 73 amino acids. The additional transcript of 1.3kb visible in the thyroid tumor sample on the Northern blot probably corresponds to the SMAP31 transcript containing exons 1, 4 and 6. Because the overall expression of SMAP31 in normal thyroid is lower this transcript does not show up in the corresponding sample on the Northern blot. Both transcripts of 1.5kb and 1.3kb encode an identical protein using ORF_A. The SMAP31 protein consists almost entirely of a single homeobox domain, indicating that SMAP31 may be a DNA-binding protein that is involved in transcriptional regulation of thyroid-specific processes. The increased expression of SMAP31 observed in the SAGE library made from a papillary thyroid carcinoma was validated in a panel of 6 papillary thyroid carcinomas (PTC) and controls. The RT-PCR results show that SMAP31 is expressed in thyroid tissue at a basic low level and that SMAP31 expression is increased in 4 out of 6 papillary thyroid tumors. Strikingly, in some pathological normal tissue dissected from tissue surrounding the tumor SMAP31 expression is also increased making it tempting to speculate that increased SMAP31 expression is an early event in thyroid carcinogenesis. The fact that all samples used for RNA isolation were laser-capture microdissected, makes it unlikely to be a result of contamination of the control tissue with tumor cells.

While preparing this manuscript, two studies describing a novel homeobox-only protein called HOP in mouse and human were published [15, 16]. HOP (homeobox only protein) is identical to a SMAP31 transcript encoding ORF_A. The HOP protein is involved in mouse cardiac morphogenesis and HOP expression is regulated by Nkx2-5, which is a transcription factor also expressed in thyroid [17]. The authors also show expression of HOP in trophoblasts which fits with our observation that SMAP31 is highly expressed in human placenta. Presently available data suggest a role for SMAP31/HOP in embryogenesis and/or carcinogenesis. Further studies to elucidate the function of the SMAP31/HOP protein are necessary to understand the role of SMAP31/HOP in thyroid (tumor) biology.

Acknowledgements

We would like to thank Prof. J.J.M. de Vijlder for critical reading of the manuscript and Mirella Tan and Linda Kruijen for technical assistance.
References