(12) PATENT
(19) AUSTRALIAN PATENT OFFICE

(11) Application No. **AU 199945024 B2**
(10) Patent No. **765515**

(54) Title
Endometriosis-associated gene

(51)[7] International Patent Classification(s)
C12N 015/12 **C12N 005/10**
A61K 031/70 **C12N 015/63**
A61K 038/17 **C12Q 001/68**
A61K 039/395 **G01N 033/574**
C07K 014/47 **G01N 033/68**
C07K 016/18

(21) Application No: **199945024** (22) Application Date: **1999.05.28**

(87) WIPO No: **WO99/63079**

(30) Priority Data

(31) Number	(32) Date	(33) Country
19824230	**1998.05.29**	**DE**

(43) Publication Date : **1999.12.20**
(43) Publication Journal Date : **2000.03.02**
(44) Accepted Journal Date : **2003.09.18**

(71) Applicant(s)
Anna Starzinski-Powitz

(72) Inventor(s)
Anna Starzinski-Powitz; Silvia Kotzian; Heike Handrow-Metzmacher

(74) Agent/Attorney
Davies Collison Cave,1 Little Collins Street,MELBOURNE VIC 3000

PCT WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro
INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation 6 :
C12N 15/12, C07K 14/47, 16/18, C12N 5/10, 15/63, G01N 33/68, 33/574, C12Q 1/68, A61K 38/17, 39/395, 31/70

A1

(11) Internationale Veröffentlichungsnummer: **WO 99/63079**

(43) Internationales Veröffentlichungsdatum: 9. Dezember 1999 (09.12.99)

(21) Internationales Aktenzeichen: PCT/EP99/03716

(22) Internationales Anmeldedatum: 28. Mai 1999 (28.05.99)

(30) Prioritätsdaten:
198 24 230.1 29. Mai 1998 (29.05.98) DE

(71)(72) Anmelder und Erfinder: STARZINSKI-POWITZ, Anna [DE/DE]; Zeisselstrasse 9, D-60318 Frankfurt am Main (DE).

(72) Erfinder; und
(75) Erfinder/Anmelder *(nur für US)*: KOTZIAN, Silvia [DE/DE]; Im Höhlchen 1, D-65795 Hattersheim am Main (DE). HANDROW-METZMACHER, Heike [DE/DE]; Alt Eschersheim 45, D-60433 Frankfurt am Main (DE).

(74) Anwälte: WEICKMANN, H. usw.; Kopernikusstrasse 9, D-81679 München (DE).

(81) Bestimmungsstaaten: AU, CA, CN, IL, JP, NZ, US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Veröffentlicht
Mit internationalem Recherchenbericht.
Vor Ablauf der für Änderungen der Ansprüche zugelassenen Frist; Veröffentlichung wird wiederholt falls Änderungen eintreffen.

(54) Title: ENDOMETRIOSIS-ASSOCIATED GENE

(54) Bezeichnung: ENDOMETRIOSE-ASSOZIIERTES GEN

(57) Abstract

The invention relates to a gene associated with invasive processes, e.g. endometriosis, to a polypeptide coded by said gene, to an antibody directed against the polypeptide, and to the pharmaceutical application of the nucleic acid, the polypeptide and the antibody.

(57) Zusammenfassung

Die vorliegende Erfindung betrifft ein mit invasiven Prozessen, z.B. Endometriose assoziiertes Gen, ein davon kodiertes Polypeptid, einen gegen das Polypeptid gerichteten Antikörper sowie die pharmazeutische Anwendung der Nukleinsäure, des Polypeptids und des Antikörpers.

Novel endometriosis-associated gene

Description

The present invention relates to a gene associated with invasive processes, for example endometriosis, to a polypeptide encoded by it, to an antibody directed against the polypeptide, and to the pharmaceutical application of the nucleic acid, the polypeptide and the antibody.

Endometriosis is the second most common disease in women and is defined as the occurrence of endometrial cells outside the womb. Endometriosis affects about one in five women of reproductive age, and as many as one in two women with fertility problems.

In normal circumstances the endometrium is only found in the womb. In endometriosis, tissue with a histological appearance resembling the endometrium is found outside the womb, for example externally on the womb, on the intestine or even in the pancreas or the lung. Although these endometriotic foci are located outside the womb, they also bleed during menstruation, thus they are influenced by hormones of the female cycle. Since endometriotic foci like the endometrium go through volume changes during the cycle, these changes may cause pain depending on location. Moreover, the body reacts to endometriotic cells with an inflammatory response which again causes pain. Furthermore, inflammation leads to adhesions in the area of the ovaries and fallopian tubes and, as a result of these, is responsible for a so-called mechanical sterility of affected women. Apparently however, in endometriosis messengers are released as well (e.g. cytokines, prostaglandins) which can reduce the fertility of affected women even in the absence of adhesions.

In view of their pathobiological properties,

endometriotic cells could be classified as being between normal cells and tumor cells: on the one hand they show no neoplastic behavior, on the other hand, however, they are, like metastasizing tumor cells, capable of moving across organ boundaries in the organism and of growing into other organs, i.e. they show invasive behavior. For this reason endometriotic cells are defined as "benign tumor cells" in the literature, although up until now no tumor-specific mutations in proto-oncogenes have been found in cells of this type.

Since the pathogenesis of endometriosis is still not clarified completely, there are as yet no effective options for the therapy or prevention of endometriosis-associated diseases.

It was the object of the invention to identify novel genes which play a role in invasive processes and which may be associated with the pathophysiological phenotype of endometriosis.

This object is achieved according to the invention by identifying, cloning and characterizing a gene which is called an endometriosis-associated gene and which codes for a polypeptide. This gene sequence was discovered with the aid of differential display RT-PCR (Liang and Pardee, Science 257 (1992), 967-971). For this, invasive and noninvasive variants of an endometriotic cell line were compared with each other. In the process a cDNA sequence was found which is specific for the invasive variant of endometriotic cells. An associated RNA of 4 kb in length was found. A corresponding cDNA isolated from a cDNA phage bank has an open reading frame (ORF) of 302 amino acids.

The present invention relates to a nucleic acid which comprises

(a) the nucleotide sequences depicted in SEQ ID NO. 1,

3 or/and 5, a combination or a protein-encoding segment thereof,

(b) a nucleotide sequence corresponding to the sequence in (a) within the scope of the degeneracy of the genetic code or

(c) a nucleotide sequence hybridizing with the sequences in (a) and/or (b) under stringent conditions.

The nucleic acids preferably code for a polypeptide associated with invasive processes or a segment thereof.

The following nucleotide sequences have been deposited in the EMBL EST database with the following accession numbers: Z98886, Ac003017, AL023586, Aa52993, Aa452856. These sequences do not represent nucleic acids according to the invention. The first two of these sequences are DNAs which were isolated from human brain and show over 90% identical bases to SEQ. ID NO. 1 in the segments from nucleotide 970 to about 2000 and from 760 to about 1450, respectively, or in the segments from nucleotide 1054 to 2084 and from 844 to about 1534 in relation to SEQ ID NO. 3 which has 84 additional bases at the 5' end. AL023586 is also a human sequence which is very similar to Z98885 and also has homology with SEQ ID NO. 1 in the region from 970 to about 2000.

Sequences Aa452993 and Aa452856 originate from mouse embryos and show base identity with the nucleotides (nt) from about 1060 to about 1450 and from about 24 to 440, respectively, of SEQ. ID NO. 1, or from about 1144 to about 1534 and from about 108 to about 524, respectively, according to the nucleotide positions in SEQ. ID NO. 3. Up until now no reading frame or function has been assigned to any of these 4 sequences.

The nucleotide sequence depicted in SEQ. ID NO. 1 contains an open reading frame which corresponds to a

polypeptide having a length of 302 amino acids. This polypeptide is indicated in the amino acid sequence depicted SEQ. ID NO. 2. SEQ. ID NO. 3 shows a nucleotide sequence as in SEQ. ID NO. 1, but it has 84 additional nucleotides at the 5' end. As a result, the positions of the nucleotides corresponding to each other shift by 84 nucleotides in each case. The polypeptide encoded by SEQ. ID NO. 3 therefore has 28 additional amino acids at the N terminus and is depicted in SEQ. ID NO. 4 with its total of 330 amino acids. SEQ. ID NO. 2 and 4 depict a C-terminal segment of the native polypeptide.

For illustration purposes reference is made to figure 1 which shows a diagrammatic representation of the cDNA of the endometriosis-associated gene according to the invention. Five exons, E1 to E5, and the position of fragment 1 (394 nt) used as a probe in DDRT-PCR are shown. The positions of the PCR primers (see example 4, table 1) used for RT-PCR are also shown.

Not shown in figure 1 is a further exon 4a whose nucleotide sequence is shown in SEQ. ID NO. 5. This exon 4a may be present. If it is present, it is found between exon 4 and exon 5. This corresponds to the position between nt1054 and nt1055 in SEQ. ID NO. 3. A combination of the sequences SEQ. ID NO. 1/3 with SEQ. ID NO. 5 is accordingly, for example, a sequence which contains the sequence of the exon 4a at said position.

Besides the nucleotide sequences shown in SEQ. ID NO. 1, 3 and 5 and combinations thereof such as the sequence of SEQ. ID NO. 3, which has the sequence of SEQ. ID NO. 5 between nt1054 and 1055 and to a nucleotide sequences which

corresponds to the sequences within the scope of the degeneracy of the genetic code, the present invention also includes nucleotide sequences which hybridise with one of the sequences mentioned before. The term "hybridisation"

5

according to the present invention is used by Sambrook et al. (Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press (1989), 1.101-1.104). Preferably a hybridization is called stringent if a positive hybridization signal is still observed after washing for one hour with 1 X SSC and 0.1% SDS at 50°C, preferably at 55°C, particularly preferably at 62°C and most preferably at 68°C, in particular for 1 h in 0.2 X SSC and 0.1% SDS at 55°C, preferably at 55°C, particularly preferably at 62°C and most preferably at 68°C. A nucleotide sequence hybridizing under these washing conditions with one or more of the nucleotide sequences depicted in SEQ ID NO. 1, 3 and 5, or with a nucleic sequence corresponding to these sequences within the scope of the degeneracy of the genetic code, is a nucleotide sequence according to the invention.

The nucleotide sequence according to the invention is preferably a DNA. However, it can also include an RNA or a nucleic acid analog such as a peptidic nucleic acid, for example. Particularly preferably the nucleic acid according to the invention includes a protein-encoding segment of the nucleotide sequences depicted in SEQ ID NO. 1, 3 and/or 5 or a sequence having a homology of more than 80%, preferably more than 90% and particularly preferably more than 95% to the nucleotide sequences depicted in SEQ ID NO. 1, 3 or 5 or a segment of preferably at least 20 nucleotides (nt) and particularly preferably at least 50 nt thereof. The same also holds for nucleic acids which have, as described above, the sequence of SEQ. ID NO. 5 in addition to those of SEQ ID NO. 1 or 3. The homology is given in percent identical positions when two nucleic acids (or peptide chains) are compared, where a 100% homology means complete identity of the compared chain molecules (Herder: Lexikon der Biochemie und Molekularbiologie [Dictionary of biochemistry and molecular biology], Spektrum Akademischer Verlag 1995).

Nucleic acids according to the invention are preferably

obtainable from mammals and in particular from humans. They may be isolated according to known techniques by using short segments of the nucleotide sequences shown in SEQ. ID NO. 1, 3 or/and 5 as hybridization probes and/or as amplification primers. Furthermore, the nucleic acids according to the invention may also be prepared by chemical synthesis, it being possible to employ modified nucleotide building blocks, for example 2'-O-alkylated nucleotide building blocks, where appropriate, instead of conventional nucleotide building blocks.

The nucleic acids according to the invention or segments thereof may therefore be used for preparing primers and probes which preferably contain markers or labeling groups. Preference is also given to intron-bridging oligonucleotide primers which are particularly suitable for identifying different mRNA species.

The present invention further relates to polypeptides encoded by the nucleic acids defined as above. These polypeptides preferably comprise

(a) the amino acid sequence depicted in SEQ ID NO. 2 or 4 or

(b) a homology of more than 70%, preferably of more than 80% and particularly preferably of more than 90% to the amino acid sequence according to (a).

Besides the polypeptides depicted in SEQ ID NO. 2 or 4, the invention also relates to muteins, variants and fragments thereof. These are sequences which differ from the amino acid sequences depicted in SEQ ID NO. 2 or 4 by substitution, deletion and/or insertion of single amino acids or of short amino acid segments.

The term "variant" includes both naturally occurring allelic variations or splicing variations of the endometriotic protein, and proteins generated by recombinant DNA technology (in particular *in vitro*

mutagenesis with the aid of chemically synthesised oligonucleotides) which correspond substantially to the proteins depicted in SEQ ID NO. 2 or 4 with respect to their biological and/or immunological activity. This term also includes chemically modified polypeptides. Polypeptides which are modified at the termini and/or in the reactive amino acid side groups by acylation, for example acetylation or amidation belong to this group. Polypeptide fragments (peptides) representing a segment of at least 10 amino acids of the amino acid sequence shown in SEQ ID NO. 2 or 4 also belong to the amino acid sequences according to the invention.

The present invention further relates to a vector containing at least one copy of a nucleic acid according to the invention. This vector may be any prokaryotic or eukaryotic vector on which the DNA sequence according to the invention, preferably linked to expression signals such as promoter, operator, enhancer etc., is located. Examples of prokaryotic vectors are chromosomal vectors such as bacteriophages and extrachromosomal vectors such as plasmids, with circular plasmid vectors being particularly preferred. Suitable prokaryotic vectors are described, for example, in Sambrook et al., supra, Chapters 1-4. Particularly preferred is the vector according to the invention, a eukaryotic vector, e.g. a yeast vector, or a vector suitable for higher cells, e.g. plasmid vector, viral vector or plant vector. Vectors of this type are well known to the skilled worker in the field of molecular biology so that there is no need for further explanation here. In particular, reference is made in this connection to Sambrook et al., supra, Chapter 16.

The invention also relates to a vector which contains a segment of at least 21 nucleotides in length of the sequences depicted in SEQ ID NO. 1, 3 or/and 5 or a combination thereof. Preferably this segment has a

nucleotide sequence which originates from the protein-encoding region of said sequences or from a region essential for the expression of the protein or polypeptide. These nucleic acids are particularly suitable for preparing therapeutically employable antisense nucleic acids preferably of up to 50 nucleotides in length.

The present invention further relates to a cell transformed with a nucleic acid according to the invention or a vector according to the invention. The cell can be both a eukaryotic and a prokaryotic cell. Methods for transforming cells with nucleic acids are general prior art and therefore need no further explanation. Examples of preferred cells are eukaryotic cells, in particular animal and particularly preferably mammalian cells.

The present invention further relates to an antibody or a fragment of such an antibody against the polypeptide(s) encoded by the endometriosis gene or against variants thereof. Antibodies of this type are particularly preferably directed against complete polypeptides encoded by it or against a peptide sequence corresponding to amino acids 1-330 of the amino acid sequence depicted in SEQ ID NO. 4.

Identification, isolation and expression of a gene according to the invention which is specifically associated with invasive processes and in particular with endometriosis provide the requirements for diagnosis, therapy and prevention of diseases based on those disorders mentioned above.

It becomes possible with the aid of a polypeptide according to the invention or fragments of this polypeptide as immunogen to prepare antibodies against those polypeptides. Preparation of antibodies may be carried out in the usual way by immunizing experimental

animals with the complete polypeptide or fragments thereof and subsequently obtaining the resulting polyclonal antisera. According to the method of Köhler and Milstein and its developments monoclonal antibodies can be obtained from the antibody-producing cells of the experimental animals by cell fusion in the known manner. In the same way, human monoclonal antibodies can be produced according to known methods. Antibodies of this type could then be used both for diagnostic tests, in particular of endometriotic cell tissue, or else for the therapy.

For example, samples such as body fluids, in particular human body fluids (e.g. blood, lymph or CSF) may be tested with the aid of the ELISA technique on the one hand for the presence of a polypeptide encoded by the endometriosis gene, on the other hand for the presence of autoantibodies against such a polypeptide. Polypeptides encoded by the endometriosis gene or fragments thereof can then be detected in such samples with the aid of a specific antibody, for example of an antibody according to the invention. For detecting autoantibodies it is preferably possible to employ recombinant fusion proteins which contain a part or a domain or even the complete polypeptide encoded by the endometriosis gene and which are fused to a protein domain which facilitates detection, for example maltose-binding protein (MBP).

Diagnostic tests may also be carried out with the aid of specific nucleic acid probes for detecting at the nucleic acid level, for example at the gene or transcript level.

Provision of the nucleotide and amino acid sequences and antibodies according to the invention further facilitates a targeted search for effectors of the polypeptides/proteins. Effectors are agents which act in an inhibitory or activating manner on the

polypeptide according to the invention and which are capable of selectively influencing cell functions controlled by the polypeptides. These may then be employed in the therapy of appropriate pathologies, such as those based on invasive processes. The invention therefore also relates to a method for identifying effectors of endometriotic proteins where cells expressing the protein are brought into contact with various potential effector substances, for example low molecular weight agents, and the cells are analyzed for modifications, for example cell-activating, cell-inhibiting, cell-proliferative and/or cell-genetic modifications. In this way it is also possible to identify binding targets of endometriotic proteins.

Since many neoplastic diseases are accompanied by invasive processes, the discovery of the gene according to the invention additionally provides possibilities for the diagnosis, prevention and therapy of cancerous diseases.

The discovery of a gene involved in the responsibility for invasive processes not only opens up possibilities for the treatment of diseases based on cellular modifications of this type, but the sequences according to the invention may also be used in order to make such processes usable. This can be of importance, for example, for the implantation of embryos.

The present invention therefore also relates to a pharmaceutical composition which includes as active components nucleic acids, vectors, cells, polypeptides, peptides and/or antibodies, as mentioned before.

The pharmaceutical composition according to the invention may further contain pharmaceutically conventional carriers, excipients and/or additives and, where appropriate, further active components. The pharmaceutical composition may be employed in

particular for the diagnosis, therapy or prevention of diseases associated with invasive processes. Furthermore the composition according to the invention may also be employed for diagnosing a predisposition for such diseases, in particular for diagnosing an endometriosis risk.

The invention is illustrated in more detail by the following figures, sequence listings and examples.

Figure 1 shows a diagrammatic representation of the cDNA of the endometriosis-associated gene where only exons E1 to E5 are shown.

SEQ ID NO. 1 represents a nucleotide sequence which contains genetic information coding for the endometriosis-associated gene, where an open reading frame extends from nucleotide 3 to 911, and

SEQ ID NO. 2 represents the amino acid sequence of the open reading frame of the nucleotide sequence shown in SEQ ID NO. 1, where the amino acid sequence of the open reading frame extends from amino acid 1 to 302.

SEQ ID NO. 3 represents a nucleotide sequence like that of SEQ ID NO. 1 but it contains an additional 84 nucleotides at the 5' end, the open reading frame extends from nucleotide 3 to 995.

SEQ ID NO. 4 represents the amino acid sequence of the open reading frame of the nucleotide sequence shown in SEQ ID NO. 3, where this amino acid sequence has 320 amino acids of which the C-terminal 302 are identical to those in SEQ ID NO. 2.

SEQ ID NO. 5 represents of the nucleotide sequence of the possibly present additional exon 4a consisting of the 218 nt shown, where exon 4a, if it is present, is located between nucleotide 1054 and 1055 (in relation to SEQ ID NO. 3).

EXAMPLES

Example 1 Cell culturing

To identify an endometriosis-associated gene, invasive and noninvasive cells of the epithelial endometriotic cell line EEC145T$^+$ were used. The cells were cultured in Dulbecco's medium (DMEM) with 10% fetal calf serum and diluted 1 : 5 2x per week (passage). For comparison of the expression patterns by means of DDRT-PCR (see below) invasive cells of passage 17 and noninvasive cells of passage 33 were used. The cells were transformed with SV40 and analyzed by differential display reverse transcription polymerase chain reaction (DDRT-PCR).

Example 2 DDRT-PCR

This method developed by Liang and Pardee is a method for distinguishing expression patterns of different cell types or the alteration in the expression pattern of one cell type under different living conditions or during altering stages of development (Liang and Pardee (1992), Science 257, 967-971). The basis of the DDRT-PCR technique is based on the idea that in each cell about 15,000 genes are expressed and that in principle each individual mRNA molecule can be prepared by means of reverse transcription and amplification with random primers.

In this example the cellular polyA$^+$ RNA was initially

transcribed into cDNA with the aid of several different $dT_{11}VX$ primers (downstream primers, anchor primers). The resulting cDNA populations were then PCR-amplified using 4 downstream and 20 upstream primers from the RNA Map™ Kit from Genhunter, Nashville (1994), with the addition of a radiolabeled nucleotide. After the amplification the reaction mixtures were concentrated in vacuo and the obtained cDNA fragments were fractionated in a six-percent native PAA (polyacrylamide) gel. DNA detection was carried out by autoradiography. PCR mixtures showing distinct differences in the band pattern for the two cell variants to be studied were repeated twice in order to test reproducibility. If the previously found differences were confirmed, the bands were eluted from the gel according to known methods, reamplified, cloned and sequenced.

By this method a 394 bp fragment (fragment 1, nucleotides 1235 to 1628 of the nucleic acid sequence depicted in SEQ ID NO. 1, see also Figure 1) was found which was specific for the invasive cell variant. This fragment 1 was used as a probe in Northern blot analysis (see below).

Example 3 Analysis of the fragment 1 expression profile in human Northern blot analyses

To test the expression pattern for DDRT-PCR fragment 1, Northern blot analyses were carried out. For this 20 µg of total RNA or 4 µg of polyA+ RNA were fractionated in 1% denaturating agarose gels and transferred onto a nylon membrane overnight. The RNA was fixed to the membrane by irradiation with UV light. Hybridization with ^{32}P-labeled probes (labeling by means of RPL kit from Amersham) took place overnight in a formamide-containing hybridization solution at 42°C. Subsequently the membrane was washed under increasing stringency until the spots of radioactive emission were of

measurable intensity. The hybridization pattern was visualized by putting on an X-ray film (NEF-NEN, DuPont) and exposing over several days. To determine the expression pattern for DDRT-PCR fragment 1, Northern blot analyses were carried out using RNA from the following cells or tissues:

- invasive cells of the epithelial endometriotic cell line EEC145T$^+$ (passage 17)
- noninvasive cells of the epithelial endometriotic cell line EEC145T$^+$ (passage 33)
- cells of the peritoneal cell line EEC143T$^+$
- endometrial tissue
- cells of the invasive human bladder carcinoma cell line EJ28
- cells of the noninvasive human bladder carcinoma cell line RT112

After hybridization with the probe for DDRT-PCR fragment 1 an mRNA of about 4 kb was detectable, and it was exclusively detectable in the invasive variant of the endometriotic cell line EEC145T$^+$.

Further human tissues were tested. In the spleen an mRNA of 4 kb in length was found which hybridized unambiguously with fragment 1, and in brain mRNAs of 4 kb and > 9 kb in length, respectively, were found.

Northern blot analyses were carried out according to the manufacturer's protocol using two human multiple tissue Northern (MTN) blots from Clontech. Expression was tested in the following tissues: colon, small intestine, heart, brain, testicles, liver, lung, spleen, kidney, ovaries, pancreas, peripheral blood leukocytes, placenta, prostate, skeletal muscle, thymus. The expression pattern obtained using the radiolabeled 3' probe "DDRT-PCR fragment 1" appears as follows:

4 kb mRNA (expected size): brain,
spleen,

pancreas

9.5 kb mRNA: brain

In the remaining tissues no specific hybridization was detectable.

In-situ hybridization

To elucidate the cellular expression pattern, mRNA in-situ hybridizations were carried out on 10 µm paraffin sections of different tissues. For this the "DDRT-PCR fragment 1" was employed as digoxigenin-labeled RNA probe. The detection reaction was carried out by means of a digoxigenin-specific antibody coupled to alkaline phosphatase (A). BM Purple served as a substrate for AP and forms a blue precipitate after dephosphorylation. The results are listed in the following table and show predominant expression in invasive/migrating cells.

Strong expression	**Weak, not quite unambiguous expression**
epithelial cells from endometriotic lesions carcinomas lymphatic infiltrates thymus germinal centers of lymph follicles (spleen) **somewhat weaker:** epithelial cells of the endometrium angiogenetic endothelial cells migrating nerve cells	skeletal muscle heart sarcomas

Example 4 RT-PCR

RT-PCR (reverse transcription PCR) provides a sensitive method for testing the expression pattern.

For this, 1 µg of the appropriate polyA$^+$ RNA was transcribed into cDNA with the aid of 400 U of M-MLV reverse transcriptase (Gibco-BRL) in a total volume of 30 µl. 1 µl of this was employed for the subsequent PCR with different primer combinations.

The PCR primers P1 to P7 used are depicted in table 1 (see figure 1).

Table 1

Number	Sequence (nucleotide position in relation to SEQ ID NO. 1
P1	5'-CCAGCTGCTGCCAAATCC-3' (36-53)
P2	5'-CATCATGGTCATAGCTGC-3' (545-562)
P3	5'-AGCGTCTCATCGGTGTAC-3' (793-776, reverse primer)
P4	5'-AACAGAAGTGGTAGGTGC-3' (1080-1063, reverse primer)
P5	5'-AAAGGGACGGGAGGAAGC-3' (1243-1260)
P6	5'-CCAAAGTAGAAAACACTG-3' (1612-1595, reverse primer)
P7	5'-GCTTGTATGACACACACG-3' (2150-2133, reverse primer)

RT-PCR experiments were carried out using polyA$^+$ RNA from different cell lines and tissues and using different primer combinations. The results are depicted in table 2.

Table 2

PC	P17	P33	Per	EM	EJ28	RT112	E	EE	PEE
P1-P4	+	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P2-P6	+	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P5+P7	+	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
P5+P6	+	-	-	+	-	-	+	+	+
P1+P3	+	-	-	+	-	-	+	+	+

PC = primer combination
P17 = endometriotic cell line EEC145T, passage 17, invasive
P33 = endometriotic cell line EEC145T, passage 33, noninvasive
Per = peritoneal cell line Per143T
EM = endometrial tissue
EJ28 = invasive bladder carcinoma cell line
RT112 = noninvasive bladder carcinoma cell line
E = endometrial tissue
EE = endometrial tissue of an endometriosis patient
PEE = peritoneal endometriosis biopsy
n.d. = not determined

The RT-PCR results confirmed the fragment 1-specific expression in the early passages (passage 17, passage 20) of the endometriotic cell line EEC145T$^+$. As a deviation from the Northern blot analyses it was possible to show in addition a weak expression in the endometrium.

RT-PCR analyses using intron-bridging primers

To test possible alternative exons, RT-PCR experiments using intron-bridging primers were carried out. In this connection it was possible to show at least one further mRNA species which exists alongside the mRNA described and which contains a further exon (4a) of 218 bp in length between the 4th and 5th exons. This exon is

located in the 3'-UTR (untranslated region), that is to say after the coding region. The sequence of exon 4a is listed below.

gcggttgtcc ggaatgccag tggctcctgg gcagatgtgc accccagatt cagcctttgt gatagattcc aacacgttct ggcctcagac caccttttgtg gtggggccag actgctctgg gcaaagtgaa gctggccttt atgctccaag gaaggggggcc tcgagagcag gcctgcattg gctctcggac taattcgcga tcatctttca tacagcag

Nucleotide sequence of the alternative exon 4a

Example 5 Preparation of the cDNA phage bank EEC14

The cDNA phage bank EEC14 was prepared according to the method of Short, J.M. et al. (1988) Nucleic Acids Res. 16: 7583-7600.

Initially, reverse transcription of polyA$^+$ RNA from invasive cells (passage 17) of the epithelial endometriotic cell line EEC145T$^+$ was carried out. The primer used here consists of an *XhoI* cleavage site and a poly(dT) sequence of 18 nucleotides in length. An adapter including an *EcoRI* cleavage site was ligated to the cDNA fragments produced. The two restriction sites permit directed insertion of the cDNA fragments into the ZAP Express™ vector. Inserts can be excised from the phage in the form of a kanamycin-resistant pBK CMV phagemid.

Example 6 Phage bank screening

The DDRT-PCR fragment 1 (394 bp) was used as a probe in order to screen 10^6 pfu (plaque forming units) of the cDNA phage bank EEC14 according to the manufacturer's protocol (Stratagene). Labeling of the probe with digoxigenin (Boehringer Mannheim) was carried out with the aid of PCR. The plaques formed after infection of the bacterial strain *XL 1blue MRF'* were transferred onto a nylon membrane and hybridized thereon with the

abovementioned probe. Detection of the hybridized, digoxigenin-labeled probe was carried out according to the chemiluminescence protocol by Boehringer Mannheim.

Positive plaques were selected and subjected to rescreening. The positive plaques from the rescreening were employed for the excision. Excising the vector portion from the phage by means of ExAssist helper phages resulted in kanamycin-resistant pBK CMV phagemids which could be isolated and sequenced after amplification in the bacterial strain XLOLR™. The isolated phagemid clone Q2A contained the longest insert of 2.3 kb in size whose sequence was determined and is shown SEQ ID NO. 1. The DDRT-PCR fragment 1 sequence is found as nucleotides 1235 to 1628 in relation to SEQ ID NO. 1.

Example 7 Southern blot analysis

10 µg of genomic DNA from female and male subjects were cleaved with various restriction endonucleases. The fragments were fractionated in an agarose gel and transferred onto a nylon membrane. Hybridization with the digoxigenin-labeled DDRT-PCR fragment 1 was carried out on this membrane.

Hybridization was detectable by chemiluminescence according to the Boehringer protocol. Using various restriction endonucleases only one band in each case was detected in both the female and male DNA samples. This result suggests that the gene on which fragment 1 is based is a single, non-sex-specific gene. Since then, two genomic clones PAC J1472 and PAC N1977 have been isolated using DDRT-PCR fragment 1.

Example 8 Fluorescence in situ hybridization (FISH)

The genomic clones obtained in Example 7 were localized on chromosome 1 (1p36) by means of fluorescence in situ

hybridization (Lichter et al. (1990), Science 247:64-69).

Example 9 Production of specific antibodies

Nucleotides 584 to 909 of the abovementioned cDNA sequence were cloned by suitable restriction cleavage sites into the expression vector pMAL cRI. To express the sequence the construct was transformed into E.coli DH5 α cells. The translated protein fragment was cut out of an SDS polyacrylamide gel and employed for immunizing rabbits.

Example 10 RACE (rapid amplification of cDNA ends)

Since the length of the cDNA clone Q2A (see Example 6) differs from the size of the detected mRNA (about 4 kb), RACE experiments were carried out to obtain further sequence information. With the aid of this method it is possible to obtain cDNA sequences from an mRNA template between a defined internal sequence and unknown sequences at the 5' or 3' end. The 3' end of clone Q2A could be confirmed by 3'RACE experiments starting from the 5th exon.

For the 5'RACE, first strand synthesis of the cDNA was carried out using a gene-specific primer which hybridizes in the 1st exon, and then a homopolymeric nucleotide tail was attached with the aid of the enzyme terminal transferase. This attached sequence permitted amplification of the sequence region located between the gene-specific primer and the homopolymeric nucleotide tail. This made it possible to obtain the following additional sequence which is located 5' from the Q2A sequence and belongs to the first exon:

```
cc cgg ccg ccc cga gtg gag cgg atc cac ggg cag atg cag atg cct    47
   Arg Pro Pro Arg Val Glu Arg Ile His Gly Gln Met Gln Met Pro
     1               5                  10                  15

cga gcc aga cgg gcc cac agg ccc cgg gac cag gcg gcc gcc ctc gtg ... 95
Arg Ala Arg Arg Ala His Arg Pro Arg Asp Gln Ala Ala Ala Leu Val ...
                20                  25                  30
```

The underlined sequence represents the first nucleotides of the Q2A sequence, the sequence in front of it corresponds to the novel sequence obtained by 5'RACE. The open reading frame fits into the one already derived for fragment and contains two putative start codons (underlined).

The nucleotide sequence which has the sequence previously obtained and is depicted in SEQ ID NO. 1 and the additional 84 nt at the 5' end is depicted in SEQ ID NO. 3.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

The reference to any prior art in this specification is not, and should not be taken as, an acknowledgment or any form of suggestion that that prior art forms part of the common general knowledge in Australia.

SEQUENCE LISTING

<110> Prof. Starzinski-Powitz, Anna
<120> Novel endometriosis-associated gene
<130> 18505PWO endometriosis gene
<140>
<141>

<150> DE 198 24 230.1
<151> 1998-05-29

<160> 5

<170> PatentIn Ver. 2.1

<210> 1
<211> 2204
<212> DNA
<213> Homo sapiens

<220>
<221> CDS
<222> (3) .. (911)

```
cc gcc ctc gtg ccc aag gca gga ctg gcc aag ccc cca gct gct gcc    47
   Ala Leu Val Pro Lys Ala Gly Leu Ala Lys Pro Pro Ala Ala Ala
     1               5                  10                  15

aaa tcc agc cct tcc ctc gcc tct tcg tcc tcg tcc tcg tcc tcc gcg   95
Lys Ser Ser Pro Ser Leu Ala Ser Ser Ser Ser Ser Ser Ser Ser Ala
                 20                  25                  30

gtg gcc ggt ggg gcc ccg gag cag cag gcc ctc ctg agg agg ggc aag   143
Val Ala Gly Gly Ala Pro Glu Gln Gln Ala Leu Leu Arg Arg Gly Lys
             35                  40                  45

agg cac ctg cag ggg gac ggt ctc agc agc ttc gac tcc aga ggc agc   191
Arg His Leu Gln Gly Asp Gly Leu Ser Ser Phe Asp Ser Arg Gly Ser
         50                  55                  60

cgg ccc acc aca gag act gag ttc atc gcc tgg ggg ccc acg ggg gac   239
Arg Pro Thr Thr Glu Thr Glu Phe Ile Ala Trp Gly Pro Thr Gly Asp
     65                  70                  75

gag gag gcc ctg gag tcc aac aca ttt ccg ggc gtt tac ggc ccc acc   287
Glu Glu Ala Leu Glu Ser Asn Thr Phe Pro Gly Val Tyr Gly Pro Thr
 80                  85                  90                  95

acg gtc tcc atc cta caa aca cgg aag aca act gtg gcc gcc acc acc   335
Thr Val Ser Ile Leu Gln Thr Arg Lys Thr Thr Val Ala Ala Thr Thr
                100                 105                 110

acc acc acc acc acg gcc acc ccc atg acg ctg cag act aag ggg ttc   383
Thr Thr Thr Thr Thr Ala Thr Pro Met Thr Leu Gln Thr Lys Gly Phe
            115                 120                 125
```

```
acc gag tcc ttg gat ccc cgg aga agg atc cca ggt ggg gtt agc aca   431
Thr Glu Ser Leu Asp Pro Arg Arg Arg Ile Pro Gly Gly Val Ser Thr
        130                 135                 140

acg gag cct tcc acc agt ccc agc aac aac ggg gaa gtc acc cag ccc   479
Thr Glu Pro Ser Thr Ser Pro Ser Asn Asn Gly Glu Val Thr Gln Pro
    145                 150                 155

cca agg att ctg ggg gag gcc tca ggt ctg gct gtc cat cag atc atc   527
Pro Arg Ile Leu Gly Glu Ala Ser Gly Leu Ala Val His Gln Ile Ile
160                 165                 170                 175

acc atc acc gtc tcc ctc atc atg gtc ata gct gct ctc atc aca act   575
Thr Ile Thr Val Ser Leu Ile Met Val Ile Ala Ala Leu Ile Thr Thr
                180                 185                 190

ctt gtc tta aaa aat tgc tgt gcc caa agc ggg aac act cgt cgg aac   623
Leu Val Leu Lys Asn Cys Cys Ala Gln Ser Gly Asn Thr Arg Arg Asn
            195                 200                 205

agc cac cag cgg aag acc aac cag cag gag gag agc tgc cag aac ctc   671
Ser His Gln Arg Lys Thr Asn Gln Gln Glu Glu Ser Cys Gln Asn Leu
        210                 215                 220

acg gac ttc ccc tcg gcc cgg gtg ccc agc agc ctg gac ata ttc acg   719
Thr Asp Phe Pro Ser Ala Arg Val Pro Ser Ser Leu Asp Ile Phe Thr
    225                 230                 235

gcc tat aac gag acc ctg cag tgt tct cac gag tgc gtc agg gca tct   767
Ala Tyr Asn Glu Thr Leu Gln Cys Ser His Glu Cys Val Arg Ala Ser
240                 245                 250                 255

gtg ccc gtg tac acc gat gag acg ctg cac tcg acg acg ggg gag tac   815
Val Pro Val Tyr Thr Asp Glu Thr Leu His Ser Thr Thr Gly Glu Tyr
                260                 265                 270

aaa tcc aca ttt aat gga aac cga ccc tcc tct tct gat cgg cat ctt   863
Lys Ser Thr Phe Asn Gly Asn Arg Pro Ser Ser Ser Asp Arg His Leu
            275                 280                 285

att cct gtg gcc ttc gtg tct gag aaa tgg ttt gaa atc tcc tgc tga   911
Ile Pro Val Ala Phe Val Ser Glu Lys Trp Phe Glu Ile Ser Cys
        290                 295                 300

ctggccgaag tctttttac ctcctggggg cagggcagac gccgtgtgtc tgtttcacgg 971

attccgttgg tgaacctgta aaaacaaaac aaacaaaaca aaacaaaaaa gacaaaacct 1031

aaaactgagc tatctaaggg ggagggtccc cgcacctacc acttctgttt gccggtggga 1091

aactcacaga gcaggacgct ctaggccaaa tctatttttg taaaaatgct catgcctatg 1151

ggtgactgcc ttctcccaga gttttctttg gagaacagaa agaagaaagg aaagaaagga 1211

accagaggca gagagacgag gatacccagc gaaagggacg ggaggaagca tccgaaacct 1271

aggattcgtc ctacgattct gaacctgtgc caataatacc attatgtgcc atgtactgac 1331

ccgaaaggct cggccacaga gccggggccc agcgaatcac gcagagaaat cttacagaaa 1391

acaggggtgg gaatctcttc cgatagagtc gctatttctg gttaatatac atatataaat 1451

atataaatac aaacacacac acacactttt tttgtactgt agcaattttt gaagatctta 1511

aatgttcctt tttaaaaaaa agaattgtgt tataggttac aaaatctgat ttatttaaca 1571

tgcttagtat gagcagaata aaccagtgtt ttctactttg gcaactcacg tcacacacat 1631
```

attacacaca tgtgcgcata cacacacaca atacacatat atgcatatag acgcatctat 1691

tggaaatgca gttccacagg tgagcatgtt ctttctggtg acctggtatt ccatcaccat 1751

tcaccccagg ggacagcctc gaccgagaca aggaggccct taaatgacag cctgcatttg 1811

ctagacggtt ggtgagtggc atcaaatgtg tgacttacta tcttgggcca gaactaagaa 1871

tgccaaggtt ttatatatgt gtgtatatat atatatatat atatatatat atgtttgtgt 1931

gtgtatatat atatatatat atatatgttt gtgtgtgtat atatatgttt gtgtatatat 1991

atacacatat gcatacatat gatttttttt ttttcattta agtgttggaa gatgctacct 2051

aacagccacg ttcacattta cgtagctggt tgcttacaaa cgggcctgag cccctggttg 2111

ggtgggtggt ggattcttgg acgtgtgtgt catacaagca tagactggat taaagaagtt 2171

ttccagttcc aaaaattaaa ggaatatatc ctt 2204

<210> 2

<211> 302

<212> PRT

<213> Homo sapiens

<400> 2

Ala Leu Val Pro Lys Ala Gly Leu Ala Lys Pro Pro Ala Ala Ala Lys
1 5 10 15

Ser Ser Pro Ser Leu Ala Ser Ser Ser Ser Ser Ser Ser Ser Ala Val
20 25 30

Ala Gly Gly Ala Pro Glu Gln Gln Ala Leu Leu Arg Arg Gly Lys Arg
35 40 45

His Leu Gln Gly Asp Gly Leu Ser Ser Phe Asp Ser Arg Gly Ser Arg
50 55 60

Pro Thr Thr Glu Thr Glu Phe Ile Ala Trp Gly Pro Thr Gly Asp Glu
65 70 75 80

Glu Ala Leu Glu Ser Asn Thr Phe Pro Gly Val Tyr Gly Pro Thr Thr
85 90 95

Val Ser Ile Leu Gln Thr Arg Lys Thr Thr Val Ala Ala Thr Thr Thr
100 105 110

Thr Thr Thr Thr Ala Thr Pro Met Thr Leu Gln Thr Lys Gly Phe Thr
115 120 125

Glu Ser Leu Asp Pro Arg Arg Arg Ile Pro Gly Gly Val Ser Thr Thr
130 135 140

Glu Pro Ser Thr Ser Pro Ser Asn Asn Gly Glu Val Thr Gln Pro Pro
145 150 155 160

Arg Ile Leu Gly Glu Ala Ser Gly Leu Ala Val His Gln Ile Ile Thr
165 170 175

Ile Thr Val Ser Leu Ile Met Val Ile Ala Ala Leu Ile Thr Thr Leu
180 185 190

Val Leu Lys Asn Cys Cys Ala Gln Ser Gly Asn Thr Arg Arg Asn Ser
195 200 205

His Gln Arg Lys Thr Asn Gln Gln Glu Glu Ser Cys Gln Asn Leu Thr
210 215 220

Asp Phe Pro Ser Ala Arg Val Pro Ser Ser Leu Asp Ile Phe Thr Ala
225 230 235 240

Tyr Asn Glu Thr Leu Gln Cys Ser His Glu Cys Val Arg Ala Ser Val
245 250 255

Pro Val Tyr Thr Asp Glu Thr Leu His Ser Thr Thr Gly Glu Tyr Lys
260 265 270

Ser Thr Phe Asn Gly Asn Arg Pro Ser Ser Ser Asp Arg His Leu Ile
275 280 285

Pro Val Ala Phe Val Ser Glu Lys Trp Phe Glu Ile Ser Cys
290 295 300

<210> 3

<211> 2288

<212> DNA

<212> Homo sapiens

<220>

<221> CDS

<222> (3) .. (995)

<400> 3

```
cc cgg ccg ccc cga gtg gag cgg atc cac ggg cag atg cag atg cct    47
   Arg Pro Pro Arg Val Glu Arg Ile His Gly Gln Met Gln Met Pro
     1               5                  10                  15

cga gcc aga cgg gcc cac agg ccc cgg gac cag gcg gcc gcc ctc gtg    95
Arg Ala Arg Arg Ala His Arg Pro Arg Asp Gln Ala Ala Ala Leu Val
                 20                  25                  30

ccc aag gca gga ctg gcc aag ccc cca gct gct gcc aaa tcc agc cct   143
Pro Lys Ala Gly Leu Ala Lys Pro Pro Ala Ala Ala Lys Ser Ser Pro
             35                  40                  45

tcc ctc gcc tct tcg tcc tcg tcc tcg tcc tcc gcg gtg gcc ggt ggg   191
Ser Leu Ala Ser Ser Ser Ser Ser Ser Ser Ser Ala Val Ala Gly Gly
         50                  55                  60

gcc ccg gag cag cag gcc ctc ctg agg agg ggc aag agg cac ctg cag   239
Ala Pro Glu Gln Gln Ala Leu Leu Arg Arg Gly Lys Arg His Leu Gln
     65                  70                  75

ggg gac ggt ctc agc agc ttc gac tcc aga ggc agc cgg ccc acc aca   287
Gly Asp Gly Leu Ser Ser Phe Asp Ser Arg Gly Ser Arg Pro Thr Thr
 80                  85                  90                  95

gag act gag ttc atc gcc tgg ggg ccc acg ggg gac gag gag gcc ctg   335
Glu Thr Glu Phe Ile Ala Trp Gly Pro Thr Gly Asp Glu Glu Ala Leu
                100                 105                 110

gag tcc aac aca ttt ccg ggc gtt tac ggc ccc acc acg gtc tcc atc   383
Glu Ser Asn Thr Phe Pro Gly Val Tyr Gly Pro Thr Thr Val Ser Ile
            115                 120                 125

cta caa aca cgg aag aca act gtg gcc gcc acc acc acc acc acc acc   431
Leu Gln Thr Arg Lys Thr Thr Val Ala Ala Thr Thr Thr Thr Thr Thr
        130                 135                 140
```

```
acg gcc acc ccc atg acg ctg cag act aag ggg ttc acc gag tcc ttg   479
Thr Ala Thr Pro Met Thr Leu Gln Thr Lys Gly Phe Thr Glu Ser Leu
    145                 150                 155

gat ccc cgg aga agg atc cca ggt ggg gtt agc aca acg gag cct tcc   527
Asp Pro Arg Arg Arg Ile Pro Gly Gly Val Ser Thr Thr Glu Pro Ser
160                 165                 170                 175

acc agt ccc agc aac aac ggg gaa gtc acc cag ccc cca agg att ctg   575
Thr Ser Pro Ser Asn Asn Gly Glu Val Thr Gln Pro Pro Arg Ile Leu
                180                 185                 190

ggg gag gcc tca ggt ctg gct gtc cat cag atc atc acc atc acc gtc   623
Gly Glu Ala Ser Gly Leu Ala Val His Gln Ile Ile Thr Ile Thr Val
            195                 200                 205

tcc ctc atc atg gtc ata gct gct ctc atc aca act ctt gtc tta aaa   671
Ser Leu Ile Met Val Ile Ala Ala Leu Ile Thr Thr Leu Val Leu Lys
        210                 215                 220

aat tgc tgt gcc caa agc ggg aac act cgt cgg aac agc cac cag cgg   719
Asn Cys Cys Ala Gln Ser Gly Asn Thr Arg Arg Asn Ser His Gln Arg
    225                 230                 235

aag acc aac cag cag gag gag agc tgc cag aac ctc acg gac ttc ccc   767
Lys Thr Asn Gln Gln Glu Glu Ser Cys Gln Asn Leu Thr Asp Phe Pro
240                 245                 250                 255

tcg gcc cgg gtg ccc agc agc ctg gac ata ttc acg gcc tat aac gag   815
Ser Ala Arg Val Pro Ser Ser Leu Asp Ile Phe Thr Ala Tyr Asn Glu
                260                 265                 270

acc ctg cag tgt tct cac gag tgc gtc agg gca tct gtg ccc gtg tac   863
Thr Leu Gln Cys Ser His Glu Cys Val Arg Ala Ser Val Pro Val Tyr
            275                 280                 285

acc gat gag acg ctg cac tcg acg acg ggg gag tac aaa tcc aca ttt   911
Thr Asp Glu Thr Leu His Ser Thr Thr Gly Glu Tyr Lys Ser Thr Phe
        290                 295                 300

aat gga aac cga ccc tcc tct tct gat cgg cat ctt att cct gtg gcc   959
Asn Gly Asn Arg Pro Ser Ser Ser Asp Arg His Leu Ile Pro Val Ala
    305                 310                 315

ttc gtg tct gag aaa tgg ttt gaa atc tcc tgc tga ctggccgaag       1005
Phe Val Ser Glu Lys Trp Phe Glu Ile Ser Cys
320                 325                 330

tctttttac ctcctggggg cagggcagac gccgtgtgtc tgtttcacgg attccgttgg 1065

tgaacctgta aaaacaaaac aaacaaaaca aaacaaaaaa gacaaaacct aaaactgagc 1125

tatctaaggg ggagggtccc cgcacctacc acttctgttt gccggtggga aactcacaga 1185

gcaggacgct ctaggccaaa tctatttttg taaaaatgct catgcctatg ggtgactgcc 1245

ttctcccaga gttttctttg gagaacagaa agaagaaagg aaagaaagga accagaggca 1305

gagagacgag gatacccagc gaaagggacg ggaggaagca tccgaaacct aggattcgtc 1365

ctacgattct gaacctgtgc caataatacc attatgtgcc atgtactgac ccgaaaggct 1425

cggccacaga gccggggccc agcgaatcac gcagagaaat cttacagaaa acaggggtgg 1485

gaatctcttc cgatagagtc gctatttctg gttaatatac atatataaat atataaatac 1545

aaacacacac acacactttt tttgtactgt agcaattttt gaagatctta aatgttcctt 1605
```

tttaaaaaaa agaattgtgt tataggttac aaaatctgat ttatttaaca tgcttagtat 1665

gagcagaata aaccagtgtt ttctactttg gcaactcacg tcacacacat attacacaca 1725

tgtgcgcata cacacacaca atacacatat atgcatatag acgcatctat tggaaatgca 1785

gttccacagg tgagcatgtt ctttctggtg acctggtatt ccatcaccat tcaccccagg 1845

ggacagcctc gaccgagaca aggaggccct taaatgacag cctgcatttg ctagacggtt 1905

ggtgagtggc atcaaatgtg tgacttacta tcttgggcca gaactaagaa tgccaaggtt 1965

ttatatatgt gtgtatatat atatatatat atatatatat atgtttgtgt gtgtatatat 2025

atatatatat atatatgttt gtgtgtgtat atatatgttt gtgtatatat atacacatat 2085

gcatacatat gatttttttt ttttcattta agtgttggaa gatgctacct aacagccacg 2145

ttcacattta cgtagctggt tgcttacaaa cgggcctgag cccctggttg ggtgggtggt 2205

ggattcttgg acgtgtgtgt catacaagca tagactggat taaagaagtt ttccagttcc 2265

aaaaattaaa ggaatatatc ctt 2288

<210> 4

<211> 330

<212> PRT

<213> Homo sapiens

<400> 4

Arg Pro Pro Arg Val Glu Arg Ile His Gly Gln Met Gln Met Pro Arg
1 5 10 15

Ala Arg Arg Ala His Arg Pro Arg Asp Gln Ala Ala Ala Leu Val Pro
20 25 30

Lys Ala Gly Leu Ala Lys Pro Pro Ala Ala Ala Lys Ser Ser Pro Ser
35 40 45

Leu Ala Ser Ser Ser Ser Ser Ser Ser Ser Ala Val Ala Gly Gly Ala
50 55 60

Pro Glu Gln Gln Ala Leu Leu Arg Arg Gly Lys Arg His Leu Gln Gly
65 70 75 80

Asp Gly Leu Ser Ser Phe Asp Ser Arg Gly Ser Arg Pro Thr Thr Glu
85 90 95

Thr Glu Phe Ile Ala Trp Gly Pro Thr Gly Asp Glu Glu Ala Leu Glu
100 105 110

Ser Asn Thr Phe Pro Gly Val Tyr Gly Pro Thr Thr Val Ser Ile Leu
115 120 125

Gln Thr Arg Lys Thr Thr Val Ala Ala Thr Thr Thr Thr Thr Thr Thr
130 135 140

Ala Thr Pro Met Thr Leu Gln Thr Lys Gly Phe Thr Glu Ser Leu Asp
145 150 155 160

Pro Arg Arg Arg Ile Pro Gly Gly Val Ser Thr Thr Glu Pro Ser Thr
165 170 175

Ser Pro Ser Asn Asn Gly Glu Val Thr Gln Pro Pro Arg Ile Leu Gly
180 185 190

```
Glu Ala Ser Gly Leu Ala Val His Gln Ile Ile Thr Ile Thr Val Ser
        195                 200                 205

Leu Ile Met Val Ile Ala Ala Leu Ile Thr Thr Leu Val Leu Lys Asn
    210                 215                 220

Cys Cys Ala Gln Ser Gly Asn Thr Arg Arg Asn Ser His Gln Arg Lys
225                 230                 235                 240

Thr Asn Gln Gln Glu Glu Ser Cys Gln Asn Leu Thr Asp Phe Pro Ser
                245                 250                 255

Ala Arg Val Pro Ser Ser Leu Asp Ile Phe Thr Ala Tyr Asn Glu Thr
            260                 265                 270

Leu Gln Cys Ser His Glu Cys Val Arg Ala Ser Val Pro Val Tyr Thr
        275                 280                 285

Asp Glu Thr Leu His Ser Thr Thr Gly Glu Tyr Lys Ser Thr Phe Asn
    290                 295                 300

Gly Asn Arg Pro Ser Ser Ser Asp Arg His Leu Ile Pro Val Ala Phe
305                 310                 315                 320

Val Ser Glu Lys Trp Phe Glu Ile Ser Cys
                325                 330
```

<210> 5

<211> 218

<212> DNA

<213> Homo sapiens

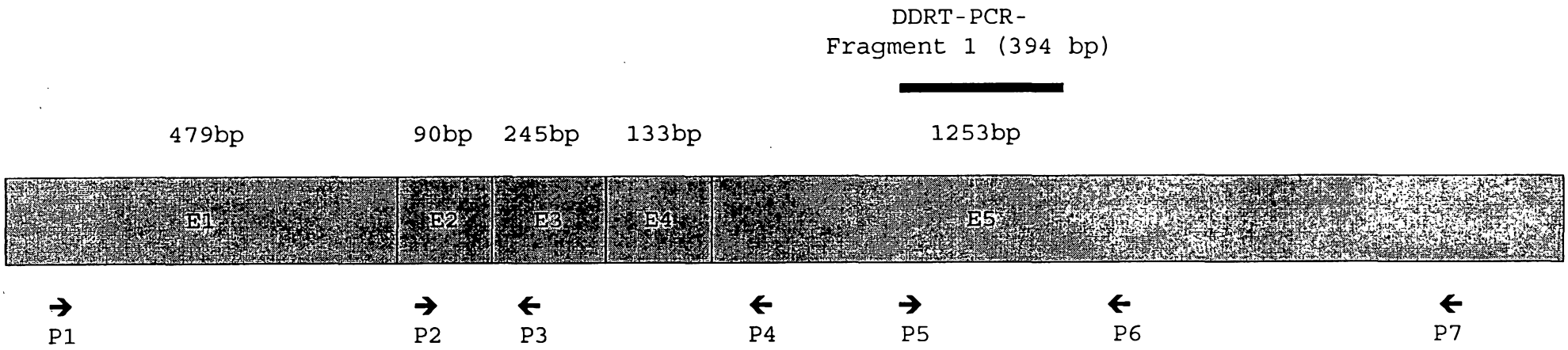
<400> 5

```
gcggttgtcc ggaatgccag tggctcctgg gcagatgtgc accccagatt cagcctttgt 60

gatagattcc aacacgttct ggcctcagac cacctttgtg gtggggccag actgctctgg 120

gcaaagtgaa gctggccttt atgctccaag gaaggggggc tcgagagcag gcctgcattg 180

gctctcggac taattcgcga tcatctttca tacagcag                          218
```



E1-E5: **Exon 1 to exon 5**
The respective exon sizes are indicated above the drawing.

P1-P7: **PCR primer**
For sequences and exact identity of the primers see table 1

The position of the DDRT-PCR fragment 1 sequence is indicated by a black bar.

THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:

1. Nucleic acid,
characterized in
that it comprises
(a) the nucleotide sequences depicted in SEQ ID NO. 1, 3 or/and 5, a combination or a protein-encoding segment thereof,
(b) a nucleotide sequence corresponding to the sequence in (a) within the scope of the degeneracy of the genetic code or
(c) a nucleotide sequence hybridizing with the sequences in (a) and/or (b) under stringent conditions,
with the proviso that the nucleic acid is different from the sequences stated with accession numbers Z98886, Ac003017, Aa453993, AL023586 and Aa452856 in the EMBL EST database.

2. Nucleic acid according to Claim 1,
characterized in
that it comprises a protein-encoding segment of the nucleotide sequences depicted in SEQ ID NO. 1, 3 or/and 5.

3. Nucleic acid according to Claim 1
characterized in
that it has a homology of more than 80% to the nucleotide sequences depicted in SEQ ID NO. 1, 3 or/and 5.

4. Nucleic acid according to any one of Claims 1 to 3,
characterized in
that it codes for a polypeptide associated with invasive processes or for a segment thereof.

5. Modified nucleic acid or nucleic acid analog which

comprises a nucleotide sequence according to any one of Claims 1 to 4.

6. Polypeptide,
characterized in
that it is encoded by a nucleic acid according to any one of Claims 1 to 4, where the proviso of Claim 1 is to be disregarded.

7. Polypeptide according to Claim 6,
characterized in
that it has
(a) the amino acid sequence depicted in SEQ ID NO. 2 or 4 or
(b) a homology of more than 70% to the amino acid sequence according to (a).

8. Modified polypeptide comprising an amino acid sequence according to Claim 6 or 7.

9. Peptide,
characterized in
that it represents a segment of at least 10 amino acids of the amino acid sequence depicted in SEQ ID NO. 2 or 4.

10. Vector,
characterized in
that it has at least one copy of a nucleic acid according to any one of Claims 1 to 4.

11. Vector according to Claim 10,
characterized in
that it facilitates expression of the nucleic acid in a suitable host cell.

12. Cell,
characterized in

that it has been transformed with a nucleic acid according to any one of Claims 1 to 4 or with a vector according to Claim 10 or 11.

13. Antibody against a polypeptide according to any one of Claims 6 to 8 or against a peptide according to Claim 9.

14. Antibody according to Claim 13
characterized in
that it is directed against the complete polypeptide or against a fragment thereof selected from a segment of amino acids 1 to 330 from SEQ ID NO. 4.

15. Composition for pharmaceutical application, characterized in
that it comprises as active component:
(a) a nucleic acid according to any one of Claims 1 to 5, where the proviso of Claim 1 is to be disregarded,
(b) a vector according to Claim 10 or 11,
(c) a cell according to Claim 12,
(d) a polypeptide according to any one of Claims 6 to 8,
(e) a peptide according to Claim 9 and/or
(f) an antibody according to Claim 13 or 14.

16. Composition according to Claim 15,
characterized in
that it additionally contains pharmaceutically conventional carriers, excipients and/or additives.

17. Use of a polypeptide according to any one of Claims 6 to 8 or of a peptide according to claim 9 for the preparation of an immunogen for producing antibodies.

18. Use of a composition according to Claim 15 or 16 for the preparation of an agent for the diagnosis of diseases related to invasive processes.

19. Use of a composition according to Claim 15 or 16 for the preparation of an agent for the diagnosis of a predisposition for diseases related to invasive processes.

20. Use of a composition according to Claim 15 or 16 for the preparation of an agent for the therapy or prevention of diseases related to invasive processes.

21. Use of a composition according to Claim 15 or 16 for the preparation of an agent for the diagnosis, therapy or prevention of endometriosis.

22. Use of a composition according to Claim 15 or 16 for the preparation of an agent for the diagnosis, therapy or prevention of neoplastic diseases.

23. Use of a composition according to Claim 15 or 16 for the preparation of an agent for gene therapy.

24. Use of a composition according to Claim 15 or 16 for the preparation of an antisense inhibitor.

25. Use of a composition according to Claim 15 or 16 for the preparation of an agent in the implantation of embryos.

26. Use of a composition according to Claim 15 or 16 for the identification of inhibitors of a polypeptide according to any one of Claims 6 to 8 and/or of inhibitors of molecules which are capable of binding to the polypeptide.

27. Use of a polypeptide according to any one of Claims 6 to 8 or of a peptide according to claim 9 for the detection of antibodies against an endometriosis-associated protein or fragments thereof in a sample.

28. Use of an antibody according to Claim 13 or 14 or of an antigen binding fragment of this antibody for the detection of endometriosis-associated proteins or fragments thereof.

DATED this 14th day of July, 2003

ANNA STARZINSKI-POWITZ
by his Patent Attorneys
DAVIES COLLISON CAVE