

Method for the Diagnosis of Thromboembolic and Coronary Heart Diseases

5

The present invention refers to a method for the *in vitro* diagnosis of thromboembolic and/or coronary heart diseases, wherein the nucleotide at position 470 of a nucleic acid coding for the human EGLN2 protein or the amino acid at position 58 of the human EGLN2 protein of a sample of a person is determined.

10

EGLN2, due to its HIF prolyl hydroxylase activity also known as prolyl hydroxylase domain-containing protein 1 (PHD1), belongs to a group of closely related proteins of the Egl-Nine gene family which has a conserved genomic structure consisting of five coding exons. HIF (hypoxia-inducible factor) is a transcriptional regulator that plays a key role in many aspects of oxygen homeostasis but the contribution of the EGLN isoforms EGLN1 (PHD1), EGLN2 (PHD2) and EGLN3 (PHD3) to the physiological regulation of HIF is still uncertain (Appelhoff, R. J. et al. (2004) J. Biol. Chem., 279, 38458-38465, No. 37). It is reported that all EGLN isoforms show a differing cell specific and inducible behaviour, which should allow flexibility in the regulation of the HIF response to hypoxia. This would mean that specific pharmacological inhibition of a particular EGLN isoenzyme could have the potential for selective modulation of the HIF response that would be useful in therapeutic applications (Appelhoff, R. J. et al. (2004), supra). EGLN2 inhibition, for example, should activate the HIF response broadly across a range of cell types under resting conditions. In contrast specific inhibition of EGLN3 should selectively augment the response to hypoxia in certain tissues that express high levels of the enzyme (Appelhoff, R. J. et al. (2004), supra). This could open the possibility to treat ischemic/hypoxic diseases. Contrary to EGLN2 and EGLN3 not much is known for the physiological role of EGLN1.

30 In order to better understand a potential involvement of EGLN2 in the occurrence and progression of coronary heart diseases, genotype-phenotype association analyses have been carried out with a well characterized patient group with respect to a variation in the EGLN2 gene in position 470 of the EGLN2 reference sequence published under the reference number NM_053046.2 in accordance with the present

invention. Different genetic variants of the EGLN2 gene are already known as SNPs (single nucleotide polymorphisms) and published under http://www.ncbi.nlm.nih.gov/SNP/snp_ref.cgi?locusId=112398).

5 Surprisingly it has been found that a variation of the nucleotide at position 470, in particular from cytosine to thymidine of a nucleic acid coding for the human EGLN2 protein or the amino acid at position 58, in particular from serine to leucine of the human EGLN2 protein correlates with the occurrence of thromboembolic and/or coronary heart diseases.

10 Therefore, a subject matter of the present invention relates to an *in vitro* or *in vivo* diagnosis of thromboembolic and/or coronary heart diseases, wherein the nucleotide at position 470 of a nucleic acid coding for the human EGLN2 protein or the amino acid at position 58 of the human EGLN2 protein of a sample of a person or patient is
15 determined.

In a preferred embodiment of the present invention the thromboembolic disease is stroke, a prolonged reversible ischemic neurological deficit (PRIND) and/or a transitoric ischemic attack (TIA) and the coronary heart disease is a myocardial
20 infarction.

In particular, if the nucleotide at position 470 is determined as thymidine in the chromosomal DNA or uracile in the mRNA or the amino acid at position 58 is determined as leucine there exists a higher risk of stroke, PRIND and/or TIA. If,
25 however, the nucleotide at position 470 is determined as cytidine or the amino acid at position 58 is determined as serine there exists a higher risk for a myocardial infarction, in particular early myocardial infarction.

According to the present invention, the term "EGLN2-C470C" refers to the group of
30 persons which have cytidine on both alleles of the gene coding for EGLN2 at position 470 of the reference sequence NM_053046.2 which leads to the amino acid serine at position 58 of the corresponding protein. These persons are homozygous with respect to this EGLN2 variant. Consequently, the term "EGLN2-C470T refers to the group of persons which have cytidine on one allele of the gene coding for EGLN2

which leads to serine at position 58 of the corresponding protein and thymidine on the other allele of the gene coding for EGLN2 which leads to leucine at position 58 of the corresponding protein. These persons are heterozygous with respect to this EGLN2 variant.

5

The nucleic acid sequence of the reference sequence coding for the human EGLN2 protein preferably has the nucleic acid sequence of SEQ ID NO: 1 and the amino acid sequence of the human EGLN2 protein preferably has the amino acid sequence of SEQ ID NO: 2. However, the present invention encompasses also other variants of human EGLN2 and the non-human homologs thereof, as for example other mammalian EGLN2 homologs or the EGLN2 homologs from *Caenorhabditis elegans*, mouse or rat, provided that there is a nucleotide exchange from cytidine to thymidine at the position corresponding to position 470 of said reference sequence and/or an amino acid exchange from serine to leucine at the position corresponding to position 58 of said reference sequence and further provided that the corresponding protein has a prolyl hydroxylase activity, in particular a HIF prolyl hydroxylase activity. Said enzyme activity can be measured for example by mass spectrometric analysis whereby the oxidization of Pro, e.g. Pro⁵⁶⁴ of HIF-1 α , can be detected, or by enzymatic assays known to a person skilled in the art.

20

Generally, the specific nucleotide at position 470 can be determined by a nucleic acid sequencing method, a mass spectrometric analysis of the nucleic acid, a hybridisation method and/or an amplification method. Examples of a nucleic acid sequencing method are pyrosequencing and/or sequencing with the help of radioactive and/or fluorescence labelled nucleotides. Examples of the hybridisation method are Southern blot analysis, Northern blot analysis and/or a hybridisation method on a DNA-microarray. Examples of an amplification method are a TaqMan analysis, a differential RNA display analysis and/or a representational difference analysis (Shi M. M. (2002) Am J Pharmacogenomics., 2(3), 197-205; Kozian & Kirschbaum (1999) Trends Biotechnol., 17(2), 73-8.)

30

Furthermore, the amino acid sequence at position 58 can be determined by a method measuring the amount of the specific protein and/or a method measuring the activity of the specific protein. Examples of a method for measuring the amount of the

specific protein are a Western blot analysis and/or an ELISA. Examples for measuring the activity of the specific protein are an *in vitro* test assay and/or an *in vitro* whole cell test assay with human cells, animal cells, bacterial cells or yeast cells, all known to a person skilled in the art.

5

Examples of a sample for the detection of the respective variant are a cell, a tissue or a body fluid, in particular in cellular components of the blood, endothelial cells or smooth muscle cells. Preferably the sample is pre-treated by conventional methods known to a person skilled in the art in order to isolate and/or purify the nucleic acids or chromosomal DNA, or the proteins of the sample for the further analysis.

10

In an optional further step the risk and/or the age of a person to suffer from a thromboembolic and/or coronary heart disease can be determined as shown in the examples.

15

In another optional further step the dosage of a pharmaceutical against a thromboembolic and/or coronary heart disease can be determined.

In general, the found genetic variation in the EGLN2 gene can be used in accordance with the present invention as a genetic marker for the risk assessment and/or the prophylactic treatment of a thromboembolic and/or coronary heart disease (also known as "cardiovascular disease"), in particular of an early myocardial infarction, stroke, PRIND, TIA and/or coronary heart diseases.

20

Furthermore, the genetic variation can be used in accordance with the present invention as a genetic marker for the adaptation of the dosage of an effective therapeutic agent for the treatment of a person, individual or patient and/or for the identification of persons, individuals or patients being under or selected to be under clinical trial studies with an increased risk for a thromboembolic and/or coronary heart disease, in particular of an early myocardial infarction, stroke, PRIND, TIA and/or coronary heart diseases. The genetic variation can also be used in accordance with the present invention for the evaluation of the tolerance, safety and efficacy of a pharmaceutically active substance for a specific person, individual or patient or for

25
30

identifying the person, individual or patient suitable for a particular treatment of said diseases.

In addition, the genetic variation can also be used in accordance with the present invention as part of a high throughput-screening assay for the detection and evaluation of pharmaceutically active compounds for the treatment of said diseases.

The present invention can also be used to identify risk factors for said diseases for each person, individual or patient to be treated or advised.

A preferred method for the diagnosis of a thromboembolic and/or coronary heart disease in accordance with the present invention contains the following steps:

- (a) obtaining a sample, in particular a cell, tissue, body fluid, a cellular component of the blood, endothelial cells or smooth muscle cells, from a person or patient that should be investigated;
- (b) isolating a nucleic acid probe, in particular a DNA probe from said sample;
- (c) amplifying the specific region encompassing position 470 of the ENGL2 gene with the help of primers, in particular the primers as specified in the Examples;
- (d) sequencing the amplified region;
- (e) analysing the sequenced region; and
- (d) assessing the risk for a thromboembolic and/or coronary heart disease, in particular for an early myocardial infarction, stroke, PRIND, TIA and/or coronary heart diseases.

An alternative method for the diagnosis of a thromboembolic and/or coronary heart disease in accordance with the present invention contains the following steps:

- (a) obtaining a sample, in particular a cell, tissue, body fluid, a cellular component of the blood, endothelial cells or smooth muscle cells, from a person or patient that should be investigated;
- (b) isolating the ENGL2 protein from said sample;
- (c) determining the amino acid at position 58 of the EGLN2 protein; and
- (d) assessing the risk for a thromboembolic and/or coronary heart disease, in particular for an early myocardial infarction, stroke, PRIND, TIA and/or coronary heart diseases.

More preferred steps are individually or collectively specified in the Examples and are incorporated hereby by reference to each step.

5 The following Figures, Tables, Sequences and Examples shall explain the present invention without limiting the scope of the invention.

10 DESCRIPTION OF THE FIGURES

Figure 1 shows the nucleic acid sequence of the human EGLN2 gene with the NCBI number NM_053046. The primers used for amplification of the genetic section with the genetic variation C→T at position 470 (bold face) are underlined.

Figure 2 shows the amino acid sequence of the human EGLN2 derived from the nucleic acid sequence with the NCBI number NM_053046. The amino acid position 58 in the EGLN2 protein is in bold face.

Figure 3 shows the influence of the genotype of EGLN2 at position 470 of the reference sequence NM_053042.2, leading to amino acid exchanges at position 58 of the EGLN2 protein, on the age of the occurrence of coronary heart diseases in the patients group. P-values less than 0.05 are statistically relevant.

DESCRIPTION OF THE SEQUENCES

30 SEQ ID NO: 1 shows the nucleic acid sequence of the human EGLN2 protein with the NCBI number NM_053046.

SEQ ID NO: 2 shows the amino acid sequence of the human EGLN2 derived from the nucleic acid sequence with the NCBI number NM_053046.

SEQ ID NO 3: shows the first primer sequence of nucleotides 444 – 463 of the reference sequence NM_053046.2.

- 5 SEQ ID NO 4: shows the second primer sequence of complementary sequence of bases 504 - 521 of the reference sequence NM_053046.2.

EXAMPLES

10

SNP DETECTION BY SEQUENCING AND ANALYSIS

Oligonucleotides (Primers) for Amplification:

15

The following primers were used for the detection of the nucleotide exchange from C to T at position 470 in the EGLN2 sequence with the reference number NM_053046.2:

- 20 Primer 1: 5'- CTGTCCAGGAGTGCCTAGTG -3' (nucleotides 444 – 463 of the reference sequence NM_053046.2; SEQ ID NO: 3);

Primer 2: 5'- GGGCTGGCAGTGGTAGAG -3' (complementary sequence of bases 504 - 521 of the reference sequence NM_053046.2; SEQ ID NO: 4).

25

PCR Protocol for Amplification:

- The reagents used were from Applied Biosystems (Foster City, USA): 20 ng of genomic DNA; 1 unit TaqGold DNA polymerase; 1x Taq polymerase buffer; 500µM dNTPs; 2,5mM MgCl₂; 200nM of each amplification primer pair as shown above; H₂O ad 5 µl.
- 30

Amplification Program of the PCR for Genotyping:

- 35 95°C for 10 min x 1 cycle

95°C for 30 sec
 70°C for 30 sec x 2 cycles;

5 95°C for 30 sec
 65°C for 30 sec x 2 cycles;

95°C for 30 sec
 60°C for 30 sec x 2 cycles;

10 95°C for 30 sec
 56°C for 30 sec
 72°C for 30 sec x 40 cycles;

15 72°C for 10 min
 4°C for 30 sec x 1 cycle;

Protocol for Minisequencing and Detection of SNPs

20 The reagents used were from Applied Biosystems (Foster City, USA). 2 µl purified PCR product, 1,5 µl BigDye-Terminator-Kit, 200nM of a sequencing primer as shown above; H₂O ad 10 µl.

25 Amplification Program for Sequencing:

96°C for 2 min x 1 cycle;

96°C for 10 sec

30 55°C for 10 sec
 65°C for 4 min x 30 cycles;

72°C for 7 min

4°C for 30 sec x 1 cycle;

Analyse of the Sequencing Products:

- 5 The sequences were analysed with the Sequenz Analyse Software (Applied Biosystems, Foster City, USA) for obtaining preliminary data first, then processed with the software Phred, Phrap, Polyphred und Consed. Phred, Phrap, Polyphred und Consed, written by Phil Green of the Washington University (<http://www.genome.washington.edu>).

10

RESULTS

Characteristics of the Group of Persons

15

Table 1 shows the characteristics of the group of persons studied.

Table 1

		n	%
Total		2074	
Sex	Female	603	29,07
	Male	1471	70,93
Age		61,8 (+/- 10,5)	
BMI (Body Mass Index)		29,1 (+/- 4,4)	
Blood Pressure		1214	58,7
Smoker		1372	66,41
Typ II Diabetes		361	17,46
Myocardial infarction		830	40,59
Stroke		145	7,01

20

Frequency and Distribution of the Variants of the EGLN2 Gene

- 25 Table 2 shows the frequency and distribution of the genetic variants of the EGLN2 gene at position 470 of the reference sequence NM_053046.2 in the patient group studied.

Table 2

	Frequency	Percentage
EGLN2-C470C (EGLN2 Ser58Ser)	1253	96,31
EGLN2-C470T (EGLN2 Ser58Leu)	47	3,61
EGLN2-T470T (EGLN2 Leu58Leu)	1	0,08
Missing values	773	

In the following only individuals with EGLN2-C470C (EGLN2 Ser58Ser) and EGLN2-C470T (EGLN2 Ser58Leu) are taken into account.

Influence of the Variants of EGLN2 on the Occurrence of Early Myocardial Infarction

Table 3 shows the influence of the genotype of EGLN2 at position 470 of the reference sequence NM_053046.2 on the occurrence of early myocardial infarction (less than 55 years old for men and less than 60 years old for women) and of stroke/PRIND (prolonged reversible ischemic neurological deficit)/TIA (transitory ischemic attack) in the patient group studied. P-values less than 0.05 are statistically relevant.

Table 3

Clinical Parameter	EGLN2		p-value
	C470C / Ser58Ser n (%)	C470T / Ser58Leu n (%)	
Patients with early myocardial infarction (<55m/60f)	215 (17,16)	2 (4,26%)	0,0199
Patients without early myocardial infarction (<55m/60f)	1038 (82,84)	45 (95,74)	
Patients with Stroke/PRIND/TIA	88 (7,23)	7 (14,89)	0,0418
Patients without Stroke/PRIND/TIA	1165 (92,77)	40 (85,11)	

Results:

1. The patients with EGLN2-C470C showed a statistically higher incidence for early myocardial infarction compared to patients with EGLN2-C470T.
2. The patients with EGLN2-C470T showed a significant increase of the risk to receive a stroke, PRIND and/or TIA compared with patients with EGLN2-C470C.

Influence of the Variants of EGLN2 on Patients' Age with Coronary Heart Diseases

Figure 3 shows the influence of the genotype of EGLN2 at position 470 of the reference sequence NM_053042.2 on the age of the occurrence of coronary heart diseases in the patients group.

Result:

A significant dependency of the age of the patients with EGLN2-C470C (EGLN2 Ser58Ser) for the early occurrence of coronary heart diseases was discovered compared to the age of patients with EGLN2-C470T (EGLN2 Ser58Leu).

Conclusion:

5 The statistically significant associations between the genetic variants of the gene coding for EGLN2 and/or the protein EGLN2 shown above are a clear indication for the involvement of said genetic variants in the occurrence of thrombotic and/or coronary heart diseases. Consequently, said genetic variants are biological markers for the prognosis of thrombotic and/or coronary heart diseases, in particular for the prognosis of early myocardial infarction and/or stroke, PRIND and/or TIA.

SEQ ID NO: 1

```

1  gctttccctt  gcctgcctgt  ctctagtttc  tctcacatcc  cttttttttt  ttcctttctc
61 tagccaccct  gaagggtccc  ttccaagcc  cttagggacc  gcagaggact  tggggaccag
5  121 caagcaaccc  ccagggcacg  agaagagctc  ttgctgtctg  ccctgcctca  ccctgccccca
181 cgccaggccc  ggtggcccc  agctgcatca  agtggaggcg  gaggaggagg  cggaggaggg
241 tggcaccatg  ggcccgggcg  gtgccctcca  tgcccgggg  atgaagacac  tgctgccatg
301 gacagcccgt  gccagccgca  gcccctaagt  caggctctcc  ctcaattacc  agggctctcg
361 tcagagccct  tggagcctga  gcctggccgg  gccaggatgg  gaggggagag  ttacctgccc
10  421 tgtcccttgc  tcccctccta  ccagtgtcca  ggagtgccta  gtgaggcctc  ggcagggagt
481 gggaccccca  gagccacagc  cacctctacc  actgccagcc  ctcttcggga  cggttttggc
541 gggcaggatg  gtggtgagct  gcggccgctg  cagagtgaag  gcgctgcagc  gctggtcacc
601 aaggggtgcc  agcgattggc  agcccagggc  gcacggcctg  agggcccca  acggaaatgg
661 gccgaggatg  gtggggatgc  cccttcaccc  agcaaacggc  cctgggccag  gcaagagaac
15  721 caggaggcag  agcgggaggg  tggcatgagc  tgcagctgca  gcagtggcag  tggtgaggcc
781 agtgctgggc  tgatggagga  ggcgctgcc  tctgcgccg  agcgctggc  cctggactat
841 atcgctccct  gcatgcggtg  ctacggcatc  tgcgtcaagg  acagcttctc  gggggcagca
901 ctgggcggcg  gcgtgctggc  cgaggtggag  gccctcaaac  ggggtggcg  cctgcgagac
961 gggcagctag  tgagccagag  ggcgatccc  ccgcgcagca  tccgtgggga  ccagattgcc
20  1021 tgggtggaag  gccatgaacc  aggtgtgcga  agcattggtg  ccctcatggc  ccattgtggc
1081 gccgtcatcc  gccactgcgc  agggcggtg  ggcagctatg  tcatcaacgg  gcgcaccaag
1141 gccatggtgg  cgtgttacc  aggcaacgg  ctgggtacg  taaggcacgt  tgacaatccc
1201 cagggcgatg  ggcgctgcat  cacctgtatc  tattacctga  atcagaactg  ggacgttaag
1261 gtgcatggcg  gcctgctgca  gatcttccct  gagggccggc  ccgtggtagc  caacatcgag
25  1321 ccactctttg  accggttgct  cattttctgg  tctgaccggc  ggaaccccca  cgaggtgaag
1381 ccagcctatg  ccaccaggta  cgccatcact  gtctggtatt  ttgatgcaa  ggagcgggca
1441 gcagccaaag  acaagtatca  gctagcatca  ggacagaaag  gtgtccaagt  acctgtatca
1501 cagccgccta  cgccaccta  gtggccagtc  ccagagccgc  atggcagaca  gcttaaataga
1561 cttcaggaga  gccctgggcc  tgtgctggct  gctccttccc  tgccaccgct  gctgcttctg
30  1621 actttgcctc  tgtcctgcct  ggtgtggagg  gctctgtctg  ttgctgagga  ccaaggagga
1681 gaagagacct  ttgctgcccc  atcatgggg  ctggggttgt  cacctggaca  gggggcagcc
1741 gtggaggcca  ccgttaccaa  ctgaagctgg  gggcctgggt  cctaccctgt  ctggtcatga
1801 cccattagg  tatggagagc  tgggaggagg  cattgtcact  tcccaccagg  atgcaggact
1861 tggggttgag  gtgagtcag  gcctcttgct  ggcaatgggg  tgggaggagt  acccccaagt
35  1921 cctctcactc  ctccagcctg  gaatgtgaag  tgactcccca  acccctttgg  ccattggcagg
1981 caccttttgg  actgggctgc  cactgcttgg  gcagagtaaa  aggtgccagg  aggagcatgg
2041 gtgtggaagt  cctgtcagcc  aagaaataaa  agtttacctc  agagctgcaa  aaaaaaaaaa
2101 aaaaaaaaaa  a

```

40

SEQ ID NO: 2

MDSPCQPQPLSQALPQLPGSSSEPLEPEPGRARMGVESYLPCLLPSYHCPGVPSEASAGSG
TPRATATSTTASPLRDGFGGQDGGELRPLQSEGAAALVTKGCQRLAAQGARPEAPKRKWAED
5 GGDAPSPSKRPWARQENQEAERE GGMSCSCSSGSGEASAGLMEEALPSAPERLALDYIVPCM
RYYGICVKDSFLGAALGGRVLA EVEALKRGGRLRDGQLVSQRAIPPRSIRGDQIAWVEGHEP
GCRSIGALMAHVDAVIRHCAGRLGSYVINGRTKAMVACYPGNGLGYVRHVDNPHGDGRCITC
IYYLNQNWDVKVHGGLLQIFPEGRPVVANIEPLFDRLLIFWSDRRNPHEVKPAYATRYAITV
10 WYFDAKERA AAKDKYQLASGQKGVQVPVVSQPPTPT

SEQ ID NO: 3

1 CTGTCCAGGA GTGCCTAGTG
15

SEQ ID NO: 4

1 GGGCTGGCAG TGGTAGAG
20

SEQ ID NO: 1

```

1  gcttttccct  gcttgcctgt  ctctagtttc  tctcacatcc  cttttttttt  ttcttttctc
61  tagccaccc  gaagggtccc  ttcccaagcc  cttagggacc  gcagaggact  tggggaccag
5  121  caagcaaccc  ccagggcacg  agaagagctc  ttgctgtctg  ccctgcctca  ccctgcccc
181  cgccaggccc  ggtggcccc  agctgcatca  agtggaggcg  gaggaggagg  cggaggaggg
241  tggcaccatg  ggcccgggcg  gtgccctcca  tgcccgggg  atgaagacac  tgctgccatg
301  gacagcccg  gccagccgca  gccctaagt  caggctctcc  ctcagttacc  agggctctcg
361  tcagagccct  tggagcctga  gcctggccgg  gccaggatgg  gagtggagag  ttacctgcc
10  421  tgtccctgc  tccctccta  ccactgtcca  ggagtgccta  gtgaggcctc  ggcagggagt
481  gggaccccca  gagccacagc  cacctctacc  actgccagcc  ctcttcggga  cggttttggc
541  gggcaggatg  gtggtgagct  gcggccgctg  cagagtgaag  gcgctgcagc  gctggtcacc
601  aaggggtgcc  agcgattggc  agcccagggc  gcacggcctg  agggcccca  acggaaatgg
661  gccgaggatg  gtggggatgc  cccttcaccc  agcaaacggc  cctgggccag  gcaagagaac
15  721  caggaggcag  agcgggagg  tggcatgagc  tgcagctgca  gcagtggcag  tggtagggc
781  agtgcctggc  tgatggagga  ggcgctgcc  tctgcgccg  agcgccctgg  cctggactat
841  atcgtgcct  gcatgcggt  ctacggcatc  tgcgtcaagg  acagcttct  gggggcagca
901  ctggcggtc  gcgtgctgg  cgagggtgag  gccctcaaac  ggggtggcg  cctgcgagac
961  gggcagctag  tgagccagag  ggcgatcccg  ccgcgcagca  tccgtgggga  ccagattgcc
20  1021  tgggtggaag  gccatgaacc  aggctgtcga  agcattggtg  ccctcatggc  ccattgtggc
1081  gccgtcatcc  gccactgcgc  agggcggtg  ggcagctatg  tcatcaacgg  ggcaccaag
1141  gccatggtgg  cgtgttacc  aggcaacgg  ctcggtacg  taaggcacgt  tgacaatcc
1201  cagggcgatg  ggcgctgcat  cacctgtatc  tattacctga  atcagaactg  ggacgttaag
1261  gtgcatggcg  gcctgctgca  gatcttccct  gagggcggc  ccgtggtagc  caacatcgag
25  1321  ccactcttg  accggttgct  cattttctgg  tctgacggc  ggaaccccca  cgaggatga
1381  ccagcctatg  ccaccaggta  cgccatcact  gtctggtatt  ttgatgcaa  ggagcgggca
1441  gcagccaaag  acaagtatca  gctagcatca  ggacagaaag  gtgtcaaagt  acctgtatca
1501  cagccgccta  cgccaccta  gtggccagtc  ccagagccgc  atggcagaca  gcttaaataga
1561  cttcaggaga  gccctggg  tgtgctggct  gctccttccc  tgccaacgct  gctgcttctg
30  1621  actttgcctc  tgcctgcct  ggtgtggagg  gctctgtctg  ttgctgagga  ccaaggagga
1681  gaagagacct  ttgctgcccc  atcatgggg  ctggggttgt  cacctggaca  gggggcagcc
1741  gtggaggcca  ccgttaccaa  ctgaagctgg  gggcctgggt  cctaccctgt  ctggtcatga
1801  cccattagg  tatggagagc  tgggaggagg  cattgtcact  tcccaccagg  atgcaggact
1861  tggggttgag  gtgagtcag  gcctcttgct  ggcaatggg  tgggaggagt  acccccaagt
35  1921  cctctcactc  ctccagcctg  gaatgtgaag  tgactcccca  acccctttgg  ccattggcagg
1981  caccttttgg  actgggctgc  cactgcttgg  gcagagtaaa  aggtgccagg  aggagcatgg
2041  gtgtggaagt  cctgtcagcc  aagaaataaa  agtttaccto  agagctgcaa  aaaaaaaaaa
2101  aaaaaaaaaa  a

```

40

SEQ ID NO: 2

5 MDSPCQPQPLSQALPQLPGSSSEPLEPEPGRARMGVESYLPCLLPSYHCPGVPSEASAGSG
TPRATATSTTASPLRDGFGGQDGGELRPLQSEGAAALVTKGCQRLAAQGARPEAPKRKWAED
GGDAPSPSKRPWARQENQEAEREAGGMS CSCSSGSEASAGLMEEALPSAPERLALDYIVPCM
RYYGICVKDSFLGAALGGRVLA EVEALKRGGRLRDGQLVSQRAIPPRSIRGDQIAWVEGHEP
GCRSIGALMAHVDAVIRHCAGRLGSYVINGRTKAMVACYPGNGLGYVRHVDNPHGDGRCITC
IYYLNQNWDVKVHGGLLQIFPEGRPVVANIEPLFDRLLI FWSDRRNPHEVKPAYATRYAITV
10 WYFDAKERA AAKDKYQLASGQKGVQVPVVSQPPTPT

SEQ ID NO: 3

1 CTGTCCAGGA GTGCCTAGTG
15

SEQ ID NO: 4

1 GGGCTGGCAG TGGTAGAG
20

Claims

- 5
1. A method for the in vitro diagnosis of thromboembolic and/or coronary heart diseases, wherein the nucleotide at position 470 of a nucleic acid coding for the human EGLN2 protein or the amino acid at position 58 of the human EGLN2 protein of a sample of a person is determined.

10

 2. The method of claim 1, wherein the thromboembolic disease is selected from the group consisting of stroke, prolonged reversible ischemic neurological deficit (PRIND) and/or transitoric ischemic attack (TIA).

15

 3. The method of claim 1, wherein the coronary heart disease is myocardial infarction.

20

 4. The method of claim 2, wherein the nucleotide at position 470 is determined as thymidine in the chromosomal DNA or uracile in the mRNA or the amino acid at position 58 is determined as leucine for a risk of stroke, PRIND and/or TIA.

25

 5. The method of claim 3, wherein the nucleotide at position 470 is determined as a cytidine or the amino acid at position 58 is determined as serine for a risk of myocardial infarction, in particular early myocardial infarction.

30

 6. The method according to any of the claims 1-5, wherein the nucleic acid coding for the human EGLN2 protein has the nucleotide sequence of SEQ ID NO: 1.
 7. The method according to any of the claims 1-5, wherein the human EGLN2 protein has the amino acid sequence of SEQ ID NO: 2.
 8. The method according to any of the claims 1-4 and 6, wherein the nucleotide at position 470 is determined by a method selected from the group consisting of a

nucleic acid sequencing method, a mass spectrometric analysis of the nucleic acid, a hybridisation method and/or an amplification method.

- 5 9. The method of claim 8, wherein the nucleic acid sequencing method is selected from the group selected from pyrosequencing and/or sequencing with the help of radioactive and/or fluorescence labelled nucleotides.
- 10 10. The method of claim 8, wherein the hybridisation method is selected from the group consisting of Southern blot analysis, Northern blot analysis and/or a hybridisation method on a DNA-microarray.
- 15 11. The method of claim 8, wherein the amplification method is selected from the group consisting of a TaqMan analysis, a differential RNA display analysis and/or a representational difference analysis.
- 20 12. The method according to any of the claims 1-3, 5 and 7, wherein the amino acid sequence at position 58 is determined by a method selected from the group consisting of a method measuring the amount of the specific protein and/or a method measuring the activity of the specific protein.
- 25 13. The method according to claim 12, wherein the amount of the specific protein is measured by a method selected from the group consisting of a western blot analysis and/or an ELISA.
- 30 14. The method according to claim 12, wherein the activity of the specific protein is measured by an in vitro test assay and/or an in vitro whole cell test assay with human cells, animal cells, bacterial cells or yeast cells.
15. The method according to any of the claims 1-14, wherein the sample is selected from the group consisting of a cell, a tissue and/or a body fluid.
16. The method according to any of the claims 1-15, wherein in a further step the risk of a person to suffer from a thromboembolic and/or coronary heart disease is determined.

17. The method according to any of the claims 1-15, wherein in a further step the dosage of a pharmaceutical is determined.

18. The method according to any of the claims 1-17 comprising the following steps:

- (a) obtaining a sample from a person that should be investigated;
- (b) isolating a nucleic acid probe, in particular a DNA probe from said sample;
- (c) amplifying the specific region encompassing position 470 of the ENGL2 gene with the help of primers;
- (d) sequencing the amplified region;
- (e) analysing the sequenced region; and
- (d) assessing the risk for a thromboembolic and/or coronary heart disease, in particular for a myocardial infarction, stroke, PRIND, TIA and/or coronary heart diseases.

19. The method according to claim 18, wherein the myocardial infarction is an early myocardial infarction.

20. The method according to claim 18 or 19, wherein the primers are selected from SEQ ID NO: 3 and/or SEQ ID NO: 4.

21. The method according to any of the claims 1-3, 5, 7 or 12-17 comprising the following steps:

- (a) obtaining a sample, in particular a cell, tissue, body fluid, a cellular component of the blood, endothelial cells or smooth muscle cells, from a person or patient that should be investigated;
- (b) isolating the ENGL2 protein from said sample;
- (c) determining the amino acid at position 58 of the EGLN2 protein; and
- (d) assessing the risk for a thromboembolic and/or coronary heart disease, in particular for an early myocardial infarction, stroke, PRIND, TIA and/or coronary heart diseases.

22. The method according to claim 20, wherein the myocardial infarction is an early myocardial infarction.

5

Summary

The present invention refers to a method for the *in vitro* diagnosis of thromboembolic and/or coronary heart diseases, wherein the nucleotide at position 470 of a nucleic acid coding for the human EGLN2 protein or the amino acid at position 58 of the human EGLN2 protein of a sample of a person is determined.

SEQUENCE LISTING

<110> Sanofi-Aventis Deutschland GmbH

<120> Method for the Diagnosis of Thromboellic and Coronary Heart Diseases

<130> DE2005/049

<140>

<141>

<160> 4

<170> PatentIn Ver. 2.1

<210> 1

<211> 2111

<212> DNA

<213> Homo sapiens

<400> 1

```

gctttccctt gcttgctgt ctctagtttc tctcacatcc cttttttttt ttcctttctc 60
tagccaccct gaaggggtccc ttcccaagcc cttagggacc gcagaggact tggggaccag 120
caagcaaccc ccagggcacg agaagagctc ttgctgtctg ccctgcctca cctgccccca 180
cgccaggccc ggtggccccc agctgcatca agtgaggcg gaggaggagg cggaggaggg 240
tggcaccatg ggcccgggcg gtgccctcca tgcccgggg atgaagacac tgctgccatg 300
gacagcccgt gccagccgca gcccctaagt caggctctcc ctcaattacc agggctctcg 360
tcagagccct tggagcctga gcttgccgg gccaggatgg gaggaggagg ttacctgccc 420
tgtcccttgc tccctccta ccactgtcca ggagtgccta gtgaggcctc gccaggaggt 480
gggaccccc gagccacagc cactcttacc actgccagcc ctcttcggga cggttttggc 540
gggcaggatg gtggtgagct gggccgctg cagagtgaag gcgctgcagc gctggtcacc 600
aaggggtgcc agcagattggc agcccagggc gcacggcctg agggccccc aaacgaaatgg 660
gccgaggatg gtggggatgc cccttcaccc agcaaaccgc cctggggccag gcaagagaac 720
caggaggcag agcgggaggg tggcatgagc tgcagctgca gcagtggcag tggtagaggc 780
agtgtgggc tgatggagga ggcgctgccc tctgcgccc agcgcctggc cctggactat 840
atcgtgcctt gcatgcggtc ctacggcatc tgcgtcaagg acagcttctt gggggcagca 900
ctgggcggtc gctgtctggc cgaggtggag gccctcaaag ggggtgggcg cctgcgagac 960
gggcagctgc ggcagcagag ggcgatcccg ccgcgcagca tccgtgggga ccagattgcc 1020
tgggtggaag gccatgaacc aggtgtgca agcattgggt ccctcatggc ccatgtggac 1080
gccgtcatcc gccactgcgc agggcggtg ggcagctatg tcatcaacgg gcgcaccaag 1140
gccatggtgg cgtgttacc aggcaacggg ctgggtacg taaggcacgt tgacaatccc 1200
cacggcgatg ggcgctgcat cactgtatc tattacctga atcagaactg ggacgttaag 1260
gtgcatggcg gctgtctgca gatcttccct gagggccggc ccgtggtagc caacatcgag 1320
ccactctttg accggttgct cttttctggt tctgacggc ggaaccccc cgaggtgaag 1380
ccagcctatg ccaccaggta cgcctcact gtctggtatt ttgatgcaa ggagcgggca 1440
gcagccaaag acaagtatca gctagcatca ggacagaaag gtgtccaagt acctgtatca 1500
cagccgccta cgcacaccta gtggccagtc ccagagccgc atggcagaca gcttaaata 1560
cttcaggaga gccctgggct tgtgtgtggt gctccttccc tggcaccgct gctgcttctg 1620
actttgcctc tgcctgccc ggtgtggagg gctctgtctg ttgctgagga ccaaggagga 1680
gaagagacct ttgtgcccc atcatggggg ctggggttgt cactggaca gggggcagcc 1740
gtggaggcca cgttaccaca ctgaagctgg gggcctgggt cctaccctgt ctggtcatga 1800
ccccattagg tatggagagc tgggaggagg cattgtcact tcccaccagg atgcaggact 1860
tggggttgag gtgagtcag gctcttggc ggcaatgggg tgggaggagt accccaagt 1920
cctctcactc ctccagcctg gaatgtgaag tgactcccca acccctttgg ccatggcagg 1980
cactttttgg actgggctgc cactgcttgg gcagagtaaa aggtgccagg aggagcatgg 2040
gtgtggaagt cctgtcagcc aagaaataaa agtttacctc agagctgcaa aaaaaaaaaa 2100

```

aaaaaaaaa a

2111

<210> 2

<211> 407

<212> PRT

<213> Homo sapiens

<400> 2

Met Asp Ser Pro Cys Gln Pro Gln Pro Leu Ser Gln Ala Leu Pro Gln
1 5 10 15

Leu Pro Gly Ser Ser Ser Glu Pro Leu Glu Pro Glu Pro Gly Arg Ala
20 25 30

Arg Met Gly Val Glu Ser Tyr Leu Pro Cys Pro Leu Leu Pro Ser Tyr
35 40 45

His Cys Pro Gly Val Pro Ser Glu Ala Ser Ala Gly Ser Gly Thr Pro
50 55 60

Arg Ala Thr Ala Thr Ser Thr Thr Ala Ser Pro Leu Arg Asp Gly Phe
65 70 75 80

Gly Gly Gln Asp Gly Gly Glu Leu Arg Pro Leu Gln Ser Glu Gly Ala
85 90 95

Ala Ala Leu Val Thr Lys Gly Cys Gln Arg Leu Ala Ala Gln Gly Ala
100 105 110

Arg Pro Glu Ala Pro Lys Arg Lys Trp Ala Glu Asp Gly Gly Asp Ala
115 120 125

Pro Ser Pro Ser Lys Arg Pro Trp Ala Arg Gln Glu Asn Gln Glu Ala
130 135 140

Glu Arg Glu Gly Gly Met Ser Cys Ser Cys Ser Ser Gly Ser Gly Glu
145 150 155 160

Ala Ser Ala Gly Leu Met Glu Glu Ala Leu Pro Ser Ala Pro Glu Arg
165 170 175

Leu Ala Leu Asp Tyr Ile Val Pro Cys Met Arg Tyr Tyr Gly Ile Cys
180 185 190

Val Lys Asp Ser Phe Leu Gly Ala Ala Leu Gly Gly Arg Val Leu Ala
195 200 205

Glu Val Glu Ala Leu Lys Arg Gly Gly Arg Leu Arg Asp Gly Gln Leu
210 215 220

Val Ser Gln Arg Ala Ile Pro Pro Arg Ser Ile Arg Gly Asp Gln Ile
225 230 235 240

Ala Trp Val Glu Gly His Glu Pro Gly Cys Arg Ser Ile Gly Ala Leu
245 250 255

Met Ala His Val Asp Ala Val Ile Arg His Cys Ala Gly Arg Leu Gly
260 265 270

Ser Tyr Val Ile Asn Gly Arg Thr Lys Ala Met Val Ala Cys Tyr Pro
275 280 285

Gly Asn Gly Leu Gly Tyr Val Arg His Val Asp Asn Pro His Gly Asp
290 295 300

Gly Arg Cys Ile Thr Cys Ile Tyr Tyr Leu Asn Gln Asn Trp Asp Val
305 310 315 320

Lys Val His Gly Gly Leu Leu Gln Ile Phe Pro Glu Gly Arg Pro Val
325 330 335

Val Ala Asn Ile Glu Pro Leu Phe Asp Arg Leu Leu Ile Phe Trp Ser
340 345 350

Asp Arg Arg Asn Pro His Glu Val Lys Pro Ala Tyr Ala Thr Arg Tyr
355 360 365

Ala Ile Thr Val Trp Tyr Phe Asp Ala Lys Glu Arg Ala Ala Ala Lys
370 375 380

Asp Lys Tyr Gln Leu Ala Ser Gly Gln Lys Gly Val Gln Val Pro Val
385 390 395 400

Ser Gln Pro Pro Thr Pro Thr
405

<210> 3
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 3
ctgtccagga gtgcctagtg

20

<210> 4
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: primer

<400> 4
gggctggcag tggtagag

18