Method for the Diagnosis of Thromboembolic and Coronary Heart Diseases

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.

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.

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).

Surprisingly it has been found that a variation of the nucleotide at position 470, in particular from cytidine 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.

SUMMARY

There is provided 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, wherein when 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 is a risk for a thromboembolic disease and/or coronary heart disease.

Therefore, a subject matter of the present disclosure relates to an *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 or patient is determined.

In a preferred embodiment of the present disclosure 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 infarction.

In particular, if the nucleotide at position 470 is determined as thymidine in the chromosomal DNA or uridine 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, 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 disclosure, the term "EGLN2-C470C" refers to the group of 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.

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 *Caenorhabdidis 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. Pro^{564} of HIF-1α, can be detected, or by enzymatic assays known to a person skilled in the art.

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.)

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.

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.

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.

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.

Furthermore, the genetic variation can be used in accordance with the present disclosure 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

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 disclosure 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 disclosure 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 disclosure 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, stoke, 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 disclosure 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, stoke, 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.

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

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

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.

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

EXAMPLES

SNP DETECTION BY SEQUENCING AND ANALYSIS

Oligonucleotides (Primers) for Amplification:

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:

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).

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 $MgCl_2$; 200nM of each amplification primer pair as shown above; H_2O ad 5 µl.

Amplification Program of the PCR for Genotyping:

95°C for 10 min x 1 cycle

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

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;

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

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

Protocol for Minisequencing and Detection of SNPs

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; H_2O ad 10 µl.

Amplification Program for Sequencing:

96°C for 2 min x 1 cycle;

96°C for 10 sec
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:

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).

RESULTS

Characteristics of the Group of Persons

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 Preasure		1214	58,7
Smoker		1372	66,41
Typ II Diabetis		361	17,46
Myocardial infarction		830	40,59
Stroke		145	7,01

Frequence and Distribution of the Variants of the EGLN2 Gene

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

	Frequeny	Percentage
EGNL2-C470C (EGNL2 Ser58Ser)	1253	96,31
EGNL2-C470T (EGNL2 Ser58Leu)	47	3,61
EGNL2-T470T (EGNL2 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 (transitoric ischemic attack) in the patient group studied. P-values less than 0.05 are statistically relevant.

Table 3

Clinical Parameter	EGLN2 C470C / Ser58Ser n (%)	EGLN2 C470T / Ser58Leu n (%)	p-value
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:

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 gctttcccct gcctgcctgt ctctagtttc tctcacatcc ctttttttct ttcctttctc
  61 tagccaccct gaagggtccc ttcccaagcc cttagggacc gcagaggact tggggaccag
 121 caagcaaccc ccagggcacg agaagagctc ttgctgtctg ccctgcctca ccctgcccca
 181 cgccaggccc ggtggccccc agctgcatca agtggaggcg gaggaggagg cggaggaggg
 241 tggcaccatg ggcccgggcg gtgccctcca tgcccggggg atgaagacac tgctgccatg
 301 gacagcccgt gccagccgca gcccctaagt caggctctcc ctcagttacc agggtcttcg
 361 tcagagccct tggagcctga gcctggccgg gccaggatgg gagtggagag ttacctgccc
 421 tgtcccctgc tcccctccta ccactgtcca ggagtgccta gtgaggcctc ggcagggagt
 481 gggaccccca gagccacagc cacctctacc actgccagcc ctcttcggga cggttttggc
 541 gggcaggatg gtggtgagct gcggccgctg cagagtgaag gcgctgcagc gctggtcacc
 601 aaggggtgcc agcgattggc agcccagggc gcacggcctg aggcccccaa acggaaatgg
 661 gccgaggatg gtggggatgc cccttcaccc agcaaacggc cctgggccag gcaagagaac
 721 caggaggcag agcgggaggg tggcatgagc tgcagctgca gcagtggcag tggtgaggcc
 781 agtgctgggc tgatggagga ggcgctgccc tctgcgcccg agcgcctggc cctggactat
 841 atcgtgccct gcatgcggta ctacggcatc tgcgtcaagg acagcttcct gggggcagca
 901 ctgggcggtc gcgtgctggc cgaggtggag gccctcaaac ggggtgggcg cctgcgagac
 961 gggcagctag tgagccagag ggcgatcccg ccgcgcagca tccgtgggga ccagattgcc
1021 tgggtggaag gccatgaacc aggctgtcga agcattggtg ccctcatggc ccatgtggac
1081 gccgtcatcc gccactgcgc agggcggctg ggcagctatg tcatcaacgg gcgcaccaag
1141 gccatggtgg cgtgttaccc aggcaacggg ctcgggtacg taaggcacgt tgacaatccc
1201 cacggcgatg ggcgctgcat cacctgtatc tattacctga atcagaactg ggacgttaag
1261 gtgcatggcg gcctgctgca gatcttccct gagggccggc ccgtggtagc caacatcgag
1321 ccactctttg accggttgct cattttctgg tctgaccggc ggaaccccca cgaggtgaag
1381 ccagcctatg ccaccaggta cgccatcact gtctggtatt ttgatgccaa ggagcgggca
1441 gcagccaaag acaagtatca gctagcatca ggacagaaag gtgtccaagt acctgtatca
1501 cagccgccta cgcccaccta gtggccagtc ccagagccgc atggcagaca gcttaaatga
1561 cttcaggaga gccctgggcc tgtgctggct gctccttccc tgccaccgct gctgcttctg
1621 actttgcctc tgtcctgcct ggtgtggagg gctctgtctg ttgctgagga ccaaggagga
1681 gaagagacct ttgctgcccc atcatggggg ctggggttgt cacctggaca gggggcagcc
1741 gtggaggcca ccgttaccaa ctgaagctgg gggcctgggt cctaccctgt ctggtcatga
1801 ccccattagg tatggagagc tgggaggagg cattgtcact tcccaccagg atgcaggact
1861 tggggttgag gtgagtcatg gcctcttgct ggcaatgggg tgggaggagt acccccaagt
1921 cctctcactc ctccagcctg gaatgtgaag tgactcccca acccctttgg ccatggcagg
1981 caccttttgg actgggctgc cactgcttgg gcagagtaaa aggtgccagg aggagcatgg
2041 gtgtggaagt cctgtcagcc aagaaataaa agtttacctc agagctgcaa aaaaaaaaaa
2101 aaaaaaaaaa a
```

SEQ ID NO: 2

MDSPCQPQPLSQALPQLPGSSSEPLEPEPGRARMGVESYLPCPLLPSYHCPGVPSEASAGSG
TPRATATSTTASPLRDGFGGQDGGELRPLQSEGAAALVTKGCQRLAAQGARPEAPKRKWAED
GGDAPSPSKRPWARQENQEAEREGGMSCSCSSGSGEASAGLMEEALPSAPERLALDYIVPCM
RYYGICVKDSFLGAALGGRVLAEVEALKRGGRLRDGQLVSQRAIPPRSIRGDQIAWVEGHEP
GCRSIGALMAHVDAVIRHCAGRLGSYVINGRTKAMVACYPGNGLGYVRHVDNPHGDGRCITC
IYYLNQNWDVKVHGGLLQIFPEGRPVVANIEPLFDRLLIFWSDRRNPHEVKPAYATRYAITV
WYFDAKERAAAKDKYQLASGQKGVQVPVSQPPTPT

SEQ ID NO: 3

1 CTGTCCAGGA GTGCCTAGTG

SEQ ID NO: 4

1 GGGCTGGCAG TGGTAGAG

SEQ ID NO: 1

```
   1 gctttccect gcctgcctgt ctctagtttc tctcacatcc ctttttttt ttcctttctc
  61 tagccaccct gaagggtccc ttcccaagcc cttagggacc gcagaggact tggggaccag
 121 caagcaaccc ccagggcacg agaagagctc ttgctgtctg ccctgcctca ccctgcccca
 181 cgccaggccc ggtggccccc agctgcatca agtggaggcg gaggaggagg cggaggaggg
 241 tggcaccatg ggcccgggcg gtgccctcca tgcccggggg atgaagacac tgctgccatg
 301 gacagcccgt gccagccgca gcccctaagt caggctctcc ctcagttacc agggtcttcg
 361 tcagagccct tggagcctga gcctggccgg gccaggatgg gagtggagag ttacctgccc
 421 tgtcccctgc tcccctccta ccactgtcca ggagtgccta gtgaggcctc ggcagggagt
 481 gggacccccа gagccacagc cacctctacc actgccagcc ctcttcggga cggttttggc
 541 gggcaggatg gtggtgagct gcggccgctg cagagtgaag gcgctgcagc gctggtcacc
 601 aaggggtgcc agcgattggc agcccagggc gcacggcctg aggcccccaa acggaaatgg
 661 gccgaggatg gtggggatgc cccttcaccc agcaaacggc cctgggccag gcaagagaac
 721 caggaggcag agcgggaggg tggcatgagc tgcagctgca gcagtggcag tggtgaggcc
 781 agtgctgggc tgatggagga ggcgctgccc tctgcgcccg agcgcctggc cctggactat
 841 atcgtgccct gcatgcggta ctacggcatc tgcgtcaagg acagcttcct gggggcagca
 901 ctgggcggtc gcgtgctggc cgaggtggag gccctcaaac ggggtgggcg cctgcgagac
 961 gggcagctag tgagccagag ggcgatcccg ccgcgcagca tccgtgggga ccagattgcc
1021 tgggtggaag gccatgaacc aggctgtcga agcattggtg ccctcatggc ccatgtggac
1081 gccgtcatcc gccactgcgc agggcggctg ggcagctatg tcatcaacgg gcgcaccaag
1141 gccatggtgg cgtgttaccc aggcaacggg ctcgggtacg taaggcacgt tgacaatccc
1201 cacggcgatg ggcgctgcat cacctgtatc tattacctga atcagaactg ggacgttaag
1261 gtgcatggcg gcctgctgca gatcttccct gagggccggc ccgtggtagc caacatcgag
1321 ccactctttg accggttgct cattttctgg tctgaccggc ggaaccccca cgaggtgaag
1381 ccagcctatg ccaccaggta cgccatcact gtctggtatt ttgatgccaa ggagcgggca
1441 gcagccaaag acaagtatca gctagcatca ggacagaaag gtgtccaagt acctgtatca
1501 cagccgccta cgcccaccta gtggccagtc ccagagccgc atggcagaca gcttaaatga
1561 cttcaggaga gccctgggcc tgtgctggct gctccttccc tgccaccgct gctgcttctg
1621 actttgcctc tgtcctgcct ggtgtggagg gctctgtctg ttgctgagga ccaaggagga
1681 gaagagacct ttgctgcccc atcatggggg ctggggttgt cacctggaca gggggcagcc
1741 gtggaggcca ccgttaccaa ctgaagctgg gggcctgggt cctaccctgt ctggtcatga
1801 ccccattagg tatggagagc tgggaggagg cattgtcact tcccaccagg atgcaggact
1861 tggggttgag gtgagtcatg gcctcttgct ggcaatgggg tgggaggagt acccccaagt
1921 cctctcactc ctccagcctg gaatgtgaag tgactcccca acccctttgg ccatggcagg
1981 caccttttgg actgggctgc cactgcttgg gcagagtaaa aggtgccagg aggagcatgg
2041 gtgtggaagt cctgtcagcc aagaaataaa agtttacctc agagctgcaa aaaaaaaaaa
2101 aaaaaaaaaa a
```

SEQ ID NO: 2

MDSPCQPQPLSQALPQLPGSSSEPLEPEPGRARMGVESYLPCPLLPSYHCPGVPSEASAGSG
TPRATATSTTASPLRDGFGGQDGGELRPLQSEGAAALVTKGCQRLAAQGARPEAPKRKWAED
GGDAPSPSKRPWARQENQEAEREGGMSCSCSSGSGEASAGLMEEALPSAPERLALDYIVPCM
RYYGICVKDSFLGAALGGRVLAEVEALKRGGRLRDGQLVSQRAIPPRSIRGDQIAWVEGHEP
GCRSIGALMAHVDAVIRHCAGRLGSYVINGRTKAMVACYPGNGLGYVRHVDNPHGDGRCITC
IYYLNQNWDVKVHGGLLQIFPEGRPVVANIEPLFDRLLIFWSDRRNPHEVKPAYATRYAITV
WYFDAKERAAAKDKYQLASGQKGVQVPVSQPPTPT

SEQ ID NO: 3

1 CTGTCCAGGA GTGCCTAGTG

SEQ ID NO: 4

1 GGGCTGGCAG TGGTAGAG

Claims

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, wherein when 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 is a risk for a thromboembolic disease and/or coronary heart disease.

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

3. 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, wherein when the nucleotide at position 470 is determined as a cytidine or the amino acid at position 58 is determined as serine, there is a risk for a thromboembolic disease and/or coronary heart disease.

4. The method of claim 3, wherein there is a risk of a coronary heart disease being myocardial infarction, in particular, early myocardial infarction.

5. The method according to any of the claims 1-4, wherein the nucleic acid coding for the human EGLN2 protein has the nucleotide sequence of SEQ ID NO: 1.

6. The method according to any of the claims 1-4, wherein the human EGLN2 protein has the amino acid sequence of SEQ ID NO: 2.

7. The method according to any of the claims 1-5, 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.

8. The method of claim 7, 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.

9. The method of claim 7, 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.

10. The method of claim 7, 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.

11. The method according to any of the claims 1-4 or 6, 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.

12. The method according to claim 11, 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.

13. The method according to claim 11, 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.

14. The method according to any of the claims 1-13, wherein the sample is selected from the group consisting of a cell, a tissue and/or a body fluid.

15. The method according to any of the claims 1-14, wherein in a further step the risk of a person to suffer from a thromboembolic and/or coronary heart disease is determined.

16. The method according to any of the claims 1-15 comprising the following steps:

(a) isolating a nucleic acid probe, in particular a DNA probe from a sample that had been obtained from a person that should be investigated;
(b) amplifying the specific region encompassing position 470 of the EGLN2 gene with the help of primers;
(c) sequencing the amplified region;
(d) analysing the sequenced region; and
(e) 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.

17. The method according to claim 16, wherein the myocardial infarction is an early myocardial infarction.

18. The method according to claim 16 or 17, wherein the primers are selected from SEQ ID NO: 3 and/or SEQ ID NO: 4.

19. The method according to any of the claims 1-4, 6 or 11-15, comprising the following steps:

(a) isolating the EGLN2 protein from a sample, in particular a cell, tissue, body fluid, a cellular component of the blood, endothelial cells or smooth muscle cells, that had been obtained from a person or patient that should be investigated;
(b) determining the amino acid at position 58 of the EGLN2 protein; and
(c) 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.

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

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 Thromboelic 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
gctttcccct gcctgcctgt ctctagtttc tctcacatcc ctttttttt ttcctttctc 60
tagccaccct gaagggtccc ttcccaagcc cttagggacc gcagaggact tggggaccag 120
caagcaaccc ccagggcacg agaagagctc ttgctgtctg ccctgcctca ccctgcccca 180
cgccaggccc ggtggccccc agctgcatca agtggaggcg gaggaggagg cggaggaggg 240
tggcaccatg ggcccgggcg gtgccctcca tgcccggggg atgaagacac tgctgccatg 300
gacagcccgt gccagccgca gcccctaagt caggctctcc ctcagttacc agggtcttcg 360
tcagagccct tggagcctga gcctggccgg gccaggatgg gagtggagag ttacctgccc 420
tgtcccctgc tcccctccta ccactgtcca ggagtgccta gtgaggcctc ggcagggagt 480
gggacccccа gagccacagc cacctctacc actgccagcc ctcttcggga cggttttggc 540
gggcaggatg gtggtgagct gcggccgctg cagagtgaag gcgctgcagc gctggtcacc 600
aaggggtgcc agcgattggc agcccagggc gcacggcctg aggcccccaa acggaaatgg 660
gccgaggatg gtggggatgc cccttcaccc agcaaacggc cctgggccag gcaagagaac 720
caggaggcag agcgggaggg tggcatgagc tgcagctgca gcagtggcag tggtgaggcc 780
agtgctgggc tgatggagga ggcgctgccc tctgcgcccg agcgcctggc cctggactat 840
atcgtgccct gcatgcggta ctacggcatc tgcgtcaagg acagcttcct gggggcagca 900
ctgggcggtc gcgtgctggc cgaggtggag gccctcaaac ggggtgggcg cctgcgagac 960
gggcagctag tgagccagag ggcgatcccg ccgcgcagca tccgtgggga ccagattgcc 1020
tgggtggaag gccatgaacc aggctgtcga agcattggtg ccctcatggc ccatgtggac 1080
gccgtcatcc gccactgcgc agggcggctg ggcagctatg tcatcaacgg gcgcaccaag 1140
gccatggtgg cgtgttaccc aggcaacggg ctcgggtacg taaggcacgt tgacaatccc 1200
cacggcgatg ggcgctgcat cacctgtatc tattacctga atcagaactg ggacgttaag 1260
gtgcatggcg gcctgctgca gatcttccct gagggccggc ccgtggtagc caacatcgag 1320
ccactctttg accggttgct cattttctgg tctgaccggc ggaaccccca cgaggtgaag 1380
ccagcctatg ccaccaggta cgccatcact gtctggtatt ttgatgccaa ggagcgggca 1440
gcagccaaag acaagtatca gctagcatca ggacagaaag gtgtccaagt acctgtatca 1500
cagccgccta cgcccaccta gtggccagtc ccagagccgc atggcagaca gcttaaatga 1560
cttcaggaga gccctgggcc tgtgctggct gctccttccc tgccaccgct gctgcttctg 1620
actttgcctc tgtcctgcct ggtgtggagg gctctgtctg ttgctgagga ccaaggagga 1680
gaagagacct ttgctgcccc atcatggggg ctggggttgt cacctggaca gggggcagcc 1740
gtggaggcca ccgttaccaa ctgaagctgg gggcctgggt cctaccctgt ctggtcatga 1800
ccccattagg tatggagagc tgggaggagg cattgtcact tcccaccagg atgcaggact 1860
tggggttgag gtgagtcatg gcctcttgct ggcaatgggg tgggaggagt acccccaagt 1920
cctctcactc ctccagcctg gaatgtgaag tgactcccca acccctttgg ccatggcagg 1980
caccttttgg actgggctgc cactgcttgg gcagagtaaa aggtgccagg aggagcatgg 2040
gtgtggaagt cctgtcagcc aagaaataaa agtttacctc agagctgcaa aaaaaaaaaa 2100

aaaaaaaaaa 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