

(19) World Intellectual Property Organization  
International Bureau



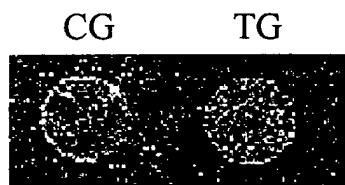
(43) International Publication Date  
18 October 2001 (18.10.2001)

PCT

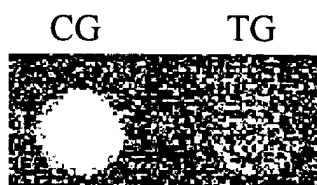
(10) International Publication Number  
**WO 01/77375 A2**

- (51) International Patent Classification<sup>7</sup>: **C12Q 1/68**, G01N 33/50, B01J 19/00
- (74) Agents: **SCHOHE, Stefan** et al.; Boehmert & Boehmert, Franz-Joseph-Strasse 38, 80801 München (DE).
- (21) International Application Number: PCT/EP01/03968
- (22) International Filing Date: 6 April 2001 (06.04.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
- |              |                               |    |
|--------------|-------------------------------|----|
| 100 19 058.8 | 6 April 2000 (06.04.2000)     | DE |
| 100 19 173.8 | 7 April 2000 (07.04.2000)     | DE |
| 100 32 529.7 | 30 June 2000 (30.06.2000)     | DE |
| 100 43 826.1 | 1 September 2000 (01.09.2000) | DE |
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant (*for all designated States except US*): **EPIGENOMICS AG** [DE/DE]; Kastanienalle 24, 10435 Berlin (DE).
- (72) Inventors; and
- (75) Inventors/Applicants (*for US only*): **OLEK, Alexander** [DE/DE]; Schröderstrasse 13/2, 10115 Berlin (DE). **PIEPENBROCK, Christian** [DE/DE]; Schwartzkopffstrasse 7 B, 10115 Berlin (DE). **BERLIN, Kurt** [DE/DE]; Marienkäferweg 4, 14532 Stahnsdorf (DE).
- Published:**
- *without international search report and to be republished upon receipt of that report*
  - *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

(54) Title: DIAGNOSIS OF DISEASES ASSOCIATED WITH GENE REGULATION



I



II

(57) Abstract: The present invention relates to the chemically modified genomic sequences of genes associated with gene regulation, to oligonucleotides and/or PNA-oligomers for detecting the cytosine methylation state of genes associated with gene regulation which are directed against the sequence, as well as to a method for ascertaining genetic and/or epigenetic parameters of genes associated with gene regulation.

## **Diagnosis of Diseases Associated with gene regulation**

### **Field of the Invention**

The levels of observation that have been well studied by the methodological developments of recent years in molecular biology, are the genes themselves, the translation of these genes into RNA, and the resulting proteins. The question of which gene is switched on at which point in the course of the development of an individual, and how the activation and inhibition of specific genes in specific cells and tissues are controlled is correlatable to the degree and character of the methylation of the genes or of the genome. In this respect, pathogenic conditions may manifest themselves in a changed methylation pattern of individual genes or of the genome.

The present invention relates to nucleic acids, oligonucleotides, PNA-oligomers and to a method for the diagnosis and/or therapy of diseases which have a connection with the genetic and/or epigenetic parameters of genes associated with gene regulation and, in particular, with the methylation status thereof.

### **Prior Art**

The completion of the sequencing of the human genome has highlighted the need to further understand the regulation and control of genes. The cell exerts multiple levels of control over genes, however for the majority of genes the most important regulatory mechanism is transcriptional control.

There are multiple mechanisms of gene regulation involving different families of proteins, genes and regulatory sequences. The transcription of eukaryotic genes by polymerases requires the prior assembly of transcription factors at the promoter region of the gene. These factors bind in a specific order, beginning with the binding of TFIID to the TATA box. This provides several stages at which gene regulation may take place, and many eukaryotic gene regulation mechanisms are thought to work by exerting positive or negative control over these processes.

- 2 -

Gene regulatory proteins recognise short stretches of double stranded DNA. Hundreds of such sequences have been identified throughout the genome, e.g. Sp1, which recognises the sequence GGGCGG (in double stranded form), GATA-1 which recognises the sequence TGATAG (in double stranded form) and Oct-1 which recognises ATGCAAAT (in double stranded form). Gene regulatory proteins contain structural motifs that are able to recognise the regulatory sequences using the unique pattern of hydrogen bond donors, hydrogen bond acceptors and hydrophobic patches in the major and minor groove of the DNA. Several families of structural motifs are known, including the helix-turn-helix motif, comprised of amino acids, the zinc finger motif, which utilises one or more molecules of zinc as a structural component, the leucine zipper and helix-loop-helix motif.

Other mechanisms of gene regulation include chromatin packaging. Chromatin that is highly compacted is incapable of transcription, and in a less condensed form, DNA is still packaged on nucleosomes. Nucleosomes positioned at the beginning of a transcription site block the assembly of transcription factors, and these must be removed. In addition gene transcription requires extensive winding and unwinding of the supercoiled DNA. Although the exact mechanisms of DNA packaging are not completely understood, they include the use of a variety of DNA helicases, gyrases and topoisomerases. Further discussion of gene regulation may be found in standard molecular biology textbooks e.g. Alberts *et. al.* 'Molecular Biology of the cell' Garland Publishing.

Disruptions to gene regulation pathways have been implicated in a wide variety of diseases including, for example, but not limited to:

- Severe combined immunodeficiency disease: Misaki *et. al.* 'Gene-Transferred Oligoclonal T Cells Predominantly Persist in Peripheral Blood from an Adenosine Deaminase-Deficient Patient during Gene Therapy.' *Mol Ther* 2001 Jan;3(1):24-27.
- Cardiac damage: Pieper *et. al.* 'Myocardial postischemic injury is reduced by polyADPri-  
pose polymerase-1 gene disruption.' *Mol Med* 2000 Apr;6(4):271-82.

- 3 -

- Inflammatory response: Lekstrom-Himes J, Xanthopoulos KG 'CCAAT/enhancer binding protein epsilon is critical for effective neutrophil-mediated response to inflammatory challenge.' *Blood* 1999 May 1;93(9):3096-105.
- Haemophilia B: Crossley M, Brownlee GG 'Disruption of a C/EBP binding site in the factor IX promoter is associated with haemophilia B.' *Nature* 1990 May 31;345(6274):444-6.
- Werner syndrome: Li B, Comai L 'Requirements for the nucleolytic processing of DNA ends by the Werner syndrome protein:Ku70/80 complex.' *J Biol Chem* 2001 Jan 4.
- Asthma: Nakamura *et. al.* 'Gene expression of the GATA-3 transcription factor is increased in atopic asthma.' *J Allergy Clin Immunol* 1999 Feb;103(2 Pt 1):215-22.
- HDR syndrome: Van Esch *et. al.* 'GATA3 haplo-insufficiency causes human HDR syndrome.' *Nature* 2000 Jul 27;406(6794):419-22.
- Congenital heart defects: Srivastava D 'HAND proteins: molecular mediators of cardiac development and congenital heart disease.' *Trends Cardiovasc Med* 1999 Jan-Feb;9(1-2):11-8.
- Saethre-Chotzen syndrome: El Ghouzzi *et. al.* 'Mutations within or upstream of the basic helix-loop-helix domain of the TWIST gene are specific to Saethre-Chotzen syndrome.' *Eur J Hum Genet* 1999 Jan;7(1):27-33.
- Renal disease: Morello *et. al.* 'Regulation of glomerular basement membrane collagen expression by LMX1B contributes to renal disease in nail patella syndrome.' *Nat Genet.* 2001 Feb;27(2):205-208.
- Preeclampsia: Rajakumar *et. al.* 'Selective Overexpression of the Hypoxia-Inducible Transcription Factor, HIF-2alpha, in Placentas from Women with Preeclampsia.' *Biol Reprod.* 2001 Feb;64(2):499-506.

- 4 -

- Ogata *et. al.* 'Inducible expression of basic transcription factor-binding protein 2 (BTEB2), a member of zinc finger family of transcription factors, in cardiac allograft vascular disease.' Transplantation. 2000 Dec 15;70(11):1653-6.
- Colorectal cancer: Rask *et. al.* 'Increased expression of the transcription factors CCAAT-enhancer binding protein-beta (C/EBBeta) and C/EBzeta (CHOP) correlate with invasiveness of human colorectal cancer.' Int J Cancer. 2000 May 1;86(3):337-43.
- Thyroid cancer: Kebebew *et. al.* 'The helix-loop-helix transcription factor, Id-1, is over-expressed in medullary thyroid cancer' Surgery. 2000 Dec;128(6):952-7.
- Esophagal cancer: Saeki *et. al.* 'Expression of ets-1 transcription factor is correlated with penetrating tumor progression in patients with squamous cell carcinoma of the esophagus' Cancer. 2000 Oct 15;89(8):1670-6.

The diversity of mechanisms involved in the regulation of genes provides an alternative level at which to target therapies and diagnosis for diseases. In particular this may be relevant to diseases where current therapies may have unwanted side effects or fail to provide effective treatment. For cancer patients such methods constitute a considerable advantage over conventional methods such as chemotherapy with their massive side effects, which sometimes result in unacceptable morbidity or lead up to the death of the patient. In practice, these unwanted side effects associated with cancer therapies frequently limit the treatment which could help a patient.

A global analysis of the status of gene regulatory mechanisms would provide a basis for the development of appropriate and specific therapies for the many diseases associated with gene regulation. Such analysis may be carried out in a gene specific manner based on the results of gene expression, e.g. DNA micro array analysis of mRNA expression or proteomic analysis. The next step would then be to look at the causal factors involved at earlier stages in the regulatory mechanisms. DNA methylation provides a novel level of information at which to analyse the genome.

- 5 -

5-methylcytosine is the most frequent covalent base modification in the DNA of eukaryotic cells. It plays a role, for example, in the regulation of the transcription, in genetic imprinting, and in tumorigenesis. Therefore, the identification of 5-methylcytosine as a component of genetic information is of considerable interest. However, 5-methylcytosine positions cannot be identified by sequencing since 5-methylcytosine has the same base pairing behavior as cytosine. Moreover, the epigenetic information carried by 5-methylcytosine is completely lost during PCR amplification.

A relatively new and currently the most frequently used method for analyzing DNA for 5-methylcytosine is based upon the specific reaction of bisulfite with cytosine which, upon subsequent alkaline hydrolysis, is converted to uracil which corresponds to thymidine in its base pairing behavior. However, 5-methylcytosine remains unmodified under these conditions. Consequently, the original DNA is converted in such a manner that methylcytosine, which originally could not be distinguished from cytosine by its hybridization behavior, can now be detected as the only remaining cytosine using "normal" molecular biological techniques, for example, by amplification and hybridization or sequencing. All of these techniques are based on base pairing which can now be fully exploited. In terms of sensitivity, the prior art is defined by a method which encloses the DNA to be analyzed in an agarose matrix, thus preventing the diffusion and renaturation of the DNA (bisulfite only reacts with single-stranded DNA), and which replaces all precipitation and purification steps with fast dialysis (Olek A, Oswald J, Walter J. A modified and improved method for bisulphite based cytosine methylation analysis. *Nucleic Acids Res.* 1996 Dec 15;24(24):5064-6). Using this method, it is possible to analyze individual cells, which illustrates the potential of the method. However, currently only individual regions of a length of up to approximately 3000 base pairs are analyzed, a global analysis of cells for thousands of possible methylation events is not possible. However, this method cannot reliably analyze very small fragments from small sample quantities either. These are lost through the matrix in spite of the diffusion protection.

An overview of the further known methods of detecting 5-methylcytosine may be gathered from the following review article: Rein, T., DePamphilis, M. L., Zorbas, H., *Nucleic Acids Res.* 1998, 26, 2255.

- 6 -

To date, barring few exceptions (e.g., Zeschnigk M, Lich C, Buiting K, Doerfler W, Horsthemke B. A single-tube PCR test for the diagnosis of Angelman and Prader-Willi syndrome based on allelic methylation differences at the SNRPN locus. *Eur J Hum Genet.* 1997 Mar-Apr;5(2):94-8) the bisulfite technique is only used in research. Always, however, short, specific fragments of a known gene are amplified subsequent to a bisulfite treatment and either completely sequenced (Olek A, Walter J. The pre-implantation ontogeny of the H19 methylation imprint. *Nat Genet.* 1997 Nov;17(3):275-6) or individual cytosine positions are detected by a primer extension reaction (Gonzalzo ML, Jones PA. Rapid quantitation of methylation differences at specific sites using methylation-sensitive single nucleotide primer extension (Ms-SNuPE). *Nucleic Acids Res.* 1997 Jun 15;25(12):2529-31, Patent application WO 95/00669) or by enzymatic digestion (Xiong Z, Laird PW. COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* 1997 Jun 15;25(12):2532-4). In addition, detection by hybridization has also been described (Olek et al., WO 99/28498).

Further publications dealing with the use of the bisulfite technique for methylation detection in individual genes are: Grigg G, Clark S. Sequencing 5-methylcytosine residues in genomic DNA. *Bioessays.* 1994 Jun;16(6):431-6, 431; Zeschnigk M, Schmitz B, Dittrich B, Buiting K, Horsthemke B, Doerfler W. Imprinted segments in the human genome: different DNA methylation patterns in the Prader-Willi/Angelman syndrome region as determined by the genomic sequencing method. *Hum Mol Genet.* 1997 Mar;6(3):387-95; Feil R, Charlton J, Bird AP, Walter J, Reik W. Methylation analysis on individual chromosomes: improved protocol for bisulphite genomic sequencing. *Nucleic Acids Res.* 1994 Feb 25;22(4):695-6; Martin V, Ribieras S, Song-Wang X, Rio MC, Dante R. Genomic sequencing indicates a correlation between DNA hypomethylation in the 5' region of the pS2 gene and its expression in human breast cancer cell lines. *Gene.* 1995 May 19;157(1-2):261-4; WO 97/46705, WO 95/15373 and WO 97/45560.

An overview of the Prior Art in oligomer array manufacturing can be gathered from a special edition of *Nature Genetics* (*Nature Genetics Supplement*, Volume 21, January 1999), published in January 1999, and from the literature cited therein.

Fluorescently labeled probes are often used for the scanning of immobilized DNA arrays. The simple attachment of Cy3 and Cy5 dyes to the 5'-OH of the specific probe are particularly

- 7 -

suitable for fluorescence labels. The detection of the fluorescence of the hybridized probes may be carried out, for example via a confocal microscope. Cy3 and Cy5 dyes, besides many others, are commercially available.

Matrix Assisted Laser Desorption Ionization Mass Spectrometry (MALDI-TOF) is a very efficient development for the analysis of biomolecules (Karas M, Hillenkamp F. Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem.* 1988 Oct 15;60(20):2299-301). An analyte is embedded in a light-absorbing matrix. The matrix is evaporated by a short laser pulse thus transporting the analyte molecule into the vapor phase in an unfragmented manner. The analyte is ionized by collisions with matrix molecules. An applied voltage accelerates the ions into a field-free flight tube. Due to their different masses, the ions are accelerated at different rates. Smaller ions reach the detector sooner than bigger ones.

MALDI-TOF spectrometry is excellently suited to the analysis of peptides and proteins. The analysis of nucleic acids is somewhat more difficult (Gut I G, Beck S. DNA and Matrix Assisted Laser Desorption Ionization Mass Spectrometry. *Current Innovations and Future Trends.* 1995, 1; 147-57). The sensitivity to nucleic acids is approximately 100 times worse than to peptides and decreases disproportionally with increasing fragment size. For nucleic acids having a multiply negatively charged backbone, the ionization process via the matrix is considerably less efficient. In MALDI-TOF spectrometry, the selection of the matrix plays an eminently important role. For the desorption of peptides, several very efficient matrixes have been found which produce a very fine crystallization. There are now several responsive matrixes for DNA, however, the difference in sensitivity has not been reduced. The difference in sensitivity can be reduced by chemically modifying the DNA in such a manner that it becomes more similar to a peptide. Phosphorothioate nucleic acids in which the usual phosphates of the backbone are substituted with thiophosphates can be converted into a charge-neutral DNA using simple alkylation chemistry (Gut IG, Beck S. A procedure for selective DNA alkylation and detection by mass spectrometry. *Nucleic Acids Res.* 1995 Apr 25;23(8):1367-73). The coupling of a charge tag to this modified DNA results in an increase in sensitivity to the same level as that found for peptides. A further advantage of charge tagging is the increased stability of the analysis against impurities which make the detection of unmodified substrates considerably more difficult.



Genomic DNA is obtained from DNA of cell, tissue or other test samples using standard methods. This standard methodology is found in references such as Fritsch and Maniatis eds., Molecular Cloning: A Laboratory Manual, 1989.

### **Description**

The object of the present invention is to provide the chemically modified DNA of genes associated with gene regulation, as well as oligonucleotides and/or PNA-oligomers for detecting cytosine methylations, as well as a method which is particularly suitable for the diagnosis and/or therapy of genetic and epigenetic parameters of genes associated with gene regulation. The present invention is based on the discovery that genetic and epigenetic parameters and, in particular, the cytosine methylation pattern of genes associated with gene regulation are particularly suitable for the diagnosis and/or therapy of diseases associated with gene regulation.

This objective is achieved according to the present invention using a nucleic acid containing a sequence of at least 18 bases in length of the chemically pretreated DNA of genes associated with gene regulation according to one of Seq. ID No.1 through Seq. ID No.224 and sequences complementary thereto and/or a sequence of a chemically pretreated DNA of genes according to table 1 and sequences complementary thereto. In the table, after the listed gene designations, the respective data bank numbers (accession numbers) are specified which define the appertaining gene sequences as unique. GenBank was used as the underlying data bank, which is located at the National Institute of Health, internet address [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov).

The chemically modified nucleic acid could heretofore not be connected with the ascertainment of genetic and epigenetic parameters.

The object of the present invention is further achieved by an oligonucleotide or oligomer for detecting the cytosine methylation state in chemically pretreated DNA, containing at least one base sequence having a length of at least 13 nucleotides which hybridizes to a chemically pretreated DNA of genes associated with gene regulation according to Seq. ID No.1 through Seq. ID No.224 and sequences complementary thereto and/or a sequence of a chemically pretreated DNA of genes according to table 1 and sequences complementary thereto. The oli-

- 9 -

gomer probes according to the present invention constitute important and effective tools which, for the first time, make it possible to ascertain the genetic and epigenetic parameters of genes associated with gene regulation. The base sequence of the oligomers preferably contains at least one CpG dinucleotide. The probes may also exist in the form of a PNA (peptide nucleic acid) which has particularly preferred pairing properties. Particularly preferred are oligonucleotides according to the present invention in which the cytosine of the CpG dinucleotide is the 5<sup>th</sup> - 9<sup>th</sup> nucleotide from the 5'-end of the 13-mer; in the case of PNA-oligomers, it is preferred for the cytosine of the CpG dinucleotide to be the 4<sup>th</sup> - 6<sup>th</sup> nucleotide from the 5'-end of the 9-mer.

The oligomers according to the present invention are normally used in so called "sets" which contain at least one oligomer for each of the CpG dinucleotides of the sequences of Seq. ID No.1 through Seq. ID No.224 and sequences complementary thereto and/or a sequence of a chemically pretreated DNA of genes according to table 1 and sequences complementary thereto. Preferred is a set which contains at least one oligomer for each of the CpG dinucleotides from one of Seq. ID No.1 through Seq. ID No.224 and sequences complementary thereto and/or a sequence of a chemically pretreated DNA of genes according to table 1 and sequences complementary thereto.

Moreover, the present invention makes available a set of at least two oligonucleotides which can be used as so-called "primer oligonucleotides" for amplifying DNA sequences of one of Seq. ID No.1 through Seq. ID No.224 and sequences complementary thereto and/or a sequence of a chemically pretreated DNA of genes according to table 1 and sequences complementary thereto, or segments thereof.

In the case of the sets of oligonucleotides according to the present invention, it is preferred that at least one oligonucleotide is bound to a solid phase.

The present invention moreover relates to a set of at least 10 n (oligonucleotides and/or PNA-oligomers) used for detecting the cytosine methylation state in chemically pretreated genomic DNA (Seq. ID No.1 through Seq. ID No.224 and sequences complementary thereto and/or a sequence of a chemically pretreated DNA of genes according to table 1 and sequences com-

- 10 -

plementary thereto). These probes enable diagnosis and/or therapy of genetic and epigenetic parameters of genes associated with gene regulation. The set of oligomers may also be used for detecting single nucleotide polymorphisms (SNPs) in the chemically pretreated DNA of genes associated with gene regulation according to one of Seq. ID No.1 through Seq. ID No.224 and sequences complementary thereto and/or a sequence of a chemically pretreated DNA of genes according to table 1 and sequences complementary thereto.

According to the present invention, it is preferred that an arrangement of different oligonucleotides and/or PNA-oligomers (a so-called "array") made available by the present invention is present in a manner that it is likewise bound to a solid phase. This array of different oligonucleotide- and/or PNA-oligomer sequences can be characterized in that it is arranged on the solid phase in the form of a rectangular or hexagonal lattice. The solid phase surface is preferably composed of silicon, glass, polystyrene, aluminum, steel, iron, copper, nickel, silver, or gold. However, nitrocellulose as well as plastics such as nylon which can exist in the form of pellets or also as resin matrices are possible as well.

Therefore, a further subject matter of the present invention is a method for manufacturing an array fixed to a carrier material for analysis in connection with diseases associated with gene regulation in which method at least one oligomer according to the present invention is coupled to a solid phase. Methods for manufacturing such arrays are known, for example, from US Patent 5,744,305 by means of solid-phase chemistry and photolabile protecting groups.

A further subject matter of the present invention relates to a DNA chip for the analysis of diseases associated with gene regulation which contains at least one nucleic acid according to the present invention. DNA chips are known, for example, from US Patent 5,837,832.

Moreover, a subject matter of the present invention is a kit which may be composed, for example, of a bisulfite-containing reagent, a set of primer oligonucleotides containing at least two oligonucleotides whose sequences in each case correspond or are complementary to an 18 base long segment of the base sequences specified in the appendix (Seq. ID No.1 through Seq. ID No.224 and sequences complementary thereto and/or a sequence of a chemically pretreated DNA of genes according to table 1 and sequences complementary thereto), oligonucleotides and/or PNA-oligomers as well as instructions for carrying out and evaluating the de-

- 11 -

scribed method. However, a kit along the lines of the present invention can also contain only part of the aforementioned components.

The present invention also makes available a method for ascertaining genetic and/or epigenetic parameters of genes associated with the cycle cell by analyzing cytosine methylations and single nucleotide polymorphisms, including the following steps:

In the first step of the method, a genomic DNA sample is chemically treated in such a manner that cytosine bases which are unmethylated at the 5'-position are converted to uracil, thymine, or another base which is dissimilar to cytosine in terms of hybridization behavior. This will be understood as 'chemical pretreatment' hereinafter.

The genomic DNA to be analyzed is preferably obtained from usual sources of DNA such as cells or cell components, for example, cell lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid, tissue embedded in paraffin such as tissue from eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver, histologic object slides, or combinations thereof.

The above described treatment of genomic DNA is preferably carried out with bisulfite (hydrogen sulfite, disulfite) and subsequent alkaline hydrolysis which results in a conversion of non-methylated cytosine nucleobases to uracil or to another base which is dissimilar to cytosine in terms of base pairing behavior.

Fragments of the chemically pretreated DNA are amplified, using sets of primer oligonucleotides according to the present invention, and a, preferably heat-stable polymerase. Because of statistical and practical considerations, preferably more than ten different fragments having a length of 100 - 2000 base pairs are amplified. The amplification of several DNA segments can be carried out simultaneously in one and the same reaction vessel. Usually, the amplification is carried out by means of a polymerase chain reaction (PCR).

In a preferred embodiment of the method, the set of primer oligonucleotides includes at least two oligonucleotides whose sequences are each reverse complementary or identical to an at least 18 base-pair long segment of the base sequences specified in the appendix (Seq. ID No.1 through Seq. ID No.224 and sequences complementary thereto and/or a sequence of a chemi-

- 12 -

cally pretreated DNA of genes according to table 1 and sequences complementary thereto). The primer oligonucleotides are preferably characterized in that they do not contain any CpG dinucleotides.

According to the present invention, it is preferred that at least one primer oligonucleotide is bonded to a solid phase during amplification. The different oligonucleotide and/or PNA-oligomer sequences can be arranged on a plane solid phase in the form of a rectangular or hexagonal lattice, the solid phase surface preferably being composed of silicon, glass, polystyrene, aluminum, steel, iron, copper, nickel, silver, or gold, it being possible for other materials such as nitrocellulose or plastics to be used as well.

The fragments obtained by means of the amplification can carry a directly or indirectly detectable label. Preferred are labels in the form of fluorescence labels, radionuclides, or detachable molecule fragments having a typical mass which can be detected in a mass spectrometer, it being preferred that the fragments that are produced have a single positive or negative net charge for better detectability in the mass spectrometer. The detection may be carried out and visualized by means of matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

The amplicates obtained in the second step of the method are subsequently hybridized to an array or a set of oligonucleotides and/or PNA probes. In this context, the hybridization takes place in the manner described in the following. The set of probes used during the hybridization is preferably composed of at least 10 oligonucleotides or PNA-oligomers. In the process, the amplicates serve as probes which hybridize to oligonucleotides previously bonded to a solid phase. The non-hybridized fragments are subsequently removed. Said oligonucleotides contain at least one base sequence having a length of 13 nucleotides which is reverse complementary or identical to a segment of the base sequences specified in the appendix, the segment containing at least one CpG dinucleotide. The cytosine of the CpG dinucleotide is the 5<sup>th</sup> to 9<sup>th</sup> nucleotide from the 5'-end of the 13-mer. One oligonucleotide exists for each CpG dinucleotide. Said PNA-oligomers contain at least one base sequence having a length of 9 nucleotides which is reverse complementary or identical to a segment of the base sequences specified in the appendix, the segment containing at least one CpG dinucleotide. The cytosine

- 13 -

of the CpG dinucleotide is the 4<sup>th</sup> to 6<sup>th</sup> nucleotide seen from the 5'-end of the 9-mer. One oligonucleotide exists for each CpG dinucleotide.

In the fourth step of the method, the non-hybridized amplicates are removed.

In the final step of the method, the hybridized amplicates are detected. In this context, it is preferred that labels attached to the amplicates are identifiable at each position of the solid phase at which an oligonucleotide sequence is located.

According to the present invention, it is preferred that the labels of the amplicates are fluorescence labels, radionuclides, or detachable molecule fragments having a typical mass which can be detected in a mass spectrometer. The mass spectrometer is preferred for the detection of the amplicates, fragments of the amplicates or of probes which are complementary to the amplicates, it being possible for the detection to be carried out and visualized by means of matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

The produced fragments may have a single positive or negative net charge for better detectability in the mass spectrometer. The aforementioned method is preferably used for ascertaining genetic and/or epigenetic parameters of genes associated with gene regulation.

The oligomers according to the present invention or arrays thereof as well as a kit according to the present invention are intended to be used for the diagnosis and/or therapy of diseases associated with gene regulation by analyzing methylation patterns of genes associated with gene regulation. According to the present invention, the method is preferably used for the diagnosis and/or therapy of important genetic and/or epigenetic parameters within genes associated with gene regulation.

The method according to the present invention is used, for example, for the diagnosis and/or therapy of severe combined immunodeficiency disease, cardiac disorders, inflammatory response, hemophilia, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease, preeclampsia, graft-versus-host disease, solid tumors and cancers.

- 14 -

The nucleic acids according to the present invention of Seq. ID No.1 through Seq. ID No.224 and sequences complementary thereto and/or a sequence of a chemically pretreated DNA of genes according to table 1 and sequences complementary thereto can be used for the diagnosis and/or therapy of genetic and/or epigenetic parameters of genes associated with gene regulation.

The present invention moreover relates to a method for manufacturing a diagnostic agent and/or therapeutic agent for the diagnosis and/or therapy of diseases associated with gene regulation by analyzing methylation patterns of genes associated with gene regulation, the diagnostic agent and/or therapeutic agent being characterized in that at least one nucleic acid according to the present invention is used for manufacturing it, possibly together with suitable additives and auxiliary agents.

A further subject matter of the present invention relates to a diagnostic agent and/or therapeutic agent for diseases associated with gene regulation by analyzing methylation patterns of genes associated with gene regulation, the diagnostic agent and/or therapeutic agent containing at least one nucleic acid according to the present invention, possibly together with suitable additives and auxiliary agents.

The present invention moreover relates to the diagnosis and/or prognosis of events which are disadvantageous to patients or individuals in which important genetic and/or epigenetic parameters within genes associated with gene regulation said parameters obtained by means of the present invention may be compared to another set of genetic and/or epigenetic parameters, the differences serving as the basis for a diagnosis and/or prognosis of events which are disadvantageous to patients or individuals.

In the context of the present invention the term "hybridization" is to be understood as a bond of an oligonucleotide to a completely complementary sequence along the lines of the Watson-Crick base pairings in the sample DNA, forming a duplex structure. To be understood by "stringent hybridization conditions" are those conditions in which a hybridization is carried out at 60°C in 2.5 x SSC buffer, followed by several washing steps at 37°C in a low buffer concentration, and remains stable.

- 15 -

The term “functional variants“ denotes all DNA sequences which are complementary to a DNA sequence, and which hybridize to the reference sequence under stringent conditions and have an activity similar to the corresponding polypeptide according to the present invention.

In the context of the present invention, “genetic parameters“ are mutations and polymorphisms of genes associated with gene regulation and sequences further required for their regulation. To be designated as mutations are, in particular, insertions, deletions, point mutations, inversions and polymorphisms and, particularly preferred, SNPs (single nucleotide polymorphisms).

In the context of the present invention, “epigenetic parameters“ are, in particular, cytosine methylations and further chemical modifications of DNA bases of genes associated with gene regulation and sequences further required for their regulation. Further epigenetic parameters include, for example, the acetylation of histones which, however, cannot be directly analyzed using the described method but which, in turn, correlates with the DNA methylation.

In the following, the present invention will be explained in greater detail on the basis of the sequences and examples with reference to the accompanying figure without being limited thereto.

#### Figure 1

Figure 1 shows the hybridisation of fluorescent labelled amplicates to a surface bound oligonucleotide. Sample I being from pilocytic astrocytoma (brain tumor) tissue and sample II being from grade II astrocytoma (brain tumor) tissue. Fluorescence at a spot shows hybridisation of the amplicate to the oligonucleotide. Hybridisation to a CG oligonucleotide denotes methylation at the cytosine position being analysed, hybridisation to a TG oligonucleotide denotes no methylation at the cytosine position being analysed.

#### Sequence ID Nos. 1 to 224

Sequence ID Nos. 1 to 224 show sequences of genes associated with gene regulation according to the invention. Sequences having odd sequence numbers (e.g., Seq. ID No. 1, 3, 5, ...) exhibit in each case sequences of the chemically pretreated genomic DNAs of different genes associated with gene regulation. Sequences having even sequence numbers (e.g., Seq. ID No.



- 16 -

2, 4, 6, ...) exhibit in each case the sequences of the chemically pretreated genomic DNAs of genes associated with gene regulation which are complementary to the preceding sequences (e.g., the complementary sequence to Seq. ID No.1 is Seq. ID No.2, the complementary sequence to Seq. ID No.3 is Seq. ID No.4, etc.)

#### Sequence ID Nos. 225 to 228

Sequence ID Nos. 225 to 228 show the sequences of oligonucleotides used in Example 1.

#### **Example 1: Methylation analysis of the gene p53 associated with gene regulation.**

This example relates to a fragment of a gene associated with gene regulation, in this case, p53 in which a specific CG-position is analyzed for its methylation status.

In the first step, a genomic sequence is treated using bisulfite (hydrogen sulfite, disulfite) in such a manner that all cytosines which are not methylated at the 5-position of the base are modified in such a manner that a different base is substituted with regard to the base pairing behavior while the cytosines methylated at the 5-position remain unchanged.

If bisulfite solution is used for the reaction, then an addition takes place at the non-methylated cytosine bases. Moreover, a denaturing reagent or solvent as well as a radical interceptor must be present. A subsequent alkaline hydrolysis then gives rise to the conversion of non-methylated cytosine nucleobases to uracil. The chemically converted DNA (Seq ID No. 221) is then used for the detection of methylated cytosines. In the second step of the method, the treated DNA sample is diluted with water or an aqueous solution. Preferably, the DNA is subsequently desulfonated (10-30 min, 90-100 °C) at an alkaline pH value. In the third step of the method, the DNA sample is amplified in a polymerase chain reaction, preferably using a heat-resistant DNA polymerase. In the present case, cytosines of the gene p53 are analyzed. To this end, a defined fragment having a length of 595 bp is amplified with the specific primer oligonucleotides GTGATAAGGGTTGTGAAGGA (sequence ID No.225) and CAAAACTTACCCAATCCAA (sequence ID No.226). This amplificate serves as a sample which hybridizes to an oligonucleotide previously bonded to a solid phase, forming a duplex structure, for example AACCCCTACGAACTCCT (sequence ID No.227), the cytosine to be detected being located at position 401 of the amplificate. The detection of the hybridization product is based on Cy3 and Cy5 fluorescently labeled primer oligonucleotides which have

- 17 -

been used for the amplification. A hybridization reaction of the amplified DNA with the oligonucleotide takes place only if a methylated cytosine was present at this location in the bisulfite-treated DNA. Thus, the methylation status of the specific cytosine to be analyzed is inferred from the hybridization product.

In order to verify the methylation status of the position, a sample of the amplificate is further hybridized to another oligonucleotide previously bonded to a solid phase. Said oligonucleotide is identical to the oligonucleotide previously used to analyze the methylation status of the sample, with the exception of the position in question. At the position to be analysed said oligonucleotide comprises a thymine base as opposed to a cytosine base i.e. sequence AACCCTACAAAACCTCCT (sequence ID No.228). Therefore, the hybridization reaction only takes place if an unmethylated cytosine was present at the position to be analysed.

The analysis was carried out on two tissue samples, Sample 1 from pilocytic astrocytoma (brain tumor) tissue, and Sample 2 from grade II astrocytoma (brain tumor) tissue. From the results (Figure 1) it can be seen that Sample 2 was methylated at position 401 of the amplificate whereas Sample 1 contained a mixture of both methylated and unmethylated cells at the same position.

### **Example 2: Diagnosis of diseases associated with gene regulation**

In order to relate the methylation patterns to one of the diseases associated with gene regulation, it is initially required to analyze the DNA methylation patterns of a group of diseased and of a group of healthy patients. These analyses are carried out, for example, analogously to example 1. The results obtained in this manner are stored in a database and the CpG dinucleotides which are methylated differently between the two groups are identified. This can be carried out by determining individual CpG methylation rates as can be done, for example, in a relatively imprecise manner, by sequencing or else, in a very precise manner, by a methylation-sensitive "primer extension reaction". It is also possible for the entire methylation status to be analyzed simultaneously, and for the patterns to be compared, for example, by clustering analyses which can be carried out, for example, by a computer.

Subsequently, it is possible to allocate the examined patients to a specific therapy group and to treat these patients selectively with an individualized therapy.

Example 2 can be carried out, for example, for the following diseases:

severe combined immunodeficiency disease, cardiac disorders, inflammatory response, hemophilia, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease, preeclampsia, graft-versus-host disease, solid tumors and cancers.

Table 1.

Listing of preferred genes associated with gene regulation according to the invention.

Gene	Database entry No. (GenBank, internet address <a href="http://www.ncbi.nlm.nih.gov">www.ncbi.nlm.nih.gov</a> )
CBFB	(L20298),
ELF1	(M82882),
ETV3	(L16464),
ETV4	(D12765),
LYL1	(M22637),
RENB	(U52112),
TAF2C2	(Y09321),
TCF3	(M31222),
H4FI	(NM_003544),
ADA	(NM_000022),
ATF3	(NM_001674),
CEBPD	(NM_005195),
CHD1L	(NM_004284),
CTPS	(NM_001905),
DCTD	(NM_001921),
EIF2B1	(NM_001414),
ELF3	(NM_004433),
ELK4	(NM_021795),
ETV5	(NM_004454),
FOXO1A	(NM_002015),
HIVEP1	(NM_002114),

- 19 -

HMG2	(NM_002129),
ID1	(NM_002165),
ID3	(NM_002167),
ID4	(NM_001546),
LAF4	(NM_0022859),
MAFG	(NM_002359),
MHC2TA	(NM_000246),
ODC1	(NM_002539),
PBX3	(NM_006195),
PCNA	(NM_002592),
POU2AF1	(NM_006235),
PRPS1	(NM_002764),
RARG	(NM_000966),
RECQL	(NM_002907),
RXRA	(NM_002957),
TIAL1	(NM_003252),
ZNF173	(NM_003449),
ATBF1	(NM_006885),
FLI1	(NM_002017),
NUMA1	(NM_006185),
POU2F2	(NM_002698),
SP100	(NM_003113)

- 20 -

### Claims

1. A nucleic acid comprising a sequence at least 18 bases in length of a segment of the chemically pretreated DNA of genes associated with gene regulation according to one of the sequences taken from the group of Seq. ID No.1 to Seq. ID No.224 and sequences complementary thereto.
2. A nucleic acid comprising a sequence at least 18 base pairs in length of a segment of the chemically pretreated DNA of genes associated with gene regulation according to one of the sequences according to one of the genes CBFB (L20298), ELF1 (M82882), ETV3 (L16464), ETV4 (D12765), LYL1 (M22637), RENBP (U52112), TAF2C2 (Y09321), TCF3 (M31222), H4FI (NM\_003544), ADA (NM\_000022), ATF3 (NM\_001674), CEBPD (NM\_005195), CHD1L (NM\_004284), CTPS (NM\_001905), DCTD (NM\_001921), EIF2B1 (NM\_001414), ELF3 (NM\_004433), ELK4 (NM\_021795), ETV5 (NM\_004454), FOXO1A (NM\_002015), HIVEP1 (NM\_002114), HMG2 (NM\_002129), ID1 (NM\_002165), ID3 (NM\_002167), ID4 (NM\_001546), LAF4 (NM\_0022859), MAFG (NM\_002359), MHC2TA (NM\_000246), ODC1 (NM\_002539), PBX3 (NM\_006195), PCNA (NM\_002592), POU2AF1 (NM\_006235), PRPS1 (NM\_002764), RARG (NM\_000966), RECQL (NM\_002907), RXRA (NM\_002957), TIAL1 (NM\_003252), ZNF173 (NM\_003449), ATBF1 (NM\_006885), FLI1 (NM\_002017), NUMA1 (NM\_006185), POU2F2 (NM\_002698), SP100 (NM\_003113), and sequences complementary thereto.
3. An oligomer, in particular an oligonucleotide or peptide nucleic acid (PNA)-oligomer, said oligomer comprising in each case at least one base sequence having a length of at least 9 nucleotides which hybridizes to or is identical to a chemically pretreated DNA of genes associated with gene regulation according to one of the Seq ID Nos 1 to 224 according to claim 1 or to a chemically pretreated DNA of genes according to claim 2 and sequences complementary thereto.
4. The oligomer as recited in Claim 3; wherein the base sequence includes at least one CpG dinucleotide

- 21 -

5. The oligomer as recited in Claim 3; characterized in that the cytosine of the CpG dinucleotide is located approximately in the middle third of the oligomer.
6. A set of oligomers, comprising at least two oligomers according to any of claims 3 to 5.
7. A set of oligomers as recited in Claim 6, comprising oligomers for detecting the methylation state of all CpG dinucleotides within one of the sequences according to Seq. ID Nos. 1 through Seq. ID 224 according to claim 1 or a chemically pretreated DNA of genes according to claim 2, and sequences complementary thereto.
8. A set of at least two oligonucleotides as recited in Claim 3, which can be used as primer oligonucleotides for the amplification of DNA sequences of one of Seq. ID No.1 through Seq. ID No.224 and sequences complementary thereto and/or sequences of a chemically pretreated DNA of genes according to claim 2, and sequences complementary thereto and segments thereof.
9. A set of oligonucleotides as recited in Claim 8, characterized in that at least one oligonucleotide is bound to a solid phase.
10. Use of a set of oligomer probes comprising at least ten of the oligomers according to any of claims 6 through 9 for detecting the cytosine methylation state and/or single nucleotide polymorphisms (SNPs) in a chemically pretreated genomic DNA according to claim 1 or a chemically pretreated DNA of genes according to claim 2.
11. A method for manufacturing an arrangement of different oligomers (array) fixed to a carrier material for analyzing diseases associated with the methylation state of the CpG dinucleotides of one of the Seq. ID No.1 through Seq. ID No.224 and sequences complementary thereto and/or chemically pretreated DNA of genes according to claim 2, wherein at least one oligomer according to any of the claims 3 through 5 is coupled to a solid phase.
12. An arrangement of different oligomers (array) obtainable according to claim 11.

- 22 -

13. An array of different oligonucleotide- and/or PNA-oligomer sequences as recited in Claim 12, characterized in that these are arranged on a plane solid phase in the form of a rectangular or hexagonal lattice.

14. The array as recited in any of the Claims 12 or 13, characterized in that the solid phase surface is composed of silicon, glass, polystyrene, aluminum, steel, iron, copper, nickel, silver, or gold.

15. A DNA- and/or PNA-array for analyzing diseases associated with the methylation state of genes, comprising at least one nucleic acid according to one of the preceding claims.

16. A method for ascertaining genetic and/or epigenetic parameters for the diagnosis and/or therapy of existing diseases or the predisposition to specific diseases by analyzing cytosine methylations, characterized in that the following steps are carried out:

a) in a genomic DNA sample, cytosine bases which are unmethylated at the 5-position are converted, by chemical treatment, to uracil or another base which is dissimilar to cytosine in terms of hybridization behavior;

b) fragments of the chemically pretreated genomic DNA are amplified using sets of primer oligonucleotides according to Claim 8 or 9 and a polymerase, the amplicates carrying a detectable label;

c) Amplicates are hybridized to a set of oligonucleotides and/or PNA probes according to the Claims 6 and 7, or else to an array according to one of the Claims 12 through 15;

d) the hybridized amplicates are subsequently detected.

17. The method as recited in Claim 16, characterized in that the chemical treatment is carried out by means of a solution of a bisulfite, hydrogen sulfite or disulfite.

18. The method as recited in one of the Claims 16 or 17, characterized in that more than ten different fragments having a length of 100 - 2000 base pairs are amplified.

- 23 -

19. The method as recited in one of the Claims 16 through 18, characterized in that the amplification of several DNA segments is carried out in one reaction vessel.

20. The method as recited in one of the Claims 16 through 19, characterized in that the polymerase is a heat-resistant DNA polymerase.

21. The method as recited in Claim 20, characterized in that the amplification is carried out by means of the polymerase chain reaction (PCR).

22. The method as recited in one of the Claims 16 through 21, characterized in that the labels of the amplicates are fluorescence labels.

23. The method as recited in one of the Claims 16 through 21, characterized in that the labels of the amplicates are radionuclides.

24. The method as recited in one of the Claims 16 through 21, characterized in that the labels of the amplicates are detachable molecule fragments having a typical mass which are detected in a mass spectrometer.

25. The method as recited in one of the Claims 16 through 21, characterized in that the amplicates or fragments of the amplicates are detected in the mass spectrometer.

26. The method as recited in one of the Claims 24 and/or 25, characterized in that the produced fragments have a single positive or negative net charge for better detectability in the mass spectrometer

27. The method as recited in one of the Claims 24 through 26, characterized in that detection is carried out and visualized by means of matrix assisted laser desorption/ionization mass spectrometry (MALDI) or using electron spray mass spectrometry (ESI).

28. The method as recited in one of the Claims 16 through 27, characterized in that the genomic DNA is obtained from cells or cellular components which contain DNA, sources of DNA



- 24 -

comprising, for example, cell lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid, tissue embedded in paraffin such as tissue from eyes, intestine, kidney, brain, heart, prostate, lung, breast or liver, histologic object slides, and all possible combinations thereof.

29. A kit comprising a bisulfite (= disulfite, hydrogen sulfite) reagent as well as oligonucleotides and/or PNA-oligomers according to one of the Claims 3 through 5.

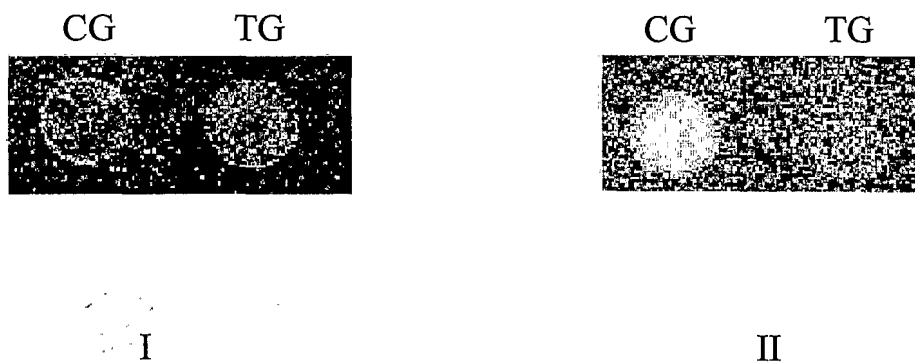
30. The use of a nucleic acid according to Claims 1 or 2, of an oligonucleotide or PNA-oligomer according to one of the Claims 3 through 5, of a kit according to Claim 29, of an array according to one of the Claims 12 through 15, of a set of oligonucleotides according to one of claims 6 through 9 for the diagnosis of severe combined immunodeficiency disease, cardiac disorders, inflammatory response, hemophilia, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease, preeclampsia, graft-versus-host disease, solid tumors and cancers.

31. The use of a nucleic acid according to Claims 1 or 2, of an oligonucleotide or PNA-oligomer according to one of Claims 3 through 5, of a kit according to Claim 29, of an array according to one of the Claims 12 through 15, of a set of oligonucleotides according to one of claims 6 through 9 for the therapy of severe combined immunodeficiency disease, cardiac disorders, inflammatory response, hemophilia, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen syndrome, renal disease, preeclampsia, graft-versus-host disease, solid tumors and cancers.

32. A kit, comprising a bisulfite (= disulfite, hydrogen sulfite) reagent as well as oligonucleotides and/or PNA-oligomers according to one of claims 3 through 5.

1/1

Figure 1



专利名称(译)	诊断与基因调控相关的疾病		
公开(公告)号	<a href="#">EP1268857A2</a>	公开(公告)日	2003-01-02
申请号	EP2001927887	申请日	2001-04-06
[标]申请(专利权)人(译)	埃皮吉諾米克斯股份公司		
申请(专利权)人(译)	AG EPIGENOMICS		
当前申请(专利权)人(译)	AG EPIGENOMICS		
[标]发明人	OLEK ALEXANDER PIEPENBROCK CHRISTIAN BERLIN KURT		
发明人	OLEK, ALEXANDER PIEPENBROCK, CHRISTIAN BERLIN, KURT		
IPC分类号	G01N27/62 A61K31/711 A61K48/00 A61P7/04 A61P9/10 A61P11/06 A61P13/12 A61P29/00 A61P35/00 B01J19/00 C07K14/46 C07K14/47 C07K14/82 C12M1/00 C12N15/09 C12Q1/48 C12Q1/68 G01N33/483 G01N33/53 G01N33/566 G01N37/00 A61K31/70		
CPC分类号	A61P11/06 A61P13/12 A61P29/00 C07K14/4703 C07K14/82 C12Q1/6883 C12Q1/6886 C12Q2600/154 C12Q2600/156		
优先权	10019058 2000-04-06 DE 10019173 2000-04-07 DE 10032529 2000-06-30 DE 10043826 2000-09-01 DE		
外部链接	<a href="#">Espacenet</a>		

#### 摘要(译)

本发明涉及与基因调节相关的基因的化学修饰的基因组序列，涉及用于检测与针对该序列的基因调节相关的基因的胞嘧啶甲基化状态的寡核苷酸和/或PNA-寡聚物，以及方法用于确定与基因调节相关的基因的遗传和/或表观遗传参数。