



(43) International Publication Date
6 February 2014 (06.02.2014)

(10) International Publication Number
WO 2014/020048 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/EP2013/066047

(22) International Filing Date:
31 July 2013 (31.07.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
12178715.4 31 July 2012 (31.07.2012) EP

(71) Applicants: **RUPRECHT-KARLS-UNIVERSITÄT HEIDELBERG** [DE/DE]; Grabengasse 1, 69117 Heidelberg (DE). **DEUTSCHES KREBSFORSCHUNGSZENTRUM** [DE/DE]; Im Neuenheimer Feld 280, 69120 Heidelberg (DE).

(72) Inventors: **BURWINKEL, Barbara**; Brückenkopfstrasse 2, 69120 Heidelberg (DE). **YANG, Rongxi**; Karlstrasse 12, 69117 Heidelberg (DE).

(74) Agent: **DICK, Alexander**; Herzog Fiesser & Partner, Patentanwälte, Dudenstrasse 46, 68167 Mannheim (DE).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

- with international search report (Art. 21(3))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: **HYAL2 METHYLATION AND EXPRESSION AS A CANCER MARKER**

(57) Abstract: The present invention is concerned with a method for diagnosing cancer in a subject, comprising the steps of: (a) determining in a sample of said subject the methylation status of at least one CpG site located in the promoter region of the *HYAL2* gene; and (b) comparing said methylation status with a reference, whereby cancer is to be diagnosed. The present invention further relates to a method for diagnosing breast cancer in a subject, comprising the steps of: (a) determining in a sample of a subject suspected to suffer from said cancer the amount of a gene product of the *HYAL2* gene; and (b) comparing said amount with a reference, whereby breast cancer is to be diagnosed. The present invention also is concerned with a device for diagnosing cancer, comprising an analyzing comprising a detection agent for determining the methylation status of at least one of the CpG sites of the present invention in a sample of a subject suspected to suffer from breast cancer, and an evaluation unit comprising a data processor having tangibly embedded an algorithm for carrying out a comparison of the amount determined by the analyzing unit with a stored reference and which is capable of generating an output file containing a diagnosis established based on the said comparison. The present invention also is concerned with a kit for carrying out a method of the present invention and with a method for therapy monitoring in a subject being treated against cancer, comprising the steps of (a) obtaining a first and a second sample from a subject, wherein the first sample is obtained at a time point before the second sample, (b) determining in said first and second sample the methylation status of at least one of the CpG sites of the present invention and (c) comparing the methylation status of said first sample to the methylation status of said second sample, thereby monitoring therapy in said subject.



WO 2014/020048 A1

HYAL2 methylation and expression as a cancer marker

5

The present invention is concerned with a method for diagnosing cancer in a subject, comprising the steps of: (a) determining in a sample of said subject the methylation status of at least one CpG site located in the promoter region of the HYAL2 gene; and (b) comparing said methylation status with a reference, whereby cancer is to be diagnosed. The present invention further relates to a method for diagnosing breast cancer in a subject, comprising the steps of: (a) determining in a sample of a subject suspected to suffer from said cancer the amount of a gene product of the HYAL2 gene; and (b) comparing said amount with a reference, whereby breast cancer is to be diagnosed. The present invention also is concerned with a device for diagnosing cancer, comprising an analyzing comprising a detection agent for determining the methylation status of at least one of the CpG sites of the present invention in a sample of a subject, or comprising a detection agent for determining the amount of a HYAL2 gene product in a sample of a subject suspected to suffer from breast cancer, and an evaluation unit comprising a data processor having tangibly embedded an algorithm for carrying out a comparison of the amount determined by the analyzing unit with a stored reference and which is capable of generating an output file containing a diagnosis established based on the said comparison. The present invention also is concerned with a kit for carrying out a method of the present invention and with a method for therapy monitoring in a subject being treated against cancer, comprising the steps of (a) obtaining a first and a second sample from a subject, wherein the first sample is obtained at a time point before the second sample, (b) determining in said first and second sample the methylation status of at least one of the CpG sites of the present invention and (c) comparing the methylation status of said first sample to the methylation status of said second sample, thereby monitoring therapy in said subject.

Breast cancer is the most common cancer among women. About one out of nine women will develop breast cancer during her life (Feuer, E.J., et al., The lifetime risk of developing breast cancer. J Natl Cancer Inst 85, 892-897 (1993)). Most of the breast cancers occur sporadic, whereas familial breast cancer accounts for about 10 % of all breast cancer cases (Fackenthal, J.D. & Olopade, O.I. Breast cancer risk associated with BRCA1 and BRCA2 in diverse populations. Nat Rev Cancer 7, 937-948 (2007)). Mutations in the main breast cancer related genes, BRCA1 and BRCA2 account for 25% and other intermediate- and low- penetrance genes for about 5% of all familial cases (Yang, R. & Burwinkel, B. (eds.). Familial risk in breast cancer, 251-256 (Springer, 2010)). Recent genome-wide association studies (GW AS) and single candidate gene approaches have been quite successful in detecting genetic low-risk variants for

breast cancer (Thomas, G., et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1 p 11.2 and 14q24.1 (RAD51 L1). Nat Genet 41, 579-584 (2009); Cox, A., et al. A common coding variant in CASP8 is associated with breast cancer risk. Nat Genet 39, 352-358 (2007); Stacey, S.N., et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 40, 703-706 (2008); Ahmed, S., et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet 41,585-590 (2009); Easton, D.F., et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 447, 1087-1093 (2007); Milne, R.L., et al. Risk of estrogen receptor-positive and -negative breast cancer and single-nucleotide polymorphism 2q35-rs13387042. J Natl Cancer Inst 101, 1012-1018 (2009); Frank, B., et al. Association of a common AKAP9 variant with breast cancer risk: a collaborative analysis. J Natl Cancer Inst 100,437-442 (2008)). However, a large number of breast cancer risk factors remain to be explored.

Epigenetic changes are defined as changes in gene expression that are not due to any alterations in the genomic DNA sequence. Aberrant epigenetic signatures have been considered as a hallmark of human cancer (Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. Nat Rev Genet 8, 286-298 (2007)). One of the most important epigenetic signatures, DNA methylation, has critical roles in the control of gene activities and in the architecture of the nucleus of the cell Weber, M., et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 37, 853-862 (2005)). Furthermore, unlike genetic markers or variants, DNA methylation is principally reversible. Therefore, the methylation profile of specific genes are considered as therapeutic targets (Mack, G.S. Epigenetic cancer therapy makes headway. J Natl Cancer Inst 98, 1443-1444 (2006)). Meanwhile, due to the variable character, DNA methylation may serve as a link between environmental factors and the genome. DNA methylation modulated by environmental factors or aging may alter the expression of critical genes of cells and consequently induce malignant transformation of cells or even a cancer (Widschwendter, M., et al. Epigenotyping in peripheral blood cell DNA and breast cancer risk: a proof of principle study. P LoS One 3, e2656 (2008)).

Previous studies have explored hypermethylation in the promoter regions of tumor suppressor genes and hypomethylation in the promoter regions of oncogenes in breast cancer compared to their normal adjacent tissues (Ito, Y., et al. Somatically acquired hypomethylation of IGF2 in breast and colorectal cancer. Hum Mol Genet 17, 2633-2643 (2008); Potapova, A., Hoffman, A.M., Godwin, A.K., Al-Saleem, T. & Cairns, P. Promoter hypermethylation of the PALB2 susceptibility gene in inherited and sporadic breast and ovarian cancer. Cancer Res 68, 998-1002 (2008); Radpour, R., et al. Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. Oncogene 28, 2969-2978 (2009); Widschwendter, M.

& Jones, P.A. DNA methylation and breast carcinogenesis. *Oncogene* 21, 5462-5482 (2002)). Very few studies have focused on the methylation signatures in the peripheral blood DNA and breast cancer risk. In these studies, only specific genes, like BRCA1 (Iwamoto, T., Yamamoto, N., Taguchi, T., Tamaki, Y. & Noguchi, S. BRCA1 promoter methylation in peripheral blood cells is associated with increased risk of breast cancer with BRCA1 promoter methylation. *Breast Cancer Res Treat* 129,69-77 (2011)), ATM (Flanagan, J.M., et al. Gene-body hypermethylation of ATM in peripheral blood DNA of bilateral breast cancer patients. *Hum Mol Genet* 18, 1332-1342 (2009)), and genes in specific pathways (Widschwendter et al. (2008), loc. cit.) have been investigated.

There is thus a need in the art for the identification of further epigenetic markers of breast cancer and other cancers, preferably allowing the identification of afflicted subjects by obtaining a sample by a means of low invasiveness, e.g. by taking a blood sample.

Thus, the present invention relates to a method for diagnosing cancer in a subject, comprising the steps of: (a) determining in a sample of said subject the methylation status of at least one CpG site located in the promoter region of the HYAL2 gene; and (b) comparing said methylation status with a reference, whereby cancer is to be diagnosed.

The method of the present invention, preferably, is an in vitro method. Moreover, it may comprise steps in addition to those explicitly mentioned above. For example, further steps may relate to sample pre-treatments or evaluation of the results obtained by the method. The method may be carried out manually or assisted by automation. Preferably, steps a) and/or b) or parts thereof may in total or in part be assisted by automation, e.g., by a suitable robotic equipment for the determining the methylation status in step a), or a computer-implemented calculation or comparison step in step b).

The method of the present invention allows assessing whether a subject suspected to be afflicted with cancer is in fact afflicted with cancer. Preferably, by carrying out the method of the present invention decisions can be made whether treatment shall be initiated or not. More preferably, the method of the present comprises the further step of initiating appropriate treatment in case cancer is diagnosed.

As used herein, the term “diagnosing” refers to assessing the probability according to which a subject is afflicted or will be afflicted with a disease or condition referred to in this specification. As will be understood by those skilled in the art, such an assessment is usually not intended to be correct for 100% of the subjects to be diagnosed. The term, however, requires that a statistically significant portion of subjects can be correctly diagnosed to be afflicted with the disease or

condition. Whether a portion is statistically significant can be determined without further ado by the person skilled in the art using various well known statistic evaluation tools, e.g., determination of confidence intervals, p-value determination, Student's t-test, Mann-Whitney test etc.. Details are found in Dowdy and Wearden, Statistics for Research, John Wiley & Sons, New York 1983. Preferred confidence intervals are at least 90%, at least 95%, at least 97%, at least 98% or at least 99 %. The p-values are, preferably, 0.1, 0.05, 0.01, 0.005, or 0.0001. Preferably, the probability envisaged by the present invention allows that the diagnosis will be correct for at least 60%, at least 70%, at least 80%, or at least 90% of the subjects of a given cohort or population.

"Cancer" in the context of this invention refers to a disease of an animal, including man, characterized by uncontrolled growth by a group of body cells ("cancer cells"). This uncontrolled growth may be accompanied by intrusion into and destruction of surrounding tissue and possibly spread of cancer cells to in the vicinity of the primary tumor or to other locations in the body. Preferably, the cancer is breast cancer or ovarian cancer, or other epithelial cancers, preferably cancers in lung, colorectal, pancreatic, melanoma, prostate, and the like, more preferably, the cancer is breast cancer with involvement of at least three lymph nodes.

The term "subject" as referred to herein encompasses animals, preferably mammals, and, more preferably, humans. More preferably, said subject was in the past afflicted with, is at present afflicted with, is suspected to be afflicted with, or is at risk to be afflicted with cancer. Subjects which are afflicted with the said disease can be identified by the accompanying symptoms known for the disease. These symptoms are known in the art and described, e.g., uncontrolled cell proliferation, division, invasion and sometimes metastasis (Kumar, V., A.K. Abbas, and A.K. Abbas, Robbins Basic Pathology. 8th ed. 2007:Saunders). The commonest presentation of breast cancer is of a painless lump; other symptoms may include dimpling of the overlying skin, nipple inversion, oedema or peau d'orange and blood-stained nipple discharge (Stewart, B.W. and P. Kleihues, World cancer report. 2003, Lyon: IARC Press. 351). However, a subject suspected to be afflicted with the aforementioned disease may also be an apparently healthy subject, e.g., investigated by routine clinical screening, or may be a subject being at risk for developing the aforementioned disease. Risk groups for the disease are known in the art and described in, e.g., genetic susceptibility, reproductive factors and hormones, nutrition and diet, environmental pollution, radiation (Stewart, B.W. and P. Kleihues, World cancer report. 2003, Lyon: IARC Press. 351). Preferably, the subject is female, more preferably from the Caucasian subpopulation.

As used herein, the term "CpG site" relates to a dinucleotide sequence 5'-CG-3' comprised in a polynucleotide, preferably comprised in DNA, more preferably comprised in genomic DNA of a subject. The CpG sites to be analyzed according to the present invention are the CpG sites

located in the promoter region of the HYAL2 gene, preferably in the region 3000 nucleotides, 2500 nucleotides, 2100 nucleotides, or 1750 nucleotides upstream of the translation start site of the HYAL2 gene. More preferably, the CpG sites to be analyzed according to the present invention are the CpG sites located in the region 1750-3000 nucleotides, 2100-3000 nucleotides, or 2500-3000 nucleotides upstream of the translation start site of the HYAL2 gene. Preferably, the subject is a human and the HYAL2 gene is the human HYAL2 gene located on human chromosome 3 (Genbank Acc No: NC_000003.11 GI:224589815). More preferably, the 3000 nucleotides upstream the HYAL2 gene are the 3000 nucleotides upstream the human HYAL2 gene; most preferably, said 3000 nucleotides upstream the human HYAL2 gene comprise the sequence of SEQ ID NO:1. It is understood by the skilled person that the sequence of SEQ ID NO:1 is displayed such that the first nucleotide of the sequence is the first nucleotide upstream the translation start site of the HYAL2 gene, the second nucleotide of SEQ ID NO:1 is the second nucleotide upstream the translation start site of the HYAL2 gene, i.e. position -2 relative to said site, and so forth. Moreover, also encompassed are variants of the aforementioned specific polynucleotide. Said variants may represent orthologs, paralogues or other homologs of the polynucleotide of the present invention. The polynucleotide variants, preferably, comprise a nucleic acid sequence characterized in that the sequence can be derived from the aforementioned specific nucleic acid sequence shown in SEQ ID NO:1 by at least one nucleotide substitution, addition and/or deletion whereby the variant nucleic acid sequence shall still be comprised in the promoter region of a HYAL2 gene as specified above. Variants also encompass polynucleotides comprising a nucleic acid sequence which is capable of hybridizing to the aforementioned specific nucleic acid sequences, preferably, under stringent hybridization conditions. These stringent conditions are known to the skilled worker and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N. Y. (1989), 6.3.1-6.3.6. A preferred example for stringent hybridization conditions are hybridization conditions in 6x sodium chloride/sodium citrate (= SSC) at approximately 45°C, followed by one or more wash steps in 0.2x SSC, 0.1% SDS at 50 to 65°C. The skilled worker knows that these hybridization conditions differ depending on the type of nucleic acid and, for example when organic solvents are present, with regard to the temperature and concentration of the buffer. For example, under "standard hybridization conditions" the temperature differs depending on the type of nucleic acid between 42°C and 58°C in aqueous buffer with a concentration of 0.1 to 5x SSC (pH 7.2). If organic solvent is present in the abovementioned buffer, for example 50% formamide, the temperature under standard conditions is approximately 42°C. The hybridization conditions for DNA:DNA hybrids are preferably for example 0.1x SSC and 20°C to 45°C, preferably between 30°C and 45°C. The hybridization conditions for DNA:RNA hybrids are preferably, for example, 0.1x SSC and 30°C to 55°C, preferably between 45°C and 55°C. The abovementioned hybridization temperatures are determined for example for a nucleic acid with approximately 100 bp (= base pairs) in length and a G + C content of 50% in the absence of formamide. The skilled worker knows how to determine the hybridization conditions required by referring to textbooks such as

the textbook mentioned above, or the following textbooks: Sambrook et al., "Molecular Cloning", Cold Spring Harbor Laboratory, 1989; Hames and Higgins (Ed.) 1985, "Nucleic Acids Hybridization: A Practical Approach", IRL Press at Oxford University Press, Oxford; Brown (Ed.) 1991, "Essential Molecular Biology: A Practical Approach", IRL Press at Oxford University Press, Oxford. Further, variants include polynucleotides comprising nucleic acid sequences which are at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98% or at least 99% identical to the nucleic acid sequences shown in SEQ ID NO: 1. The percent identity values are, preferably, calculated over the entire nucleic acid sequence region. A series of programs based on a variety of algorithms is available to the skilled worker for comparing different sequences. In this context, the algorithms of Needleman and Wunsch or Smith and Waterman give particularly reliable results. To carry out the sequence alignments, the program PileUp (J. Mol. Evolution., 25, 351-360, 1987, Higgins et al., CABIOS, 5 1989: 151-153) or the programs Gap and BestFit [Needleman and Wunsch (J. Mol. Biol. 48; 443-453 (1970)) and Smith and Waterman (Adv. Appl. Math. 2; 482-489 (1981))], which are part of the GCG software packet [Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 53711 (1991)], are to be used. The sequence identity values recited above in percent (%) are to be determined, preferably, using the program GAP over the entire sequence region with the following settings: Gap Weight: 50, Length Weight: 3, Average Match: 10.000 and Average Mismatch: 0.000, which, unless otherwise specified, shall always be used as standard settings for sequence alignments.

Preferably, the methylation status of at least one of the CpG sites located between position 50334760 and position 50335700 on human chromosome 3 is determined. More preferably, preferably referring to build 36.1/hg18 of the human genome, the methylation status of at least one of the CpG sites located at position 50335694 (cg27091787), 50335584 (HYAL_CpG_1), 50335646 (HYAL_CpG_2), or 50335671 (HYAL_CpG_3), 50335166 (HYAL-is-310 CpG_1), 50335180 (HYAL-is-310 CpG_2), 50335192 (HYAL-is-310 CpG_3), 50335195 (HYAL-is-310 CpG_4), 50335227 (HYAL-is-310 CpG_5), 50335233 (HYAL-is-310 CpG_6), 50335300 (HYAL-is-310 CpG_7), 50335315 (HYAL-is-310 CpG_8), 50335375 (HYAL-is-310 CpG_9), 50335392 (HYAL-is-310 CpG_10), 50335401 (HYAL-is-310 CpG_11), 50334744 (HYAL2-is-325_CpG_1), 50334761 (HYAL2-is-325_CpG_2), 50334804 (HYAL2-is-325_CpG_3), 50334844 (HYAL2-is-325_CpG_4), 50334853 (HYAL2-is-325_CpG_5), 50334862 (HYAL2-is-325_CpG_6), 50334880 (HYAL2-is-325_CpG_7), 50334906 (HYAL2-is-325_CpG_8), 50334913 (HYAL2-is-325_CpG_9), 50334917 (HYAL2-is-325_CpG_10), 50334928 (HYAL2-is-325_CpG_11), 50334944 (HYAL2-is-325_CpG_12), 50334956 (HYAL2-is-325_CpG_13), 50334980 (HYAL2-is-325_CpG_14), 50334982 (HYAL2-is-325_CpG_15), 50335010 (HYAL2-is-325_CpG_16) 50335014 (HYAL2-is-325_CpG_17) is determined. Most preferably, at least one CpG site is selected from the list consisting of cg27091787 at position 50335694, HYAL_CpG_1 at position 50335584, HYAL_CpG_2 at position 50335646, and HYAL_CpG_3

at position 50335671. Preferably, the methylation status of at least two, at least three, at least four, at least five, at least six, at least seven, at least eight, at least nine, at least ten, at least eleven, at least twelve, or at least fifteen CpG sites of the present invention is determined. It is understood by the skilled person that the exact numbering of said CpG sites may depend on the specific genomic sequence and on the specific sequence of the HYAL2 promoter region comprised in the sample to be analyzed. E.g the HYAL2 gene is located on Chromosome 3: positions 50,355,221-50,360,337 in build37/hg19, but on Chromosome 3: positions 50,330,244-50,335,146 in build36/hg18. Thus, analysis of a CpG site corresponding to a CpG site of the present invention is also encompassed by the present invention. The skilled person knows how to determine the CpG sites in a sample corresponding to the CpG sites detailed herein above, e.g. by determining the translation start site of the HYAL2 gene and / or by aligning said sequence from a sample to the sequence of SEQ ID NO:1. Further, it is also envisaged by the present invention that the methylation status of other CpG sites is determined in addition to determining the methylation status of a CpG site of the present invention.

In a preferred embodiment, the promoter region of the HYAL2 gene includes sequences downstream of the translation start site contributing to regulation of the HYAL2 gene. Thus, preferably, the term "CpG site" also relates to CpG sites located up to 2000 nucleotides upstream or up to 2000 nucleotides downstream of the last nucleotide of the translational stop codon of the HYAL2 gene. More preferably, the HAYL2 gene is the human HYAL2 gene and the up to 2000 nucleotides upstream or up to 2000 nucleotides downstream of the last nucleotide of the translational stop codon of the HYAL2 gene are the up to 2000 nucleotides upstream or up to 2000 nucleotides downstream of the last nucleotide of the translational stop codon of the human HYAL2 gene; even more preferably, the up to 2000 nucleotides upstream or up to 2000 nucleotides downstream of the last nucleotide of the translational stop codon of the HYAL2 gene are nucleotides 50328231 to 50332163 of human chromosome 3 in build36/hg18. Most preferably, said up to 2000 nucleotides upstream or up to 2000 nucleotides downstream of the last nucleotide of the translational stop codon of the human HYAL2 gene comprise the sequence of SEQ ID NO: 12. Accordingly, preferred CpG sites are cg08776109 (chr3:50331237, build36/hg18), located in the human HYAL2 gene in the last intron, and cg06721473 (chr3:50330420, build36/hg18), located in the human HYAL2 gene in the 3'UTR.

In a preferred embodiment, the CpG site is 50335694 (cg27091787), 50335584 (HYAL_CpG_1), 50335646 (HYAL_CpG_2), 50335671 (HYAL_CpG_3), 50335166 (HYAL-is-310 CpG_1), 50335180 (HYAL-is-310 CpG_2), 50335192 (HYAL-is-310 CpG_3), 50335195 (HYAL-is-310 CpG_4), 50335227 (HYAL-is-310 CpG_5), 50335233 (HYAL-is-310 CpG_6), 50335300 (HYAL-is-310 CpG_7), 50335315 (HYAL-is-310 CpG_8), 50335375 (HYAL-is-310 CpG_9), 50335392 (HYAL-is-310 CpG_10), 50335401 (HYAL-is-310 CpG_11), 50334744 (HYAL2-is-325_CpG_1), 50334761 (HYAL2-is-325_CpG_2), 50334804 (HYAL2-is-

325_CpG_3), 50334844 (HYAL2-is-325_CpG_4), 50334853 (HYAL2-is-325_CpG_5), 50334862 (HYAL2-is-325_CpG_6), 50334880 (HYAL2-is-325_CpG_7), 50334906 (HYAL2-is-325_CpG_8), 50334913 (HYAL2-is-325_CpG_9), 50334917 (HYAL2-is-325_CpG_10), 50334928 (HYAL2-is-325_CpG_11), 50334944 (HYAL2-is-325_CpG_12), 50334956 (HYAL2-is-325_CpG_13), 50334980 (HYAL2-is-325_CpG_14), 50334982 (HYAL2-is-325_CpG_15), 50335010 (HYAL2-is-325_CpG_16) 50335014 (HYAL2-is-325_CpG_17), 50331237 (cg08776109), or 50330420 (cg06721473). As detailed above, preferably, analysis of more than one CpG site is envisaged.

In a further preferred embodiment, the CpG site(s) is/are selected from the list consisting of 50335694 (cg27091787), 50335584 (HYAL_CpG_1), 50335646 (HYAL_CpG_2), 50335671 (HYAL_CpG_3), 50335166 (HYAL-is-310 CpG_1), 50335180 (HYAL-is-310 CpG_2), 50335192 (HYAL-is-310 CpG_3), 50335195 (HYAL-is-310 CpG_4), 50335227 (HYAL-is-310 CpG_5), 50335233 (HYAL-is-310 CpG_6), 50335300 (HYAL-is-310 CpG_7), 50335315 (HYAL-is-310 CpG_8), 50335375 (HYAL-is-310 CpG_9), 50335392 (HYAL-is-310 CpG_10), 50335401 (HYAL-is-310 CpG_11), 50334744 (HYAL2-is-325_CpG_1), 50334761 (HYAL2-is-325_CpG_2), 50334804 (HYAL2-is-325_CpG_3), 50334844 (HYAL2-is-325_CpG_4), 50334853 (HYAL2-is-325_CpG_5), 50334862 (HYAL2-is-325_CpG_6), 50334880 (HYAL2-is-325_CpG_7), 50334906 (HYAL2-is-325_CpG_8), 50334913 (HYAL2-is-325_CpG_9), 50334917 (HYAL2-is-325_CpG_10), 50334928 (HYAL2-is-325_CpG_11), 50334944 (HYAL2-is-325_CpG_12), 50334956 (HYAL2-is-325_CpG_13), 50334980 (HYAL2-is-325_CpG_14), 50334982 (HYAL2-is-325_CpG_15), 50335010 (HYAL2-is-325_CpG_16) 50335014 (HYAL2-is-325_CpG_17), 50331237 (cg08776109) and 50330420 (cg06721473), and cg06721473.

In a particularly preferred embodiment, the CpG site is 50335694 (cg27091787), 50335584 (HYAL_CpG_1), 50335646 (HYAL_CpG_2), 50335671 (HYAL_CpG_3), 50331237 (cg08776109), or 50330420 (cg06721473). As detailed above, preferably, analysis of more than one CpG site is envisaged.

In a further particularly preferred embodiment, the CpG site(s) is/are selected from the list consisting of 50335694 (cg27091787), 50335584 (HYAL_CpG_1), 50335646 (HYAL_CpG_2), 50335671 (HYAL_CpG_3), 50331237 (cg08776109) and 50330420 (cg06721473), and cg06721473.

The term "determining the methylation status" relates to determining if a methyl group is present at the 5 position of the pyrimidine ring of a cytosine in a polynucleotide. Preferably, the cytosine residue is followed in 3' direction by a guanosine residue, the two residues forming a CpG site. The presence of said methyl group can be determined by various methods well known to the

skilled person, including, e.g., methylation-specific PCR (MSP), whole genome bisulfite sequencing or other sequencing based methods, real-time PCR based methods of bisulfite treated DNA, e.g. MethyLight, restriction with a methylation-sensitive restriction enzyme, e.g. in the HpaII tiny fragment enrichment by ligation-mediated PCR (HELP)-Assay, pyrosequencing of bisulfite treated DNA, or the like AIMS, amplification of inter-methylated sites; BC-seq, bisulphite conversion followed by capture and sequencing; BiMP, bisulphite methylation profiling; BS, bisulphite sequencing; BSPP, bisulphite padlock probes; CHARM, comprehensive high-throughput arrays for relative methylation; COBRA, combined bisulphite restriction analysis; DMH, differential methylation hybridization; HELP, HpaII tiny fragment enrichment by ligation-mediated PCR; MCA, methylated CpG island amplification; MCAM, MCA with microarray hybridization; MeDIP, mDIP and mCIP, methylated DNA immunoprecipitation; MIRA, methylated CpG island recovery assay; MMAS, microarray-based methylation assessment of single samples; MS-AP-PCR, methylation-sensitive arbitrarily primed PCR; MSCC, methylation-sensitive cut counting; MSP, methylation-specific PCR; MS-SNuPE, methylation-sensitive single nucleotide primer extension; NGS, next-generation sequencing; RLGS, restriction landmark genome scanning; RRBS, reduced representation bisulphite sequencing; -seq, followed by sequencing; WGSBS, whole-genome shotgun bisulphite sequencing. (Manel Esteller, Cancer epigenomics: DNA methylomes and histone-modification maps, *Nature*, 2007, 8:286-298; Peter W. Laird, Principles and challenges of genome-wide DNA methylation analysis. *Nature Review Genetics*, 2010, 11: 191-203). Preferably, the methylation status is determined by the methods described in the examples herein below, e.g. the sequencing-based Infinium 27K methylation assay or the mass spectrometry based method of MALDI-TOF mass spectrometry. As such, the methylation status of a specific cytosine residue in a specific polynucleotide molecule can only be "unmethylated" (meaning 0% methylation) or "methylated" (meaning 100% methylation). In the case of a CpG site in a double-stranded DNA molecule, which comprises two cytosine residues, the methylation status can be "unmethylated" (meaning 0% methylation, i.e. none of the two cytosine residues methylated), "hemimethylated" (meaning 50% methylation, i.e. one of the two cytosine residues methylated), or "methylated" or "fully methylated" (meaning 100% methylation, i.e. both cytosine residues methylated) It is, however, understood by the person skilled in the art that if polynucleotides from a multitude of cells are obtained and the methylation status of a specific cytosine residue within said multitude of polynucleotides is determined, an average methylation status is determined, which can e.g. preferably, be expressed as a percentage (% methylation), and which can assume any value between 0% and 100%. It is also understood by the skilled person, that the methylation status can be expressed as a percentage in case the average methylation of different cell populations is determined. E.g. the blood cells according to the present invention are a mixture of variant cell types. It is possible that certain cell types have high methylation levels whereas other cell types have lower methylation levels, and finally reach an average methylation of e.g. 50 %.

The term “sample”, as used herein, refers to a sample of a body fluid, to a sample of separated cells or to a sample from a tissue or an organ or to a sample of wash/rinse fluid obtained from an outer or inner body surface. Samples can be obtained by well known techniques and include, preferably, scrapes, swabs or biopsies from the digestive tract, liver, pancreas, anal canal, the oral cavity, the upper aerodigestive tract and the epidermis. Such samples can be obtained by use of brushes, (cotton) swabs, spatula, rinse/wash fluids, punch biopsy devices, puncture of cavities with needles or surgical instrumentation. Tissue or organ samples may be obtained from any tissue or organ by, e.g., biopsy or other surgical procedures. More preferably, samples are samples of body fluids, e.g., preferably, blood, plasma, serum, urine, saliva, lacrimal fluid, and fluids obtainable from the breast glands, e.g. milk. Most preferably, the sample of a body fluid comprises cells of the subject. Separated cells may be obtained from the body fluids or the tissues or organs by separating techniques such as filtration, centrifugation or cell sorting. Preferably, samples are obtained from those body fluids described herein below. More preferably, cells are isolated from said body fluids as described herein below. Most preferably, cells are T-cells enriched or purified from a blood sample.

“Comparing” as used herein encompasses comparing the methylation status referred to herein which determined in the sample to be analyzed with the methylation status in a suitable reference sample as specified elsewhere herein in this description. It is to be understood that comparing as used herein refers to a comparison of corresponding parameters or values, e.g., a value of % methylation as referred to herein is compared to a value of % methylation in a reference; a number of methylated sequences per a defined amount of molecules evaluated as referred to herein is compared to a number of methylated sequences per the same defined amount of molecules; or an intensity signal obtained from determining the methylation status as referred to herein in a test sample is compared to the same type of intensity signal obtained from a reference sample. The comparison referred to in the methods of the present invention may be carried out manually or computer assisted. For a computer assisted comparison, the value of the determined amount or ratio may be compared to values corresponding to suitable references which are stored in a database by a computer program. The computer program may further evaluate the result of the comparison by means of an expert system. Accordingly, the result of the identification referred to herein may be automatically provided in a suitable output format.

The terms “reference”, “reference amount”, and “reference value” as used herein refer to a measure of the methylation status according to the present invention, preferably a value of % methylation, which allows assessing if being afflicted with cancer or not being afflicted with cancer is to be assumed for the subject from which the sample is derived. A suitable reference value may be determined from a reference sample to be analyzed together, i.e. simultaneously or subsequently, with the sample.

Reference amounts can, in principle, be calculated for a group or cohort of subjects as specified herein based on the average or mean values for a methylation status by applying standard methods of statistics. In particular, accuracy of a test such as a method aiming to diagnose an event, or not, is best described by its receiver-operating characteristics (ROC) (see especially
5 Zweig 1993, Clin. Chem. 39:561-577). The ROC graph is a plot of all of the sensitivity versus specificity pairs resulting from continuously varying the decision threshold over the entire range of data observed. The clinical performance of a diagnostic method depends on its accuracy, i.e. its ability to correctly allocate subjects to a certain prognosis or diagnosis. The ROC plot
10 indicates the overlap between the two distributions by plotting the sensitivity versus 1-specificity for the complete range of thresholds suitable for making a distinction. On the y-axis is sensitivity, or the true-positive fraction, which is defined as the ratio of number of true-positive test results to the product of number of true-positive and number of false-negative test results. This has also been referred to as positivity in the presence of a disease or condition. It is
15 calculated solely from the affected subgroup. On the x-axis is the false-positive fraction, or 1-specificity, which is defined as the ratio of number of false-positive results to the product of number of true-negative and number of false-positive results. It is an index of specificity and is calculated entirely from the unaffected subgroup. Because the true- and false-positive fractions are calculated entirely separately, by using the test results from two different subgroups, the
20 ROC plot is independent of the prevalence of the event in the cohort. Each point on the ROC plot represents a sensitivity/-specificity pair corresponding to a particular decision threshold. A test with perfect discrimination (no overlap in the two distributions of results) has an ROC plot that passes through the upper left corner, where the true-positive fraction is 1.0, or 100% (perfect sensitivity), and the false-positive fraction is 0 (perfect specificity). The theoretical plot for a test
25 with no discrimination (identical distributions of results for the two groups) is a 45° diagonal line from the lower left corner to the upper right corner. Most plots fall in between these two extremes. If the ROC plot falls completely below the 45° diagonal, this is easily remedied by reversing the criterion for "positivity" from "greater than" to "less than" or vice versa. Qualitatively, the closer the plot is to the upper left corner, the higher the overall accuracy of the
30 test. Dependent on a desired confidence interval, a threshold can be derived from the ROC curve allowing for the diagnosis or prediction for a given event with a proper balance of sensitivity and specificity, respectively. Accordingly, the reference to be used for the methods of the present invention can be generated, preferably, by establishing a ROC for said cohort as described above and deriving a threshold amount there from. Dependent on a desired sensitivity and specificity
35 for a diagnostic method, the ROC plot allows deriving suitable thresholds.

Preferably, the reference amount as used herein is derived from samples of subjects known or suspected not to be afflicted with cancer. This reference amount level may be a discrete figure or may be a range of figures. Evidently, the reference level or amount may vary between individual

CpG sites or between age groups (see Examples below). The measuring system therefore, preferably, is calibrated with a sample or with a series of samples comprising known amounts of methylated and / or unmethylated CpG sites. It is understood by the skilled person that in such case the methylation status can preferably be expressed as arbitrary units (AU).

5 The reference amount applicable for an individual subject may vary depending on various physiological parameters such as age, gender, or subpopulation. Thus, a suitable reference amount may be determined by the methods of the present invention from a reference sample to be analyzed together, i.e. simultaneously or subsequently, with the test sample. Moreover, a
10 threshold amount can be preferably used as a reference amount. Preferably, a methylation status which is below the threshold is indicative of a subject afflicted with cancer; and a methylation status which is equal to or below the threshold amount will be indicative a subject not afflicted with cancer. It is to be understood that the aforementioned amounts may vary due to statistics and errors of measurement.

15 The definitions made above apply mutatis mutandis to the following:

The present invention also relates to a method for diagnosing breast cancer in a subject, comprising the steps of: (a) determining in a sample of a subject suspected to suffer from said
20 cancer the amount of a gene product of the HYAL2 gene; and (b) comparing said amount with a reference, whereby breast cancer is to be diagnosed.

The method for diagnosing breast cancer, preferably, is an in vitro method. Moreover, it may comprise steps in addition to those explicitly mentioned above. For example, further steps may
25 relate to sample pre-treatments or evaluation of the results obtained by the method. The method may be carried out manually or assisted by automation. Preferably, steps a) and/or b) or parts thereof may in total or in part be assisted by automation, e.g., by a suitable robotic equipment for the determining the methylation status in step a), or a computer-implemented calculation or comparison step in step b).

30 The sample of the method for diagnosing breast cancer, preferably, is the sample as described herein above. More preferably, the sample of the method for diagnosing breast cancer is not a tissue sample and / or is not a tumor sample and / or is not a sample derived from a suspected tumor site and / or does not comprise cancer cells. Most preferably, the sample of the method for
35 diagnosing breast cancer is a sample of cells obtained from peripheral blood, e.g. peripheral blood leukocytes.

As used herein, the term "gene product" relates to a, preferably macromolecular, physical entity, the presence of which in a cell depends on the expression of said gene in said cell. The

mechanisms of gene expression are well-known to the one skilled in the art to include the basic mechanisms of transcription, i.e. formation of RNA corresponding to the said gene or parts thereof, and translation, i.e. production of polypeptide molecules having an amino acid sequence encoded by said RNA according to the genetic code; it is well-known to the one skilled in the art that other cellular processes may be involved in gene expression as well, e.g. RNA processing, RNA editing, proteolytic processing, protein editing, and the like. The term gene product thus includes RNA, preferably mRNA, as well as polypeptides expressed from said gene. It is clear from the above that the term gene product also includes fragments of said RNA(s), preferably with a length of at least ten, at least twelve, at least 20, at least 50, or at least 100 nucleotides, and fragments (peptides) from said polypeptides, preferably with a length of at least eight, at least ten, at least twelve, at least 15, at least 20 amino acids.

"Determining" the amount of a gene product relates to measuring the amount of said gene product, preferably semi-quantitatively or quantitatively. Measuring can be done directly or indirectly. Preferably, measuring is performed on a processed sample, said processing comprising extraction of polynucleotides or polypeptides from the sample. It is, however, also envisaged by the present invention that the gene product is determined in situ, e.g. by immunohistochemistry (IHC)

The amount of the polynucleotides of the present invention can be determined with several methods well-known in the art. Quantification preferably is absolute, i.e. relating to a specific number of polynucleotides or, more preferably, relative, i.e. measured in arbitrary normalized units. Preferably, a normalization is carried out by calculating the ratio of a number of specific polynucleotides and total number of polynucleotides or a reference amplification product. Methods allowing for absolute or relative quantification are well known in the art. E.g., quantitative PCR methods are methods for relative quantification; if a calibration curve is incorporated in such an assay, the relative quantification can be used to obtain an absolute quantification. Other methods known are, e.g. nucleic acid sequence-based amplification (NASBA) or the Branched DNA Signal Amplification Assay method in combination with dot blot or luminex detection of amplified polynucleotides. Preferably, the polynucleotide amounts are normalized polynucleotide amounts, i.e. the polynucleotide amounts obtained are set into relation to at least one reference amplification product, thereby, preferably, setting the polynucleotide amounts into relation to the number of cells in the sample and/or the efficiency of polynucleotide amplification. Thus, preferably, the reference amplification product is a product obtained from a polynucleotide known to have a constant abundancy in each cell, i.e. a polynucleotide comprised in most, preferably all, cells of a sample in approximately the same amount. More preferably, the reference amplification product is amplified from a chromosomal or mitochondrial gene or from the mRNA of a housekeeping gene.

The amount of peptides or polypeptides of the present invention can be determined in various ways. Direct measuring relates to measuring the amount of the peptide or polypeptide based on a signal which is obtained from the peptide or polypeptide itself and the intensity of which directly correlates with the number of molecules of the peptide present in the sample. Such a signal -
5 sometimes referred to as intensity signal -may be obtained, e.g., by measuring an intensity value of a specific physical or chemical property of the peptide or polypeptide. Indirect measuring includes measuring of a signal obtained from a secondary component (i.e. a component not being the peptide or polypeptide itself) or a biological read out system, e.g., measurable cellular responses, ligands, labels, or enzymatic reaction products.

10 In accordance with the present invention, determining the amount of a peptide or polypeptide can be achieved by all known means for determining the amount of a peptide in a sample. Said means comprise immunoassay and / or immunohistochemistry devices and methods which may utilize labeled molecules in various sandwich, competition, or other assay formats. Said assays
15 will develop a signal which is indicative for the presence or absence of the peptide or polypeptide. Moreover, the signal strength can, preferably, be correlated directly or indirectly (e.g. reverse- proportional) to the amount of polypeptide present in a sample. Further suitable methods comprise measuring a physical or chemical property specific for the peptide or polypeptide such as its precise molecular mass or NMR spectrum. Said methods comprise,
20 preferably, biosensors, optical devices coupled to immunoassays, biochips, analytical devices such as mass- spectrometers, NMR- analyzers, or chromatography devices. Further, methods include micro-plate ELISA-based methods, fully-automated or robotic immunoassays, Cobalt Binding Assays, and latex agglutination assays.

25 Also preferably, determining the amount of a peptide or polypeptide comprises the step of measuring a specific intensity signal obtainable from the peptide or polypeptide in the sample. As described above, such a signal may be the signal intensity observed at an m/z variable specific for the peptide or polypeptide observed in mass spectra or a NMR spectrum specific for the peptide or polypeptide.

30 Determining the amount of a peptide or polypeptide may, preferably, comprise the steps of (a) contacting the peptide with a specific ligand, (b) (optionally) removing non-bound ligand, (c) measuring the amount of bound ligand. The bound ligand will generate an intensity signal. Binding according to the present invention includes both covalent and non-covalent binding. A
35 ligand according to the present invention can be any compound, e.g., a peptide, polypeptide, nucleic acid, or small molecule, binding to the peptide or polypeptide described herein. Preferred ligands include antibodies, nucleic acids, peptides or polypeptides such as receptors or binding partners for the peptide or polypeptide and fragments thereof comprising the binding domains for the peptides, and aptamers, e.g. nucleic acid or peptide aptamers. Methods to prepare such

ligands are well-known in the art. For example, identification and production of suitable antibodies or aptamers is also offered by commercial suppliers. The person skilled in the art is familiar with methods to develop derivatives of such ligands with higher affinity or specificity. For example, random mutations can be introduced into the nucleic acids, peptides or polypeptides. These derivatives can then be tested for binding according to screening procedures known in the art, e.g. phage display. Antibodies as referred to herein include both polyclonal and monoclonal antibodies, as well as fragments thereof, such as Fv, Fab and F(ab)₂ fragments that are capable of binding antigen or hapten. The present invention also includes single chain antibodies and humanized hybrid antibodies wherein amino acid sequences of a non-human donor antibody exhibiting a desired antigen-specificity are combined with sequences of a human acceptor antibody. The donor sequences will usually include at least the antigen-binding amino acid residues of the donor but may comprise other structurally and/or functionally relevant amino acid residues of the donor antibody as well. Such hybrids can be prepared by several methods well known in the art. Preferably, the ligand or agent binds specifically to the peptide or polypeptide. Specific binding according to the present invention means that the ligand or agent should not bind substantially to ("cross-react" with) another peptide, polypeptide or substance present in the sample to be analyzed. Preferably, the specifically bound peptide or polypeptide should be bound with at least 3 times higher, more preferably at least 10 times higher and even more preferably at least 50 times higher affinity than any other relevant peptide or polypeptide. Nonspecific binding may be tolerable, if it can still be distinguished and measured unequivocally, e.g. according to its size on a Western Blot, or by its relatively higher abundance in the sample. Binding of the ligand can be measured by any method known in the art. Preferably, said method is semi-quantitative or quantitative. Suitable methods are described in the following.

First, binding of a ligand may be measured directly, e.g. by NMR or surface plasmon resonance. Second, if the ligand also serves as a substrate of an enzymatic activity of the peptide or polypeptide of interest, an enzymatic reaction product may be measured (e.g. the amount of a protease can be measured by measuring the amount of cleaved substrate, e.g. on a Western Blot). Alternatively, the ligand may exhibit enzymatic properties itself and the "ligand/peptide or polypeptide" complex or the ligand which was bound by the peptide or polypeptide, respectively, may be contacted with a suitable substrate allowing detection by the generation of an intensity signal. For measurement of enzymatic reaction products, preferably the amount of substrate is saturating. The substrate may also be labeled with a detectable label prior to the reaction. Preferably, the sample is contacted with the substrate for an adequate period of time. An adequate period of time refers to the time necessary for a detectable, preferably measurable, amount of product to be produced. Instead of measuring the amount of product, the time necessary for appearance of a given (e.g. detectable) amount of product can be measured. Third, the ligand may be coupled covalently or non-covalently to a label allowing detection and

measurement of the ligand. Labelling may be done by direct or indirect methods. Direct labelling involves coupling of the label directly (covalently or non-covalently) to the ligand. Indirect labelling involves binding (covalently or non-covalently) of a secondary ligand to the first ligand. The secondary ligand should specifically bind to the first ligand. Said secondary ligand may be coupled with a suitable label and/or be the target (receptor) of tertiary ligand binding to the secondary ligand. The use of secondary, tertiary or even higher order ligands is often used to increase the signal intensity. Suitable secondary and higher order ligands may include antibodies, secondary antibodies, and the well-known streptavidin-biotin system (Vector Laboratories, Inc.). The ligand or substrate may also be "tagged" with one or more tags as known in the art. Such tags may then be targets for higher order ligands. Suitable tags include biotin, digoxigenin, His-Tag, Glutathion-S-Transferase, FLAG, GFP, myc-tag, influenza A virus haemagglutinin (HA), maltose binding protein, and the like. In the case of a peptide or polypeptide, the tag is preferably at the N-terminus and/or C-terminus. Suitable labels are any labels detectable by an appropriate detection method. Typical labels include gold particles, latex beads, acridan ester, luminol, ruthenium, enzymatically active labels, radioactive labels, magnetic labels ("e.g. magnetic beads", including paramagnetic and superparamagnetic labels), and fluorescent labels. Enzymatically active labels include e.g. horseradish peroxidase, alkaline phosphatase, beta-Galactosidase, Luciferase, and derivatives thereof. Suitable substrates for detection include di-amino-benzidine (DAB), 3,3'-5,5'-tetramethylbenzidine, NBT-BCIP (4-nitro blue tetrazolium chloride and 5-bromo-4-chloro-3-indolyl-phosphate), CDP-Star™ (Amersham Biosciences), ECFT™ (Amersham Biosciences). A suitable enzyme-substrate combination may result in a colored reaction product, fluorescence or chemo luminescence, which can be measured according to methods known in the art (e.g. using a light-sensitive film or a suitable camera system). As for measuring the enzymatic reaction, the criteria given above apply analogously. Typical fluorescent labels include fluorescent proteins (such as GFP and its derivatives), Cy3, Cy5, Texas Red, Fluorescein, and the Alexa dyes (e.g. Alexa 568). Further fluorescent labels are available e.g. from Molecular Probes (Oregon). Also the use of quantum dots as fluorescent labels is contemplated. Typical radioactive labels include 35S, 125I, 32P, 33P and the like. A radioactive label can be detected by any method known and appropriate, e.g. a light-sensitive film or a phosphor imager. Suitable measurement methods according the present invention also include precipitation (particularly immunoprecipitation), electrochemiluminescence (electro-generated chemiluminescence), RIA (radioimmunoassay), ELISA (enzyme-linked immunosorbent assay), sandwich enzyme immune tests, electrochemiluminescence sandwich immunoassays (ECLIA), dissociation-enhanced lanthanide fluoro immuno assay (DELFLIA), scintillation proximity assay (SPA), turbidimetry, nephelometry, latex-enhanced turbidimetry or nephelometry, or solid phase immune tests, like e.g. reverse phase protein arrays or antibody arrays. Further methods known in the art (such as gel electrophoresis, 2D gel electrophoresis, SDS polyacrylamid gel electrophoresis (SDS-PAGE), Western Blotting, and mass spectrometry),

can be used alone or in combination with labelling or other detection methods as described above.

The amount of a peptide or polypeptide may be, also preferably, determined as follows: (a) contacting a solid support comprising a ligand for the peptide or polypeptide as specified above with a sample comprising the peptide or polypeptide and (b) measuring the amount peptide or polypeptide which is bound to the support. The ligand, preferably chosen from the group consisting of nucleic acids, peptides, polypeptides, antibodies and aptamers, is preferably present on a solid support in immobilized form. Materials for manufacturing solid supports are well known in the art and include, inter alia, commercially available column materials, polystyrene beads, latex beads, magnetic beads, colloid metal particles, glass and/or silicon chips and surfaces, nitrocellulose strips, membranes, sheets, duracytes, wells and walls of reaction trays, plastic tubes etc. The ligand or agent may be bound to many different carriers. Examples of well-known carriers include glass, polystyrene, polyvinyl chloride, polypropylene, polyethylene, polycarbonate, dextran, nylon, amyloses, natural and modified celluloses, polyacrylamides, agaroses, and magnetite. The nature of the carrier can be either soluble or insoluble for the purposes of the invention. Suitable methods for fixing/immobilizing said ligand are well known and include, but are not limited to ionic, hydrophobic, covalent interactions and the like. It is also contemplated to use "suspension arrays" as arrays according to the present invention (Nolan 2002, Trends Biotechnol. 20(1):9-12). In such suspension arrays, the carrier, e.g. a microbead or microsphere, is present in suspension. The array consists of different microbeads or microspheres, possibly labeled, carrying different ligands. Methods of producing such arrays, for example based on solid-phase chemistry and photo-labile protective groups, are generally known (US 5,744,305).

The method of the present invention comprises determining the amounts of gene product of at least the genes coding for HYAL2. Said gene and its preferred products are known to the skilled person and the respective sequences have been deposited in databases; relevant accession numbers and SEQ ID NOs are Genbank Acc No: NM_003773.4 GI:289802998, SEQ ID NO:2 for transcript variant 1, Genbank Acc No: NM_033158.4 GI:289802999, SEQ ID NO:3; Genbank Acc No: NP_003764.3 GI:15022801, SEQ ID NO:4 for the HYAL2 polypeptide encoded by transcript variant 1, and Genbank Acc No: NP_149348.2 GI:34304377, SEQ ID NO:5 for the HYAL2 polypeptide encoded by transcript variant 2. It is understood by the skilled person that the gene products are referenced as biomarkers, not as specific polynucleotides or polypeptides. Accordingly, the aforementioned polynucleotides and polypeptides having the specific sequences deposited under the Genbank accession numbers are to be understood as exemplary sequences representing a biomarker. Encompassed as gene products according to the present invention are also variant polynucleotides which vary due to at least one nucleotide addition, substitution and/or deletion from the polynucleotide having the specific sequence as

long as they are also suitable as biomarkers for expression of one of the HYAL2 gene as discussed above. Preferably, the variant polynucleotides are at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98% or at least 99% identical to the specific polynucleotides. The term "identical" as used herein refers to sequence identity characterized by determining the number of identical nucleotides between two nucleic acid sequences or amino acid sequences wherein the sequences are aligned so that the highest order match is obtained. It can be calculated using published techniques or methods codified in computer programs such as, for example, BLASTP, BLASTN or FASTA (Altschul 1990, J Mol Biol 215, 403). The percent identity values are, in one aspect, calculated over the entire nucleic acid or amino acid sequence. A series of programs based on a variety of algorithms is available to the skilled worker for comparing different sequences. In this context, the algorithms of Needleman and Wunsch or Smith and Waterman give particularly reliable results. To carry out the sequence alignments, the program PileUp (Higgins 1989, CABIOS 5, 151) or the programs Gap and BestFit (Needleman 1970, J Mol Biol 48; 443; Smith 1981, Adv Appl Math 2, 482), which are part of the GCG software packet (Genetics Computer Group 1991, 575 Science Drive, Madison, Wisconsin, USA 53711), may be used. The sequence identity values recited above in percent (%) are to be determined, in another aspect of the invention, using the program GAP over the entire sequence region with the following settings: Gap Weight: 50, Length Weight: 3, Average Match: 10.000 and Average Mismatch: 0.000, which, unless otherwise specified, shall always be used as standard settings for sequence alignments. If a variant polynucleotide is suitable as a biomarker for expression of one of the genes can be assessed by determining according to the methods specified herein if the variant polynucleotide has essentially the same expression pattern as the biomarker it is a variant of. Also encompassed according to the present invention are variant polypeptides which vary due to at least one amino acid addition, substitution and/or deletion from the polypeptide having the specific sequence as long as they are also suitable as biomarkers for expression of one of the genes as discussed above. Preferably, the variant polypeptides are at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98% or at least 99% identical to the specific polypeptides. The term "identical" as used herein refers to sequence identity characterized by determining the number of identical amino acids between two nucleic acid sequences or amino acid sequences according to the methods specified herein above.

The sensitivity of mammography in women aged over 50 has been estimated to range from 68% to over 90%, with most trials and programmes achieving sensitivities of around 85%. In women aged 40-49 the sensitivity has been reported to be lower, with estimates between 62% and 76%. The specificity of breast screening by mammography ranges between 82% and 97% (IARC, Screening Techniques. IARC Handbooks of Cancer Prevention: Breast Cancer Screening, ed. B.F. Vainio H. Vol. 7. 2002, Lyon: IARC Press). Using HYAL2 methylation levels in peripheral blood, different levels of sensitivity and specificity can be reached. According to the ROC

analysis, Sensitivity of 10 %, specificity of 99 % (cut off of cg27091787 methylation level at 0.42); sensitivity of 25 %, specificity of 97 % (cut off of cg27091787 methylation level at 0.46); sensitivity of 50 %, specificity of 95 % (cut off of cg27091787 methylation level at 0.50); sensitivity of 75 %, specificity of 90 % (cut off of cg27091787 methylation level at 0.54);
5 sensitivity of 90 %, specificity of 75 % (cut off of cg27091787 methylation level at 0.59); sensitivity of 95 %, specificity of 66 % (cut off of cg27091787 methylation level at 0.61); sensitivity of 97 %, specificity of 58 % (cut off of cg27091787 methylation level at 0.64); sensitivity of 99 %, specificity of 30 % (cut off of cg27091787 methylation level at 0.70). Sensitivity of 84 % and specificity of 84 % (cut off of cg27091787 methylation level at 0.56) can
10 be reached compared to the mammography. Further, since the young healthy women have even higher HYAL2 methylation (age-related HYAL2 methylation in controls but not in cases), better estimation of breast cancer risk in young people can be expected, whereas the mammography has lower sensitivity in younger women.

- 15 The present invention further relates to a device for diagnosing cancer, comprising
- (a) an analyzing unit comprising
 - (i) a detection agent for determining the methylation status of at least one of the CpG sites of claims 1 to 4 in a sample of a subject, or comprising
 - (ii) a detection agent for determining the amount of a HYAL2 gene product in a
20 sample of a subject suspected to suffer from breast cancer; and
 - (b) an evaluation unit comprising a data processor having tangibly embedded an algorithm for carrying out a comparison of the amount determined by the analyzing unit with a stored reference and which is capable of generating an output file containing a diagnosis established based on the said comparison.

25 The term “device” as used herein relates to a system of means comprising at least the aforementioned means operatively linked to each other as to allow the diagnosis. Preferred means for determining the methylation status or the amount of said HAYL2 gene product and means for carrying out the comparison are disclosed above in connection with the methods of the
30 invention. How to link the means in an operating manner will depend on the type of means included into the device. For example, where means for automatically determining the methylation status or the amount of said HAYL2 gene product are applied, the data obtained by said automatically operating means can be processed by, e.g., a computer program in order to establish a diagnosis. Preferably, the means are comprised by a single device in such a case. Said
35 device may accordingly include an analyzing unit for determining the methylation status or the amount of said HAYL2 gene product in a sample and an evaluation unit for processing the resulting data for the diagnosis. Preferred means for detection are disclosed in connection with embodiments relating to the methods of the invention above. In such a case, the means are operatively linked in that the user of the system brings together the result of the determination of

the amount and the diagnostic value thereof due to the instructions and interpretations given in a manual. The means may appear as separate devices in such an embodiment and are, preferably, packaged together as a kit. The person skilled in the art will realize how to link the means without further inventive skills. Preferred devices are those which can be applied without the particular knowledge of a specialized clinician, e.g., test stripes or electronic devices which merely require loading with a sample. The results may be given as output of parametric diagnostic raw data, preferably, as absolute or relative amounts. It is to be understood that these data will need interpretation by the clinician. However, also envisaged are expert system devices wherein the output comprises processed diagnostic raw data the interpretation of which does not require a specialized clinician. Further preferred devices comprise the analyzing units/devices (e.g., biosensors, arrays, solid supports coupled to ligands specifically recognizing the polypeptides, Plasmon surface resonance devices, NMR spectro-meters, mass- spectrometers etc.) or evaluation units/devices referred to above in accordance with the methods of the invention.

The present invention also relates to a kit for carrying out a method for diagnosing breast cancer, wherein said kit comprises instructions for carrying out said method, a detection agent for a HYAL2 gene product in a sample of a subject suspected to suffer from breast cancer, and standards for a reference.

The term "kit" as used herein refers to a collection of the aforementioned components, preferably, provided separately or within a single container. The container, also preferably, comprises instructions for carrying out the method of the present invention. Examples for such the components of the kit as well as methods for their use have been given in this specification.

The kit, preferably, contains the aforementioned components in a ready-to-use formulation. Preferably, the kit may additionally comprise instructions, e.g., a user's manual for adjusting the components, e.g. concentrations of the detection agents, and for interpreting the results of any determination(s) with respect to the diagnoses provided by the methods of the present invention. Particularly, such manual may include information for allocating the amounts of the determined HYAL2 gene product to the kind of diagnosis. Details are to be found elsewhere in this specification. Additionally, such user's manual may provide instructions about correctly using the components of the kit for determining the amount(s) of the respective biomarker. A user's manual may be provided in paper or electronic form, e.g., stored on CD or CD ROM. The present invention also relates to the use of said kit in any of the methods according to the present invention.

The present invention further relates to a method for therapy monitoring in a subject being treated against cancer, comprising the steps of (a) obtaining a first and a second sample from a subject, wherein the first sample is obtained at a time point before the second sample, (b)

determining in said first and second sample the methylation status of at least one of the CpG sites of the present invention; and (c) comparing the methylation status of said first sample to the methylation status of said second sample, thereby monitoring therapy in said subject.

5 As used herein, the term "therapy monitoring" relates to obtaining an indication on the effect of a treatment against cancer on the cancer status of a subject afflicted with said cancer. Preferably, therapy monitoring comprises application of a method of the present invention on two samples from the same subject, wherein a first sample is obtained at a time point before the second sample. Preferably, the time point of obtaining the first sample is separated from the time point
10 of obtaining the second sample by about one week, about two weeks, about three weeks, about four weeks, about five weeks, about six weeks, about seven weeks, about two months, about three months, about five months, about six months, or more than about six months. It is, however, also envisaged by the present invention that the method of therapy monitoring is used for long-term monitoring of subjects, e.g. monitoring the time of relapse-free survival or the like. In such case,
15 the time point of obtaining the first sample is separated from the time point of obtaining the second sample, preferably, by at least six months, at least one year, at least two years, at least three years, at least four years, at least five years, or at least six years. It is known to the person skilled in the art that the first sample is preferably obtained before cancer therapy is started, while the second sample is preferably obtained after therapy is started. It is, however, also
20 envisaged by the present invention that both samples are obtained after therapy is started. The skilled artisan also understands that more than two successive samples may be obtained according to the method for therapy monitoring of the present invention and that in such case the sample obtained at the first point in time may be used as the first sample relative to the second sample as well as for a third sample. Mutatis mutandis, the sample obtained at the second point
25 in time may nonetheless be used as a first sample relative to a third sample, and the like.

The term "comparing", as used herein, relates to bringing the methylation status of two samples into relation, either directly or indirectly. Directly comparing two samples relates to determining if the methylation status in the first sample is higher than in the second sample (i.e. the
30 methylation status decreases), if the methylation status of the first sample is lower than the methylation status of the second sample (i.e. the methylation status increases), or if it is the same for both samples (i.e. the methylation status is unchanged). In such case, a decrease in methylation status or an unchanged methylation status are indicative of a cancer therapy not being effective, while an increasing methylation status is indicative of a cancer therapy being
35 effective. Indirectly comparing two samples relates to comparing both samples to a reference value as described herein above. In such case, the first sample being indicative of a subject being afflicted with cancer and the second sample being indicative of said subject not being afflicted with cancer, is indicative of a cancer therapy being effective. Conversely, the first and the second

sample being indicative of a subject being afflicted with cancer is indicative of a cancer therapy not being effective.

5 All references cited in this specification are herewith incorporated by reference with respect to their entire disclosure content and the disclosure content specifically mentioned in this specification.

10 Figure Legends:

Figure 1. Amplicon and primer design for HYAL2 methylation analysis.

Panel A shows sequence of three HYAL2 amplicons examined by MassArray covering the main locus cg27091787 and the adjacent CpG island (chr3:50334844-50335046, build 36.1/hg18, as
15 defined by the UCSC Genome Browser and marked by dotted underline). The main-hit amplicon (HYAL2) for the cg27091787 locus covers the region between the fifth and sixth pairs of vertical lines. Amplicon HYAL2-is-310 covers the region between the third and fourth pairs of vertical lines. Amplicon HYAL2-is-325 covers the region between the first and second pairs of vertical lines. In together, MassArray assay determined methylation levels of 32 CpGs in the three
20 amplicons, and yielded 26 distinguishable peaks. The CpG sites that could be measured are underlined. All the four SNPs located in the main-hit amplicon are in italic; cg27091787 is underlined, italic and bold. Panel B provides the list of Bisulfite-specific primers for amplicons. Uppercase letters indicate the sequence specific primer regions, and non-specific tags are shown in lower case letters. There are no SNPs located in the primer regions, and none of the covered
25 CpGs in the three amplicons are overlapped with known SNPs.

Figure 2. Significant 2,125 methylation loci comparing FBC cases and healthy controls in normalized data from Illumina 27K Array ($p < 0.05$). Data are presented as $-\log_{10}(p)$ versus
30 methylation intensity difference between FBC cases and controls. The cycle indicates the most significant loci cg27091787. The line indicates the threshold of $p = 0.05$.

Figure 3. cg27091787 methylation in the discovery and further verification rounds.

Panel A shows location of the four measured CpG loci in the main-hit amplicon of HYAL2. Panel B shows methylation levels of cg27091787 locus in the discovery round. Methylation
35 levels of cg27091787 in the same sample measured by Illumina and MassArray separately are presented in X-axis and Y-axis respectively. Panel C and D show methylation levels of the four CpG loci in the main-hit amplicon of HYAL2 in the replication round and the third verification round by MassArray, respectively. The dots and box plot indicate the distribution of methylation levels of each CpG locus.

Figure 4. The definition of BC-associated methylation region of HYAL2.

Panel A shows a schematic diagram of cg27091787, CpG islands and designed amplicons in HYAL2. There are two CpG islands (CpG island-1 and CpG island-2) located before the translation start site (TSS) of HYAL2. Main-hit amplicon HYAL2 covers the CpG cg27091787, whereas amplicon HYAL2-is-310 covers the region between the main-hit amplicon and the CpG island-1 and amplicon HYAL2-is-325 covers most of the region of the CpG island-1. The distances between cg27091787 and the CpG islands and TSS are marked in the figure. Panel B shows the methylation levels of all the 32 measureable CpG loci in the HYAL2 main-hit amplicon, amplicon HYAL2-is-310 and amplicon HYAL2-is-325 determined by MassArray in 96 FBC cases and 96 healthy controls. The means of methylation levels and error of each CpG locus in cases and controls are presented separately as the two lines with error bars. The differences of average methylation levels between cases and controls in each CpG locus are presented as the columns at the bottom. * indicates p-value < 0.05 (by logistic regression). The three short lines at the bottom indicate the covering regions of each amplicon.

Figure 5. The inverse correlation between the methylation levels of cg27091787 and HYAL2 expression.

The methylation level of cg27091789 CpG locus and the expression of HYAL2 in the leukocyte from peripheral blood were quantified by MassArray and real-time PCR respectively. The Spearman rho and p-value for the correlation between cg27091787 methylation levels and HYAL2 expression are presented.

Figure 6. Decreased cg27091787 methylation in T cells of sporadic BC cases.

The methylation levels of cg27091787 were measured in triplicates in the DNA of whole blood and different cell populations in leucocytes from 7 sporadic BC case and 14 healthy controls. The t-test was performed to analyse the methylation difference between cases and controls. The methylation levels of cg27091787 were presented by box and whisker plot. The circle indicates an outlier.

Figure 7. ROC analysis for BC risk evaluation.

Panel A and B show the logistic regression model-based ROC analysis for the discriminatory power of cg27091787 methylation to distinguish FBC cases or sporadic BC cases from healthy controls.

Figure 8. Methylation level of cg27091787 is related with age in controls but not in BC cases.

Panel A-C show cg27091787 methylation levels and the age distributions in all the controls, in FBC cases and in sporadic BC cases, respectively. The red line shows the regression line, the

two black lines represent 95 % prediction intervals. The regression equation and p-value by logistic regression is presented on each panel.

Figure 9. Schematic diagram of *HYAL2* gene and investigated CpG sites by Illumina 450K array.

- 5 Four transcript variants of *HYAL2* gene are presented. The second exon could also be an intron due to variant splicing. CpG island-1 and CpG island-2 are located before the translation start site. CpG island-3 is located after translation start site in the third exon. The HYAL2-A amplicon covers the CpG cg27091787, HYAL2-B amplicon covers the region between HYAL2-A amplicon and the CpG island-1, and HYAL2-C amplicon covers most region of the CpG island-1. 26 CpG sites investigated by 450K are labeled by short lines (In together 27 CpG sites were investigated by 450K, cg24335984 was located at the position of 50336558 which is 1kb upstream of cg27091787, and thus not been labeled out in this figure). The CpG sites which showed significant difference between sporadic BC cases and controls by 450K are in short bars, whereas the non-significant CpG sites are in short dashed bars. The most significant hit by 27K, cg27091787 is pointed out by the arrow.
- 10
- 15

The following Examples shall merely illustrate the invention. They shall not be construed, whatsoever, to limit the scope of the invention.

20 METHODS

Example 1: SAMPLES

This study was approved by the Ethics Committee of the University of Heidelberg (Germany). All the BC cases and controls shared the same ethnic background (Caucasian/German population) and gave informed consent for the study. The peripheral blood was kept at 4°C up to 24h after collection and stored at -20°C or -80°C before use. DNA was isolated from peripheral blood using the DNA isolation kits from Qiagen. DNA and RNA were isolated from leucocytes and different cell populations in leucocytes using AllPrep DNA/RNA/Protein Mini Kit from Qiagen. Details for all the samples are shown in Table 1.

25

30 BC Cases: Blood samples of 410 BRCA1/2 mutation-negative index familial breast cancer (FBC) patients were collected after operation by the centers of the German Consortium for Hereditary Breast and Ovarian Cancer in Heidelberg and Cologne. The blood samples of 232 sporadic BC were collected at the time point of first BC diagnosis before any BC treatment and operation at the University Hospital of Heidelberg. Controls: Blood samples were collected from 720 healthy and unrelated female blood donors by the German Red Cross Blood Service of Baden-Württemberg-Hessen (Mannheim, Germany). Controls were randomly selected and no further inclusion criteria were applied during recruitment.

35

Leucocytes: peripheral blood was collected freshly from 36 sporadic BC patients and 40 female healthy blood donors at the University Hospital of Heidelberg. Leucocytes were isolated using red blood cell lysis buffer as described previously (Alan H L, Andrew R C, Davor B, John F, Seng S, Vinayakumar S. A method for ameliorating autoimmune disease by passive transfer of IVIg-primed leukocytes. 2006.) Blood cell fractionation: leucocytes were obtained as described from 7 sporadic BC patients and 14 female healthy blood donors at the University Hospital of Heidelberg. First, B cells were positively isolated using a Dynal® CD19 positive isolation kit (Invitrogen). Subsequently these B cell-depleted leucocytes were used for T cell purification with a Dynal® CD3 positive isolation kit (Invitrogen). Then the leftover cells were collected as 'B/T-lymphocytes-depleted leucocytes'. The cell pellets were immediately frozen in liquid nitrogen after purification and kept at -80°C before use.

Example 2: INFINIUM 27K METHYLATION ASSAY

In the discovery round, 500 ng genomic DNA from each sample was treated by EZ-96 DNA Methylation Kit (Zymo Research) for bisulfite conversion and subjected to genome-wide methylation screening by Human Methylation27 BeadChip (Illumina) according to the manufacturer recommendations (Steemers FJ, Chang W, Lee G, Barker DL, Shen R, Gunderson KL. Whole-genome genotyping with the single-base extension assay. Nat Methods 2006;3:31-3. Bork S, Pfister S, Witt H, et al. DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. Aging Cell 2009;9:54-63). All the 96 samples passed the quality control according to manufacturer instructions.

Example 3: METHYLATION ANALYSIS VIA MALDI-TOF MASS SPECTROMETRY

MALDI-TOF mass spectrometry (Sequenom) described by Breitling et al. (Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. Am J Hum Genet 2011;88:450-7.) was used in various verification rounds. DNA was bisulfite converted by EZ-96 DNA Methylation Gold Kit (Zymo Research) and amplified by bisulfite-specific primers (Fig.1). The PCR products were treated according to the standard protocol of Sequenom EpiTyper Assay and dispensed to a 384 SpectroCHIP by a Nanodispenser. The chips were read by a Sequenom Mass Spectrometer system. Data were collected by SpectroACQUIRE v3.3.1.3 software and visualized with MassArray EpiTyper v1.0 software. 5% samples were randomly chosen for the duplication analysis of cg27091787 methylation and achieved a Spearman correlation larger than 0.8.

Example 4: QUANTITATIVE REAL-TIME PCR FOR RNA EXPRESSION

100 ng of total RNA from each sample was transcribed to cDNA by TaqMan® Reverse Transcription Reagents (Applied Biosystems). Quantitative real-time PCR was performed using a LightCycler 480 (Roche) in combination with TaqMan gene expression assays (Applied Biosystems) for HYAL2 gene and housekeeping gene HPRT1 (hypoxanthine-guanine

phosphoribosyltransferase1) as endogenous control. Crossing point values were calculated using the second-derivative maximum method by the LightCycler 480 basic software (Roche). Relative expression of HYAL2 for each sample was calculated according to $\Delta\Delta C_t$ method (Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2(-\Delta\Delta C(T))$ Method. Methods 2001;25:402-8) by normalization to HPRT1.

Example 5: STATISTICAL ANALYSIS

The Illumina 27K Array data were processed by Illumina BeadStudio software with default settings. Probes with detection p-value > 0.01 were removed and samples were quantile-normalized. Association of probes with case/control status was assessed by beta regression models with logistic link from R package betareg v2.2-3 (Simas AB, Barreto-Souza W, Rocha AV. Improved Estimators for a General Class of Beta Regression Models. Computational Statistics and Data Analysis 2010;54(2):348-66.) and Wald tests. To remove false hits due to chip confounding, likelihood ratio (LR) tests were used to compare the case/control model with the nested model for chip differences, resulting in a reduced list of 2,125 hits after removing all hits with significant LR test results. Multiple testing adjustments were done with the Benjamini-Hochberg method controlling the false discovery rate at 0.05. All analysis was performed with the statistical software R v2.11.

MassArray data were all processed by SPSS Statistics 17.0 software. Logistic regression model and nonparametric tests (Mann-Whitney U test) were used for the comparisons between two groups. ROC analysis was performed to assess the discriminatory power of methylation levels. The age of onset and different bathes for measurement were adjusted in logistic regression when necessary.

Example 6: DISCOVERY ROUND: INITIAL INVESTIGATION BY ILLUMINA 27K ARRAY

In the discovery round, methylation profiles of 72 FBC cases and 24 controls were investigated by Illumina 27K Array (Table 1). Out of the 27,578 assayed CpG sites, 95.1% had complete data in all the 96 samples, 96.9% and 97.9% had ≤ 1 and ≤ 2 missing values, and only 0.5% had 10 or more missing values (detection p-values ≥ 0.01). Fig. 2 showed 2,125 methylation loci which had significant difference comparing FBC cases and controls ($p < 0.05$ after false discovery rate-based multiple testing adjustment using the method of Benjamin-Hochberg, see statistical analysis in the section of Methods). CpG site cg27091787 in HYAL2 showed by far the most significant difference (adjusted $p = 3.37 \times 10^{-15}$, Fig. 2). In our study, CpG loci in SFRP1, PTGS2 and BRCA1 showed lower methylation in BC patients compared to controls (adjusted $p < 0.05$), which is in agreement with the results from previous candidate gene approaches (Iwamoto T, Yamamoto N, Taguchi T, Tamaki Y, Noguchi S. BRCA1 promoter methylation in

peripheral blood cells is associated with increased risk of breast cancer with BRCA1 promoter methylation. *Breast Cancer Res Treat* 2011;129:69-77. Widschwendter M, Apostolidou S, Raum E, et al. Epigenotyping in peripheral blood cell DNA and breast cancer risk: a proof of principle study. *PLoS One* 2008;3:e2656.). We did not observe altered methylation in the ATM gene in BC patients (Flanagan JM, Munoz-Alegre M, Henderson S, et al. Gene-body hypermethylation of ATM in peripheral blood DNA of bilateral breast cancer patients. *Hum Mol Genet* 2009;18:1332-42; Brennan K, Garcia-Closas M, Orr N, et al. Intragenic ATM Methylation in Peripheral Blood DNA as a Biomarker of Breast Cancer Risk. *Cancer Res* 2012;72:2304-13), probably due to different investigating regions.

Example 7: DISCOVERY ROUND: VALIDATION OF THE ILLUMINA 27K RESULTS USING MASSARRAY

For validation, 70 FBC cases and 20 controls that had been investigated by the Illumina 27K Array in the discovery round were further examined by Sequenom MALDI-TOF mass spectrometry assay (Table 1). The main-hit amplicon of HYAL2 harboring cg27091787 site and three additional flanking CpG loci was amplified and analysed by MassArray (Fig. 1A, and Fig. 3). Methylation data from MassArray at the site referring to cg27091787 had a proper correlation with normalized Illumina 27K measurements (Fig. 3B and Table 2). In line with the Illumina 27K results, the MassArray data showed significantly different cg27091787 methylation levels between FBC cases and controls ($p = 5.66 \times 10^{-5}$ by logistic regression, Fig. 3B, and Table 2). Moreover, MassArray data presented a wider range of methylation levels than the Illumina 27K measurements (Fig. 3B), which is in agreement with the observation by Breitling et al.(Breitling LP, Yang R, Korn B, Burwinkel B, Brenner H. Tobacco-smoking-related differential DNA methylation: 27K discovery and replication. *Am J Hum Genet* 2011;88:450-7.). The methylation levels of the three additional CpG loci in the main-hit amplicon were highly correlated with the methylation levels of cg27091787 (Spearman $\rho \geq 0.58$ and $p < 2 \times 10^{-9}$ for all, Table 2).

Example 8: INDEPENDENT REPLICATION ROUND

For further verification, additional 338 FBC cases and 507 controls were investigated by MassArray (Table 1) and showed significant cg27091787 methylation difference between FBC cases and controls ($p = 1.38 \times 10^{-37}$ by logistic regression, Fig. 3C, and Table 3). An inter-quartile strategy (Flanagan JM, Munoz-Alegre M, Henderson S, et al. Gene-body hypermethylation of ATM in peripheral blood DNA of bilateral breast cancer patients. *Hum Mol Genet* 2009;18:1332-42.) was performed for the methylation analysis by taking the cg27091787 methylation levels of all detectable samples in the replication round into quartile calculation. The highest quartile (Q4: methylation level $> 69\%$, Table 4) was considered as reference. Compared to the highest quartile, all the other three quartiles showed highly significant difference between FBC cases and controls (Table 4). Especially, the lowest quartile (Q1: methylation level $\leq 54\%$)

was associated with a 41-folds (95% confidence interval [CI], 22.2 to 74.4) increased risk of BC referring to the highest quartile ($p = 1.43 \times 10^{-31}$ by logistic regression, Table 4).

In the independent replication round, the methylation levels of all the three additional CpG loci in the main-hit amplicon were highly correlated with the methylation levels of cg27091787 (all Spearman $\rho \geq 0.63$, Table 3). Moreover, the three additional CpG sites also showed lower methylation in FBC cases compared to controls (all $p < 1 \times 10^{-14}$ by logistic regression, Fig. 3C, and Table 3).

Example 9: THIRD VERIFICATION ROUND: CG27091787 METHYLATION IN SPORADIC BC

We further explored the methylation status of HYAL2 in 189 sporadic BC cases and another 189 controls by MassArray (Table 1). In order to exclude any effects of treatment, all the patient samples were obtained at the time point of BC diagnosis before any BC treatment and operation.

Similar to the comparison between FBC cases and controls, all the four CpG loci in the main-hit amplicon of HYAL2 showed significantly lower methylation in sporadic BC cases than in controls (all $p < 5 \times 10^{-9}$ by logistic regression, Figure 3D, and Table 5). We further performed an inter-quartile analysis and the highest quartile (Q4: methylation intensity $> 65\%$) was considered as reference (Table 6). Compared to the highest quartile, all the other quartiles showed significant difference between cases and controls (Table 6). Especially, the lowest quartile (Q1: methylation intensity $\leq 49\%$) was associated with a 133-fold (95% CI, 40.7 to 434.5) increased risk of BC referring to the highest quartile ($p = 5.73 \times 10^{-16}$ by logistic regression, Table 6). The methylation levels of the three additional CpG loci also showed high correlation with the methylation levels of cg27091787 in the third verification round (all Spearman $\rho \geq 0.55$), and had lower methylation levels in sporadic BC cases than in controls (Figure 3D, and Table 5).

Example 10: METHYLATION ANALYSIS FOR CG27091787 FLANKING CpG ISLAND

Two additional amplicons, HYAL2-is-310 and HYAL2-is-325, were designed to investigate the methylation status of the CpG loci located between the closest CpG island (CpG island-1) to the main-hit amplicon in 96 FBC cases and 96 healthy controls by MassArray (Table 1 and Fig. 4A). All the 11 measureable CpG loci in the amplicon HYAL2-is-310 showed significant methylation difference between cases and controls (all $p < 0.05$ by logistic regression), and 10 out of the 11 CpG were strongly correlated with cg27091787 (all Spearman $\rho > 0.50$). In contrast, most of the CpG loci in the amplicon HYAL2-is-325 showed no significant methylation difference between cases and controls, and had weaker correlation with cg27091787 (all Spearman $\rho < 0.50$) (Fig. 4B, and Table 7). Meanwhile, the ratio of methylation levels comparing cases and controls is lower in the main-hit amplicon and amplicon HYAL2-is-310 than in and amplicon

HYAL2-is-325 (Fig. 4B, and Table7). Thus, we defined a breast cancer related alternative HYAL2 methylation region, which extends about 650 bp, locating between cg27091787 and the beginning of the most closed CpG island.

5 Example 11: CG27091787 METHYLATION IS INVERSELY CORRELATED TO THE EXPRESSION OF HYAL2 IN LEUKOCYTES

The correlation between cg27091787 methylation and HYAL2 expression was further analysed in the leukocytes from an independent sample set comprising 36 sporadic BC patients and 40 healthy controls (Table 1). We observed inverse correlation between the expression of HYAL2
10 and methylation level of cg27091787 (Spearman rho = - 0.323, p = 0.006, Fig. 5 and Table 8), and proposed that the methylation level of cg27091787 affects the expression of HYAL2, and subsequently influences the HYAL2 involved pathways.

15 Example 12: DECREASED HYAL2 METHYLATION IN T CELLS AND Other NON-B-Zells OF BC CASES

In order to understand the origination of altered HYAL2 methylation in peripheral blood, we further investigated the HYAL2 methylation in different cell populations in leucocytes. In comparison with whole blood, the methylation of HYAL2 was higher in B and T lymphocytes, but lower in the other cell types (Fig. 6). Even with limited sample size (7 sporadic BC cases
20 versus 14 healthy controls, Table 1), significantly decreased cg27091787 methylation in whole blood and in T cells, slightly in B/T-lymphocytes-depleted leucocytes', but not in B cells, was observed in sporadic BC cases compared to controls (p < 0.05 by t-test, Fig. 6). Furthermore, in T cells, the methylation levels of all the three additional CpG loci in the HYAL2 main hit amplicon were strongly correlated with the methylation levels of cg27091787 (all Spearman rho
25 ≥ 0.70 , all p < 0.0005).

Example 13: HYAL2 METHYLATION IS ASSOCIATED WITH LYMPHNODE INVOLVEMENT AND STAGE IN BREAST CANCER

30 The methylation level of cg27091787 was also analysed referring to the clinical data of the sporadic breast cancer patients (samples from the third verification round). These patients were stratified by their hormone status, primary tumor size, lymphnode involvement status, stage, grading, ER (estrogen receptor) status, PR (progesterone receptor) status, Her2 (human epidermal growth factor receptor-2) status, and the level of KI-67, Bcl-2 and p53 (data not
35 shown). The status lymphnode involvement was stratified into two subgroups, less than or equal to three lymphnode involvement, and more than 3 lymphnode involvement. The subgroup with more than 3 lymphnode involvement had significantly lower methylation in two CpG loci (HYAL2_CpG_2 and HYAL2_CpG_3) of the HYAL2 main-hit amplicon than the other subgroups (all p < 0.005 by logistic regression, Table 9). The patient of Stage III showed lower

methylation levels in two CpG loci (HYAL2_CpG_2 and HYAL2_CpG_3) of the HYAL2 main-hit amplicon than the patients with other stages (all $p < 0.02$ by Kruskal Wallis Test, Table 9). The patients with large tumor size (T3 and T4) showed slightly lower methylation at CpG site HYAL2_CpG_3 of the HYAL2 main-hit amplicon than the other subgroups ($p = 0.024$ by Kruskal Wallis Test, Table 9). Considering the hormone status, patients with ER and PR negative showed higher methylation at CpG site HYAL2_CpG_1 ($p = 0.028$ and $p = 0.068$ respectively by logistic regression, Table 9) than the ER and PR positive patients, whereas Her2 positive patients showed no different methylation compared to the Her2 negative patients (Table 9). The other clinical factors, like menopause status, grading, family history showed no significant difference among subgroups (Table 9).

Example 14: BC RISK EVALUATION AND AGE RELATED METHYLATION OF CG27091787

By ROC analysis, we evaluated the discriminatory power of cg27091787 methylation in differentiating controls from FBC cases (all 397 FBC cases vs. all 696 controls, area under curve-AUC: 0.82) and sporadic BC cases (all 183 sporadic cases vs. 696 controls, AUC: 0.91) (Fig.7). Particularly, a sensitivity of 90% can be reached with a specificity of 75% when distinguishing sporadic BC cases and controls. In addition, we also observed a significant age-related decrease of cg27091787 methylation in controls (considering all the 696 controls, $p = 3.87 \times 10^{-11}$ by logistic regression), but neither in FBC cases nor sporadic BC cases (Fig. 8).

Example 15: HYAL2 methylation in peripheral blood as a promising marker for ovarian cancer diagnosis

SAMPLES

This study was approved by the Ethics Committee of University of Heidelberg (Germany). All the cases and controls share the same ethnic background and geographical region (Caucasian/Southwest German population) and gave informed consent for the study. The blood samples from 97 sporadic ovarian cancer patients and 149 healthy and unrelated female blood donors were collected by the university hospital of Heidelberg. The peripheral blood were freshly taken and centrifuged at 1300 g for 20 minutes. The blood pellet after centrifugation were taken for DNA isolation by the DNA isolation kits from Qiagen. Details for all the samples please see Table 10.

METHYLATION ANALYSIS VIA MALDI-TOF MASS SPECTROMETRY

Please refer to the previous description for the breast cancer study

STATISTICAL ANALYSIS

MassArray data were all processed by SPSS Statistics 17.0 software. Logistic regression model and nonparametric tests (Mann-Whitney U test) were used for the comparisons between two

groups. The age of onset and different bathes for measurement were adjusted in logistic regression when necessary.

RESULTS AND DISCUSSION

We investigated the methylation status of HYAL2 in 97 sporadic ovarian cases and 149 healthy controls by MassArray (Table 10). All the four CpG loci in the main-hit amplicon of HYAL2 showed significantly lower methylation in sporadic ovarian cases than in controls (all $p < 4 \times 10^{-5}$ by logistic regression, Table 11). Especially, the CpG site HYAL2_CpG_1 but not cg27091787 exert the largest difference between ovarian cancer cases and controls (Table 11). We further performed an inter-quartile analysis regarding the methylation level of HYAL2_CpG_1. The highest quartile (Q4: methylation intensity $\geq 35\%$) was considered as reference (Table 12). Compared to the highest quartile, the lowest quartile (Q1: methylation intensity $< 26\%$) was associated with a 35-fold (95% CI, 7.2 to 172.8) increased risk of ovarian cancer referring to the highest quartile ($p = 1.08 \times 10^{-5}$ by logistic regression, Table 12). Here Somehow, we didn't observe significantly increased risk of ovarian cancer in the other quartiles. The methylation levels of the three additional CpG loci in the main-hit amplicon of HYAL2 showed high correlation with the methylation levels of cg27091787 (all Spearman rho ≥ 0.68 , Table 11).

Therefore, the decreased methylation of HYAL2 in peripheral blood is not only in breast cancer patients but also in ovarian cancer patients, or maybe even in another cancer patients as well. Notably, the most significant CpG locus (HYAL2_CpG_1) distinguishing ovarian cancer patients from controls is different from the most significant CpG locus (cg27091787) distinguishing BC patients from controls. Thus, it is possible that the alternative methylation of HYAL2 might be tumor-specific through the association with different CpG locus.

Example 16:

METHOD

Illumina 450K methylation array was performed on the peripheral blood DNA from 48 sporadic breast cancer cases and 48 matched healthy controls. 27 CpG sites covering the HYAL2 gene have been analysed by the 450K array. Statistical analyses adjusted the covariants, like group, batch and age.

RESULTS

The 450K methylation array contains probes for 27 CpG sites in the HYAL2 gene, covering all three CpG islands, CpG island shore and some CpG loci in exons and introns (Figure 9). According to the results from the Illumina 450K methylation array, four CpG sites located at the CpG island shore (cg27091787, cg26460678, cg03051392 and cg00840516) and two CpG sites located at the C-terminal of HYAL2 gene (cg08776109 in the last intron and cg06721473 in the

3'UTR) showed significant differently lower methylation levels in sporadic BC cases than controls (pvalWald_Group < 0.01 and adjWaldPval < 0.1, Table 10).

These results are consistent with our previous observation concluded from 27K methylation array and following validation by sequenom. The 450K data confirmed that methylation levels of cg27091787, cg26460678 (HYAL_CpG_3) and cg03051392 (HYAL-is-310_CPG_2) are associated with breast cancer. Of note, the 450K data indicate a significant methylation difference in the CpG sites located at the C-terminal of HYAL2 gene (cg08776109 in the last intron and cg06721473 in the 3'UTR) between BC cases and controls (pvalWald_Group < 0.001 and adjWaldPval < 0.05, Table10).

Tables:

Table 1. Description of analysed samples in different rounds

Case-control association study									
Rounds	Sample types	Assays	Groups	Target N	Assayed N	call rate	mean of age	median of age	range of age
Discovery round	Peripheral blood DNA from FBC cases and controls	Illumina 27K assay	control	24	24	100.00%	48.0	44	32-66
			case	72	72	100.00%	47.0	46	26-78
		MassARRAY	control	20	20	100.00%	47.0	42	32-66
			case	70	69	98.57%	47.2	46	26-78
Independent replication round	Peripheral blood DNA from FBC cases and controls	MassARRAY	control	507	491	96.84%	44.7	42	30-67
			case	338	328	97.04%	46.1	46	24-77
Third verification round	Peripheral blood DNA from sporadic BC cases and controls	MassARRAY	control	189	185	97.88%	61.2	64	31-69
			case	189	183	96.83%	59.6	60	32-87
Further exploration of <i>HYAL2</i> methylation									
Rounds	Sample types	Assays	Groups	Target N	Assayed N	call rate	mean of age	median of age	range of age
CpG island of <i>HYAL2</i>	Peripheral blood DNA from FBC cases and controls (a subset of samples from discovery and replication rounds)	MassARRAY	control	96	95	98.96%	46.0	43	30-67
			case	96	96	100.00%	48.0	47	24-78
Correlation between <i>HYAL2</i> methylation and expression	DNA and RNA in leukocytes from sporadic BC case and controls (independent samples)	MassARRAY, real-time PCR	control	40	38	95.00%	38.2	39	21-63
			case	36	34	94.44%	58.2	57	29-81
<i>HYAL2</i> methylation in different blood cell types	DNA from different blood cell types of sporadic BC cases and controls (independent samples)	MassARRAY	control	14	14	100.00%	38.3	42	21-51
			case	7	7	100.00%	50.3	49	34-76

Table 2. The analysis of the main-hit amplicon in the discovery round by MassArray

		cg27091787 (Illumina after normalization)	HYAL2_CpG_1	HYAL2_CpG_2	HYAL2_CpG_3	cg27091787 (MassArray)
mean	control	0.56	0.41	0.24	0.41	0.66
	assayed N in controls	24	20	20	19	20
	case	0.46	0.34	0.17	0.33	0.51
	assayed N in cases	72	69	69	67	69
	absolute difference (case-control)	-0.11	-0.08	-0.07	-0.08	-0.15
	relative difference (case/control)	81.2%	81.3%	69.9%	80.7%	77.8%
p-value	adjusted P (BH)	3.37E-15	—	—	—	—
	logistic regression *	—	0.211	0.411	3.25E-04	5.66E-05
	Mann-Whitney U	—	0.006	0.001	5.63E-05	3.79E-08
Correlations to		Spearman rho	0.634	0.584	0.746	1.000
cg27091787 in MassArray		p-value	2.63E-11	1.90E-09	4.35E-11	1.76E-16

* adjusted for age of onset and different batches for the measurements

Table 3. The analysis of the main-hit amplicon in the independent replication round by MassArray

		HYAL2_CpG_1	HYAL2_CpG_2	HYAL2_CpG_3	cg27091787
mean	control	0.40	0.26	0.41	0.66
	assayed N in controls	491	490	480	491
	case	0.32	0.19	0.32	0.54
	assayed N in cases	328	326	321	328
	abosulte difference (case-control)	-0.08	-0.07	-0.09	-0.11
	relative difference (case/control)	79.4%	72.9%	77.7%	82.7%
p-value	logistic regression *	9.78E-15	2.54E-23	1.69E-30	1.38E-37
	Mann-Whitney U	2.84E-35	4.98E-31	4.03E-44	3.98E-53
Correlations to cg27091787	Spearman rho	0.648	0.628	0.703	1.000
	p-value	8.75E-99	8.85E-91	2.71E-120	—

* adjusted for age of onset and different batches for the measurements

Table 4. Inter-quartile analysis of cg27091787 methylation in the independent replication round

	Control (n= 491)	Case (n= 328)	OR (95 % CI) *	p-value *
Q4 (69 % - 90 %)	161 (32.8 %)	19 (5.8 %)	1	—
Q3 (61 % - 69 %)	173 (35.2 %)	45 (13.7 %)	2.42 (1.33-4.40)	0.0038
Q2 (54 % - 61 %)	114 (23.2 %)	96 (29.3 %)	8.21 (4.63-14.57)	6.10E-13
Q1 (25 % - 54 %)	43 (8.76 %)	168 (51.2 %)	41.47 (22.21-74.43)	1.43E-31

* logistic regression, adjusted for age of onset and different batches for the measurements

Table 5. The analysis of the main-hit amplicon in the third verification round by MassArray

		HYAL2_CpG_1	HYAL2_CpG_2	HYAL2_CpG_3	cg27091787
mean	control	0.41	0.24	0.41	0.64
	assayed N in controls	185	185	184	185
	case	0.33	0.17	0.32	0.50
	assayed N in cases	183	183	183	183
	abosulte difference (case-control)	-0.08	-0.07	-0.09	-0.13
	relative difference (case/control)	81.4%	70.5%	78.3%	79.2%
p-value	logistic regression *	4.71E-09	5.27E-15	1.51E-17	4.80E-23
	Mann-Whitney U	2.85E-11	1.55E-19	1.69E-23	9.30E-37
Correlations to cg27091787	Spearman rho	0.551	0.649	0.693	1.000
	p-value	1.40E-30	2.31E-45	7.89E-54	—

* adjusted for age of onset and different batches for the measurements

Table 6. Inter-quartile analysis of cg27091787 methylation in the third verification round

	Control (n= 185)	Case (n= 183)	OR (95 % CI) *	p-value *
Q4 (65 % - 89 %)	78 (42.2 %)	4 (2.2 %)	1	—
Q3 (56 % - 65 %)	72 (38.9 %)	25 (13.7 %)	6.81 (2.25-20.58)	0.0007
Q2 (49 % - 56 %)	21 (11.4 %)	75 (41.0 %)	76.03 (24.6-234.96)	5.34E-14
Q1 (31 % - 49 %)	14 (7.6 %)	79 (43.2 %)	132.98 (40.70-434.54)	5.73E-16

* logistic regression, adjusted for age of onset and different batches for the measurements

Table 7. The analysis of methylation in *HYAL2* CpG island by MassArray

		HYAL2-is- 325_CpG_1	HYAL2-is- 325_CpG_2	HYAL2-is- 325_CpG_3	HYAL2-is- 325_CpG_4	HYAL2-is- 325_CpG_5	HYAL2-is- 325_CpG_6
mean	control	0.65	0.54	0.63	0.63	0.73	0.63
	assayed N in controls	94	91	94	94	94	94
	case	0.61	0.53	0.61	0.61	0.72	0.61
	assayed N in cases	94	89	94	94	94	94
	absolute difference (case-control)	-0.04	-0.01	-0.02	-0.02	-0.01	-0.02
	relative difference (case/control)	94.2%	98.8%	97.4%	97.4%	98.8%	97.4%
p-value	logistic regression *	0.007	0.623	0.113	0.113	0.505	0.113
	Mann-Whitney U	0.003	0.637	0.019	0.019	0.533	0.019
Correlations to	Spearman rho	0.39	0.25	0.36	0.36	0.25	0.36
	p-value	2.78E-08	0.001	5.88E-07	5.88E-07	4.98E-04	5.88E-07

* logistic regression, adjusted for age of onset and different batches for the measurements

HYAL2-is- 325_CpG_7	HYAL2-is- 325_CpG_8	HYAL2-is- 325_CpG_9.10	HYAL2-is- 325_CpG_11	HYAL2-is- 325_CpG_12.13	HYAL2-is- 325_CpG_14.1	HYAL2-is- 325_CpG_16.17	HYAL2-is- 310_CpG_1
0.40	0.82	0.63	0.73	0.89	0.81	0.75	0.65
93	94	94	94	94	94	94	94
0.39	0.78	0.64	0.70	0.86	0.79	0.72	0.62
91	94	94	94	94	94	94	95
-0.01	-0.04	0.01	-0.03	-0.02	-0.02	-0.02	-0.03
97.6%	95.1%	102.1%	96.2%	97.4%	97.6%	96.7%	95.2%
0.291	0.007	0.484	0.010	0.258	0.103	0.011	0.017
0.656	0.006	0.669	0.002	0.045	0.020	0.001	3.77E-06
0.29	0.41	0.20	0.46	0.27	0.38	0.35	0.53
8.58E-05	6.24E-09	0.007	2.70E-11	1.98E-04	8.25E-08	1.15E-06	3.92E-15

5

HYAL2-is- 310_CpG_2	HYAL2-is- 310_CpG_3.4	HYAL2-is- 310_CpG_5.6	HYAL2-is- 310_CpG_7	HYAL2-is- 310_CpG_8	HYAL2-is- 310_CpG_9	HYAL2-is- 310_CpG_10	HYAL2-is- 310_CpG_11
0.55	0.63	0.78	0.56	0.65	0.36	0.63	0.54
94	94	94	94	94	94	94	94
0.49	0.60	0.71	0.50	0.62	0.32	0.58	0.47
95	95	95	95	95	95	95	95
-0.07	-0.03	-0.06	-0.06	-0.03	-0.04	-0.05	-0.06
88.1%	95.0%	92.2%	89.1%	95.2%	88.6%	92.3%	88.6%
2.22E-05	8.92E-05	3.19E-07	3.66E-05	0.017	0.001	3.65E-04	1.27E-06
2.62E-06	5.08E-06	5.23E-08	3.54E-06	3.77E-06	1.49E-04	6.51E-05	6.27E-08
0.62	0.57	0.65	0.57	0.53	0.48	0.63	0.60
3.45E-21	3.76E-17	1.30E-23	2.21E-17	3.92E-15	3.16E-12	2.57E-22	9.51E-20

HYAL2_C pG_1	HYAL2_Cp G_2	HYAL2_Cp G_3	cg27091787
0.38	0.24	0.41	0.65
93	93	92	93
0.34	0.17	0.33	0.52
95	95	93	95
-0.04	-0.07	-0.08	-0.13
88.4%	71.4%	81.1%	80.0%
0.013	8.33E-08	3.06E-08	2.23E-12
7.17E-04	1.31E-10	1.39E-11	8.06E-20
0.58	0.69	0.77	1.00
1.45E-18	4.84E-28	3.84E-38	—

Table 8. The correlation between HYAL2 mRNA expression and methylation in the main-hit amplicon

		HYAL2_CpG_1	HYAL2_CpG_2	HYAL2_CpG_3	cg27091787	Relative HYAL2 mRNA expression (HYAL2/HPRT1) *
mean	control	0.39	0.21	0.37	0.56	0.15
	assayed N in controls	38	38	38	38	38.00
	case	0.34	0.17	0.32	0.51	0.20
	assayed N in cases	34	33	34	34	34.00
	absolute difference (case-control)	-0.05	-0.04	-0.05	-0.05	0.05
	relative difference (case/control)	87.2%	81.1%	87.8%	90.6%	132.7%
p-value	logistic regression	0.238	0.009	0.023	0.001	0.098
	Mann-Whitney U	0.009	0.001	4.59E-04	4.15E-04	0.024
Correlations to HYAL2 expression	Spearman rho	-0.415	-0.278	-0.452	-0.323	1.000
	p-value	2.87E-04	0.019	6.60E-05	0.006	—

* the relative expression of HYAL2 was presented as the folds to the endogenous control HPRT1

Table 9. The HYAL2 methylation in sporadic BC patients with different clinical status

Status (N)	Group (N) and p-values	Mean of age	HYAL2_CpG	HYAL2_CpG	HYAL2_CpG	cg27091787
Menopause status (176)	perimenopause (54)	46.54 ± 0.71	0.32	0.17	0.33	0.49
	postmenopause (122)	66.16 ± 0.72	0.33	0.17	0.32	0.51
	p-value (logistic regression *)	1.03E-06	0.336	0.846	0.922	0.156
	p-value (Mann-Whitney U)	1.80E-24	0.646	0.730	0.479	0.139
Tumor size (183)	Tis (12)	57.02 ± 3.20	0.37	0.16	0.34	0.50
	T1 (106)	59.44 ± 1.07	0.33	0.17	0.33	0.51
	T2 (55)	60.37 ± 1.58	0.33	0.17	0.32	0.51
	T3 and T4 (10)	63.58 ± 5.20	0.31	0.13	0.27	0.48
	p-value (Kruskal Wallis Test)	0.537	0.354	0.072	0.024	0.796
Lymphnode involvement (180)	lymphnode ≤ 3 (159)	59.53 ± 0.91	0.33	0.17	0.33	0.51
	lymphnode > 3 (21)	61.81 ± 2.57	0.31	0.14	0.28	0.49
	p-value (logistic regression *)	0.393	0.427	0.003	0.002	0.180
	p-value (Mann-Whitney U)	0.277	0.331	0.001	0.003	0.311
Stage (183)	Stage 0 (12)	57.03 ± 3.20	0.37	0.16	0.34	0.50
	Stage I (89)	58.95 ± 1.19	0.33	0.17	0.33	0.51
	Stage II (58)	60.48 ± 1.53	0.33	0.18	0.32	0.51
	Stage III (24)	62.59 ± 2.57	0.32	0.14	0.29	0.49
	p-value (Kruskal Wallis Test)	0.396	0.809	0.013	0.010	0.746
Grading (182)	Grade 1 (35)	60.26 ± 1.72	0.35	0.17	0.32	0.51
	Grade 2 (110)	60.03 ± 1.13	0.32	0.17	0.33	0.50
	Grade 3 (37)	58.93 ± 2.01	0.36	0.17	0.32	0.51
	p-value (Kruskal Wallis Test)	0.831	0.248	0.656	0.926	0.872
ER status (180)	ER negative (22)	56.74 ± 2.50	0.37	0.16	0.33	0.51
	ER positive (158)	60.02 ± 0.91	0.32	0.17	0.32	0.50
	p-value (logistic regression *)	0.213	0.028	0.630	0.787	0.373
	p-value (Mann-Whitney U)	0.233	0.023	0.548	0.914	0.601
PR status (181)	PR negative (36)	59.33 ± 1.91	0.36	0.16	0.33	0.52
	PR positive (145)	59.69 ± 0.96	0.32	0.17	0.32	0.50
	p-value (logistic regression *)	0.867	0.068	0.386	0.282	0.169
	p-value (Mann-Whitney U)	0.746	0.029	0.331	0.372	0.218
Her2 status (180)	Her2 negative (158)	59.89 ± 0.93	0.33	0.17	0.32	0.50
	Her2 positive (22)	57.65 ± 2.17	0.31	0.16	0.32	0.50
	p-value (logistic regression *)	0.393	0.500	0.355	0.383	0.725
	p-value (Mann-Whitney U)	0.381	0.819	0.677	0.281	0.602
Family history (181)	no 1st degree relative with BC or ovarian cancer (152)	59.98 ± 0.94	0.33	0.17	0.33	0.51
	1st degree relative with BC or ovarian cancer (29)	57.88 ± 2.09	0.31	0.16	0.31	0.50
	p-value (logistic regression *)	0.368	0.376	0.206	0.190	0.550
	p-value (Mann-Whitney U)	0.399	0.431	0.293	0.295	0.378

* adjusted for age of onset and different batches for the measurements

Table 10. The methylation levels of HYAL2 CpG sites by Illumina 450K

SNP ID	pvalWald_Group	meanDiff	adjWaldPval	CHR	MAPINFO	location-1	location-2
cg27091787	0.0014	-0.0261	0.0657	3	50335694	S_Shore	promoter
cg26460678	0.0028	-0.0226	0.0968	3	50335671	S_Shore	promoter
cg03051392	0.0003	-0.0171	0.0260	3	50335180	S_Shore	promoter
cg08776109	0.0003	-0.0201	0.0253	3	50331237		last intron
cg06721473	0.0010	-0.0203	0.0523	3	50330420		last exon 3'UTR

Claims

1. A method for diagnosing cancer in a subject, comprising the steps of:
 - (a) determining in a sample of said subject the methylation status of at least one CpG site located in the promoter region of the HYAL2 gene; and
 - (b) comparing said methylation status with a reference, whereby cancer is to be diagnosed.
2. The method of claim 1, wherein the subject is a human and wherein in step (a) the CpG site is one of the CpG sites located between position 50334760 and position 50335700 on human chromosome 3.
3. The method of claim 1 or 2, wherein the subject is a human and wherein in step (a) the CpG site is selected from the list of CpG sites on the human chromosome 3 consisting of cg27091787 at position 50335694, HYAL_CpG_1 at position 50335584, HYAL_CpG_2 at position 50335646, HYAL_CpG_3 at position 50335671, HYAL-is-310 CpG_1 at position 50335166, HYAL-is-310 CpG_2 at position 50335180, HYAL-is-310 CpG_3 at position 50335192, HYAL-is-310 CpG_4 at position 50335195, HYAL-is-310 CpG_5 at position 50335227, HYAL-is-310 CpG_6 at position 50335233, HYAL-is-310 CpG_7 at position 50335300, HYAL-is-310 CpG_8 at position 50335315, HYAL-is-310 CpG_9 at position 50335375, HYAL-is-310 CpG_10 at position 50335392, HYAL-is-310 CpG_11 at position 50335401, HYAL2-is-325_CpG_1 at position 50334744, HYAL2-is-325_CpG_2 at position 50334761, HYAL2-is-325_CpG_3 at position 50334804, HYAL2-is-325_CpG_4 at position 50334844, HYAL2-is-325_CpG_5 at position 50334853, HYAL2-is-325_CpG_6 at position 50334862, HYAL2-is-325_CpG_7 at position 50334880, HYAL2-is-325_CpG_8 at position 50334906, HYAL2-is-325_CpG_9 at position 50334913, HYAL2-is-325_CpG_10 at position 50334917, HYAL2-is-325_CpG_11 at position 50334928, HYAL2-is-325_CpG_12 at position 50334944, HYAL2-is-325_CpG_13 at position 50334956, HYAL2-is-325_CpG_14 at position 50334980, HYAL2-is-325_CpG_15 at position 50334982, HYAL2-is-325_CpG_16 at position 50335010, and HYAL2-is-325_CpG_17 at position 50335014.
4. The method of claim 1, wherein the subject is a human and wherein in step (a) the CpG site is one of the CpG sites located between position 50328231 to 50332163 of human chromosome 3.

5. The method of claim 4, wherein in step (a) the CpG site is selected from the list of CpG sites on human chromosome 3 consisting of cg08776109 at position 50331237 and cg06721473 at position 50330420.
- 5 6. The method of any one of claims 1 or 5, wherein in step (a) the methylation status of at least two, at least three, at least four, or at least five of said CpG sites is determined.
7. The method of any one of claims 2 to 6, wherein in step (a) at least one CpG site is selected from the list consisting of cg27091787 at position 50335694, HYAL_CpG_1 at
10 position 50335584, HYAL_CpG_2 at position 50335646, and HYAL_CpG_3 at position 50335671 or selected from cg08776109 at position 50331237 and cg06721473 at position 50330420.
8. The method of any one of claims 1 to 7, wherein said cancer is breast cancer or ovarian
15 cancer.
9. The method of any one of claims 1 to 8, wherein the cancer is breast cancer with involvement of at least three lymph nodes.
- 20 10. The method of any one of claims 1 to 9, wherein the sample is a sample of peripheral blood leukocytes.
11. The method of any one of claims 1 to 10, wherein the methylation status is determined by a sequencing-based methylation assay, a real-time PCR based methylation assay, a
25 methylation sensitive restriction enzyme based assay or a mass spectrometry-based methylation assay.
12. The method of any one of claims 1 to 11, wherein the subject is a human.
- 30 13. The method of any one of claims 1 to 12, wherein the subject is a Caucasian.
14. A method for diagnosing breast cancer in a subject, comprising the steps of:
(a) determining in a sample of a subject suspected to suffer from said cancer the amount of a gene product of the HYAL2 gene; and
35 (b) comparing said amount with a reference, whereby breast cancer is to be diagnosed.
15. The method of claim 14, wherein in step (a) the amount of HYAL2 transcript or HYAL2 protein is determined.

16. The method of claim 14 or 15, wherein the sample is a sample of peripheral blood or of peripheral blood leukocytes.
17. A device for diagnosing cancer, comprising
- 5 (a) an analyzing unit comprising
- (i) a detection agent for determining the methylation status of at least one of the CpG sites of claims 1 to 3 or 4 to 5 in a sample of a subject, or comprising
- (ii) a detection agent for determining the amount of a HYAL2 gene product in a sample of a subject suspected to suffer from breast cancer; and
- 10 (b) an evaluation unit comprising a data processor having tangibly embedded an algorithm for carrying out a comparison of the amount determined by the analyzing unit with a stored reference and which is capable of generating an output file containing a diagnosis established based on the said comparison.
- 15 18. A kit for carrying out a method of any one of claims 14 to 16, wherein said kit comprises instructions for carrying out said method, a detection agent for a HYAL2 gene product in a sample of a subject suspected to suffer from breast cancer, and standards for a reference.
- 20 19. A method for therapy monitoring in a subject being treated against cancer, comprising the steps of
- (a) obtaining a first and a second sample from a subject, wherein the first sample is obtained at a time point before the second sample,
- (b) determining in said first and second sample the methylation status of at least one of
- 25 the CpG sites of any one of claims 1 to 3 or 4 to 5; and
- (c) comparing the methylation status of said first sample to the methylation status of said second sample, thereby monitoring therapy in said subject.

A

CCCTGAGATCCCGTCCTTTT | CTGCCCAGTCTCTGAGGTCCCTTTGGGAGCCCGGCTCCCTCTCA
ACCCCGAGGAGGCCTCCATTCCCTTGACAGTCTTCTGGCTGGACCTCCGAGAAAAGCCAGCAAGA
GGGGTCTCCCAGGGCCTCCATCCCGTATCCAGCGCCCCCTACGCTCCACAGGCCTCTCCCGGT
GGGGGGCAGGCTTAGACCTGGCGAGCCCCGTGCGAAACACCTCGGCTCTTCCCTAGGTCCGG
TCCCTTGTGCGAGACTCAACTCCCAAAGGGCCCGCGTCTGCACCACCAGCCCCTAACCTCCC
GGGCGCCAGGAGAGGGAGGGAGTGCAAC | TGCGAGCGCATCTGTGGGGGTGCAGACCAGCCG
TCTGCTCTGGGCTCACCTGTGTGGGAGTCCGTGCGCCAGTAGGTGCGGTGGCCTCCCT | CCTT
CCTGGGGTGAGCCTCTCCAGCTAGGCTACTTGCGCCAGCTGCTGCCCGTTGGGCAGCTCGCCG
AGGACTTGATAAACCCATTCCCTCCCATCCGCTCCCGCCTCCCATCCAGCCCTTTCTCCTCAC
TTCCTGCTGCCCCCTGAGAGGAAAGGGCCAGGAACAGGCGTGGGGGAGGAGATCGCCTTGACC
CTGTTTCCCTCCAGAATGCCCTGCTGGCTGGCTGCCCTCCCCACCCTTCCCGAGTAGGATCTGG
CCCCGCCAGCTGCGCCAGGTGACAGGAAGGCTATGGGCTCAGGATCAGGTG | CCAATCTCCCC
CACACCCACAGCTTCCAACACTAGGGAACCTCTTCACTCCTTCCCTAGGATTGGGGCCCAT | CC
TCAAATCCAGTAGGGTGTGAGAGGATGGGGTCAGGTGGTGGTGCTTTTGGGAGTCAAATACTT
GGGGGTCGTTTACAGCTGATGGTCCCCCAGAGCAGGTGCCCAAGAAGGGAAGTACCTTGGGGGG
AGGGTCCGGGGGACTTCCAGTAGCTGAGTCCGTTTTTTTTTCCACTGAGAGCTTCCCGCATCCTGTGT
GACAGGACAATGTGTTAGTGAGTGTGACTGTGTGGGTGGGGAGGAGGGGATGGACTGGGAGG
GTTCTGCAGAGTGCTTTCCTGCTGGAGTTCATTCTGGATGGATTTTCAGAGTCCACTATACAGAG
TAGGGGGCTCCTTACAACATGGATGAG | GAAAGGGGTATACTCCTCTGGGAGGGGTGTTCTGT
GGCTGGCCCTAGGGAGGACTGCAGCAAGGTTCCAGCAGACCTCTCCAGGGGTGCAAA

Fig. 1

B**Primers for different amplicons**

Amplicon: HYAL2

HYAL2 Sense (SEQ ID NO: 6): 5'-

aggaagagagTTTTAAATTTAGTAGGGTGTGAGAGGA-3'

HYAL2 Antisense (SEQ ID NO: 7):

5'-cagtaatacgactcactatagggaaggctCTCATCCATATTATAAAAAACCCCC-3'

Amplicon: HYAL2-is-310

HYAL2-is-310 Sense(SEQ ID NO: 8): 5'-

aggaagagagTTTTTTTGGGGTGAGTTTTTTTAGT-3'

HYAL2-is-310 Antisense(SEQ ID NO: 9):

5'-cagtaatacgactcactatagggaaggctCACCTAATCCTAAACCCATAACCTT-3'

Amplicon: HYAL2-is-325

HYAL2-is-325 Sense(SEQ ID NO: 10): 5'-

aggaagagagTTGTTTAGTTTTTGAGGTTTTTTTGG-3'

HYAL2-is-325 Antisense(SEQ ID NO: 11):

5'-cagtaatacgactcactatagggaaggctATTACACTCCCTCCCTCTCCTAAC-3'

Fig. 1 (continued)

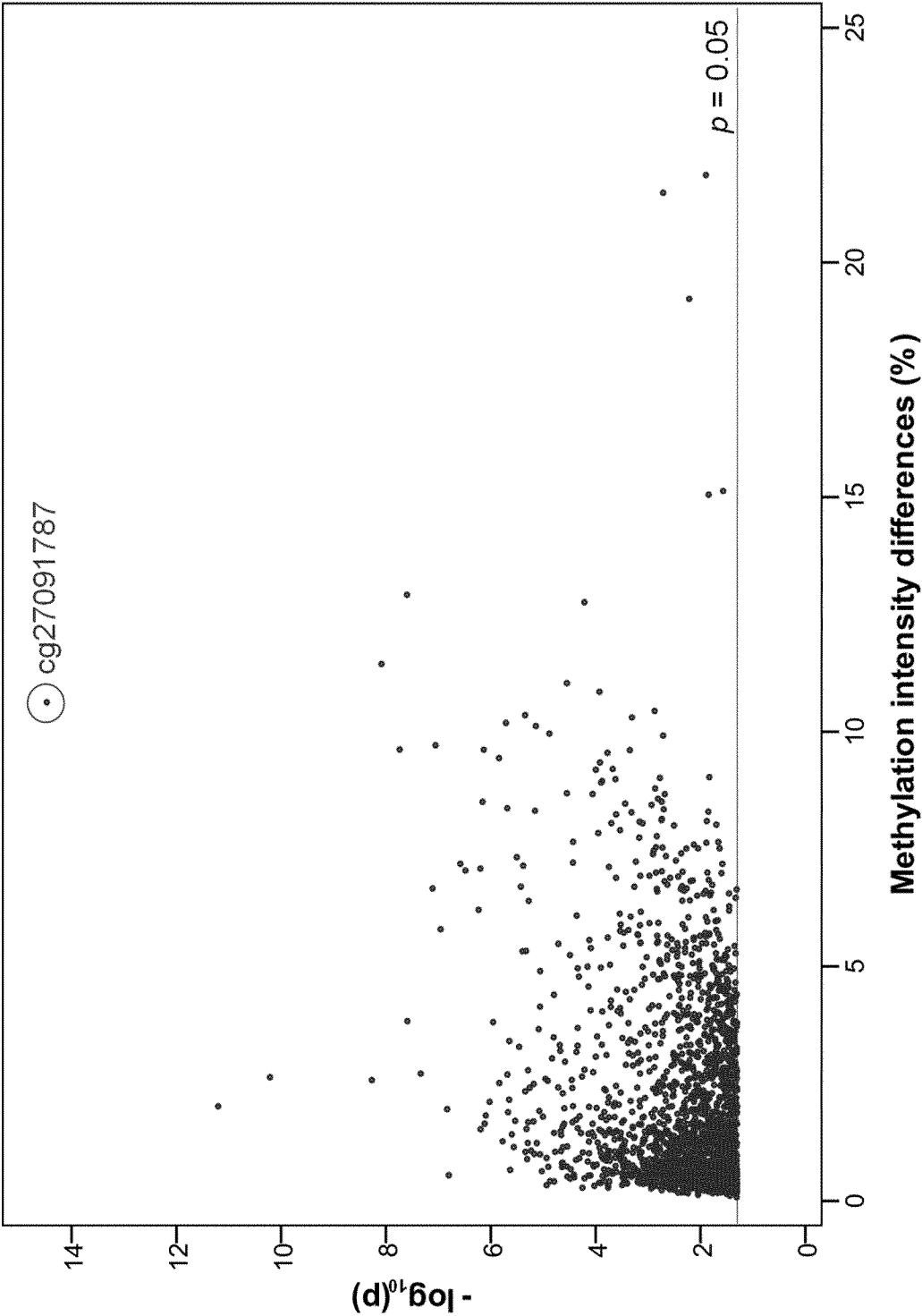


Fig. 2

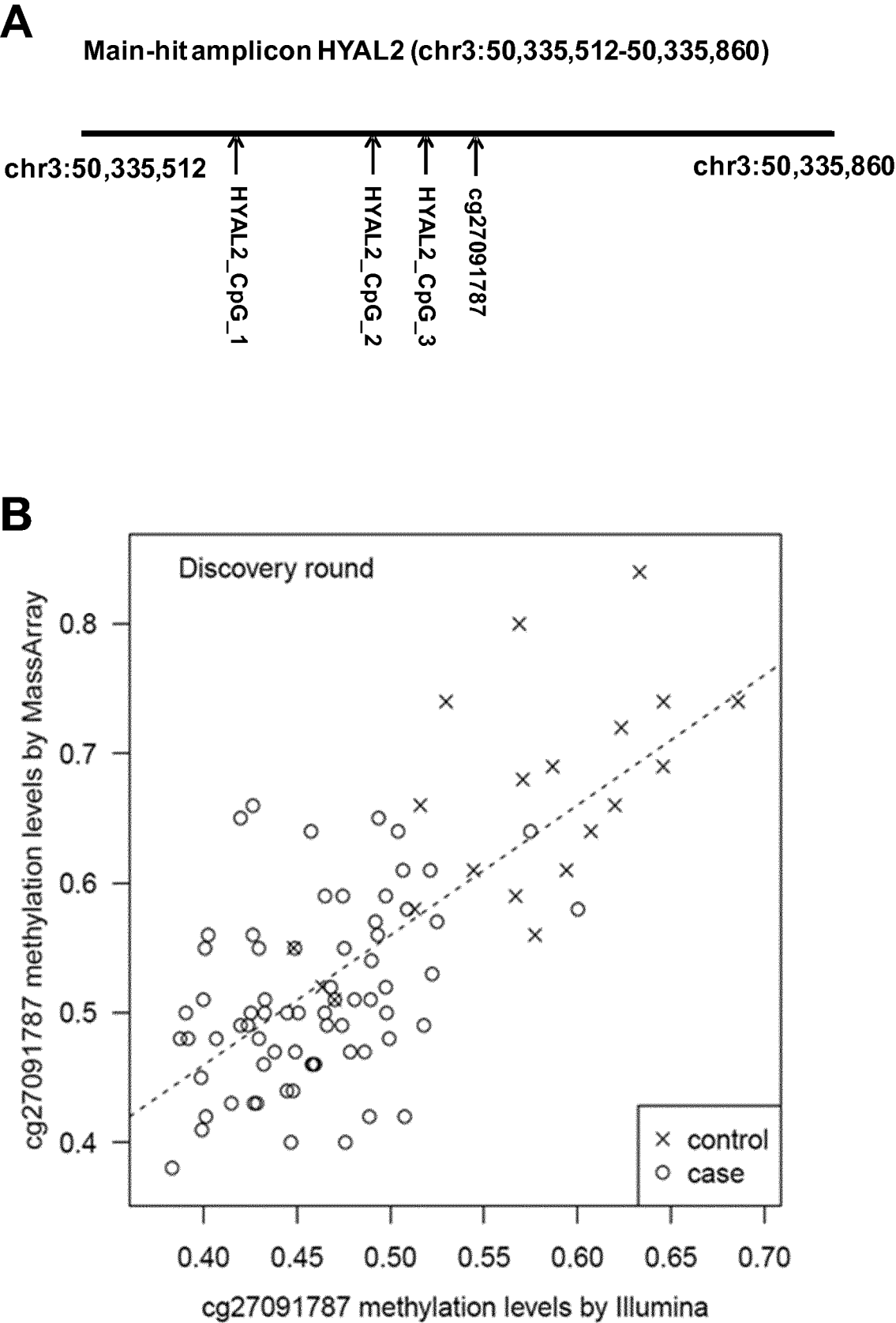
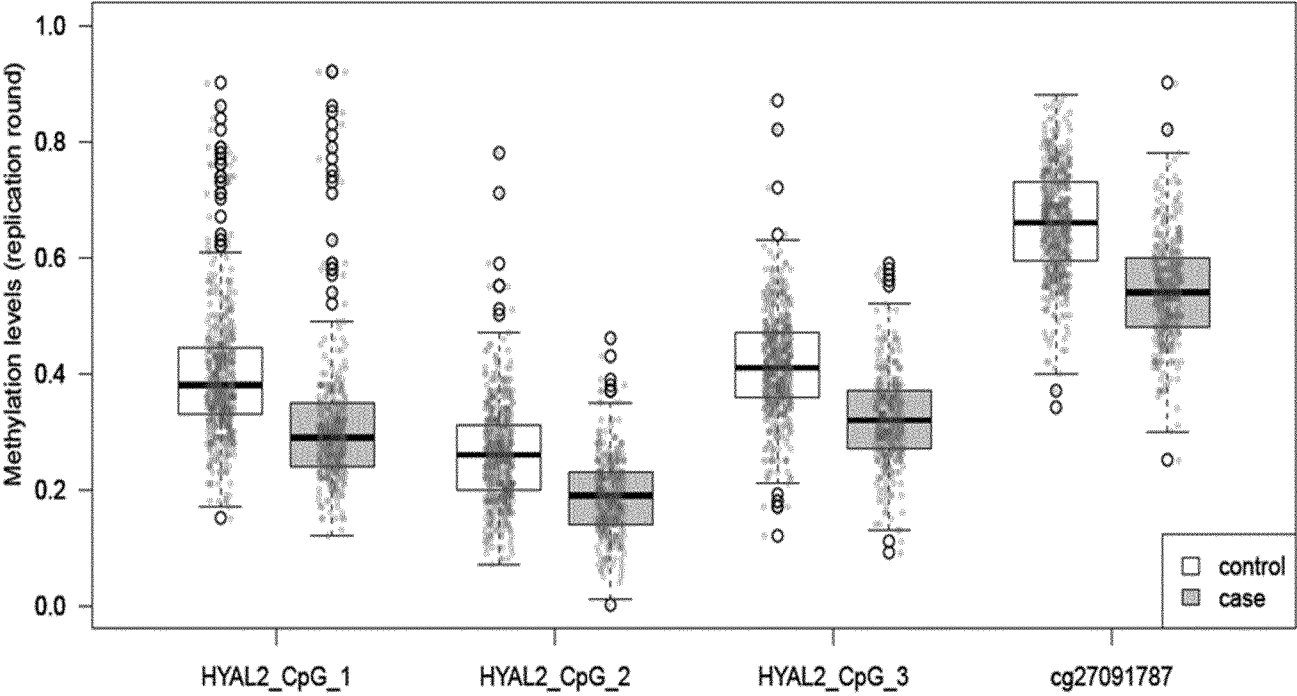


Fig. 3

C



D

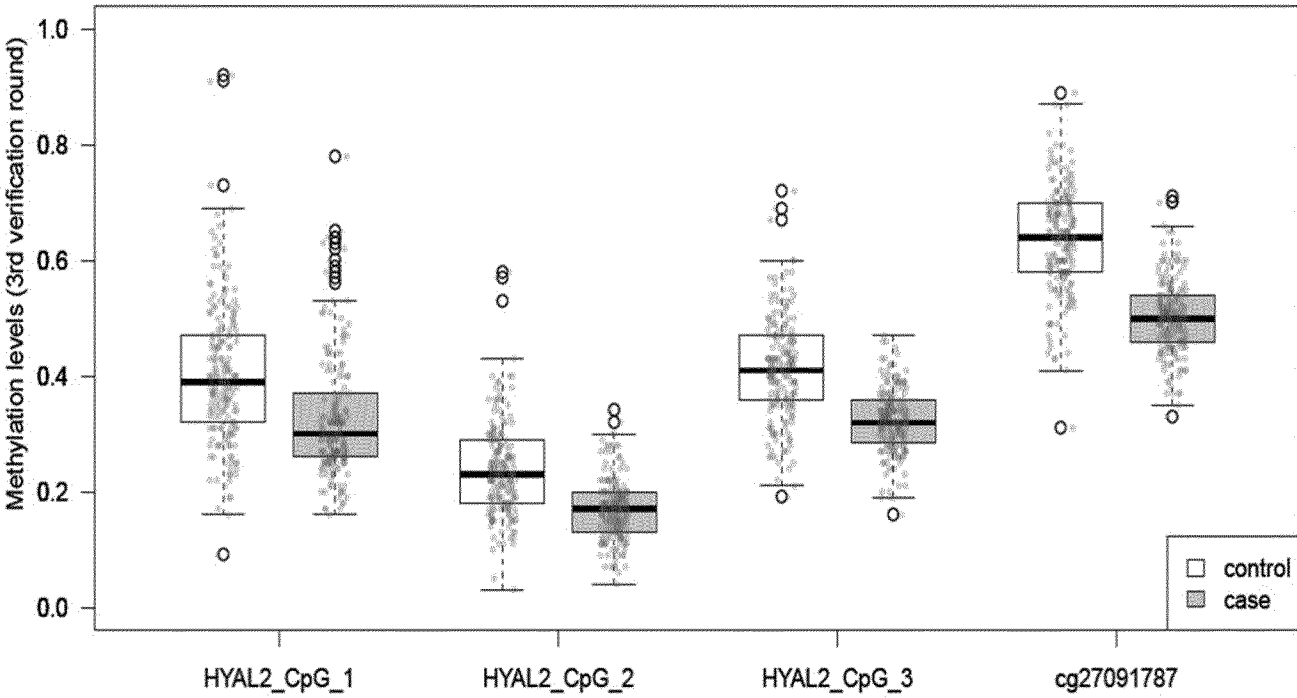


Fig. 3 (continued)

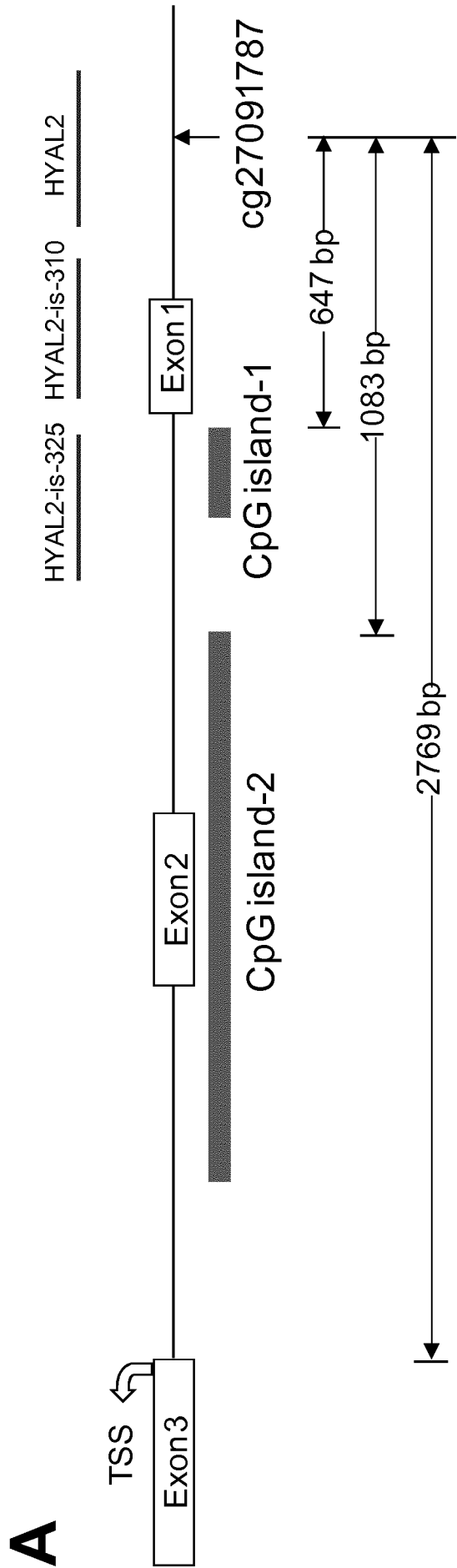


Fig.4

B

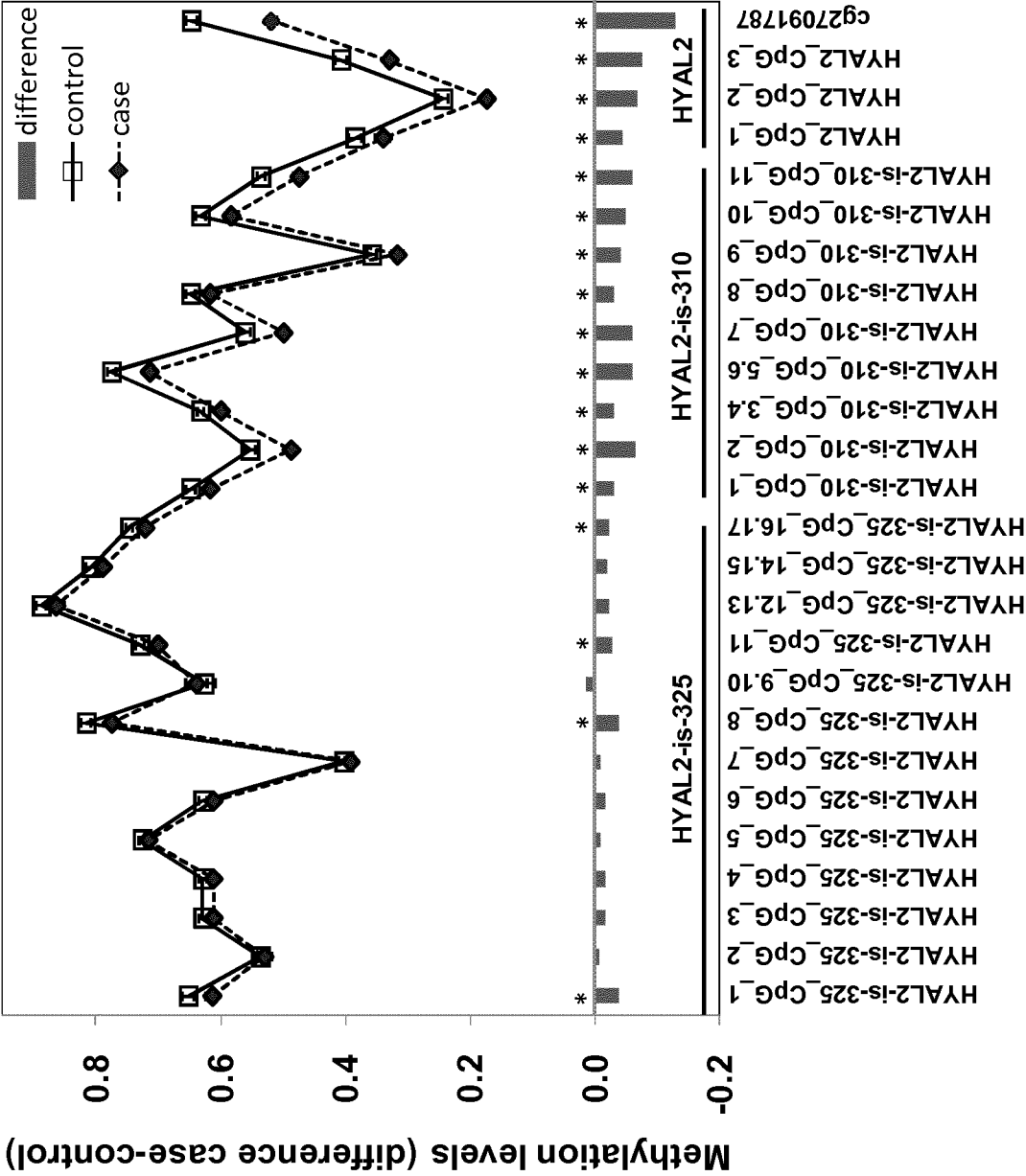


Fig 4 (continued)

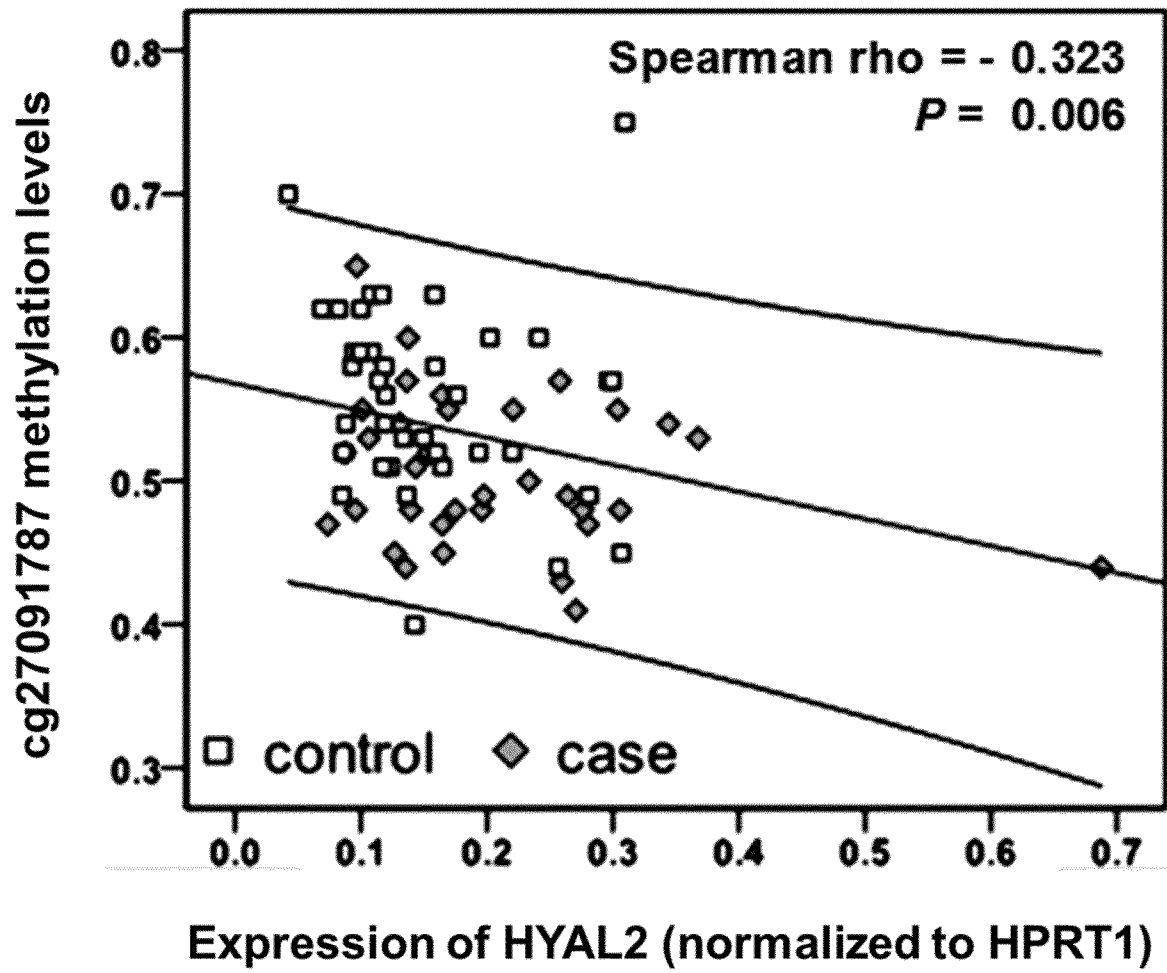


Fig. 5

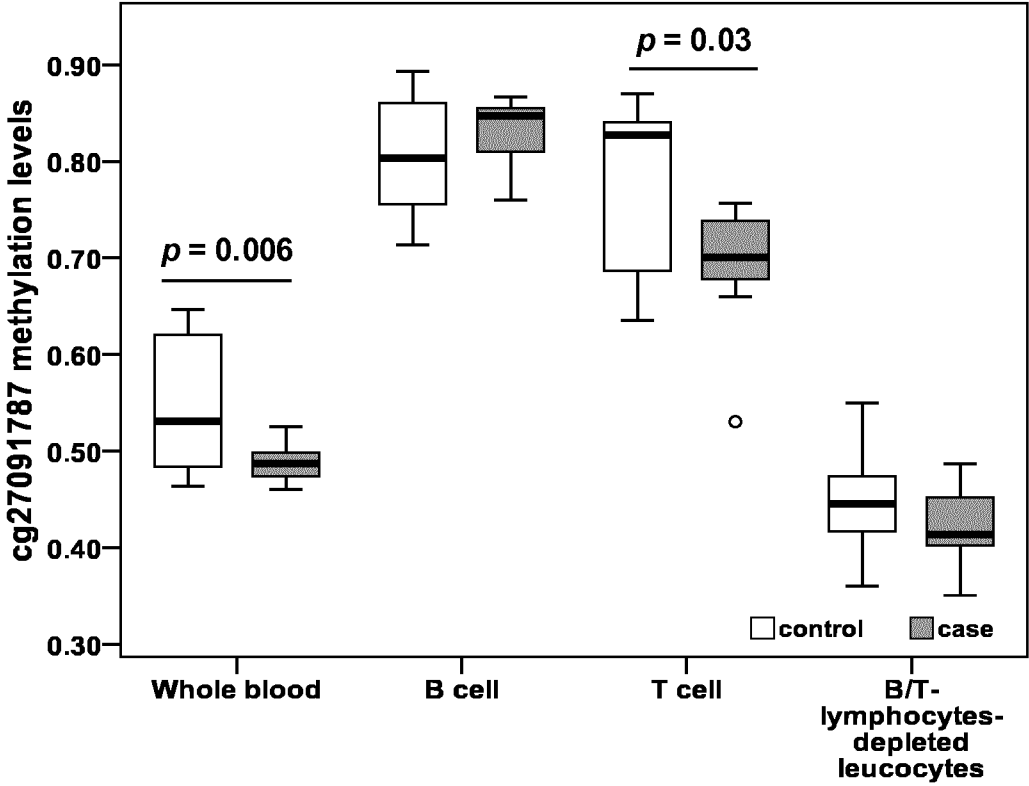


Fig. 6

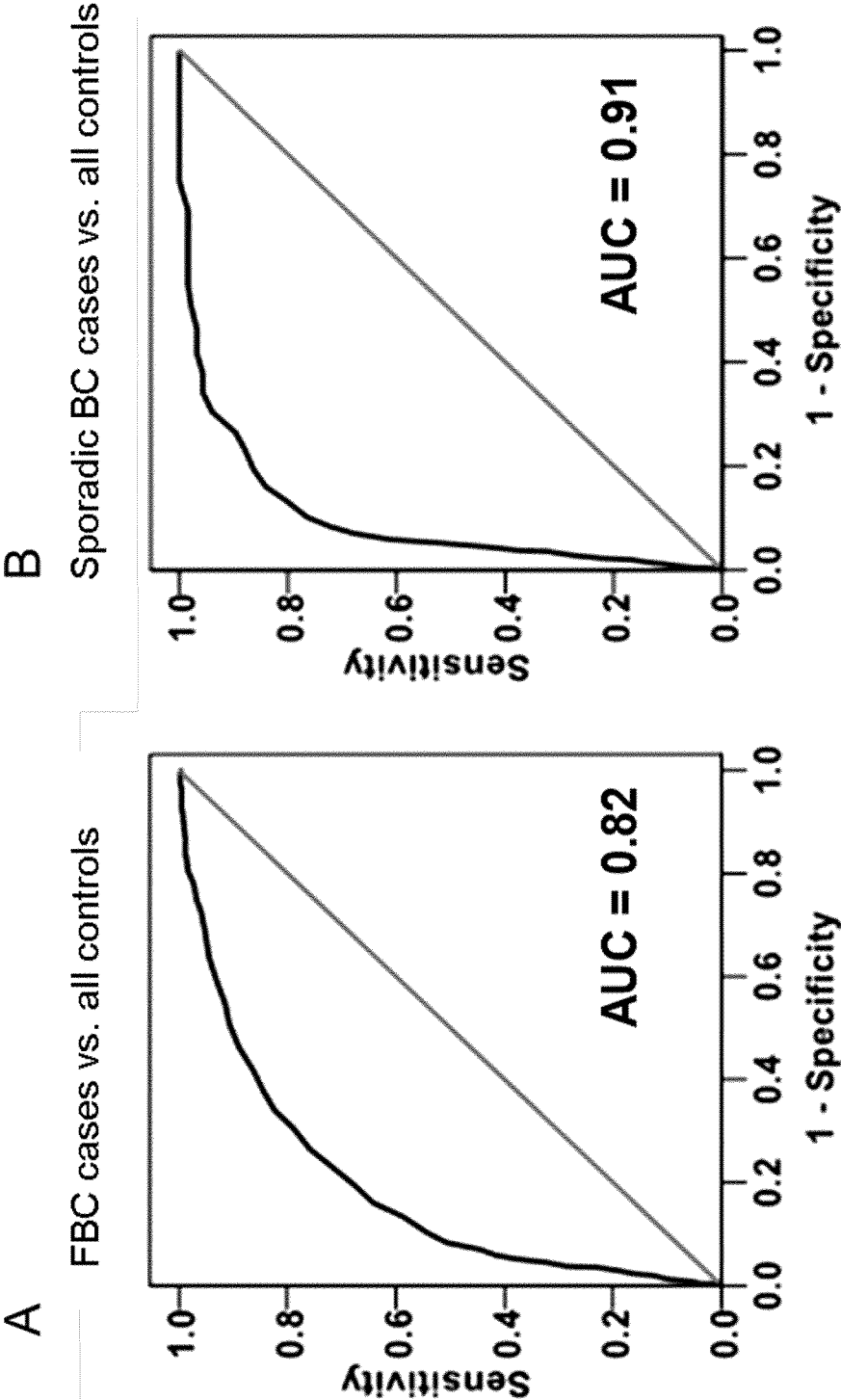


Fig. 7

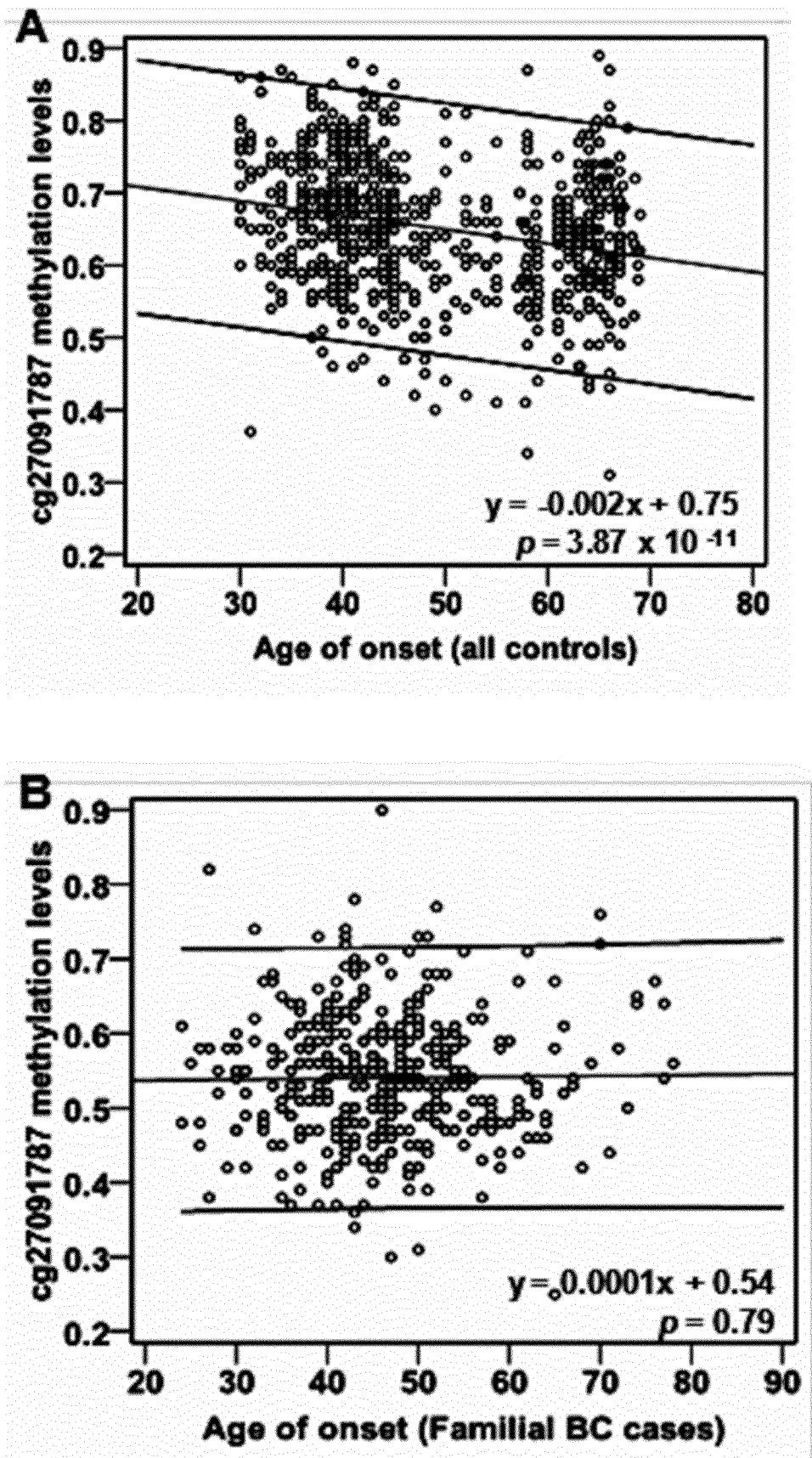


Fig. 8

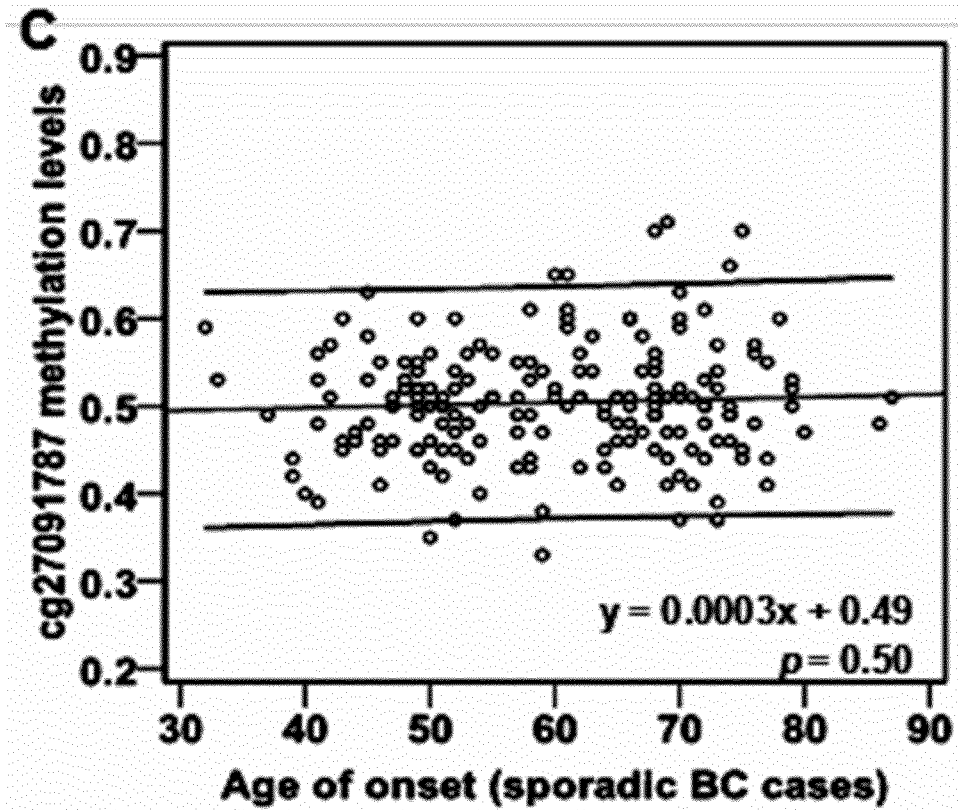


Fig. 8 (continued)

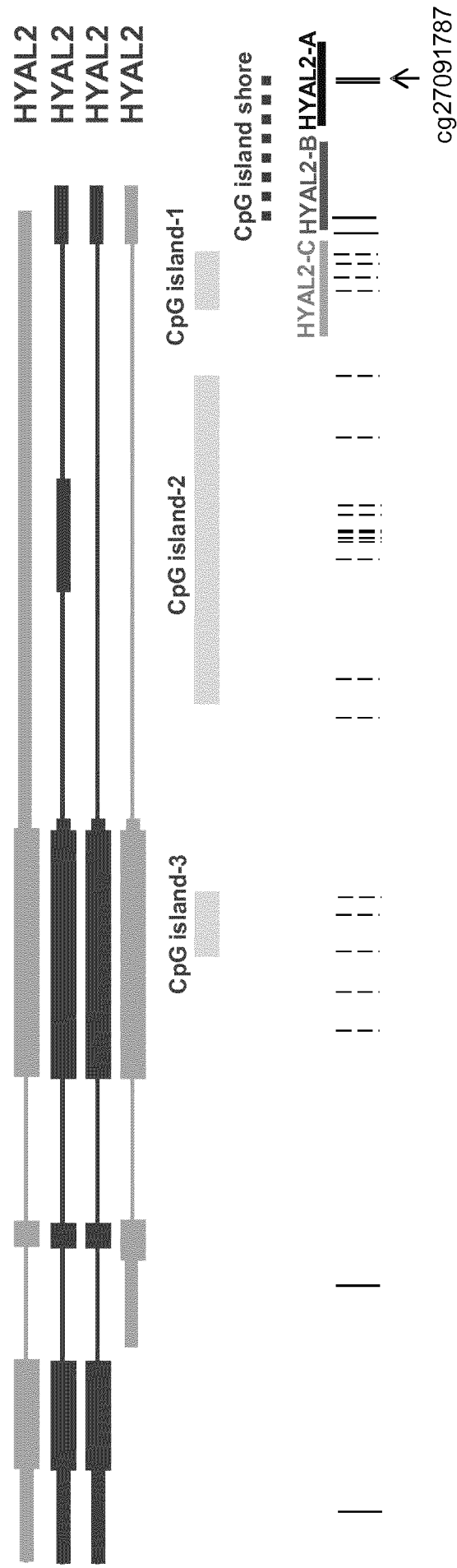


Fig. 9

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2013/066047

A. CLASSIFICATION OF SUBJECT MATTER
 INV. C12Q1/68
 ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ANNA POTAPOVA ET AL: "Promoter Hypermethylation of the PALB2 Susceptibility Gene in Inherited and Sporadic Breast and Ovarian Cancer", CANCER RESEARCH, AMERICAN ASSOCIATION FOR CANCER RESEARCH, US, vol. 68, no. 4, 15 February 2008 (2008-02-15), pages 998-1002, XP007910295, ISSN: 0008-5472, DOI: 10.1158/0008-5472.CAN-07-2418 the whole document	1-19
A	WO 2009/086472 A2 (WAYNE JOHN CANCER INST [US]; HOON DAVE S B [US]; TANEMURA ATSUSHI [JP]) 9 July 2009 (2009-07-09) the whole document ----- -/--	1-19



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

19 September 2013

Date of mailing of the international search report

02/10/2013

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
 NL - 2280 HV Rijswijk
 Tel. (+31-70) 340-2040,
 Fax: (+31-70) 340-3016

Authorized officer

Mueller, Frank

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2013/066047

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LEPPERDINGER G ET AL: "Hyal2--less active, but more versatile?", MATRIX BIOLOGY, ELSEVIER, NL, vol. 20, no. 8, 1 December 2001 (2001-12-01), pages 509-514, XP002507270, ISSN: 0945-053X, DOI: 10.1016/S0945-053X(01)00170-6 see whole doc. esp. p.512, 1.col -----	14,15,18
X	UDABAGE ET AL: "The over-expression of HAS2, Hyal-2 and CD44 is implicated in the invasiveness of breast cancer", EXPERIMENTAL CELL RESEARCH, ACADEMIC PRESS, US, vol. 310, no. 1, 15 October 2005 (2005-10-15), pages 205-217, XP005093320, ISSN: 0014-4827, DOI: 10.1016/J.YEXCR.2005.07.026 see whole doc. esp. p.206, 2.col., 2. par.ff -----	14,15,18
X	YUEJUAN LI ET AL: "Silencing of hyaluronan synthase 2 suppresses the malignant phenotype of invasive breast cancer cells", INTERNATIONAL JOURNAL OF CANCER, vol. 120, no. 12, 1 January 2007 (2007-01-01), pages 2557-2567, XP055038445, ISSN: 0020-7136, DOI: 10.1002/ijc.22550 see whole doc. esp. Fig.1 and p.2565, 2.col.,2.par. -----	14,15,18
X	PRANDO ERIKA DA COSTA ET AL: "Evidence of epigenetic regulation of the tumor suppressor gene cluster flanking RASSF1 in breast cancer cell lines", EPIGENETICS, LANDES BIOSCIENCE, US, vol. 6, no. 12, 1 December 2011 (2011-12-01), pages 1413-1424, XP009162951, ISSN: 1559-2294, DOI: 10.4161/EPI.6.12.18271 the whole document -----	1-19

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2013/066047

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2009086472 A2	09-07-2009	AU 2008345036 A1	09-07-2009
		CA 2710520 A1	09-07-2009
		EP 2238265 A2	13-10-2010
		US 2009220980 A1	03-09-2009
		US 2011212444 A1	01-09-2011
		WO 2009086472 A2	09-07-2009
