



(51) International Patent Classification:

C12Q 1/6886 (2018.01)

(21) International Application Number:

PCT/EP2017/076076

(22) International Filing Date:

12 October 2017 (12.10.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

PA 2016 70811 14 October 2016 (14.10.2016) DK

(71) Applicant: AARHUS UNIVERSITET [DK/DK]; Nordre Ringgade 1, 8000 Aarhus C (DK).

(72) Inventors: KROG, DAUGAARD, Iben; Skelagervej 384, 8200 Aarhus N (DK). HANSEN, Lise Lotte; Ørbyvej 11, 8240 Risskov (DK). WOJDACZ, KAZIMIERZ, Tomasz; Jens Baggesens Vej 33, 3, 13, 8200 Aarhus N (DK).

(74) Agent: HØIBERG P/S; Adelgade 12, 1304 Copenhagen K (DK).

(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, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, 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, SA, 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, ST, 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).

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

— with international search report (Art. 21(3))

— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

(54) Title: METHYLATION BIOMARKERS FOR LUNG CANCER

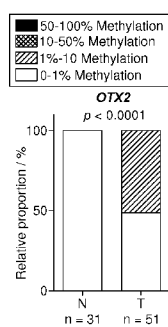


FIG. 1A

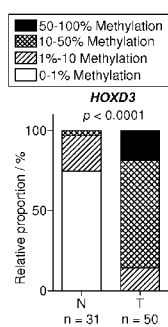


FIG. 1B

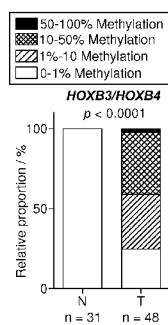


FIG. 1C

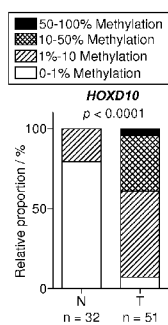


FIG. 1D

(57) Abstract: A method is provided for determining non-small cell lung cancer, a predisposition to non-small cell lung cancer, or the prognosis of a non-small cell lung cancer in a subject, said method comprising in a sample from said subject determining the methylation status of at least one gene locus including regulatory sequences of said gene locus, wherein said gene locus is selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHRSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

METHYLATION BIOMARKERS FOR LUNG CANCER**Field of invention**

The present invention relates to DNA methylation biomarkers for non-small cell lung cancer.

Background of invention

Lung cancer is the most common type of cancer and each year, the disease is responsible for approximately 1.5 million deaths worldwide. There are two major types of lung cancer; small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC) accounting for 10% and 85%, of all newly diagnosed lung cancers, respectively. Lung adenocarcinoma (LAC) is the most common subtype of NSCLC, which account for approximately 40% of all lung cancers. The overall 5-year survival rate for lung cancer is 15%, but the prognosis is highly dependent on the stage, at which the disease is diagnosed. If the disease is localized at the time of diagnosis, the 5-year survival rate is approximately 50%, compared to approximately 25% for cases with regional disease, and less than 5% for patients that already suffer from metastatic disease. Most early stage lung cancers are asymptomatic and consequently, only 15% of lung cancers are diagnosed at a local stage and more than 50% are diagnosed at an advanced stage. Thus, new efficient diagnostic tools for early and accurate disease detection are needed in order to improve the poor prognosis of lung cancer.

Methylation of the carbon-5 position of cytosine residues within CpG dinucleotides is a well-established epigenetic mechanism involved in the regulation of gene expression. Most CpG dinucleotides cluster in CpG rich regions in the genome, known as CpG islands (CGI), and these regions are often located within gene regulatory elements. In fact, the promoter region of more than half of all protein encoding genes contain a CGI and the methylation status of this sequence is instrumental in regulating the transcriptional activity of the gene. Consequently, disruption of the cell's normal methylation pattern can have severe consequences and contribute to neoplastic transformation. Genome-wide studies have shown that aberrant DNA methylation is a common feature in human cancer and hundreds of tumor suppressor genes have been shown to be subject to DNA-methylation mediated silencing. Gene expression changes as a consequence of aberrant methylation have also been reported for multiple genes in lung cancer, especially hypermethylation-mediated silencing of tumor suppressor

genes such as RASSF1A, APC, RAR β , DAPK and MGMT, and to a lesser extent, hypomethylation-mediated overexpression of proto-oncogenes such as ELMO3.

5 The utility of DNA methylation biomarkers has already been established in all aspects of clinical cancer management, including risk assessment, early disease detection, prognostication and treatment personalization. However, the development of biomarkers for clinical implementation is a challenging process that includes biomarker candidate discovery and evaluation of biomarker specificity and sensitivity in large-scale validation studies. Now, based on a genome-wide methylation screening novel
10 methylation biomarkers have been identified for use in clinical lung cancer management. The biomarkers display cancer-specific methylation changes and have been validated and evaluated for sensitivity and specificity.

Summary of invention

15 The invention relates to methylation biomarkers for non-small cell lung cancer. The invention provides a number of methylation markers, which can be used to distinguish between lung tumour tissue and healthy tissue. A plurality of individual methylation biomarkers are identified, which show high sensitivity and specificity.

20 In one aspect, a method is provided for determining non-small cell lung cancer, a predisposition to non-small cell lung cancer, the prognosis of a non-small cell lung cancer, and/or monitoring a non-small cell lung cancer in a subject, said method comprising in a sample from said subject determining the methylation status of at least one gene locus selected from the group consisting of SIM1, Chr6(p22.1),
25 HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

Another aspect pertains to a method of assessing whether a human subject is likely to
30 develop non-small cell lung cancer, said method comprising

i) providing a sample from said human subject,

ii) determining in said sample the methylation status of at least one gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E,
35 HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1), and

iii) on the basis of said methylation status identifying a human subject that is more likely to develop non-small cell lung cancer.

A third aspect pertains to a method for categorizing or predicting the clinical outcome of a non-small cell lung cancer of a subject, said method comprising in a sample from said subject determining the methylation status of at least one gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

A fourth aspect relates to a method of evaluating the risk for a subject of contracting cancer, said method comprising in a sample from said subject determining the methylation status of a gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

In a fifth aspect, a method is provided of treating a non-small cell lung cancer in a human subject, said method comprising the steps of

- i. determining non-small cell lung cancer, a predisposition to non-small cell lung cancer, or the prognosis of a non-small cell lung cancer in a subject by a method according to the present disclosure, which involve determining the methylation status of a gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1),
- ii. selecting human subjects having non-small cell lung cancer, a predisposition to non-small cell lung cancer, or a negative or positive prognosis of a non-small cell lung cancer,
- iii. subjecting said subjects identified in step ii. to a suitable treatment for non-small cell lung cancer.

In the methods of the invention, the sample may be a lung tissue sample, or a bodily fluid such as blood or plasma (for example peripheral blood), and methylation status may be determined by any method selected from the group consisting of Methylation-

Sensitive High Resolution Melting (MS-HRM), Methylation - Sensitive Melting Curve Analysis (MS-MCA), EpiTyper, Methylation-Specific PCR (MSP), DNA methylation specific qPCR (qMSP), Pyrosequencing, Methyl Light, Amplicon bisulfite sequencing (AmpliconBS), Enrichment bisulfite sequencing (EnrichmentBS), Whole genome
5 bisulfite sequencing (BS-Seq), HELP assays, and Methyl Sensitive Southern Blotting, and determination may involve Methylated DNA immunoprecipitation (MeDIP).

In one embodiment, the methylation status is determined by a method comprising the steps of

- 10 i) providing a lung tissue sample from said subject comprising nucleic acid material comprising said gene,
- ii) processing said nucleic acid sequence using one or more methylation-sensitive restriction endonuclease enzymes,
- iii) optionally, amplifying said processed nucleic acid sequence in order to
15 obtain an amplification product, and
- iv) analyzing said processed nucleic acid sequence or said amplification product for the presence of processed and/or unprocessed nucleic acid sequences, thereby inferring the presence of methylated and/or unmethylated nucleic acid sequences.

20

In a preferred embodiment of the methods, the methylation status is determined by a method comprising the steps of

- i) providing a lung tissue sample from said subject comprising nucleic acid material comprising said gene,
- 25 ii) modifying said nucleic acid using an agent which modifies unmethylated cytosine or cleaves nucleic acid sequences in a methylation-dependent manner,
- iii) amplifying at least one portion of said gene using primers, which span or comprise at least one CpG dinucleotide in said gene in order to obtain an amplification product, and
- 30 iv) analyzing said amplification product for the presence of modified and/or unmodified cytosine residues, wherein the presence of modified cytosine residues are indicative of methylated cytosine residues.

The amplified CpG-containing nucleic acid is preferably analyzed by melting curve analysis. However, methylation status may also be determined by methylation specific

PCR, bisulfite sequencing, COBRA, endonucleolytic digestion, or DNA methylation arrays.

A kit is also provided in a sixth aspect for determining non-small cell lung cancer, predisposition to non-small cell lung cancer, or categorizing or predicting the clinical outcome of a non-small cell lung cancer, said kit comprising in a package

i. an agent that (a) modifies methylated cytosine residues but not non-methylated cytosine residues; or (b) modifies non-methylated cytosine residues but not methylated cytosine residues; or (c) modifies a nucleic acid sequence in a methylation-dependent manner,

ii. and at least one pair of oligonucleotide primers that specifically hybridizes under amplification conditions to a region of a gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1), such as at least one primer selected from the group consisting of SEQ ID NO: 37 to 72.

In a seventh aspect of the present disclosure provides a use of oligonucleotide primers comprising a sequence, which is a subsequence of a gene loci selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1) or the complement thereof for diagnosing non-small cell lung cancer in any of the methods provided in the present disclosure.

In yet an eighth aspect, a method is provided for identifying a therapeutically effective agent for treatment of non-small cell lung cancer, said method comprising

i. providing a non-small cell lung cancer cell line comprising one or more genetic loci selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1),

ii. providing one or more potential therapeutic agents,

iii. treating said non-small cell lung cancer cells by bringing said agents in contact with said non-small cell lung cancer cells,

iv. determining methylation status of said one or more genetic loci

v. comparing said methylation status of said treated non-small cell lung cancer cells with the methylation status of said non-small cell lung cancer cells, when untreated, wherein a decreased level of methylation positive alleles is indicative of a therapeutic agent.

5

Description of Drawings

Figure 1: Differentially methylated regions in LAC.

The methylation level of 18 DMRs was investigated in 52 LAC primary tumors and 32 tumor-adjacent normal lung samples using MS-HRM analysis. The results of the methylation assessment are shown as stacked bar percentage plots for each DMR in a-r). The relative proportion of samples in each category with 0-1% methylated templates are shown in white, 1-10% methylated templates in white with light grey stripes, 10-50% methylated templates in dark grey and 50-100% methylated templates in black. The statistical significance of the detected differences in methylation between groups was assessed using a Mann-Whitney test of ranks and two-tailed p-values ≤ 0.05 were considered statistically significant.

Figure 2: Examples of melting profiles observed for the HOXD3, OSR1 and HIST1H3E regions.

The methylation level of the 18 DMRs was determined using MS-HRM analysis. Representative normalized melting profiles for 10 tumor-adjacent normal lung samples and 10 LAC tumors are shown in black in a-b) for HOXD3, c-d) for OSR1 and in e-f) for HIST1H3E. The DNA methylation standards were generated as a serial dilution of fully methylated DNA into an unmethylated background. The 100% methylated standard is shown in red, 50% methylated standard in light blue, 10% methylated standard in green, 1% methylated standard in dark blue and the 0% methylated standard in orange.

Figure 3: Methylation levels in metastasizing and non-metastasizing LAC tumors.

The methylation levels of the 18 DMRs were compared between 26 metastasizing and 26 non-metastasizing LAC primary tumors. Three regions demonstrated a tendency towards increased methylation in the metastasizing tumors and a similar increase was observed in the 24 paired metastases (20 brain and 4 adrenal gland). The results are shown as stacked bar percentage plots in a) for HOXB3/HOXB4, b) for LOC648987 and in c) for HOXA5. The relative proportion of samples in each category with 0-1%

5 methylated templates are shown in white, 1-10% methylated templates in white with light grey stripes, 10-50% methylated templates in dark grey and 50-100% methylated templates in black. The statistical significance of the detected differences in methylation between groups was assessed using a Mann-Whitney test of ranks and two-tailed p-values ≤ 0.05 were considered statistically significant.

Figure 4: MS-HRM Assays

10 The technical specifications for 18 MS-HRM assays. Genomic location (UCSC Genome Browser on Human Dec. 2013 (GRCh38/hg38) Assembly), primer sequences and assay-specific PCR cycling and HRM protocols are shown in the sequences section. For each assay, 180 bp of regional genomic sequence (top strand) and corresponding bisulfite-modified sequence (bottom strand) is shown. The locations of the primers are indicated in red. Normalized melting profiles for the DNA methylation standards are shown for each assay in technical duplicates. The DNA methylation
15 standards were generated as a serial dilution of fully methylated DNA into an unmethylated background. The 100% methylated standard is indicated in red, 50% methylated standard in light blue, 10% methylated standard in green, 1% methylated standard in dark blue and the 0% methylated standard in orange.

20 Detailed description of the invention

The present invention relates to methylation biomarkers for use in the diagnosis and treatment of non-small cell lung cancer. Generally, the methylation markers of the invention can be used in methods for identifying subjects, which are predisposed to non-small cell lung cancer; i.e. subjects having an increased likelihood of developing
25 non-small cell lung cancer. The methylation markers of the invention can also be used in methods for identifying subjects having non-small cell lung cancer, and in this case, the markers allow early diagnosis. Further, the markers provide prognostic information with respect to non-small cell lung cancer, and this, the markers can be used to identify a subject having non-small cell lung cancer, and the cancer DNA can be tested for
30 predictive prognostic information based on the methylation markers of the invention, as well as information on which curative and/or ameliorative treatment to provide for the non-small cell lung cancer. The methylation status of the methylation markers of the invention may also be used to monitor a treatment provided for the curing and/or ameliorating a non-small cell lung cancer. Additionally, the marker methylation status

can be used to monitor relapse of non-small cell lung cancer for subject previously cured for non-small cell lung cancer.

Thus, aspects of the present invention relates to i) methods for identifying subjects, which are predisposed to non-small cell lung cancer, and/or which have a non-small cell lung cancer, including early stages, such as asymptomatic stages of non-small cell lung cancer, ii) methods for providing prognostic information of a non-small cell lung cancer and/or inferring a suitable treatment based thereupon, iii) methods of monitoring a treatment of a non-small cell lung cancer, and/or monitoring relapse of a non-small cell lung cancer.

In order to facilitate the understanding of the invention a number of definitions are provided below.

Definitions

Amplification according to the present invention is the process wherein a plurality of exact copies of one or more gene loci or gene portions (template) is synthesised. In one preferred embodiment of the present invention, amplification of a template comprises the process wherein a template is copied by a nucleic acid polymerase or polymerase homologue, for example a DNA polymerase or an RNA polymerase. For example, templates may be amplified using reverse transcription, the polymerase chain reaction (PCR), ligase chain reaction (LCR), in vivo amplification of cloned DNA, isothermal amplification techniques, and other similar procedures capable of generating a complementing nucleic acid sequence.

Amplified copies of a targeted genetic region are sometimes referred to as an amplicon.

The term "PCR bias" as used herein refers to conditions, wherein PCR more efficiently amplifies templates with a specific methylation status. It has been reported that at least some unmethylated nucleic acid templates are more efficiently amplified than methylated nucleic acid template.

A double stranded nucleic acid contains two strands that are complementary in sequence and capable of hybridizing to one another. In general, a gene is defined in terms of its coding strand, but in the context of the present invention, an oligonucleotide

primer, which hybridize to a gene as defined by the sequence of its coding strand, also comprise oligonucleotide primers, which hybridize to the complement thereof.

5 A nucleotide is herein defined as a monomer of RNA or DNA. A nucleotide is a ribose or a deoxyribose ring attached to both a base and a phosphate group. Both mono-, di-, and tri-phosphate nucleosides are referred to as nucleotides.

10 The term oligonucleotide comprises oligonucleotides of both natural and/or non-natural nucleotides, including any combination thereof. The natural and/or non-natural nucleotides may be linked by natural phosphodiester bonds or by non-natural bonds. Preferred oligonucleotides comprise only natural nucleotides linked by phosphodiester bonds. The oligomer or polymer sequences of the present invention are formed from the chemical or enzymatic addition of monomer subunits. The term "oligonucleotide" as used herein includes linear oligomers of natural or modified monomers or linkages, 15 including deoxyribonucleotides, ribonucleotides, anomeric forms thereof, peptide nucleic acid monomers (PNAs), locked nucleotide acid monomers (LNA), and the like, capable of specifically binding to a single stranded polynucleotide tag by way of a regular pattern of monomer-to-monomer interactions, such as Watson-Crick type of base pairing, base stacking, Hoogsteen or reverse Hoogsteen types of base pairing, or 20 the like. Usually monomers are linked by phosphodiester bonds or analogs thereof to form oligonucleotides ranging in size from a few monomeric units, e.g. 3-4, to several tens of monomeric units, e.g. 40-60. Whenever an oligonucleotide is represented by a sequence of letters, such as "ATGCCTG," it will be understood that the nucleotides are in 5' → 3' order from left to right and the "A" denotes deoxyadenosine, "C" denotes 25 deoxycytidine, "G" denotes deoxyguanosine, and "T" denotes thymidine, unless otherwise noted. When a double stranded DNA molecule is shown, the nucleotides of the top strand are in 5' → 3' order from left to right and the nucleotides of the bottom strand are then in 3' → 5' order from left to right. Usually, oligonucleotides of the invention comprise the four natural nucleotides; however, they may also comprise 30 methylated or non-natural nucleotide analogs.

The term "dinucleotide" as used herein refers to two sequential nucleotides. The dinucleotide may be comprised in an oligonucleotide or a nucleic acid sequence. In particular, the dinucleotide CpG, which denotes a cytosine linked to a guanine by a 35 phosphodiester bond, may be comprised in an oligonucleotide according to the present

invention, and also comprised in a targeted gene locus sequence according to the present invention. A CpG dinucleotide is also herein referred to as a CpG site. CpG sites are targets for methylation of the cytosine residue.

5 Methylation status: the term “methylation status” as used herein, refers to the presence or absence of methylation in a specific nucleic acid region. In particular, the present invention relates to detection of methylated cytosine (5-methylcytosine). A nucleic acid sequence, e.g. a gene locus of the invention, may comprise one or more CpG methylation sites. The nucleic acid sequence of the gene locus may be methylated on
10 all methylation sites (i.e. 100% methylated), or unmethylated on all methylation sites (i.e. 0% methylated). However, the nucleic acid sequence may also be methylated on a subset of its potential methylation sites (CpG-sites). In this latter case, the nucleic acid molecule is heterogeneously methylated.

The gene loci methylation markers of the present invention can be used to infer non-
15 small cell lung cancer based on the relative amount of methylation positive (fully methylated) and methylation negative (fully unmethylated) alleles in a sample comprising in a mixture of nucleic acid molecules from a subject. For example, the methylation status of a specific gene locus marker of the present invention may be that at least 50%, such as on at least 60%, such as on at least 70%, for example on at least
20 80%, such as on at least 90%, such as on at least 95%, for example on at least 99%, such as least 99.9% of the nucleic acid sequence molecules (alleles) in a sample are methylation positive (fully methylated).

Gene locus: The term “gene locus” as used herein, such as the gene loci defined by
25 the genes SIM1, HOXB3/HOXB4, OSR1, GHRSR, OTX2, LOC648987, HIST1H3E, HIST1H3G/HIST1H2BI, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5, Chr1(q21.1) and Chr6(p22.1) is meant to include all regions relevant for expression of a given gene, both the coding region and upstream and downstream regions, which may comprise cis-acting activating signals. A gene locus, specifically is
30 meant to include at least 1000 bp upstream and/or downstream of the open reading frame of an encoded gene, such as at least 900 bp, such as at least 800 bp, such as at least 700 bp, such as at least 600 bp, such as at least 500 bp, such as at least 400 bp, such as at least 300 bp, such as at least 200 bp, such as at least 100 bp upstream and/or downstream of the open reading frame of an encoded gene. A gene locus is
35 also meant to include any intronic sequences in the open reading frame. It is also

understood that specific subregions of a gene locus specified herein can be of particular importance for the methods described herein. In particular CG-rich regions also known as CpG islands are particularly relevant, because CG-dinucleotides are targets for methylation.

5

Method of determining non-small cell lung cancer

A number of methods are provided herein for analysing a human subject with respect to non-small cell lung cancer. In particular, methods are provided for determining non-small cell lung cancer in a human subject, methods for determining a predisposition to non-small cell lung cancer for a human subject, methods for determining the prognosis of a non-small cell lung cancer in a subject and/or inferring a suitable treatment, methods for categorizing or staging a non-small cell lung cancer of a human subject, methods for monitoring a non-small cell lung cancer, such as monitoring the treatment of a non-small cell lung cancer and/or relapse of a non-small cell lung cancer. The methylation biomarkers for non-small cell lung cancer are described in more detailed herein below. Generally, the one or more methylation biomarkers for non-small cell lung cancer according to the methods herein are selected from a gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

Thus in one aspect, a method is provided for determining non-small cell lung cancer, a predisposition to non-small cell lung cancer, the prognosis of a non-small cell lung cancer, and/or monitoring a non-small cell lung cancer in a subject, said method comprising in a sample from said subject determining the methylation status of at least one gene including regulatory sequences of said gene, wherein said gene locus is selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

30

In another aspect, a method is provided for categorizing or predicting the clinical outcome of a non-small cell lung cancer of a subject, said method comprising in a sample from said subject determining the methylation status of at least one gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI,

HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

5 In another aspect, a method is provided for evaluating the risk for a human subject of developing non-small cell lung cancer, or for monitoring relapse of a non-small cell lung cancer, said method comprising in a sample from said subject determining the methylation status of a gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, 10 HOXA5 and Chr1(q21.1).

A further aspect relates to a method for assessing whether a human subject is likely to develop non-small cell lung cancer, said method comprising

- i) providing a sample from said human subject,
- 15 ii) determining in said sample the methylation status of at least one gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1),
- iii) on the basis of said methylation status identifying a human subject that is 20 more likely to develop non-small cell lung cancer.

The methods thus involve determining the methylation status of one or more gene loci as defined herein. Thus, methylation status may be determined for multiple gene loci, for example methylation status for at least two gene loci are determined, such as at 25 least three gene loci, such as at least four gene loci, or five or more gene loci. The plurality of gene loci is preferably selected from a marker gene loci of the invention, i.e. a gene loci selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and 30 Chr1(q21.1).

Generally, increased levels of methylation positive alleles of the respective marker gene locus relative to methylation levels of a predetermined control sample of non-cancer cells is indicative of the presence of a non-small cell lung cancer, higher

likelihood of developing cancer, decreased overall survival, negative outcome, different stage cancer, different grade cancer, and/or higher risk of contracting cancer.

Thus, the provided methods may preferably comprise steps of comparing the methylation status of the respective gene locus determined for a subject with a predetermined methylation status for the corresponding gene of a reference sample comprising non-cancer cells, for example tumour adjacent normal lung tissue cells, and/or comprising a different stage cancer cells. The predetermined status is preferably determined from non-cancer cells of other subjects, which do not have non-small cell lung cancer and/or are not predisposed to non-small cell lung cancer. The predetermined methylation status differs between the different methylation markers of the invention.

Table A: Overview of sensitivities and specificities for each marker for preferred cut-off values.

ID	Cutoff Value	Sensitivity	Specificity
	Methylation Level		
<i>OSR1</i>	>1%	0.98	0.97
<i>SIM1</i>	>1%	0.92	0.90
<i>HOXD10</i>	>1%	0.90	0.81
<i>HIST1H3E</i>	>50%	0.90	0.91
<i>HOXD3</i>	>10%	0.86	0.97
<i>GHSR</i>	>10%	0.83	0.97
Chr1(q21.1).A	>10%	0.79	1.00
<i>HOXB3/HOXB4</i>	>1%	0.75	1.00
Chr6(p22.1)	>1%	0.67	1.00
<i>HIST1H3G/HIST1H2BI</i>	>1%	0.63	1.00
<i>OTX2</i>	>1%	0.51	1.00
<i>LOC648987</i>	>1%	0.24	1.00
<i>HIST1H2AJ/HIST1H2BM</i>	>1%	0.22	1.00
<i>HOXA3</i>	>50%	0.14	1.00
<i>HOXA5</i>	>50%	0.12	1.00

Table A shows an overview of cut-off values, which are preferred for each of the representative marker loci.

For example, for loci ID SIM1, any level of methylation positive alleles above 1 % is indicative of a non-small cell lung cancer, higher likelihood of developing cancer, decreased overall survival, negative outcome, different stage cancer, different grade cancer, and/or higher risk of contracting cancer for a human subject. The same applies to the other genetic loci listed in table A above.

Thus, for the methylation marker loci identified as SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, HOXD10, OTX2, LOC648987 and HIST1H2AJ/HIST1H2BM, a level of methylation positive alleles above 1%, such as above 2%, 3%, 4%, 5%, such as above 10%, such as above 15%, such as above 20%, such as above 25%, such as above 30%, such as above 35%, such as above 40%, such as above 45%, such as above 50%, such as above 55%, such as above 60%, such as above 65%, such as preferably above 65.3%, such as above 70%, such as above 75%, such as above 80%, such as above 85%, such as above 90%, such as above 95%, such as above 96%, 97%, 98%, or 99%, such as 100% is indicative of non-small cell lung cancer, a predisposition to non-small cell lung cancer, increased risk of non-small cell lung cancer, the prognosis of non-small cell lung cancer, and/or relapse of non-small cell lung cancer, and thus indicates that a given treatment being monitored is inefficient.

Specifically, a methylation level of the methylation marker locus OSR1 above 1%, such as above 2%, 3%, 4%, 5%, such as above 10%, such as above 15%, such as above 20%, such as above 25%, such as above 30%, such as above 35%, such as above 40%, such as above 45%, such as above 50%, such as above 55%, such as above 60%, such as above 65%, such as preferably above 65.3%, such as above 70%, such as above 75%, such as above 80%, such as above 85%, such as above 90%, such as above 95%, such as above 96%, 97%, 98%, or 99%, such as 100% is indicative of non-small cell lung cancer, a predisposition to non-small cell lung cancer, increased risk of non-small cell lung cancer, the prognosis of non-small cell lung cancer, and/or relapse of non-small cell lung cancer, and thus indicates that a given treatment being monitored is inefficient.

Specifically, a methylation level of the methylation marker locus SIM1 above 1%, such as above 2%, 3%, 4%, 5%, such as above 10%, such as above 15%, such as above 20%, such as above 25%, such as above 30%, such as above 35%, such as above 40%, such as above 45%, such as above 50%, such as above 55%, such as above 60%, such as above 65%, such as preferably above 65.3%, such as above 70%, such as above 75%, such as above 80%, such as above 85%, such as above 90%, such as above 95%, such as above 96%, 97%, 98%, or 99%, such as 100% is indicative of non-small cell lung cancer, a predisposition to non-small cell lung cancer, increased risk of non-small cell lung cancer, the prognosis of non-small cell lung cancer, and/or relapse of non-small cell lung cancer, and thus indicates that a given treatment being monitored is inefficient.

Specifically, a methylation level of the methylation marker locus HOXD10 above 1%, such as above 2%, 3%, 4%, 5%, such as above 10%, such as above 15%, such as above 20%, such as above 25%, such as above 30%, such as above 35%, such as above 40%, such as above 45%, such as above 50%, such as above 55%, such as above 60%, such as above 65%, such as preferably above 65.3%, such as above 70%, such as above 75%, such as above 80%, such as above 85%, such as above 90%, such as above 95%, such as above 96%, 97%, 98%, or 99%, such as 100% is indicative of non-small cell lung cancer, a predisposition to non-small cell lung cancer, increased risk of non-small cell lung cancer, the prognosis of non-small cell lung cancer, and/or relapse of non-small cell lung cancer, and thus indicates that a given treatment being monitored is inefficient.

Specifically, a methylation level of the methylation marker locus HOXB3/HOXB4 above 1%, such as above 2%, 3%, 4%, 5%, such as above 10%, such as above 15%, such as above 20%, such as above 25%, such as above 30%, such as above 35%, such as above 40%, such as above 45%, such as above 50%, such as above 55%, such as above 60%, such as above 65%, such as preferably above 65.3%, such as above 70%, such as above 75%, such as above 80%, such as above 85%, such as above 90%, such as above 95%, such as above 96%, 97%, 98%, or 99%, such as 100% is indicative of non-small cell lung cancer, a predisposition to non-small cell lung cancer, increased risk of non-small cell lung cancer, the prognosis of non-small cell lung cancer, and/or relapse of non-small cell lung cancer, and thus indicates that a given treatment being monitored is inefficient.

Specifically, a methylation level of the methylation marker locus Chr6(p22.1) above 1%, such as above 2%, 3%, 4%, 5%, such as above 10%, such as above 15%, such as above 20%, such as above 25%, such as above 30%, such as above 35%, such as above 40%, such as above 45%, such as above 50%, such as above 55%, such as above 60%, such as above 65%, such as preferably above 65.3%, such as above 70%, such as above 75%, such as above 80%, such as above 85%, such as above 90%, such as above 95%, such as above 96%, 97%, 98%, or 99%, such as 100% is indicative of non-small cell lung cancer, a predisposition to non-small cell lung cancer, increased risk of non-small cell lung cancer, the prognosis of non-small cell lung cancer, and/or relapse of non-small cell lung cancer, and thus indicates that a given treatment being monitored is inefficient.

Specifically, a methylation level of the methylation marker locus HIST1H3G/HIST1H2BI above 1%, such as above 2%, 3%, 4%, 5%, such as above 10%, such as above 15%, such as above 20%, such as above 25%, such as above 30%, such as above 35%, such as above 40%, such as above 45%, such as above 50%, such as above 55%, such as above 60%, such as above 65%, such as preferably above 65.3%, such as above 70%, such as above 75%, such as above 80%, such as above 85%, such as above 90%, such as above 95%, such as above 96%, 97%, 98%, or 99%, such as 100% is indicative of non-small cell lung cancer, a predisposition to non-small cell lung cancer, increased risk of non-small cell lung cancer, the prognosis of non-small cell lung cancer, and/or relapse of non-small cell lung cancer, and thus indicates that a given treatment being monitored is inefficient.

Specifically, a methylation level of the methylation marker locus OTX2 above 1%, such as above 2%, 3%, 4%, 5%, such as above 10%, such as above 15%, such as above 20%, such as above 25%, such as above 30%, such as above 35%, such as above 40%, such as above 45%, such as above 50%, such as above 55%, such as above 60%, such as above 65%, such as preferably above 65.3%, such as above 70%, such as above 75%, such as above 80%, such as above 85%, such as above 90%, such as above 95%, such as above 96%, 97%, 98%, or 99%, such as 100% is indicative of non-small cell lung cancer, a predisposition to non-small cell lung cancer, increased risk of non-small cell lung cancer, the prognosis of non-small cell lung cancer, and/or

relapse of non-small cell lung cancer, and thus indicates that a given treatment being monitored is inefficient.

Specifically, a methylation level of the methylation marker locus LOC648987 above 1%, such as above 2%, 3%, 4%, 5%, such as above 10%, such as above 15%, such as above 20%, such as above 25%, such as above 30%, such as above 35%, such as above 40%, such as above 45%, such as above 50%, such as above 55%, such as above 60%, such as above 65%, such as preferably above 65.3%, such as above 70%, such as above 75%, such as above 80%, such as above 85%, such as above 90%, such as above 95%, such as above 96%, 97%, 98%, or 99%, such as 100% is indicative of non-small cell lung cancer, a predisposition to non-small cell lung cancer, increased risk of non-small cell lung cancer, the prognosis of non-small cell lung cancer, and/or relapse of non-small cell lung cancer, and thus indicates that a given treatment being monitored is inefficient.

Specifically, a methylation level of the methylation marker locus HIST1H2AJ/HIST1H2BM above 1%, such as above 2%, 3%, 4%, 5%, such as above 10%, such as above 15%, such as above 20%, such as above 25%, such as above 30%, such as above 35%, such as above 40%, such as above 45%, such as above 50%, such as above 55%, such as above 60%, such as above 65%, such as preferably above 65.3%, such as above 70%, such as above 75%, such as above 80%, such as above 85%, such as above 90%, such as above 95%, such as above 96%, 97%, 98%, or 99%, such as 100% is indicative of non-small cell lung cancer, a predisposition to non-small cell lung cancer, increased risk of non-small cell lung cancer, the prognosis of non-small cell lung cancer, and/or relapse of non-small cell lung cancer, and thus indicates that a given treatment being monitored is inefficient.

For the methylation marker loci identified as HOXD3, GHSR and Chr1(q21.1).A, a level of methylation positive alleles above 10%, such as above 15%, such as above 20%, such as above 25%, such as above 30%, such as above 35%, such as above 40%, such as above 45%, such as above 50%, such as above 55%, such as above 60%, such as above 65%, such as preferably above 65.3%, such as above 70%, such as above 75%, such as above 80%, such as above 85%, such as above 90%, such as above 95%, such as above 96%, 97%, 98%, or 99%, such as 100% is indicative of non-small cell lung cancer, a predisposition to non-small cell lung cancer, increased

risk of non-small cell lung cancer, the prognosis of non-small cell lung cancer, and/or relapse of non-small cell lung cancer, and thus indicates that a given treatment being monitored is inefficient.

- 5 Indeed for methylation marker loci identified as HIST1H3E, HOXA3 and HOXA5, a level of methylation positive alleles above 50%, such as above 55%, such as above 60%, such as above 65%, such as above 70%, such as above 75%, such as above 80%, such as above 85%, such as above 90%, such as above 95%, such as above 96%, 97%, 98%, or 99%, such as 100% is indicative of non-small cell lung cancer, a
10 predisposition to non-small cell lung cancer, increased risk of non-small cell lung cancer, the prognosis of non-small cell lung cancer, and/or relapse of non-small cell lung cancer, and thus indicates that a given treatment being monitored is inefficient.

Method for treatment of non-small cell lung cancer

- 15 Aspects provided herein also relates to methods for determining the prognosis of a non-small cell lung cancer in a subject and/or inferring a suitable treatment, as well as for monitoring a non-small cell lung cancer, and in particular monitoring the treatment of a non-small cell lung cancer and/or monitoring relapse of a non-small cell lung cancer.

20

So in one aspect, a method is provided for treatment of non-small cell lung cancer in a human subject, the method comprises the steps of

- i. determining non-small cell lung cancer, a predisposition to non-small cell lung cancer, or the prognosis of a non-small cell lung cancer in a subject by a method
25 of the present invention, as defined elsewhere herein,
- ii. selecting human subjects having non-small cell lung cancer, a predisposition to non-small cell lung cancer, or a relapse of a non-small cell lung cancer,
- iii. subjecting said subjects identified in step ii. to a suitable treatment for
30 non-small cell lung cancer.

The step of determining non-small cell lung cancer by a method of the present invention allows early detection of non-small cell lung cancer, and therefore allows treatment of the cancer to be initiated before developing into later stages and/or before
35 forming metastases. This allows the use of less serious types of therapeutic

interventions, and may for example avoid the need for surgery, such as surgical removal of the entire lung. In one embodiment, the selected human subject is subjected to a treatment selected from surgery, chemotherapy, immunotherapy and/or radiotherapy, however, in a preferred embodiment, the treatment is radiotherapy. In one embodiment, the treatment is a combination of surgery, chemotherapy and radiotherapy, for example surgery followed by chemotherapy and/or radiotherapy.

The methylation markers also allow monitoring relapse of non-small cell lung cancer, as well as offering a personalized treatment of non-small cell lung cancer by surveillance and quality of control of the treatment offered, thereby allowing terminating ineffective treatments and offering alternative treatments. Thus, in another aspect, the invention provides a method for personalized treatment of a non-small cell lung cancer of a human subject, said method comprising

- i) in a sample from said human subject, determining the methylation status of a gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1)
- ii) providing a treatment of non-small cell lung cancer to said human subject,
- iii) after a sufficient amount of time having provided the treatment, in a sample from said human subject, determining the methylation status of said gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1),
- iv) comparing the methylation status of said gene locus before and after treatment, and
- v) if methylation of said genetic locus is similar to the methylation before treatment, terminating said provided treatment and preferably offering an alternative treatment, or
- vi) if methylation of said genetic locus is reduced relative to the methylation before treatment, continuing said provided treatment or even terminating the treatment.

Methylation biomarkers for non-small cell lung cancer

As described herein above, a number of different methods are provided herein for evaluating non-small cell lung cancer in a human subject based on methylation status of specific gene loci. The invention also provides specific oligonucleotide primers and kits for use in determining methylation status of specific gene loci, which are methylation biomarkers for non-small cell lung cancer according to the present

invention. These gene loci include SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5, Chr1(q21.1), such as those identified by SEQ ID NO: 1-36.

5

Generally, in the methods of the invention, the methylation status is determined for at least one gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and
10 Chr1(q21.1). In one embodiment, the methylation status is determined for a gene locus selected from the group consisting of OSR1, SIM1, HOXD10, HIST1H3E, HOXD3, GHSR, Chr1(q21.1) and/or HOXB3/HOXB4. In another embodiment, the methylation status is determined for a gene locus selected from the group consisting of SIM1, HIST1H3G/HIST1H2BI, Chr6(p22.1) and/or HOXB3/HOXB4. In another embodiment,
15 the methylation status is determined for a gene locus selected from the group consisting of SIM1, HIST1H3G/HIST1H2BI and/or Chr6(p22.1).

In one embodiment, the methylation status is determined for a gene locus selected from the group consisting of SIM1 and HIST1H3G/HIST1H2BI; or the group consisting
20 of SIM1 and Chr6(p22.1); or the group consisting of HIST1H3G/HIST1H2BI and Chr6(p22.1).

In one embodiment, the methylation status is determined for SIM1 and/or Chr6(p22.1) and/or HIST1H3G/HIST1H2BI and/or HOXB3/HOXB4 and/or OSR1 and/or HOXD10
25 and/or OTX2 and/or LOC648987 and/or HIST1H2AJ/HIST1H2BM.

In one embodiment, the methylation status is determined for one gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E,
30 HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1). In further embodiments, the methylation status is determined for that one gene locus and at least one gene locus selected from the remainder of the group.

Thus, in specific one embodiment, methylation status is determined in the SIM1 gene
35 locus and at least one additional gene locus selected from the group consisting of

Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

5 In another specific embodiment, methylation status is determined in the Chr6(p22.1) gene locus and at least one additional gene locus selected from the group consisting of SIM1, HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

10

In another specific embodiment, methylation status is determined in the HIST1H3G/HIST1H2BI gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

15

In another specific embodiment, methylation status is determined in the HOXB3/HOXB4 gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

20

In another specific embodiment, methylation status is determined in the OSR1 gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

25

In another specific embodiment, methylation status is determined in the GHSR gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

30

In another specific embodiment, methylation status is determined in the OTX2 gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

In another specific embodiment, methylation status is determined in the LOC648987 gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

In another specific embodiment, methylation status is determined in the HIST1H3E gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

In another specific embodiment, methylation status is determined in the HIST1H2AJ/HIST1H2BM gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

In another specific embodiment, methylation status is determined in the HOXD10 gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

In another specific embodiment, methylation status is determined in the HOXD3 gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXA3, HOXA5 and Chr1(q21.1).

In another specific embodiment, methylation status is determined in the HOXA3 gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHRSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA5 and Chr1(q21.1).

In another specific embodiment, methylation status is determined in the HOXA5 gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHRSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, and Chr1(q21.1).

In another specific embodiment, methylation status is determined in the Chr1(q21.1) gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHRSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3 and HOXA5.

DNA sequences of specific gene loci are provided herein below. Thus, in a preferred embodiment, the methylation status is determined in a gene locus identified by SEQ ID NO: 1-18. The corresponding sequences of the same loci after bisulfite modification and amplification are identified by SEQ ID NO: 19-36, respectively.

In a preferred embodiment, the methylation status is determined by a method comprising amplifying a gene locus of the invention using at least one primer selected from the group consisting of SEQ ID NO: 37-72. Methylation status is preferably determined for a gene locus mentioned in table B using the respective forward primer and/or reverse primer identified in table B; i.e.

30

SIM1: forward primer SEQ ID NO: 51 and/or reverse primer SEQ ID NO: 52;
Chr6(p22.1): forward primer SEQ ID NO: 57 and/or reverse primer SEQ ID NO: 58;
etc.

Thus, in one preferred embodiment, the methylation status is determined in a genetic region of a gene locus of the invention, wherein said region is delineated by the primer pairs identified in table B for each respective gene; i.e.

for OTX2: primers SEQ ID NO: 37 and/or 38;

5 for HOXD3: primers SEQ ID NO: 39 and/or 40;...

For SIM1: primers SEQ ID NO: 51 and/or 52;

For Chr6(p22.1): primers SEQ ID NO: 57 and/or 58;

For HIST1H3G/HIST1H2BI: primers SEQ ID NO: 47 and/or 48;

etc...; and

10 for HIST1H3E: primers SEQ ID NO: 71 and/or 72.

Table B. Markers Sequence table

Marker gene/Loci ID	Non-modified Gene/loci SEQ ID NO	Modified Gene/loci SEQ ID NO	Forward primer SEQ ID NO	Reverse primer SEQ ID NO
OTX2	1	19	37	38
HOXD3	2	20	39	40
HOXB3/HOXB4	3	21	41	42
HOXD10	4	22	43	44
Chr1(q21.1)	5	23	45	46
HIST1H3G/HIST1H2BI	6	24	47	48
GHSR	7	25	49	50
SIM1	8	26	51	52
OSR1	9	27	53	54
FRG1B	10	28	55	56
Chr6(p22.1)	11	29	57	58
HOXA3	12	30	59	60
LY75-CD302	13	31	61	62
CTAGE15	14	32	63	64
LOC648987	15	33	65	66
HIST1H2AJ/HIST1H2BM	16	34	67	68
HOXA5	17	35	69	70
HIST1H3E	18	36	71	72

Sample

According to the present invention, the methylation status of one or more gene loci is determined in a sample from a human subject. Thus, the sample of the invention comprises biological material, in particular genetic material comprising nucleic acid molecules. The nucleic acid molecules may be extracted from the sample prior to the analysis. The sample may be obtained or provided from any human source. In one embodiment, determination of methylation status of a gene locus or genetic region of the invention is performed on samples selected from the group consisting of lung tissue, hematopoietic tissue, bone marrow, expiration air, stem cells, including cancer stem cell, and body fluids, such as sputum, bronchial lavage, urine, blood and sweat.

In preferred embodiments the sample is or comprises lung tissue, such as lung cells and/or genetic material of lung cells.

It is well-known that tumor DNA may leak to the blood stream or other bodily fluids, so in one preferred embodiment, the sample is a body fluid, such as sputum, urine, blood and sweat. In particular, it is preferred that the sample is a blood or plasma sample. Body fluids are often retrievable by less invasive methods than lung tissue, which must be obtained surgically for example by biopsies.

The provided sample is in one embodiment a formalin-fixed paraffin-embedded (ffpe) sample, for example an ffpe sample, wherein prestages to non-small cell lung cancer can be seen. In particular, the sample used for predetermining methylation status can be an ffpe sample. Many ffpe samples may be provided, which can give rise to statistically strong predetermined values with respect to evaluation of non-small cell lung cancer risk, categorizing or staging a non-small cell lung cancer of a human subject, methods for monitoring a non-small cell lung cancer, such as monitoring the treatment of a non-small cell lung cancer and/or relapse of a non-small cell lung cancer.

The nucleic acid to be analysed for the presence of methylated CpG may be extracted from the samples by a variety of techniques such as that described by Maniatis, et al (Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, N.Y., pp 280, 281, 1982). However, the sample may be used directly.

Any nucleic acid, in purified or nonpurified form, can be utilized as the starting nucleic acid or acids, provided it contains, or is suspected of containing, the specific nucleic acid sequence containing the methylation target site (e.g., CpG). The specific nucleic acid sequence which is to be amplified may be a part of a larger molecule or is present initially as a discrete molecule. The nucleic acid sequence to be amplified need not to be present in a pure form, it may for example be a fraction of a complex mixture of other DNA molecules, and/or RNA. In one example, the nucleic acid sequence is a fraction of a genomic nucleic acid preparation.

Extremely low amounts of nucleic acid may be used as target sequence according to the methods of the present invention. It is appreciated by the person skilled in the art that in practical terms no upper limit for the amount of nucleic acid to be analysed exists. The problem that the skilled person may encounter is that the amount of sample to be analysed is limited. Therefore, it is beneficial that the method of the present invention can be performed on a small amount of sample and thus a limited amount of nucleic acid in said sample. The present methods allow the detection of only very few nucleic acid copies. The amount of the nucleic acid to be analysed is in one embodiment at least 0.01 ng, such as 0.1 ng, such as 0.5 ng, for example 1 ng, such as at least 10 ng, for example at least 25 ng, such as at least 50 ng, for example at least 75 ng, such as at least 100 ng, for example at least 125 ng, such as at least 150 ng, for example at least 200 ng, such as at least 225 ng, for example at least 250 ng, such as at least 275 ng, for example at least 300 ng, 400 ng, for example at least 500 ng, such as at least 600 ng, for example at least 700 ng, such as at least 800, ng, for example at least 900 ng or such as at least 1000 ng.

In one preferred embodiment the amount of nucleic acid as the starting material for the method of the present invention is approximately 50 ng, alternatively 100 ng or 200 ng.

Methylation status

The methods of the present invention for determining non-small cell lung cancer in a human subject, methods for determining a predisposition to non-small cell lung cancer for a human subject, methods for determining the prognosis of a non-small cell lung cancer in a subject and/or inferring a suitable (personalized) treatment, methods for categorizing or staging a non-small cell lung cancer of a human subject and methods for monitoring a non-small cell lung cancer, all include a step of providing or obtaining a

sample from the human subject, and in that sample determining the methylation status of at least one genetic locus selected from the group consisting of SIM1, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H3G/HIST1H2BI, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5, Chr1(q21.1) and Chr6(p22.1), as well as subregions thereof, in particular GC-rich subregions, such as those delineated by the respective primer pairs identified in table B.

Methylation status of the target gene loci or genetic regions of the present invention may be determined by any suitable method available to the skilled person for detecting methylation status. However, in a preferred embodiment, methylation status is determined by a quantitative method, such as a semi-quantitative method, which is capable of detecting levels of methylation positive alleles and/or methylation negative alleles in a population of target molecules present in a sample. For example, the quantitative method is preferably capable of detecting different levels of methylation positive alleles of a given target locus sequence, such as detecting whether 0%, less than 1%, more than 1%, such as approximately 10%, 25%, 50%, 75% or 100% of the alleles of a given marker locus are methylation positive. A semi-quantitative method provides categorical data, such as the level of 1-10% methylated templates or 10-50% methylated templates. Some techniques in the art merely detect the presence of one or more methylation positive and/or methylation negative alleles of a given target sequence without providing quantitative data, and without providing information of the relative levels of methylation positive and methylation negative alleles. However, preferred methods of the present invention provide a quantitative measure of the relative level of methylation positive alleles of a specific target region.

The term "methylation status" as used herein, refers to the extent to which a nucleic acid region and/or in particular a CpG methylation site is methylated or unmethylated, which may be expressed as the methylation level of a given sample. The methylation status of a single CpG methylation site can be either methylated or unmethylated. A nucleic acid sequence comprising multiple potential methylation (CpG) sites, may be methylated on only a subset of those CpG sites. Such nucleic acid molecules/alleles are heterogeneous methylated. The term "methylation status", thus, refers to whether a nucleic acid sequence is methylation positive (methylated on all CpG sites), is

methylation negative (all CpG sites of the sequence is unmethylated), or is heterogeneous methylated (a subset of CpG sites of the sequence is methylated). The methods for inferring non-small cell lung cancer of the present invention, thus determine methylation status of specific methylation markers by determining whether a specific methylation marker in a sample obtained or provided from a subject is methylation positive, methylation negative or heterogeneously methylated, as well as detecting the relative level of methylated alleles of a given locus. The methods may also include detecting marker sequences with low methylation, which defines methylation of less than 1% of the alleles of a sample.

Methylation status may be determined by any suitable method available to those of skill in the art. For example, method may be selected from the group consisting of Methylation-Sensitive High Resolution Melting (MS-HRM), Methylation - Sensitive Melting Curve Analysis (MS-MCA), EpiTyper, Methylation-Specific PCR (MSP), DNA methylation specific qPCR (qMSP), Pyrosequencing, Methyl Light, Amplicon bisulfite sequencing (AmpliconBS), Enrichment bisulfite sequencing (EnrichmentBS), Whole genome bisulfite sequencing (BS-Seq), HELP assays, and Methyl Sensitive Southern Blotting, and determination may involve Methylated DNA immunoprecipitation (MeDIP). In one embodiment, the method is selected from the group consisting of AmpliconBS 1, AmpliconBS 2, AmpliconBS 3, AmpliconBS 4, EnrichmentBS 1, EnrichmentBS 2, EpiTyper 1, EpiTyper 3, Infinium, Pyroseq 1, Pyroseq 1 (replicate), Pyroseq 2, Pyroseq 3, Pyroseq 4, Pyroseq 5, MethyLight, MS-HRM, MS-MCA, qMSP (preamp), qMSP (standard), DNA-methylation-specific amplification by qPCR, HPLC-MS, Immunoquant, Pyroseq AluYb8, Bisulfite pyrosequencing using primers that amplify AluYb8 repetitive DNA, Pyroseq D4Z4, Pyroseq LINE1, Pyroseq NBL2 and ClonalBS; cf. Bock et al, 2016, nature biotech. 34 (7).

In another embodiment, the method is selected from the group consisting of High-performance liquid chromatography (HPLC), High-performance capillary electrophoresis (HPCE), Sssl assay, Gene specific Methylation-specific PCR (MSP-PCR), Methyl-sensitive restriction enzyme PCR (MSRE-PCR), MethyLight, Pyrosequencing, Methylation-sensitive single nucleotide Primer extension (MS-SNuPE), Combined bisulfite restriction analysis (COBRA), Methylation sensitive-high resolution melting (MS-HRM), Methylation-specific multiplex ligationdependent probe amplification (MS-MLPA), Mass ARRAY EpiTYPER, Restriction landmark genomic

scanning (RLGS), Differential methylation hybridization (DMH), Methylated DNA immunoprecipitation and microarray chip (MeDIPchip), Bead arrays (Illumina) Bisulfite, Whole-genome bisulfite sequencing, Single molecule real time (SMRT) sequencing and MethylCap sequencing; cf. Syedmoradi et al, 2016, Royal Soc of Chem. (DOI: 10.1039/c6an01649a).

In one embodiment of methods of the present invention for determining non-small cell lung cancer in a human subject, for determining a predisposition to non-small cell lung cancer for a human subject, for determining the prognosis of a non-small cell lung cancer in a subject and/or inferring a suitable treatment, for categorizing or staging a non-small cell lung cancer of a human subject, and/or for monitoring a non-small cell lung cancer, such as monitoring the treatment of a non-small cell lung cancer and/or relapse of a non-small cell lung cancer, the methylation status is determined by use of methylation-sensitive restriction enzymes. Many restriction enzymes are sensitive to the DNA methylation states. Cleavage can be blocked or impaired when a particular base in the recognition site is modified. For example, the MspJI family of restriction enzymes has been found to be dependent on methylation and hydroxymethylation for cleavage to occur. These enzymes excise ~ 32 base pair fragments containing a centrally located 5-hmC or 5-mC modified residue that can be extracted and sequenced. Due to the known position of this epigenetic modification, bisulfite conversion is not required prior to downstream analysis.

Methylation-sensitive enzymes are well-known in the art and include:

AatII, AccII, Aor13HI, Aor51HI, BspT104I, BssHII, Cfr10I, ClaI CpoI, Eco52I, HaeII, HapII, HhaI, MluI, NaeI, NotI, NruI, NsbI, PmaCI, Psp1406I, PvuI, SacII, Sall, SmaI and SnaBI.

The digested nucleic acid sample is subsequently analysed by for example gel electrophoresis.

So, in one embodiment of the methods of the invention, methylation status is determined by a method comprising the steps of

- i) providing a sample, such as a lung tissue sample or a blood or plasma sample from said subject comprising nucleic acid material comprising said gene,
- ii) processing said nucleic acid sequence using one or more methylation-sensitive restriction endonuclease enzymes,

- iii) optionally, amplifying said processed nucleic acid sequence in order to obtain an amplification product, and
- iv) analyzing said processed nucleic acid sequence or said amplification product for the presence of processed and/or unprocessed nucleic acid sequences, thereby inferring the presence of methylated and/or unmethylated nucleic acid sequences.

In a preferred embodiment of methods of the present invention, the methodology employed for determining methylation status is determined by a method, which comprises at least the steps of modifying the DNA with an agent which targets either methylated or unmethylated sequences, amplifying the DNA, and analysing the amplification products.

For example, amplification product is analysed by detecting the presence or absence of amplification product, wherein the presence of amplification product indicates that the target nucleic acid has not been cleaved by the restriction enzymes, and wherein the absence of amplification product indicates that the target nucleic acid has been cleaved by the restriction enzymes.

Thus, generally, the in the methods of the invention methylation status is determined by a method comprising the steps of

- i) providing a sample, such as a lung tissue sample or a blood or plasma sample from said subject comprising nucleic acid material comprising a gene locus of the invention,
- ii) modifying said nucleic acid material using an agent, which modifies nucleic acid sequences in a methylation-dependent manner,
- iii) amplifying at least one portion of said gene locus using primers, which span or comprise at least one CpG dinucleotide in said gene locus in order to obtain an amplification product, and
- iv) analyzing said amplification product for the presence of modified and/or unmodified cytosine residues, wherein the presence of modified cytosine residues are indicative of methylated cytosine residues.

For example, the method comprises the steps of

- i) providing a sample, such as a lung tissue sample, from said subject comprising nucleic acid material comprising said gene locus,
- ii) modifying said nucleic acid using an agent which modifies unmethylated cytosine,
- iii) amplifying at least one portion of said gene locus using primers, which span or comprise at least one CpG dinucleotide in said gene locus in order to obtain an amplification product, and
- iv) analyzing said amplification product.

The amplification product can be analysed for nucleic acid substitutions resulting from conversion of modified cytosine residues, preferably wherein the presence of converted cytosine residues are indicative of unmethylated cytosine residues, and presence of unconverted cytosine residues is indicative of methylated cytosine residues. Typically, unmethylated cytosine is converted to thymidine after bisulphite treatment and amplification, while methylated cytosine is left unchanged after same treatment.

In a preferred embodiment, the amplification product is analysed by melting curve analysis; cf. herein below.

The amplification product, the amplicon, is in a preferred embodiment a genetic region of a gene of the invention, wherein said region is delineated by the primer pairs identified in table B for each respective gene; i.e.

for OTX2: primers SEQ ID NO: 37 and/or 38;

for HOXD3: primers SEQ ID NO: 39 and/or 40;

...

For SIM1: primers SEQ ID NO: 51 and/or 52;

For Chr6(p22.1): primers SEQ ID NO: 57 and/or 58;

For HIST1H3G/HIST1H2BI: primers SEQ ID NO: 47 and/or 48;

etc...; and

for HIST1H3E: primers SEQ ID NO: 71 and/or 72.

Modification of DNA

The method for determining methylation status in the present invention preferably comprise a step of modifying the nucleic acids comprised in the sample, or extracted from the sample, using an agent which specifically modifies unmethylated cytosine in the nucleic acid. As used herein the term "modifies" refers the specific modification of

either an unmethylated cytosine or a methylated cytosine, for example the specific conversion of an unmethylated cytosine to another nucleotide which will distinguish the modified unmethylated cytosine from a methylated cytosine. In one preferred embodiment, an agent modifies unmethylated cytosine to uracil. Such an agent may be any agent conferring said conversion, wherein unmethylated cytosine is modified, but not methylated cytosine. In one preferred embodiment the agent for modifying unmethylated cytosine is sodium bisulfite. Sodium bisulfite (NaHSO_3) reacts readily with the 5,6-double bond of cytosine, but only poorly with methylated cytosine. The cytosine reacts with the bisulfite ion, forming a reaction intermediate in the form of a sulfonated cytosine which is prone to deamination, eventually resulting in a sulfonated uracil. Uracil can subsequently be formed under alkaline conditions which removes the sulfonate group.

During a nucleic acid amplification process, uracil will by the Taq polymerase be recognised as a thymidine. The product upon PCR amplification of a Sodium bisulfite modified nucleic acid contains cytosine at the position where a methylated cytosine (5-methylcytosine) occurred in the starting template DNA of the sample. Moreover, the product upon PCR amplification of a Sodium bisulfite modified nucleic acid contains thymidine at the position where an unmethylated cytosine (5-methylcytosine) occurred in the starting template DNA of the sample. Thus, an unmethylated cytosine is converted into a thymidine residue upon amplification of a bisulfite modified nucleic acid.

In a preferred embodiment of the present invention, the nucleic acids are modified using an agent which modifies unmethylated cytosine in the nucleic acid. In a specific embodiment, such an agent is a bisulfite, hydrogen sulfite, and/or disulfite reagent, for example sodium bisulfite.

However, in another embodiment, an agent is used, which specifically modifies methylated cytosine in the nucleic acid and does not modify unmethylated cytosine.

Amplifying step

After modification of the nucleic acids of the sample, the specific genetic region selected for determination of methylation status is preferably amplified in order to generate and thereby obtain multiple copies (amplicons) of the respective genetic

regions, which can allow its further analysis with respect to methylation status. The amplification is preferably preformed using at least one oligonucleotide primer, which targets the specific genetic region comprising methylation markers for non-small cell lung cancer according to the present invention. Most preferably amplification is performed using two oligonucleotide primers, which delineates the analysed region. The skilled person may use his common general knowledge in designing suitable primers. However, in a preferred embodiment, at least one, and preferably two methylation-independent oligonucleotide primers are employed for amplification of the modified nucleic acid. The nature of methylation-independent primers is described on more detail herein below.

The amplifying step is a polymerisation reaction wherein an agent for polymerisation is involved, effecting an oligonucleotide primer extension. The agent for polymerization may be any compound or system which will function to accomplish the synthesis of primer extension products, including enzymes. Enzymes that are suitable for this purpose include, for example, E. coli DNA polymerase I, Klenow fragment of E. coli DNA polymerase I, T4 DNA polymerase, other available DNA polymerases, polymerase muteins, reverse transcriptase, and other enzymes, including heat-stable enzymes (i.e., those enzymes which perform primer extension after being subjected to temperatures sufficiently elevated to cause denaturation also known as Taq polymerases). Suitable enzymes will facilitate combination of the nucleotides in the proper manner to form the primer extension products which are complementary to each locus nucleic acid strand. Generally, the synthesis will be initiated at the 3' end of each primer and proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths. There may be agents for polymerization, however, which initiate synthesis at the 5' end and proceed in the other direction, using the same process as described above.

A preferred method for amplifying the modified nucleic acid by means of at least one methylation-independent oligonucleotide primer is by the polymerase chain reaction (PCR), as described herein and as is commonly used by those skilled in the art. It is appreciated that PCR amplification requires a set of oligonucleotide primers, one forward primer and one reverse primer. According to the present invention, the forward primer is a methylation independent primer. The reverse primer is in another embodiment a methylation independent primer. However, both reverse and forward

primer may be methylation independent oligonucleotide primers according to the definitions herein.

The amplification product (amplicon) may be of any length, however in one preferred embodiment, the amplification product comprise between 15 and 1000 nucleotides, such as between 15 and 500 nucleotides, such as between 50 and 120 nucleotides, preferably between 80 and 100 nucleotides. In a preferred embodiment, the amplicon is delineated by the primers identified in table B for each respective gene, cf. herein above.

- 10 The PCR reaction is characterised by three steps a) melting a nucleic acid template, b) annealing at least one methylation-independent oligonucleotide primer to said nucleic acid template, and c) elongating said at least one methylation-independent oligonucleotide primer.

15 *Melting*

The melting of a CpG-containing nucleic acid template may also be referred to as strand separation. Melting is necessary where the target nucleic acid contains two complementary strands bound together by hydrogen bonds. This strand separation can be accomplished using various suitable denaturing conditions, including physical, chemical, or enzymatic means. One physical method of separating nucleic acid strands involves heating the nucleic acid until it is denatured. The denaturation by heating is the preferred procedure for melting in the present invention. Heat denaturation involves temperatures ranging from about 60 degrees Celsius to 100 degrees Celsius. The time for melting may be in the range of 5 seconds to 10 minutes or even longer for initial melting of the template.

The melting temperature is typically between 80 and 90 degrees Celsius, such as at least 81, for example at least 82, such as at least 84, preferably at least 85, at least 86, such as at least 87, for example at least 88 degrees Celsius. The PCR reaction mixture is incubated at the melting temperature for at least 5 seconds, alternatively at least 10, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or at least 100 seconds.

Annealing

Separated strands are used as a template for the synthesis of additional nucleic acid strands. It is understood that the separated strands may result from the separation of complementary strands in an originally double stranded nucleic acid. However, separated strands originally single stranded are also used as templates according to the present invention. The synthesis of additional nucleic acid strands is performed under conditions that allow the hybridisation of oligonucleotide primers to templates. Such a step is herein referred to as annealing. The oligonucleotide primers form hydrogen bonds with the template.

The annealing temperature is between 40 and 75 degrees Celsius, such as at least 40, at least 45, for example at least 50, at least 52, at least 54, at least 56, at least 57, at least 58, at least 59 preferably at least 60, at least 61, at least 62, at least 63, at least 64, at least 65, at least 66, at least 67, for example at least 68, at least 69, at least 70, at least 72, at least 73, at least 75 degrees Celsius. The PCR reaction mixture is incubated at the annealing temperature for 1 to 100 seconds, such as at least 1, at least 2, at least 3, at least 4, preferably at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, alternatively at least 11, at least 13, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or at least 100 seconds.

In a specific embodiment of the present invention, the annealing temperature is between at least 15 degrees Celsius above the optimal annealing temperature, such as at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15 degrees Celsius above the optimal annealing temperature.

The optimal annealing temperature can be calculated by standard algorithms, as known to people skilled within the art. In one embodiment, the optimal primer annealing temperature (T_m) is calculated as: $T_m = 4(G + C) + 2(A + T)$, wherein G, C, A, T designates the number of the respective nucleotides. In another embodiment, the optimal primer annealing temperature (T_m) is calculated as:
 $T_m = 64.9^{\circ}\text{C} + 41^{\circ}\text{C} \times (\text{number of G's and C's in the primer} - 16.4)/N$, where N is the length of the primer. However, the annealing temperature should be empirically determined in respect of each specific primer. The modulation of the annealing

temperature is used to adjust hybridization stringency as described elsewhere herein. Thus, the optimal annealing temperature should be set at a level, wherein the PCR bias towards amplification of unmethylated nucleic acid template is balanced by the less efficient annealing of methylation-independent oligonucleotide primer according to the present invention to unmethylated nucleic acid target sequence.

Generally, the choice of annealing temperature depends on the sensitivity of the assay, and the composition of the sample with respect to the relative levels of methylation positive and methylation negative alleles. Thus, optimal annealing temperatures should preferably be determined for each sample. However, in one embodiment, the annealing temperature in respect of specific methylation-independent oligonucleotide primer according to the present invention is as specified below for each assay; cf. sequences.

Thus, in one embodiment of the methods and uses of the invention, the oligonucleotide primer is e.g. SEQ ID NO: 51 and 52 (assay 8), and the annealing temperature is 55°C. Similarly, specific annealing temperatures as well as PCR cycling conditions for preferred oligonucleotide primers of the invention is indicated below for each assay.

Elongation

The oligonucleotide primers annealed to the template is elongated to form an amplification product. The elongating temperature depends on optimum temperature for the polymerase, and is usually between 30 and 80 degrees Celsius. Typically, the elongating temperature is between 60 and 80 degrees Celsius, such as at least 60, at least 65, at least 68, at least 69, at least 70, preferably at least 71, at least 72, at least 73, at least 74, alternatively at least 75, at least 76, at least 77, at least 78, at least 79, at least 80 degrees Celsius. The PRC reaction mixture is incubated at the elongating temperature for 1 to 100 seconds, such as at least 1, at least 2, at least 3, at least 4, preferably at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, alternatively at least 11, at least 13, at least 15, at least 20, at least 25, at least 30, at least 35, at least 40, at least 45, at least 50, at least 55, at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, or at least 100 seconds.

Elongation occurs in a buffered aqueous solution, preferably at a pH of 7-9.

The two oligonucleotide primers are added to the reaction mixture in a molar excess of primer: template especially when the template is genomic DNA which will ensure an improved efficiency. Deoxyribonucleoside triphosphates dATP, dCTP, dGTP, and

dTTP are added to the reaction mixture, either separately or together with the primers. An appropriate agent for effecting the primer extension reaction, referred to and described elsewhere herein as an agent for polymerization is added to the reaction mixture. It is appreciated by a person skilled in the art that for PCR the agent for polymerisation preferable is a heat-stable polymerase enzyme, such as Taq polymerase.

Cycling

The PCR method comprises incubating the nucleic acid at a cycle of different specific temperatures in order to control the steps of amplification. The amplification buffer and polymerase required for PCR are well known to people of skill within the art.

The PCR reaction mixture is incubated sequentially at the melting temperature, the annealing temperature and the elongating temperature, respectively, for a number of cycles. The PCR reaction may run between 10 and 70 cycles. Typically, the PCR reaction run between 25 and 55 cycles, such as at least 25, at least 30, at least 35, at least 40, preferably at least 45, at least 50 or at least 55 cycles.

In one embodiment, cycles of melting, annealing and elongation consist of 10-80, 10-80 and 10-80 seconds, respectively. Optimal cycling intervals are easily determined by those of skill in the art. Specific embodiments of cycle intervals for melting, annealing and elongation are indicated in for each assay below together with preferred annealing temperature for the respective primers; i.e.:

Assay 1: Primers SEQ ID NO: 37 and 38: 1,1,1 minutes, respectively;

....

Assay 18: Primers SEQ ID NO: 71 and 72: 1,1,1 minutes, respectively;

PCR can be performed on a PCR machine, which is also known as a thermal cycler. Specifically, the thermal cycler may be coupled to a fluorometer, thus allowing the monitoring of the nucleic acid amplification in real time by use of intercalating fluorescent dyes, or other fluorescent probes. Applicable dyes according to the present invention include any DNA intercalating dye.

Suitable dyes include ethidium bromide, EvaGreen, LC Green, Syto9, SYBR Green, SensiMix HRM™ kit dye, however many dyes are available for this same purpose.

Real-time PCR allows for easy performance of quantitative PCR (qPCR), which is usually aided by algorithms comprised in the software, which is usually supplied with the PCR machines.

5

The fluorometer can furthermore be equipped with software that will allow interpretation of the results. Such software for data analyses may also be supplied with the kit of the present invention.

10 Another variant of the PCR technique, multiplex PCR, enables the simultaneous amplification of many targets of interest in one reaction by using more than one pair of primers.

15 PCR according to the present invention comprise all known variants of the PCR technique known to people of skill within the art. Thus, the PCR technology comprise real-time PCR, qPCR, multiplex PCR.

Oligonucleotide primers

20 The oligonucleotide primer employed for amplification of modified nucleic acid is preferably a methylation-independent primer. The term "methylation-independent primer" refers to an oligonucleotide primer, which is capable of hybridizing to both methylated and unmethylated nucleic acid alleles and modified as well as unmodified alleles. A methylation-independent primer may not anneal with the exact same affinity to methylated/unmethylated nucleic acid alleles or modified/unmodified alleles.

25

The oligonucleotide primers of the present invention are capable of being employed in amplification reactions, wherein the primers are used in amplification of template DNA originating from either a methylation positive or amethylation negative strand. The preferred methylation-independent primers of the present invention comprise at least
30 one CpG dinucleotide, as described below. Accordingly, in a methylation positive and bisulfite modified nucleic acid target sequence, the primer sequence will anneal to the nucleic acid template with a perfect match, wherein all of the nucleotides in a consecutive region of the primer forms base pairs with a complementary region in the nucleic acid target. However, in a methylation negative nucleic acid target after bisulfite
35 modification, the methylation-independent primers of the present invention will anneal

to the nucleic acid template with an imperfect match, wherein the primer sequence comprise a mis-match (i.e. the primer and template does not form base pairs) at the position of the unmethylated Cytosine at a CpG site in the nucleic acid template.

Nonetheless, as the primers of the present invention are methylation-independent, the primers will hybridize to both methylation negative and methylation positive nucleic acid sequences after bisulfite modification, and the primers will form a perfect match with the target sequence of a methylated nucleic acid target and an imperfect match, where the primers and target nucleic acid sequence does not form base pairing at the positions of unmethylated Cytosine (which is converted by bisulfite to Uracil) at CpG sites.

The methylation-independent primers of the present invention will, due to the mis-match after bisulfite modification at positions of unmethylated cytosine of a CpG-site in the nucleic acid target sequence, hybridize less efficiently to a methylation negative nucleic acid sequence. However, by reducing the stringency of hybridization, the methylation-independent primers of the present invention are able to anneal to the nucleic acid target, also when the nucleic acid target comprise unmethylated CpG-sites, which have been modified by for example bisulfite treatment. In one example, the stringency is reduced by reducing the annealing temperature as described elsewhere herein.

The design of oligonucleotide primers suitable for nucleic acid amplification techniques, such as PCR, is known to people skilled within the art. The design of such primers involves analysis of the primer's melting temperatures and ability to form duplexes, hairpins or other secondary structures. Both the sequence and the length of the oligonucleotide primers are relevant in this context. The oligonucleotide primers according to the present invention comprise between 10 and 200 consecutive nucleotides, such as at least 60, at least 65, at least 70, at least 75, at least 80, at least 85, at least 90, at least 95, at least 100, at least 110, at least 120, at least 130, at least 140, at least 150, at least 160, at least 180 or at least 200 nucleotides. In a specific embodiment, the oligonucleotide primers comprise between 15 and 60 consecutive nucleotides, such as 15, 16, 17, 18, 19, 20, preferably 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, such as 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, alternatively at least 41, at least 42, at least 44, at least 46, at least 48, at least 50, at least 52, at least 54, at least 56, at least 58, or at least 60 consecutive nucleotides.

- The methods employed for determining the methylation status of a nucleic acid according to the present invention, preferably comprise amplification of a modified nucleic acid by use of a methylation independent oligonucleotide primer. In one
5 embodiment, the oligonucleotide primers of the present invention are able to hybridize to a nucleic acid sequence comprising CpG islands. In a preferred embodiment, at least one of the oligonucleotide primers according to the present invention comprises at least one CpG dinucleotide. In another embodiment of the present invention, the oligonucleotide primers comprise 2, alternatively 3, 4, 5, 6, 7, 8, 9 or 10 CpG
10 dinucleotides. In even further embodiments, the oligonucleotide primers of the present invention comprise at least 10 CpG dinucleotides. In one preferred embodiment the at least one methylation-independent oligonucleotide primer comprises one CpG dinucleotide at the 5'-end of the primer.
- 15 The CpG dinucleotide may be located anywhere within the oligonucleotide primer sequence. However, in a preferred embodiment of the present invention, the at least one CpG dinucleotide is located in the 5'-end of the oligonucleotide primer. In another preferred embodiment, the at least one CpG dinucleotide constitute the first two nucleotides of the 5'-end. In an even further preferred embodiments of the present
20 invention, the at least one CpG dinucleotide is located within the first 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of the 5'-terminus. In alternative embodiments, the at least one CpG dinucleotide is located within the first 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or 120 nucleotides of the 5'-terminus. In yet another embodiment, at least two CpG dinucleotides are located
25 within the first 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides of the 5'-terminus, or at least two CpG dinucleotides are located within the first 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or 120 nucleotides of the 5'-terminus.
- 30 The primers of the present invention may in one preferred embodiment comprise at least one CpG site, whereby annealing with a higher efficiency to a methylated than to an unmethylated template upon modification of unmethylated cytosine is achieved. The primers of the present invention comprise at least one CpG site. However, the primers comprise also for example two CpG sites.

5 The at least one CpG site is positioned in the 5' end of the primer. For example within the first 10 nucleotides in the 5' end of the primer, within the first 9 nucleotides in the 5' end of the primer, within the first 8 nucleotides in the 5' end of the primer, within the first 7 nucleotides in the 5' end of the primer, within the first 6 nucleotides in the 5' end of the primer, within the first 5 nucleotides in the 5' end of the primer, within the first 4 nucleotides in the 5' end of the primer or within the first 3 nucleotides in the 5' end of the primer. In a preferred embodiment the CpG site is introduced immediately after the first nucleotide of the 5' end of the primer.

10 Specific hybridization typically is accomplished by a primer having at least 10, for example at least 12, such as at least 14, for example at least 16, such as at least 18, for example at least 20, such as at least 22, for example at least 24, such as at least 26, for example at least 28, or such as at least 30 contiguous nucleotides, which are complementary to the target template. Often the primer will be close to 100% identical
15 to the target template. However, the primer may also be 98% identical to the target template or for example at least 97%, such as at least 96%, for example at least 95%, such as at least 94%, for example at least 93%, such as at least 92%, for example at least 91%, such as at least 90%, for example at least 89%, such as at least 88%, for example at least 87%, such as at least 86%, for example at least 85%, such as at least
20 84%, for example at least 83%, such as at least 82%, for example at least 81%, such as at least 80%, for example at least 79%, such as at least 78%, for example at least 77%, such as at least 76%, for example at least 75%, such as at least 74%, for example at least 73%, such as at least 72%, for example at least 71%, such as at least 70%, for example at least 68%, such as at least 66%, for example at least 64%, such
25 as at least 62% or for example at least 60% identical to the target template. If there is a sufficient region of complementary nucleotides, e.g., at least 10, such as at least 12, for example at least 15, such as at least 18, or for example at least 20, for example at least 30, such as at least 40, for example at least 50, such as at least 60, for example at least 70 nucleotides, then the primer may also contain additional nucleotide residues
30 that do not interfere with hybridization but may be useful for other manipulations. Examples of such other residues may be sites for restriction endonuclease cleavage, for ligand binding or for factor binding or linkers.

The methylation-independent oligonucleotide primer of the present invention is
35 designed to hybridize to nucleic acids in a sample. Importantly, the nucleic acids in that

sample are treated with an agent, which modifies unmethylated cytosine in said nucleic acid. Thereby, any unmethylated Cytosine of CpG dinucleotides comprised in the nucleic acid are converted to Uracil as explained elsewhere herein. Consequently, in primers comprising a CpG dinucleotide, designed to hybridize with the complementary CpG dinucleotide of the nucleic acid of the sample, the CpG dinucleotide will only hybridize to the methylated CpG dinucleotide fraction of the nucleic acid. In the unmethylated fraction of CpG dinucleotides comprised in the nucleic acid of the sample, Cytosine are modified to uracil which does not hybridize with the CpG dinucleotide of the oligonucleotide primer.

The methylation-independent oligonucleotide primers according to the present invention are designed to comprise sufficient nucleotides for specific hybridization to the target nucleic acid sequence regardless of its original methylation status. In some embodiments the oligonucleotide primers also comprise one or more CpG dinucleotides, as specified elsewhere herein. These CpG dinucleotides only hybridize with the originally methylated alleles of the nucleic acids. Nevertheless, the oligonucleotide primers can still be functionally used for amplification of both originally methylated and unmethylated nucleic acids. The CpG dinucleotides are typically comprised in the 5'-terminus of the oligonucleotide primers, as described elsewhere herein. A primer-template mismatch within the 5'-terminus of the primer usually allow the primers to hybridize with the target nucleic acid, and still function as primers in an amplification reaction.

The presence of one or more mismatches between the primer and template affects the optimal annealing temperature of said oligonucleotide primer for use in amplification reactions. The more hybridizing nucleotides comprised on the oligonucleotide primers, the higher is the optimal annealing temperature. Consequently, amplification of methylated alleles of nucleic acids by CpG-containing oligonucleotide primers according to the present invention is favoured by increased annealing temperature. Conversely, amplification of unmethylated alleles is favoured by decreased annealing temperature. In the present invention, the PCR bias towards amplification of unmethylated alleles of a nucleic acid template is reversed by amplification of said nucleic acid template at a relatively higher annealing temperature, which favours oligonucleotide primer binding and priming of the methylated allele. By modulation of the primer annealing temperature, the priming of either the unmethylated modified

allele or the methylated allele of the nucleic acid can be favoured. By increasing the annealing temperature below the theoretical optimum, the amplification of the methylated allele is favoured, while a decrease of the annealing temperature will tend to favour amplification of the unmethylated allele.

5

Besides annealing temperature, other factors also affect hybridisation to a target sequence of a methylation-independent primer. At highly stringent conditions, hybridization between perfect matching primer and target sequences are favoured, such as hybridization between a methylation-independent primer according to the present invention and a methylated target sequence upon cytosine modification. Less stringent conditions will tend to favour oligonucleotide primer binding, priming and amplification of the unmethylated allele. Modulation of temperature is one way of adjusting the stringency of hybridization, but the stringency of hybridization may also be modulated by adjusting buffer composition, and/or salt concentrations in the hybridization mixture, which is known to those of skill within the art. The present invention comprises any such method of modulating hybridization stringency to balance the PCR bias towards amplification of unmethylated template. However, modulation of temperature is preferred.

20 In one embodiment, the oligonucleotide primer of the present invention is selected from the group consisting of SEQ ID NO: 37-72. Methylation status is preferably determined for a gene mentioned in table B using the respective forward primer and reverse primer identified in table B; i.e.

for OTX2: primers SEQ ID NO: 37 and/or 38;

25 for HOXD3: primers SEQ ID NO: 39 and/or 40;

...

For SIM1: primers SEQ ID NO: 51 and/or 52;

For Chr6(p22.1): primers SEQ ID NO: 57 and/or 58;

For HIST1H3G/HIST1H2Bl: primers SEQ ID NO: 47 and/or 48;

30 etc...; and

for HIST1H3E: primers SEQ ID NO: 71 and/or 72.

In one embodiment, an oligonucleotide primer of the present invention specifically hybridizes to regions within 1 kb of the gene loci of the present invention. In one embodiment, the oligonucleotide primers hybridize to a target nucleic acid sequence of

35

a gene loci selected from the group consisting of SIM1, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H3G/HIST1H2BI, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5, Chr1(q21.1) and Chr6(p22.1), or the complement thereof.

5

In another embodiment of the present invention the oligonucleotide primer hybridizes to a target nucleic acid sequence of a gene loci selected from the group consisting of SIM1, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H3G/HIST1H2BI, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5, Chr1(q21.1) and Chr6(p22.1), or the complement thereof.

10

In another embodiment of the present invention the oligonucleotide primer hybridizes to a target nucleic acid sequence of a gene loci selected from the group consisting of OSR1, SIM1, HOXD10, HIST1H3E, HOXD3, GHSR, Chr1(q21.1) and/or HOXB3/HOXB4, or the complement thereof.

15

In another embodiment of the present invention the oligonucleotide primer hybridizes to a target nucleic acid sequence of a gene loci selected from the group consisting of SIM1, HIST1H3G/HIST1H2BI, Chr6(p22.1) and/or HOXB3/HOXB4, or the complement thereof.

20

In another embodiment of the present invention the oligonucleotide primer hybridizes to a target nucleic acid sequence of a gene loci selected from the group consisting of SIM1, HIST1H3G/HIST1H2BI and/or Chr6(p22.1), or the complement thereof.

25

In another embodiment of the present invention the oligonucleotide primer hybridizes to a target nucleic acid sequence of a gene loci selected from the group consisting of SIM1 and HIST1H3G/HIST1H2BI; or the group consisting of SIM1 and Chr6(p22.1); or the group consisting of HIST1H3G/HIST1H2BI and Chr6(p22.1), or the complement thereof.

30

In another embodiment of the present invention the oligonucleotide primer hybridizes to a target nucleic acid sequence of SIM1 and/or Chr6(p22.1) and/or HIST1H3G/HIST1H2BI and/or HOXB3/HOXB4 and/or OSR1 and/or HOXD10 and/or OTX2 and/or LOC648987 and/or HIST1H2AJ/HIST1H2BM, or the complement thereof.

35

In a preferred embodiment of the present invention the at least one oligonucleotide primer hybridizes to a target nucleic acid sequence selected from the group consisting of SEQ ID NO: 1-18, or the complement thereof (non-modified strand); and/or the
5 oligonucleotide prime hybridizes to a target nucleic acid sequence selected from the group consisting of SEQ ID NO: 19-36, or the complement thereof (modified strand).

In one embodiment, an oligonucleotide primer of the present invention specifically comprises or consists of 5-50, such as 5-30, such as 10-20 consecutive nucleotides of
10 a subsequence of a gene loci selected from the group consisting of SIM1, HOXB3/HOXB4, OSR1, GHRSR, OTX2, LOC648987, HIST1H3E, HIST1H3G/HIST1H2BI, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5, Chr1(q21.1) and Chr6(p22.1), or the complement thereof.

15 In particular, the present invention relates to oligonucleotide primer pairs, which span or comprise at least one CpG dinucleotide in a gene locus of the invention. The term "span" as used in this context is meant to indicated the at least one CpG site is located in the nucleic acid region between the primer pairs; i.e. the amplified nucleic acid region comprise at least one CpG dinucleotide. The term "comprising" as used in
20 connection with "primers comprising at least one CpG dinucleotide is meant to specify that the oligonucleotide primer itself comprise a CpG site.

In a specific embodiment of the present invention the oligonucleotide primers comprises or consists of 5-50, such as 5-30, such as 10-20 consecutive nucleotides of
25 a nucleic acid sequence of a gene loci selected from the group consisting of SIM1, HOXB3/HOXB4, OSR1, GHRSR, OTX2, LOC648987, HIST1H3E, HIST1H3G/HIST1H2BI, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5, Chr1(q21.1) and Chr6(p22.1), or the complement thereof.

30 In another embodiment of the present invention the oligonucleotide primer comprises or consists of 5-50, such as 5-30, such as 10-20 consecutive nucleotides of a nucleic acid sequence of a gene loci selected from the group consisting of OSR1, SIM1, HOXD10, HIST1H3E, HOXD3, GHRSR, Chr1(q21.1) and/or HOXB3/HOXB4, or the complement thereof.

35

In another embodiment of the present invention the oligonucleotide primer comprises or consists of 5-50, such as 5-30, such as 10-20 consecutive nucleotides of a nucleic acid sequence of a gene loci selected from the group consisting of SIM1, HIST1H3G/HIST1H2BI, Chr6(p22.1) and/or HOXB3/HOXB4, or the complement thereof.

In another embodiment of the present invention the oligonucleotide primer comprises or consists of 5-50, such as 5-30, such as 10-20 consecutive nucleotides of a nucleic acid sequence of a gene loci selected from the group consisting of SIM1, HIST1H3G/HIST1H2BI, Chr6(p22.1) and/or the complement thereof.

In a preferred embodiment of the present invention the at least one oligonucleotide primer comprises or consists of 5-50, such as 5-30, such as 10-20 consecutive nucleotides of a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1-18, or the complement thereof (non-modified strand); and/or the oligonucleotide prime comprises or consists of 5-50, such as 5-30, such as 10-20 consecutive nucleotides of a nucleic acid sequence selected from the group consisting of SEQ ID NO: 19-36, or the complement thereof (modified strand).

Thus, in the methods of the present invention for determining or prognosing non-small cell lung cancer, determining a predisposition to non-small cell lung cancer, categorizing or predicting non-small cell lung cancer, or evaluating the risk of contracting a non-small cell lung cancer, methylation status is preferably determined by amplifying at least one portion of a gene loci selected from SIM1, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H3G/HIST1H2BI, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5, Chr1(q21.1) and Chr6(p22.1), using at least one primer pair selected from the nucleic acid sequences set forth in table B (SEQ ID NO: SEQ ID NO: 37/38, 39/40, 41/42, 43/44, 45/46, 47/48, 49/50, 51/52, 53/54, 55/56, 57/58, 59/60, 61/62, 63/64, 65/66, 67/68, 69/70 and 71/72, respectively).

Detection of an amplification product can be performed by hybridizing the amplification product to an oligonucleotide probe, as described below. In a preferred embodiment, methylation status is determined by amplifying at least one portion of the respective at least one gene loci, and further employing at least one oligonucleotide probe which

hybridizes to an amplification product selected from the group consisting SEQ ID NO: 1-18 and/or the complement thereof (non-modified strand) or SEQ ID NO: 19-36 and/or the complement thereof (modified strand).

5 In a preferred embodiment, the oligonucleotide probe comprise 10-100 consecutive nucleic acids selected from the group of sequences consisting SEQ ID NO: 1-18 and/or the complement thereof (non-modified strand) or SEQ ID NO: 19-36 and/or the complement thereof (modified strand).

10 One aspect of the invention also relates to the use of oligonucleotide primers of the present invention for determining or prognosing a non-small cell lung cancer, determining a predisposition to non-small cell lung cancer, categorizing or predicting non-small cell lung cancer, or evaluating the risk of contracting a non-small cell lung cancer. Thus, in one aspect, the present invention provides a use of oligonucleotide primers comprising a subsequence of a loci selected from the group consisting of
15 SIM1, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H3G/HIST1H2BI, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5, Chr1(q21.1) and Chr6(p22.1) or the complement thereof for diagnosing non-small cell lung cancer in a method of the invention as defined elsewhere herein.

In a preferred embodiment of the use of the invention, the primers are selected from
20 the group set forth in table B (SEQ ID NO: SEQ ID NO: 37-72). In a preferred embodiment, the oligonucleotide primers comprise a sequence selected from the group consisting of SEQ ID NO: 1-18 and/or the complement thereof (non-modified strand) or the group consisting of SEQ ID NO: 19-36 and/or the complement thereof (modified strand). In a preferred embodiment, the oligonucleotide primers comprising a
25 subsequence selected from a gene loci selected from the group consisting of OSR1, SIM1, HOXD10, HIST1H3E, HOXD3, GHSR, Chr1(q21.1) and/or HOXB3/HOXB4, or the group consisting of SIM1, HIST1H3G/HIST1H2BI, Chr6(p22.1) and/or HOXB3/HOXB.

30 Analysis of amplified CpG-containing nucleic acids

According to the present invention the nucleic acid (target) sample is subjected to an agent that converts an unmethylated cytosine to another nucleotide which will distinguish the unmethylated from the methylated cytosine. In a preferred embodiment the agent modifies unmethylated cytosine to uracil. The modifying agent can be sodium
35 bisulphite. During the amplification process uracil will be converted to thymidine.

Thus, after conversion of unmethylated cytosines to uracils in the nucleic acid (target) sample, the subsequent PCR amplification converts uracils to thymine. As a consequence of the sodium bisulfite and PCR-mediated specific conversion of unmethylated cytosines to thymines, G:C base pairs are converted to A:T base pairs at positions, where the cytosine was methylated.

The difference in nucleic acid sequence at previously methylated (methylation positive) or unmethylated (methylation negative) cytosines allows for the analysis of methylation status in a sample. This analysis can comprise identifying cytosine residues, which have been converted to thymidine after amplification, as unmethylated cytosine residues, and identifying cytosine residues, which has not been converted under as methylated cytosine residues.

By this method, analysis of the amplified nucleic acid after treatment with a modifying agent such as sodium bisulphite and subsequent PCR amplification can reveal the methylation status of the target nucleic acid sequence. Thus, in one embodiment, the method for determining methylation status of a nucleic acid according to the present invention further comprises a step of analyzing the amplified nucleic acids.

Specifically, the subsequent analysis can be selected from the group consisting of melting curve analysis, high resolution melting analysis, nucleic acid sequencing, primer extension, denaturing gradient gel electrophoresis, southern blotting, restriction enzyme digestion, methylation-sensitive single-strand conformation analysis (MS-SSCA) and denaturing high performance liquid chromatography (DHPLC).

In one embodiment, the methylation status of the amplified containing nucleic acid is determined by any method selected from the group consisting of Methylation-Specific PCR (MSP), Whole genome bisulfite sequencing (BS-Seq), HELP assays, ChIP-on-chip assays, Restriction landmark genomic scanning, Methylated DNA immunoprecipitation (MeDIP), Pyrosequencing of bisulfite treated DNA, Molecular break light assays, and Methyl Sensitive Southern Blotting.

In another embodiment, the methylation status of the amplified containing nucleic acid is determined by a method selected from the group consisting methylation specific PCR, bisulfite sequencing, COBRA, melting curve analysis, or DNA methylation arrays.

In a preferred embodiment of the present invention, the analysis of the amplified nucleic acid region is melting curve analysis. In another preferred embodiment of the present invention, the analysis of the amplified nucleic acid is high resolution melting analysis (HRM).

Melting curve analysis

Melting curve analysis or high resolution melting analysis exploits the fact that methylated and unmethylated alleles are predicted to differ in thermal stability because of the difference in GC contents after bisulphite treatment and PCR-amplification, which converts methylated C:G base pairs to A:T base pairs. This means that the melting curve profile of methylated (methylation positive) and unmethylated (methylation negative) alleles of PCR products originating from bisulfite modified methylated and unmethylated can be distinguished. Thus, the level of fluorescence changes, depending on the relative amount template; i.e. the relative amount of methylation positive and methylated negative alleles.

By comparing the melting curve profile of an unknown sample with different mixes of controls having known relative amounts of methylation positive and methylation negative alleles, the relative amount of methylation positive and methylation negative alleles of the unknown sample can be determined.

The melting curve profile of an amplification product according to the present invention is determined by the composition of methylated and unmethylated alleles in the nucleic acid sample. If the nucleic acid molecules of a sample are all methylation negative, all cytosines are converted to thymines, and the resulting PCR product will have a relatively low melting temperature compared to a methylated nucleic acid, which can be seen in its melting curve. If on the other hand, the nucleic acids comprised in the sample are methylation positive, the melting temperature of the PCR product will be relatively higher, and the melting curve is shifted, as fluorescence is observed at higher temperatures. If the nucleic acid sample comprises a mixture of methylated and unmethylated alleles, bisulphite treatment followed by amplification will result in two distinct amplification products. The unmethylated alleles will display a low melting temperature and the methylated alleles a high melting temperature, and the melting curve profile of such a sample shows fluorescence from both PCR products (methylated and unmethylated). If only a subset of the CpG dinucleotides of the target

sequence contain a methylated cytosine (heterogeneous methylation), the amplification product represents a pool of molecules, which are present in different cells of the tumor, with different melting temperatures, which leads to an overall intermediate melting temperature.

5

Melting curve analysis is performed by incubating the nucleic acid amplification product at a range of increasing temperatures. The temperature is increased from a starting temperature of at least 50 degrees Celsius, alternatively at least 55, at least 60, at least 62, at least 64, preferably at least 65, at least 66, at least 67, at least 68, at least 69, at least 70, at least 71, at least 72, at least 73, at least 74, at least 75, for example at least 76, at least 78, at least 80, at least 82, at least 84 degrees Celsius. The temperature is then increased to a final temperature of at least 70, at least 72, at least 74, at least 76, at least 78, at least 80, at least 82, at least 84, at least 86, preferably at least 88, at least 89, at least 90, at least 91, at least 92, at least 93, at least 94, at least 95, at least 96, at least 97, at least 98, at least 99, at least 100 degrees Celsius. In one embodiment, the temperature transitions from the starting temperature to the final temperature are a linear function of time. In a specific embodiment of the present invention, the linear transitions are at least 0.05 degrees Celsius per second, alternatively at least 0.01, at least 0.02, at least 0.03, at least 0.04, at least 0.06, at least 0.07, at least 0.08, at least 0.09, at least 0.1, at least 0.2, at least 0.3, at least 0.4, at least 0.5, at least 0.6, at least 0.7, at least 0.8, at least 0.9, at least 1.0, at least 1.1, at least 1.2, at least 1.3, at least 1.4, at least 1.5, at least 1.6, at least 1.7, at least 1.8, at least 1.9, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10 degrees Celsius per second. In a preferred embodiment, the melting curve analysis is performed by incubating the nucleic acid amplification product at increasing temperatures, from 70 to 95 degrees Celsius, wherein the temperature increases by 0.05 degrees per second.

The melting of the nucleic acid can be measured by a number of methods, which are known to people within skill of the art. One method involves use of agents, which fluoresce when bound to a nucleic acid in its double stranded conformation. Such agents include fluorescent probes or dyes, such as ethidium bromide, EvaGreen, LC Green, Syto9, SYBR Green, SensiMix HRM™ kit dye. Thus, in one embodiment, the melting curve analysis is performed by measurement of fluorescence. The melting of the nucleic acid amplification product according to the present invention can then be

monitored as a decrease in the level of fluorescence from the sample. After measurement of the fluorescence the melting curves can be generated by plotting fluorescence as a function of temperature.

5 For direct comparison of melting curves from samples that have different starting fluorescence levels, the melting curves for data collected in HRM can be normalized, as described in the examples of the present invention. Such normalization methods are known to people of skill in the art. One preferred means of normalization include calculation of the 'line of best fit' in between two normalization regions before and after
10 the major fluorescence decrease representing the melting of the amplification product. The 'line of best fit' is a statistical measure, designating a line plotted on a scatter plot of data (using a least-squares method) which is closest to most points of the plot. Calculation of the line of best fit is performed differentially on LightCycler and LightScanner, as illustrated in the examples of the present invention.

15 A platform with a combined thermal cycler and a fluorescence detector is ideal to perform in-tube melting analyses. Thus, in one embodiment, the melting curve analysis is performed on a thermal cycler coupled to a fluorometre, such as the Ligthcycler, LC480 (Roche) or the Rotorgene 6000 (Corbett Research). Thereby, the measurement
20 of fluorescence, corresponding to the melting of the double stranded nucleic acid template, can be monitored in real time. In a specific embodiment, the melting curve analysis is performed immediately after amplification. This allows an in-tube methylation assay, wherein the amplification and melting curve analysis is performed sequentially without transferring the sample from the tube. This procedure reduces the
25 risk of contamination of the sample as a result from handling during the methylation assay.

Melting curve analysis allows the determination of the relative amount of methylated nucleic acid in a sample. By comparison of the melting curve of an amplification
30 product of nucleic acid for which methylation status in unknown sample with the melting curve of at least one standard sample comprising the corresponding amplification product for which methylation status is known, the relative amount of methylated CpG-containing nucleic acid can be estimated. Thus, the present invention relates to a method, wherein the relative amount of methylated nucleic acid is estimated by
35 comparison the melting curve of at least one standard sample comprising said nucleic

acid with a control level of methylation. In one embodiment of the present invention, said standard sample comprise any combination of methylated and unmethylated nucleic acid. In a specific embodiment, said standard sample comprise 100% methylated nucleic acid. In another specific embodiment, said standard sample
5 comprise 100% unmethylated nucleic acid. In yet another specific embodiment, said standard sample comprise 50% methylated nucleic acid and 50% unmethylated nucleic acid. In even another specific embodiment, said standard sample comprise 0.1% 0.5%, 1%, 2%, 3%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100% methylated nucleic acid.

10

In one embodiment of the present invention, the relative amount of methylated nucleic acid in the nucleic acid sample is between 40-60%. In another embodiment, the relative amount of methylated nucleic acid in the nucleic acid sample is below 50%. In yet another embodiment, the relative amount of methylated nucleic acid in the nucleic acid
15 sample is below 10%, below 1% or below 0.1%. Thus, the term "presence of methylation" and/or the term "methylation status" as used herein includes the relative amount of methylated nucleic acid in the nucleic acid sample of at least 0.1%, such as at least 1 %, for example at least 10%, such as at least 20%, for example at least 30%, such as at least 50%, for example at least 70%, such as at least 90%, or for example at
20 least 99%.

If the fluorescence level at an allele-specific peak melting temperature in the melting curve of an unknown sample is higher than that of a standard sample, then the relative amount of that specific allele (methylation positive or methylation negative) in the
25 unknown sample is also higher than the relative amount of that allele in the standard sample. Thus, if a standard sample comprise 80% methylation positive alleles, for which the peak melting temperature is 70°C, and the fluorescence of an unknown sample at 70°C is the same as that of the standard sample, then the relative level of methylation positive alleles in the unknown sample can be inferred to be around 80%.

30

If the fluorescence of the unknown sample is less than the standard, the amount of methylation positive alleles can be inferred to be less than 80%, and if the fluorescence is higher than the standard, then the unknown sample comprise more than 80% methylation positive alleles. Thus, by comparing a melting curve fluorescence profile of an unknown sample with the profiles of standard samples with different compositions of

methylation positive and methylation negative alleles, the level of methylated alleles of the unknown sample can be inferred.

The more standard samples, the more precise the relative amount of nucleic acids can be determined.

5

Thus, in one embodiment of the present invention, a higher fluorescence level at the peak melting temperature of the amplified nucleic acid sample than of a standard sample comprising a specific allele is indicative of a higher relative amount of that specific allele in that sample than in the standard sample. Conversely, a lower

10 fluorescence level at the peak melting temperature of the amplified nucleic acid sample than of the standard sample comprising a specific allele is indicative of a lower relative amount of that specific allele in that sample than in the standard sample.

The “peak melting temperature” is mathematical a derivative of melting curve and
15 refers to the temperature at which the largest discrete melting step occurs. The top of the peak corresponds to the major drop in fluorescence on the melting curve. The level of fluorescence at the peak melting temperature reflects the level of methylation for a given amplicon. Thus, two amplicons (derived from two different samples) may have peak melting temperatures of for example be 70°C, while having different fluorescence
20 at this temperature, which then reflects that the amplicons have different methylation levels. The peak melting temperature corresponds to the highest level of the negative derivative of fluorescence ($-dF/dT$) over temperature versus temperature (T). A nucleic acid sample subjected to melting curve analysis may display more than one peak melting temperature. In a preferred embodiment of the present invention, the melting
25 curve analysis display at least 1, 2 or 3 peak melting temperatures.

Melting curve analysis is illustrated in the examples herein below, and figures 1-3.

Nucleic acid sequencing

30 In another embodiment of the present invention, the method for analysis of the amplified nucleic acid is sequencing of the nucleic acid. By nucleic acid sequencing the order of nucleotides (base sequences) in the nucleic acid is determined. Sequencing is usually performed by extending a primer, which anneals to the nucleic acid sequence of interest. The primer is extended by a polymerase in the presence of
35 deoxynucleonucleotides. Sequencing may also be performed by pyrosequencing.

In the dideoxy sequencing method 2,3-Dideoxyribose - a deoxyribose sugar lacking the 3 hydroxyl group is incorporated into the extended nucleic acid chain. When 2,3-Dideoxyribose is incorporated into a nucleic acid chain, it blocks further chain
5 elongation. This method is also known as the Sanger method or chain termination method. The primer is extended in the presence of the normal dNTPs (A, T, G, C) and a small amount of 2,3-DideoxyriboseNTPs (ddNTP). The reactions are either performed in four separate reactions, one for each of the ddNTPs (ddATP, ddTTP, ddCTP, ddGTP), or in a joint reaction, wherein ddATP, ddTTP, ddCTP and ddGTP are
10 coupled to different fluorescent dyes. The primers are then extended to variable lengths, each transcript being terminated upon incorporation of a ddNTP. The sequence of the nucleic acid of interest can then be read after denaturing polyacrylamid gel electrophoresis. Such sequencing techniques are known to people skilled within the art. Additionally, a number of different commercial kits are available
15 for sequencing of nucleic acids.

Primer extension

In yet another embodiment of the present invention, the method for analysis of the amplified nucleic acid is primer extension. The primer extension method uses primers
20 designed to hybridize with a target. The primers may end one base upstream of the position of the putative single nucleotide polymorphism, in this method, the C of a CpG dinucleotide. In the single nucleotide primer extension technique a single chain-ending nucleotide, such as a ddNTP, is added. The only one of the four nucleotides that will extend the primer is the one that is complementary. The identity of the added
25 nucleotide, which reflects the methylation status, is determined in a variety of ways known to people of general skill within the art. For example, the chain-ending nucleotide may be radioactively labelled or coupled to a fluorescent dye, which can subsequently be identified.

Restriction enzyme digestion

In a further embodiment, the method according to the present invention for analysis of the amplified nucleic acid is restriction enzyme digestion. Restriction enzymes can be divided into exonucleases and endonucleases. In a specific embodiment, the analysis
30 of the amplified nucleic acid is restriction endonuclease digestion.

The method of the present invention results in the specific conversion of unmethylated cytosines to thymines, i.e. G:C base pairs are converted to A:T base pairs at positions, where a cytosine was methylated. This means that the nucleic acid sequence is changed, which may lead to disruption of a restriction endonuclease site or the change of a site specific for one restriction endonuclease to another restriction endonuclease. In a preferred embodiment of the present invention, the modified and amplified nucleic acid is analyzed for disruption of a site specific for the endonuclease Acil, BstUI, HhaI, HinP1I, HpaII, HpyCH4IV, MspI, TaqI, Fnu4HI, Hpy188I, HpyCH4III, NciI, ScrFI, BssKI, Hpy99I, Nt.CviPII, StyD4I, AatII, AccI, AclI, AfeI, AflIII, AgeI, Aval, BanI, BmgBI, BsaAI, BsaHI, BsaJI, BsaWI, BsiEI, BsiWI, BsoBI, BspDI, BspEI, BsrBI, BsrFI, BssHII, BssSI, BstBI, BtgI, Cac8I, ClaI, EaeI, EagI, FspI, HaeII, HincII, Hpy188III, KasI, MluI, MspAII, NaeI, NarI, NgoMIV, NlaIV, NruI, PaeR7I, PmlI, PvuI, SacII, SalI, SfoI, SmaI, SmlI, SnaBI, TliI, TspMI, XhoI, XmaI, ZraI, RsrII, AscI, AsiSI, FseI, NotI, PspXI, SgrAI, AlwNI, DraIII, PfiFI, Tth111I, AleI, BsaBI, MsiI, PshAI, XmnI, AhdI, BglI, BstI, BstAPI, EcoNI, MwoI, PfiMI, BsmBI, FauI, BstXI, DrdI, SfiI, XcmI, HgaI, EciI, BceAI, BtgZI, MmeI, NmeAIII, BsaXI, BcgI, CspCI, BaeI, AccII, AspLEI, Bsh1236I, BsiSI, BstFNI, BstHHI, CfoI, HapII, Hin6I, HspAI or MaeII. The digested nucleic acid sample is subsequently analysed by for example gel electrophoresis.

20 *Denaturing gradient gel electrophoresis*

In another embodiment of the present invention, the method for analysis of the amplified nucleic acid is denaturing gradient gel electrophoresis (DGGE). In this technique, the modified and amplified nucleic acid is loaded on a denaturing gel. This techniques allows the resolution of nucleic acids with different melting temperatures, which is based on the conversion of C:G base pairs to A:T base pairs, explained elsewhere herein. For DGGE analysis the nucleic acid is subjected to denaturing polyacrylamide gel electrophoresis, wherein the gel contain an increasing gradient of denaturants, such as for example a combination of urea and formamide. The increasing denaturant concentration corresponds to increased temperature, and therefore, a gradient of denaturants mimics a temperature gradient within the gel. The concentrations of denaturants alone, however, are not sufficient to induce DNA melting. Therefore, the gel is immersed in an electrophoresis buffer kept at 54–60 degrees Celsius. When a nucleic acid molecule reaches a level of denaturant that matches the melting temperature of the lowest melting domain, a partially melted intermediate will be formed that moves very slowly. Small shifts in the melting temperature of the low

melting domain induced by differences in G:C content will cause the domain to unwind at different concentrations of denaturant. Accordingly, the modified and amplified nucleic acid of the present invention will be retarded at different positions in the gel, providing the basis for physical separation between species with different G:C contents, which is reflective of methylation status.

Southern blotting

In another embodiment of the present invention, the method for analysis of the amplified nucleic acid is Southern blotting. In this procedure, the nucleic acid to be analysed are separated by gel electrophoresis and transferred to a nitrocellulose filter, where to it is immobilized. After immobilization, the transferred nucleic acids can be identified by hybridization with specific probes comprising a complementary nucleic acid. After hybridization and removal of excess unbound probe, the amount of hybridized indicate whether the sequence of interest was represented in the nucleic acids immobilized on the nitrocellulose membrane. The probes are usually radioactively labelled for subsequent detection by radiography. The details of the southern blotting technique are well known to people of skill within the art.

Methylation-sensitive, single-strand conformation analysis (MS-SSCA)

MS-SSCA is a method of screening for methylation changes. MS-SSCA uses single-strand conformation analysis for the screening of an amplified region of bisulfite-modified nucleic acid. The amplified products are denatured and electrophoresed on a nondenaturing polyacrylamide gel, whereby the sequence differences between unmethylated and methylated sequences lead to the formation of different secondary structures (conformers) with different mobilities. Once the normal mobility pattern is established, any variation would indicate some degree of methylation.

Denaturing high performance liquid chromatography (DHPLC)

DHPLC is yet another technique for methylation screening of bisulfite-modified PCR products. As for other techniques mentioned herein, DHPLC identifies single nucleotide polymorphisms, which are arise after bisulfite treatment of unmethylated alleles of the CpG containing nucleic acid. The optimum temperature for DHPLC can be predicted by the sequence of the fully methylated product. Subsequently, the temperature is verified to obtain tight peaks. The retention time of the peak reflects methylation status,

because the more unmethylated the target is, the less GC rich the PCR product is and the lower the retention time is.

Kit

- 5 One aspect of the present invention relates to a kit for the detection of methylation status of a nucleic acid in a sample. A kit will typically comprise both a forward and a reverse primer to be used in the amplifying step of the present invention. The forward primer, the reverse primer or both may be a methylation-independent oligonucleotide primer as described herein. Thus, one aspect of the invention relates to a kit for
- 10 determining non-small cell lung cancer, predisposition to non-small cell lung cancer, or categorizing or predicting the clinical outcome of a non-small cell lung cancer, or monitoring the treatment of a non-small cell lung cancer, and/or monitoring relapse of a previously treated non-small cell lung cancer.
- The kit of the invention comprise
- 15 i. an agent that (a) modifies methylated cytosine residues but not non-methylated cytosine residues; or (b) modifies non-methylated cytosine residues but not methylated cytosine residues; or (c) modifies a nucleic acid sequence in a methylation-dependent manner, and
- ii. at least one pair of oligonucleotide primers that specifically hybridizes
- 20 under amplification conditions to a region of a gene locus selected from the group consisting of SIM1, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H3G/HIST1H2BI, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5, Chr1(q21.1) and Chr6(p22.1).
- 25 The agent is preferably a methylation-dependent endonuclease as described elsewhere herein, and/or an agent capable of modifying non-methylated cytosine residues but not methylated cytosine residues, such as a bisulphite compound as described elsewhere herein, for example sodium bisulphite.
- 30 Generally, the kit preferably comprises at least one oligonucleotide primer or probe of the present invention, as defined herein above. In a preferred embodiment, the kit comprise at least one oligonucleotide primer selected from the group consisting of SEQ ID NO: 37-72. In a more preferred embodiment, the kit comprises at least one primer pair selected from table B; i.e. at least one primer pair identified as SEQ ID NO: 37/38,

39/40, 41/42, 43/44, 45/46, 47/48, 49/50, 51/52, 53/54, 55/56, 57/58, 59/60, 61/62, 63/64, 65/66, 67/68, 69/70 and 71/72.

5 The kit may also comprise one or more reference sample, in particular reference samples comprising a nucleic acid sequence selected from a gene locus selected from the group consisting of SIM1, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H3G/HIST1H2BI, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5, Chr1(q21.1) and Chr6(p22.1). Thus, the kit may comprise a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1-18 and/or the
10 complement thereof (non-modified strand) or the group consisting of SEQ ID NO: 19-36 and/or the complement thereof (modified strand).

For example, the at least one reference sample comprises 100% methylation positive reference nucleic acid, and/or 100% methylation negative reference nucleic acid. In a
15 preferred embodiment the kit comprises at least two reference samples, wherein one of said reference samples comprises 100% methylation positive reference nucleic acid and a second reference sample comprises 100% methylation negative reference nucleic acid. The methylation positive and methylation negative reference nucleic acids may be mixed, by a person employing the kit, in ratios that are suitable for the
20 detection of methylation in a particular sample. It is understood that reference samples in different ratios of methylation positive to methylation negative CpG-containing nucleic acids may be comprised in the kit. For example the kit may comprise at least one reference sample comprising 50% methylated and 50% non-methylated nucleic acid alleles of the respective genetic locus marker.

25 In particular, the nucleic acid comprised on the reference sample of the kit is preferably methylated (methylation positive) or non-methylated (methylation negative), and the kit preferably comprise two or more reference samples with different methylation status; i.e. different levels of methylation positive and methylation negative alleles. Thus, the specific nucleic acid alleles (e.g. alleles of the gene locus OSR1) of the reference
30 sample may be unmethylated (0% methylated), 1%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 100% methylated. The kit may thus comprise one reference sample with a nucleic acid sequence as defined above, which is unmethylated; and another reference sample with the same nucleic acid sequence, which is 100% methylated; and one or more samples comprising the same nucleic acid

sequence having different levels of intermediate methylation status, e.g. 10%, 50% and/or 90% methylation.

5 The kit preferably comprises the following combinations of one or more reference samples and primer pairs:

Reference sequence comprising gene locus OTX2: forward primer = SEQ ID NO: 37 and reverse primer = SEQ ID NO: 38;

Reference sequence comprising gene locus HOXD3: forward primer = SEQ ID NO: 39 and reverse primer = SEQ ID NO: 40;

10 Reference sequence comprising gene locus HOXB3/HOXB4: forward primer = SEQ ID NO: 41 and reverse primer = SEQ ID NO: 42;

Reference sequence comprising gene locus HOXD10: forward primer = SEQ ID NO: 43 and reverse primer = SEQ ID NO: 44;

15 Reference sequence comprising gene locus Chr1(q21.1): forward primer = SEQ ID NO: 45 and reverse primer = SEQ ID NO: 46;

Reference sequence comprising gene locus HIST1H3G/HIST1H2BI: forward primer = SEQ ID NO: 47 and reverse primer = SEQ ID NO: 48;

Reference sequence comprising gene locus GHSR: forward primer = SEQ ID NO: 49 and reverse primer = SEQ ID NO: 50;

20 Reference sequence comprising gene locus SIM1: forward primer = SEQ ID NO: 51 and reverse primer = SEQ ID NO: 52;

Reference sequence comprising gene locus OSR1: forward primer = SEQ ID NO: 53 and reverse primer = SEQ ID NO: 54;

25 Reference sequence comprising gene locus FRG1B: forward primer = SEQ ID NO: 55 and reverse primer = SEQ ID NO: 56;

Reference sequence comprising gene locus Chr6(p22.1): forward primer = SEQ ID NO: 57 and reverse primer = SEQ ID NO: 58;

Reference sequence comprising gene locus HOXA3: forward primer = SEQ ID NO: 59 and reverse primer = SEQ ID NO: 60;

30 Reference sequence comprising gene locus LY75-CD302: forward primer = SEQ ID NO: 61 and reverse primer = SEQ ID NO: 62;

Reference sequence comprising gene locus CTAGE15: forward primer = SEQ ID NO: 63 and reverse primer = SEQ ID NO: 64;

35 Reference sequence comprising gene locus LOC648987: forward primer = SEQ ID NO: 65 and reverse primer = SEQ ID NO: 66;

Reference sequence comprising gene locus HIST1H2AJ/HIST1H2BM: forward primer = SEQ ID NO: 67 and reverse primer = SEQ ID NO: 68;

Reference sequence comprising gene locus HOXA5: forward primer = SEQ ID NO: 69 and reverse primer = SEQ ID NO: 70

5 Reference sequence comprising gene locus HIST1H3E: forward primer = SEQ ID NO: 71 and reverse primer = SEQ ID NO: 72

The kit may also comprise at least one probe. Probes of the invention are defined herein above, and in a preferred embodiment, the kit comprise at least one
10 oligonucleotide probe comprising 10-100 consecutive nucleic acids selected from the group of sequences consisting of SEQ ID NO: 1-18 and/or the complement thereof (non-modified strand) or the group of sequences consisting of SEQ ID NO: 19-36 and/or the complement thereof (modified strand). Thus, the kit of the invention preferably comprise at least one oligonucleotide probe which hybridizes to a nucleic
15 acid sequence selected from the group consisting SEQ ID NO: 1-18 and/or the complement thereof (non-modified strand) or the group of sequences consisting of SEQ ID NO: 19-36 and/or the complement thereof (modified strand).

The kit of the invention preferably comprise at least one oligonucleotide probe which
20 hybridizes to an amplification product generated by a primer pair selected from the group consisting of SEQ ID NO: SEQ ID NO: 37/38, 39/40, 41/42, 43/44, 45/46, 47/48, 49/50, 51/52, 53/54, 55/56, 57/58, 59/60, 61/62, 63/64, 65/66, 67/68, 69/70 and 71/72.

The kit may also comprise additional reagents used in the amplifying step of the
25 detection method as disclosed herein. Thus, the kit may further comprise deoxyribonucleoside triphosphates, DNA polymerase enzyme and/or nucleic acid amplification buffer. In another embodiment the kit further comprises an agent that modifies unmethylated cytosine nucleotides. Such an agent may for example be bisulfite, hydrogen sulfite, and/or disulfite reagent.

30

The kit may also comprise other components suitable for detection of methylation status. For example, the kit may comprise a methylation-sensitive restriction enzyme.

The kit may in preferred embodiments further comprise instructions for the
35 performance of the detection method of the kit and for the interpretation of the results.

The instructions for performing the method of the kit comprises for example information of particular annealing temperatures to be used for the at least one methylation-independent primers, as well as for example information on PCR cycling parameters. The kit may further comprise instructions for the interpretation of the results obtained by the method. For example how to interpret the amplified products subsequently analysed by melting curve analysis or methods as described elsewhere herein. Information of the interpretation of melting curve analysis is described elsewhere herein.

The kit may in preferred embodiments further comprise software comprising an algorithm for calculation of primer annealing temperature and interpretation of results. Preferred embodiments for the CpG-containing nucleic acid for which the methylation

It is appreciated that the kit may be used for evaluating a non-small cell lung cancer in a human subject based on methylation status of specific genes as specified elsewhere herein.

Examples

The invention relates to methylation biomarkers for non-small cell lung cancer. Candidate marker genes were first identified by micro array analysis. Then, the ability of each of these methylation markers to distinguish between lung tumour tissue and healthy tissue was evaluated. From this analysis, 19 highly sensitive and specific methylation biomarkers were identified.

Example 1

This example shows a genome-wide methylation screening, which identifies novel methylation biomarkers that can be used in clinical lung cancer management. Cancer-specific methylation changes are identified and the sensitivity and specificity of the most promising biomarkers are validated and evaluated.

Results

Identification of differentially methylated regions (DMRs) in LAC

In order to identify novel genomic regions with disease-specific changes in methylation, we performed a genome-wide methylation screening of tumor and tumor-adjacent normal lung tissue from four LAC patients using the NimbleGen Human DNA Methylation 3×720K CpG Island Plus RefSeq Promoter Array, which interrogates

15,980 CpG islands and 20,404 reference gene promoter regions²³. After data processing, a total of 346 probes (oligonucleotides of 60 bp spotted on the array) demonstrated a significant change in methylation levels between the tumor and tumor-adjacent normal lung tissue with 288 (83.2%) and 58 (16.7%) reporting hyper- and hypomethylation in the tumor tissue, respectively. The mapping of the probes to genomic regions revealed no enrichment bias in any specific parts of the genome. Out of the 346 probes, 131 (37.9%) were located within known CGIs, 164 (47.4%) at CGI shores, 6 (1.7%) in CGI shelves and 45 (13.0%) were not associated with any known CGIs. When mapping to genes, 176 (50.9%) probes were intragenic, 128 (37.0%) were located <5kb upstream of known genes, 3 (0.9%) were located <5kb downstream of known genes, and 39 (11.2%) were not associated with any known gene. The 346 probes were sorted by differential methylation score, which indicates the magnitude of the detected change in methylation (See Methods section: Microarray analyses). To locate the most informative differentially methylated regions (DMRs) in LAC, we grouped the probes with a maximum inter-probe distance of 5 kb. Using this approach, we identified a total of 74 DMRs of which 63 (85.1%) showed hypermethylation and 11 (14.9%) hypomethylation in LAC. Each DMR was on average targeted by 4.1 probes (Range: 2-13 probes) and spanned 547.3 bp (Range: 126 - 4245 bp). Of the 74 DMRs, 65 (87.8%) and 66 (89.2%) were located in association with known genes and CGIs, respectively. Several of the identified DMRs, including *OTX2*, *OSR1* and *GHSR*, have previously been reported differentially methylated in LAC. The DMRs were sorted according to the probe with highest differential methylation score in each region and the asterisks denote the DMRs that were selected for further validation.

25 *Validation of candidate DNA methylation biomarkers*

To evaluate our findings potential for clinical application, we selected 18 DMRs based upon highest differential methylation score and the number of probes targeting the region to undergo validation in a LAC validation cohort, comprising 52 primary lung tumors, 24 paired distant metastases (20 brain and 4 adrenal gland) and 32 tumor-adjacent normal lung samples. The histological and clinical characteristics of the patients are shown in Table 4. The array data showed that 16 of the 18 selected DMRs were hypermethylated and two regions, *FRG1BP* and *CTAGE15*, were hypomethylated in primary tumors compared to tumor-adjacent normal lung tissue. We then assessed the methylation status of each of the 18 candidate regions in our patient samples using Methylation-Sensitive High Resolution Melting (MS-HRM) analysis. The genomic

location of the MS-HRM assays is shown in Table 1 and the technical specifications for each assay are indicated for each of the oligonucleotide sequences below. The results of the MS-HRM-based methylation assessment are summarized in Table 2 and displayed as stacked bar percentage plots in Fig. 1. Using this approach, we were able to confirm a significant increase in methylated templates in the tumor samples for 15 of the 18 selected DMRs corresponding to the hypermethylation indicated by the array analysis. Normalized melting curves for representative tumor and normal lung samples are shown in Fig. 2 for the *HOXD3*, *OSR1* and *HIST1H3E* MS-HRM assays, where the gain in methylation is seen as a relative shift in the melting curves towards the 100% methylated standard. We were unable to confirm the array results for three DMRs, including one hypermethylated region, *LY75-CD302*, and both hypomethylated regions, *FR1GB* and *CTAGE15*, as shown in Fig. 1j, 1m and 1n. For 12 of the 15 DMRs with concordant MS-HRM and array results, the difference in methylation frequency between tumor and normal lung tissue was very pronounced as indicated by the *p*-values ($p < 0.0001$) in Fig. 1. As an example, an elevated methylation level was detected in 75% of the tumor samples and 0% of the normal lung samples for the *HOXB3/HOXB4* MS-HRM assay, as shown in Fig. 1c, and in 87.9% of the tumor samples and only 3.2% of the normal lung samples for the *OSR1* MS-HRM assay shown in Fig. 1i. A high methylation frequency was also observed in the brain and adrenal gland metastases for all 15 confirmed DMRs. For the majority of the DMRs, the detected increase in methylation was even more prominent in the metastases compared to primary tumors, but due to the considerable difference in average tumor content between the primary tumors and metastases samples, these groups are not directly comparable. In conclusion, we have identified and validated 15 DMRs that can be targeted as novel biomarkers in LAC.

The candidate DNA methylation biomarkers and metastases formation in LAC

The presence of metastatic disease greatly reduces the overall survival of LAC patients. To determine if hypermethylation of the 15 identified candidate biomarkers are predictive of metastases formation in LAC, we compared the methylation status of the 15 DMRs between metastases-free patients with a minimum of 5 years recurrence-free survival, and patients that suffered from distant metastatic disease at the time of diagnosis. Three of the 15 DMRs, *HOXB3/HOXB4*, *LOC648987* and *HOXA5*, showed a trend towards an increase in methylation in the metastasizing tumors as illustrated in Fig. 3a-c. At the *HOXB3/HOXB4* region, we detected increased methylation in 15 of 24

(62.5%) non-metastasizing tumors compared to in 21 of 24 (87.5%) metastasizing tumors. Similarly, increased methylation in 3 out of 21 (14.3%) non-metastasizing tumors and in 8 out of 24 (33.3%) metastasizing tumors were detected for the *LOC648987* region, as well as in 19 out of 26 (73.1%) non-metastasizing tumors compared to 23 out of 26 (88.4%) metastasizing tumors for the *HOXA5* region. For all three DMRs, a similar increase in methylation was detected in the paired brain and adrenal gland metastases, as shown in Table 2.

Evaluation of the clinical potential of the candidate DNA methylation biomarkers

Sensitivity and specificity are the most important parameters when describing the potential diagnostic applicability of a biomarker. In order to calculate these values, we determined an unambiguous consensus for when a sample was considered methylation positive or negative for each assay. MS-HRM is a semi-quantitative method capable of determining the relative amount of methylated alleles in a sample, and we therefore determined a specific cutoff value for each of the assays based on the relative amount of methylated alleles that is detected. For each potential cutoff value, we calculated the corresponding sensitivity and specificity, and the cutoff was then set to achieve maximal sensitivity without compromising a specificity limit of 0.8. The determined cutoff value, sensitivity and specificity for each candidate biomarker are shown in Table 3. For 9/15 assays, the cutoff was set at 1% methylation and all samples containing more than 1% methylated templates were therefore considered positive. Similarly, the cutoff was set at >10% methylation for the *HOXD3*, Chr1(q21.1).A and *GHSR* assays and at >50% methylation for the *HOXA3*, *HOXA5* and *HIST1H3E* assays. Using this approach, we achieved a specificity of ≥ 0.90 for all assays, except *HOXD10*, and 9/15 assays reached a specificity of 1.00, which translates into a false positive rate of 0%. The sensitivity ranged from 0.12 to 0.98 with 8/15 assays demonstrating a sensitivity of ≥ 0.75 .

While it is possible to successfully employ cutoff values when using MS-HRM analysis, it is still preferential to use biomarkers that do not show any methylation in the corresponding normal tissue, as this allows for more accurate and stringent analyses. To identify the most promising candidate biomarkers, we therefore applied a lower sensitivity limit of 0.75 and a lower specificity limit of 0.90 to the assays with a cutoff of >1% methylation. Three candidate biomarkers fulfilled these criteria, *OSR1*, *SIM1* and *HOXB3/HOXB4*, which are indicated in bold print in Table 3. The *OSR1* assay

demonstrated a sensitivity of 0.98 and a specificity of 0.97 and correctly identified 97% of all tested samples. The *SIM1* assay correctly identified 91% of the samples and showed a sensitivity of 0.92 and a specificity of 0.90 and similarly, the *HOXB3/HOXB4* assay provided a sensitivity of 0.75 and a specificity of 1.00 and therefore correctly identified 85% of the tested samples. The *OSR1*, *SIM1*, *HOXB3/HOXB4* DMRs therefore show high clinical potential as biomarkers in LAC.

Discussion

Lung cancer has the highest mortality rates among cancers, but the prognosis for the individual patient varies considerably depending on the stage at which the disease is diagnosed. Efficient diagnostic tools that allow early and accurate disease detection are therefore of critical importance in clinical lung cancer management. Compelling evidence supporting the utility of methylation biomarkers in various aspects of cancer management, such as risk assessment, disease detection and personalization of treatment, has accumulated during the last decades. In this study, we aimed to identify and validate novel DMRs in LAC that can be targeted as biomarkers. Using a microarray-based genome-wide methylation screening approach, we identified 74 genomic regions that demonstrated differential methylation in tumor and tumor-adjacent normal lung tissue. Eighteen DMRs were selected for validation by MS-HRM analysis and we were able to confirm differential methylation in 15/18 DMRs. This yields a true positive rate of 83.3%, which indicates that the obtained microarray data is of high quality, but also emphasizes the importance of a thorough validation process when performing a microarray-based genome-wide methylation screening study. The validity of the data is furthermore supported by the fact that we were able to confirm differential methylation for the *HIST1H3E* region, which is number 61 out of the 74 identified DMRs when sorted by highest differential methylation score. The *HIST1H3E* region was only targeted by 3 probes and only showed a differential methylation score of 2.961 and we still confirmed a significant increase in methylation ($p < 0.0001$) in the tumor samples as illustrated in Fig. 1r and Fig. 2e and f. Moreover, Rauch et al. recently published a similar methylation screening study using 8 LAC patient samples and there are several overlaps in the identified DMRs, which serves to confirm the validity of both studies.

Most of the DMRs identified in our study showed hypermethylation in LAC. In fact, 85.1% of all of the 74 identified DMRs and all 15 confirmed DMRs were

hypermethylated. This overrepresentation of hypermethylated DMRs can be explained by the fact that the microarray used in our study is designed to specifically target CpG islands and reference gene promoter regions, which are known to frequently undergo *de novo* DNA methylation during tumorigenesis.

5

In order for a biomarker to be clinically relevant, it needs to be capable of distinguishing cancerous from healthy tissue with high sensitivity and specificity, as well as deliver unambiguous results. MS-HRM analysis allows implementation of assay-specific cutoff values, which can be useful when investigating methylation changes in regions with frequent low-level methylation in the surrounding non-cancerous tissue. However, the use of assay-specific cutoff values is challenging for clinical purposes, as a tumor-related increase in methylation can be easily masked by the normal methylation level in contaminating normal cells, which are inevitably present in surgical resections and biopsies. The tumor cell content in clinical specimens vary extensively between samples and the biomarker assessment assays therefore require a high dynamic range in order to successfully test samples with both high and low tumor content and this is difficult to achieve when introducing higher cutoff values. While this can be overcome through macro- or microdissection of each specimen prior to biomarker assessment, it greatly reduces the time-efficiency and increases the cost of the individual experiment and thus limits the clinical potential of a candidate biomarker. It is therefore highly preferential to target regions that do not show methylation in normal tissue, as any increase in methylation, regardless of the magnitude, can be attributed to the presence of cancerous cells regardless of the tumor content in the clinical specimen.

25 We have evaluated the sensitivity and specificity for the MS-HRM assays targeting the 15 confirmed DMRs and identified the *OSR1*, *SIM1* and *HOXB3/HOXB4* regions, which all showed minor to no methylation in the surrounding normal tissue, as the most promising biomarkers in LAC. The *OSR1* region demonstrated a remarkably high sensitivity and specificity, as we detected hypermethylation in 97.9% of the LAC tumors (n=48) and only in 3.2% of the tumor-adjacent normal lung samples (n=31). The odd-skipped related 1 (*OSR1*) gene encodes a zinc-finger transcription factor that was recently shown to function as a tumor suppressor in gastric cancer by activating *TP53* transcription ENREF_22. Furthermore, *OSR1* was shown to be silenced by promoter hypermethylation in 51.8% (n=164) of gastric cancer patients and was identified as an independent predictor of poor survival. Rauch et al. also reported *OSR1*

35

hypermethylation in 100% of the LACs (n=8) tested in their study, which underlines the potential of the region as a diagnostic biomarker in LAC. The single-minded homolog 1 (*SIM1*) region also showed high potential as a biomarker in LAC. *SIM1* is frequently methylated in astrocytoma and breast cancer, but this study is the first to describe

5 hypermethylation in lung cancer. A substantial subset of the 74 identified DMRs, including 5/15 of the validated DMRs, *HOXD3*, *HOXB3/HOXB4*, *HOXD10*, *HOXA3* and *HOXA5*, were associated with homeobox genes. Hypermethylation of homeobox genes is a common observation in genome-wide methylation screening studies and have been reported in several cancers, including lung cancer. While the homeobox genes

10 lack tumor subtype specificity, they may still be useful in combination with other diagnostic biomarkers in LAC. The *HOXB3/HOXB4* region showed a tendency towards increased methylation in metastasizing compared to non-metastasizing tumors. We detected higher methylation levels in the metastases compared to the primary tumors for the majority of the DMRs. These results show that the hypermethylation observed in

15 the primary tumors is maintained during the metastatic process, and suggests that it may play an important role in LAC development and progression.

This study was performed using tumor-adjacent normal lung tissue as a control due to the limited availability of lung tissue from healthy individuals. We were able to validate

20 differential methylation in 15 DMRs, but it is possible that the low-level methylation that is observed in a small subset of the normal samples for several of the regions, e.g. *HOXD10*, *HOXD3* and *SIM1*, is a result of the use of tissue, which have been exposed to the same environmental factors as the tumor tissue. It would therefore be highly relevant to investigate if the low-level methylation observed in these regions is present

25 in normal lung tissue from healthy individuals as well. If this is not the case, then several of the DMRs that were excluded as a result of too low specificity, in particular *HOXD3* and *HOXD10*, will hold a high clinical potential as well.

The Chr6(p22.1), *HIST1H3G/HIST1H2BI* and *OTX2* regions all demonstrated a

30 specificity of 1.0, but they were excluded due to their lower sensitivity of 0.67, 0.63 and 0.51, respectively. It would therefore be interesting to investigate the clinical potential of these regions in combination with the three most promising regions, *OSR1*, *SIM1* and *HOXB3/HOXB4*, in a larger LAC cohort.

This study has focused on the discovery of differentially methylated regions between

35 primary tumor and tumor-adjacent normal lung tissue and the identified candidate

biomarkers can therefore potentially be applied diagnostically to separate malignant tumors from benign conditions of the lung where biopsy is indicated. Similarly, all primary tumors used in this study were adenocarcinomas and it would therefore be interesting to determine if the candidate biomarkers are specific for this subtype of lung cancer, as any candidate biomarkers with such specificity may be useful diagnostic tools for tumor sub classification. Furthermore, it would be highly relevant to investigate if the candidate biomarkers can be detected in non-invasive patient samples, such as blood or expectorates, as this would allow them to be used in screening programs of high-risk individuals, such as patients suffering from chronic obstructive pulmonary disease (COPD) and thus enable early disease detection. In conclusion, this study has identified 74 DMRs in LAC through a genome-wide methylation screening and confirmed significant changes in methylation for 15 selected regions in a LAC patient cohort using MS-HRM analysis. These 15 DMRs can be targeted as novel diagnostic biomarkers in LAC.

Table 1: DMRs selected for validation by MS-HRM analysis

ID	DMR			MS-HRM assay location relative to known genes and CpG Islands (CGI)				
	No.	Diff. Methylation Score	Diff. Methylation Hypo (-) / Hyper (+)	Upstream	Intragenic	Downstream	CGI	Location (Hg38)
<i>OTX2</i>	1	9.130	+		<i>OTX2</i>		Island	chr14: 56809871- 56809962
<i>HOXD3</i>	2	8.830	+	<i>HOXD3</i>			Shore	chr2: 17616267 6- 17616274 7
<i>HOXB3/HOXB4</i>	3	7.625	+	<i>HOXB3</i>	<i>HOXB4</i>		Shore	chr17: 48577876- 48577988
<i>HOXD10</i>	4	7.318	+		<i>HOXD10</i>		Island	chr2:1761 17574- 17611765 1
Chr1(q21.1).A	5	6.765	+				Island	chr1:1470 80162- 14708026 2
<i>HIST1H3G/HIST1H2BI</i>	6	6.749	+	<i>HIST1H3G, HIST1H2BI</i>			Island	chr6:2627 2252- 26272379

<i>GHSR</i>	7	6.585	+	<i>GHSR</i>	Island	chr3:1724 48360- 17244844 7
<i>SIM1</i>	8	6.191	+	<i>SIM1</i>	Island	chr6:1004 65031- 10046512 5
<i>OSR1</i>	9	6.117	+	<i>OSR1</i>	Island	chr2:1935 7150- 19357251
<i>FRG1 BP</i>	10	5.838	-	<i>FRG1B P</i>	Island	chr20: 30377475- 30377573
Chr6(p22.1)	13	5.653	+		Island	chr6:2820 7550- 28207669
<i>HOXA 3</i>	15	5.435	+	<i>HOXA3</i> <i>HOXA 4</i>	Island	chr7: 27124232- 27124343
<i>LY75- CD30 2</i>	17	5.223	+	<i>LY75- CD302, CD302</i> <i>LY75</i>	Island	chr2: 15979775 6- 15979783 5
<i>CTAG E15</i>	18	5.217	-	<i>CTAG E15</i>		chr7: 14357168 5- 14357179 7
<i>LOC6 48987</i>	19	4.790	+	<i>ANXA 2R</i> <i>LOC64 8987</i>	Island	chr5: 43040396- 43040521
<i>HIST1 H2AJ/ HIST1 H2BM</i>	21	4.635	+	<i>HIST1 H2BM</i> <i>HIST1H 2AJ</i>	Island	chr6:2781 4577- 27814732
<i>HOXA 5</i>	22	4.522	+	<i>HOXA5</i> <i>HOXA 6</i>	Island	chr7: 27143474- 27143563
<i>HIST1 H3E</i>	61	2.961	+	<i>HIST1H 3E</i>	Island	chr6:2622 5275- 26225421

Table 2: DNA methylation frequencies in tumor-adjacent normal lung and primary lung tumors.

ID	Normal Lung					Lung Tumor					<i>P</i>
	Methylation Level					Methylation Level					
	0-1% n (%)	1-10% n (%)	10-50% n (%)	50-100% n (%)	<i>N</i> (%)	0-1% n (%)	1-10% n (%)	10-50% n (%)	50-100% n (%)	<i>N</i> (%)	
<i>OTX2</i>	31 (100)	0 (0)	0 (0)	0 (0)	31 (100)	25 (49.0)	26 (51.0)	0 (0)	0 (0)	51 (100)	<0.0001
<i>HOXD3</i>	23 (74.2)	7 (22.6)	1 (3.2)	0 (0)	31 (100)	0 (0)	7 (14.0)	34 (68.0)	9 (18.0)	50 (100)	<0.0001
<i>HOXB3/HOXB4</i>	31 (100)	0 (0)	0 (0)	0 (0)	31 (100)	12 (25.0)	16 (33.3)	19 (39.6)	1 (2.1)	48 (100)	<0.0001
<i>HOXD10</i>	26 (81.3)	6 (18.7)	0 (0)	0 (0)	32 (100)	5 (9.8)	27 (52.9)	18 (35.3)	1 (2.0)	51 (100)	<0.0001
Chr1(q21.1).A <i>HIST1H3G/HIST1H2B</i>	15 (50.0)	15 (50.0)	0 (0)	0 (0)	30 (100)	0 (0)	10 (20.8)	32 (66.7)	6 (12.5)	48 (100)	<0.0001
<i>GHSR</i>	11 (35.5)	19 (61.3)	1 (3.2)	0 (0)	31 (100)	2 (4.2)	6 (12.5)	29 (60.4)	11 (22.9)	48 (100)	<0.0001
<i>SIM1</i>	28 (90.3)	3 (9.7)	0 (0)	0 (0)	31 (100)	4 (8.2)	22 (44.9)	21 (42.8)	2 (4.1)	49 (100)	<0.0001
<i>OSR1</i>	30 (96.8)	1 (3.2)	0 (0)	0 (0)	31 (100)	1 (2.1)	12 (25.0)	30 (62.5)	5 (10.4)	48 (100)	<0.0001
<i>FRG1BP</i>	0 (0)	0 (0)	32 (100)	0 (0)	32 (100)	0 (0)	4 (7.7)	48 (92.3)	0 (0)	52 (100)	0.159
Chr6(p22.1)	32 (100)	0 (0)	0 (0)	0 (0)	32 (100)	16 (32.6)	13 (26.6)	16 (32.6)	4 (8.2)	49 (100)	<0.0001
<i>HOXA3</i>	0 (0)	14 (48.3)	15 (51.7)	0 (0)	29 (100)	0 (0)	3 (6.0)	40 (80.0)	7 (14.0)	50 (100)	<0.0001
<i>LY75-CD302</i>	30 (96.8)	1 (3.2)	0 (0)	0 (0)	31 (100)	47 (97.9)	1 (2.1)	0 (0)	0 (0)	48 (100)	>0.999
<i>CTAGE15</i>	0 (0)	0 (0)	0 (0)	32 (100)	32 (100)	0 (0)	0 (0)	0 (0)	52 (100)	52 (100)	>0.999
<i>LOC648987</i>	29 (100)	0 (0)	0 (0)	0 (0)	29 (100)	34 (75.5)	7 (15.6)	4 (8.9)	0 (0)	45 (100)	0.005
<i>HIST1H2AJ/HIST1H2</i>	32 (100)	0 (0)	0 (0)	0 (0)	32 (100)	40 (78.4)	11 (21.6)	0 (0)	0 (0)	51 (100)	0.005

BM

<i>HOX</i> <i>A5</i>	0 (0)	11 (35.5)	20 (64.5)	0 (0)	31 (100)	0 (0)	10 (19.2)	36 (69.2)	6 (11.6)	52 (100)	0.03 5
<i>HIST</i> <i>1H3E</i>	0 (0)	0 (0)	29 (90.6)	3 (9.4)	32 (100)	0 (0)	0 (0)	5 (10.0)	45 (90.0)	50 (100)	<0. 000 1

Table 3: Performance of MS-HRM assays

ID	Cutoff Value	Normal Lung			Lung Tumor			Sensitivity	Specificity
		Methylation Level	True Negative n (%)	False positive n (%)	N (%)	False Negative n (%)	True Positive n (%)		
OSR1	>1%		30 (96.8)	1 (3.2)	31 (100)	1 (2.1)	47 (97.9)	0.98	0.97
SIM1	>1%		28 (90.3)	3 (9.7)	31 (100)	4 (8.2)	45 (91.8)	0.92	0.90
<i>HOXD</i> <i>10</i>	>1%		26 (81.3)	6 (18.7)	32 (100)	5 (9.8)	46 (90.2)	0.90	0.81
<i>HIST1</i> <i>H3E</i>	>50%		29 (90.6)	3 (9.4)	32 (100)	5 (10.0)	45 (90.0)	0.90	0.91
<i>HOXD</i> <i>3</i>	>10%		30 (96.8)	1 (3.2)	31 (100)	7 (14.0)	43 (86.0)	0.86	0.97
<i>GHSR</i>	>10%		30 (96.8)	1 (3.2)	31 (100)	8 (16.7)	40 (83.3)	0.83	0.97
Chr1(q 21.1).A	>10%		30 (100)	0 (0)	30 (100)	10 (20.8)	38 (79.2)	0.79	1.00
HOXB 3/HOX B4	>1%		31 (100)	0 (0)	31 (100)	12 (25.0)	36 (75.0)	0.75	1.00
Chr6(p 22.1)	>1%		32 (100)	0 (0)	32 (100)	16 (32.6)	33 (67.4)	0.67	1.00
<i>HIST1</i> <i>H3G/HI</i> <i>ST1H2</i> <i>BI</i>	>1%		32 (100)	0 (0)	32 (100)	19 (36.5)	33 (63.5)	0.63	1.00
<i>OTX2</i>	>1%		31 (100)	0 (0)	31 (100)	25 (49.0)	26 (51.0)	0.51	1.00
<i>LOC64</i> <i>8987</i>	>1%		29 (100)	0 (0)	29 (100)	34 (75.5)	11 (24.5)	0.24	1.00
<i>HIST1</i> <i>H2AJ/</i> <i>HIST1</i> <i>H2BM</i>	>1%		32 (100)	0 (0)	32 (100)	40 (78.4)	11 (21.6)	0.22	1.00
<i>HOXA</i> <i>3</i>	>50%		29 (100)	0 (0)	29 (100)	43 (86.0)	7 (14.0)	0.14	1.00
<i>HOXA</i> <i>5</i>	>50%		31 (100)	0 (0)	31 (100)	46 (88.4)	6 (11.6)	0.12	1.00

Table 4: Clinical characteristics for the patient cohorts

Clinical Characteristics	LAC w/o metastases	LAC w/ metastases
Patients (n)		
Cases	26	26
Cases with available metastatic tissue	0	24
Gender (n (%))		
Male	10 (38.5%)	12 (46.2%)
Female	16 (61.5%)	14 (53.8%)
Age (years)		
Min-Max (Average)	45-76 (62.1)	38-76 (61.2)
Smoking status (n (%))		
Current Smoker	18 (69.2%)	16 (61.5%)
Previous Smoker	8 (30.8%)	8 (30.8%)
Unknown	0 (0%)	2 (7.7%)
TNM Classification (n (%))		
TX	0 (0%)	1 (3.8%)
T1	9 (34.7%)	9 (34.7%)
T2	16 (61.5%)	16 (61.5%)
T3	0 (0%)	0 (0%)
T4	1 (3.8%)	0 (0%)
NX	0 (0%)	2 (7.7%)
N0	21 (80.8%)	12 (46.2%)
N1	5 (19.2%)	3 (11.4%)
N2	0 (0%)	9 (34.7%)
N3	0 (0%)	0 (0%)
M0	26 (100%)	0 (0%)
M1	0 (0%)	26 (100%)
Tumor content (%)		
Tumors, Min-Max (Average)	5-60% (27.0%)	5-80% (39.4%)
Metastases, Min-Max (Average)		5-90% (68.7%)

Methods

5 Patient samples

The regional ethical committee (De Videnskabsetiske Komitéer Region Midtjylland, Permission No.: 1-10-72-20-14) approved this study and all experiments were performed in accordance with the approved guidelines and regulations. Formalin-fixed, paraffin embedded (FFPE) blocks of surgical resections from lung adenocarcinoma (LAC) patients were selected from the archives at the Institute of Pathology, Aarhus

University Hospital. Primary lung tumor and tumor-adjacent normal lung tissue from four LAC patients were used for the microarray analysis. For validation, 52 LAC patients were selected. These patients were divided into two cohorts matched on gender, age, smoking status, histology, T-stage and proportion of tumor cells. The first cohort comprises FFPE primary tumor and paired metastatic tissue (20 brain and 4 adrenal gland) from 26 patients that suffered from distant metastatic disease at the time of diagnosis. The second cohort comprises FFPE primary tumor tissue from 26 distant metastases-free patients with a minimum of 5 years recurrence-free survival following surgical resection. The histological and clinical characteristics for the 52 patients are shown in Table 4. FFPE tumor-adjacent normal lung tissue was selected by an experienced pathologist from 32 LAC patients and used as a control cohort. Peripheral blood samples obtained from healthy medical students of both sexes were used to generate unmethylated control DNA. Written informed consent was obtained from the subjects.

DNA extraction and Sodium Bisulfite treatment

For each FFPE sample, DNA was extracted from $5 \times 10 \mu\text{m}$ sections using the QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. DNA was extracted from the peripheral blood (PB) samples using a modified salt precipitation protocol. In brief, 10 ml blood was incubated for 30 min at 4°C with 40 ml Triton lysis buffer (1% Triton X-100, 10 mM Tris, 0.32 M sucrose, 5 mM MgCl_2) and spun for 30 min at 3-4000 rpm (4°C). The supernatant was then removed and the nuclei were washed using 0.9% NaCl. After a 10 min spin 2300 rpm, the remaining supernatant was discarded and the nuclei were lysed using 3 ml nuclei lysis buffer (24 mM EDTA, 75 mM NaCl), 230 μl 10% SDS and 25 μl pronase (20mg/ml) and left shaking at room temperature over night. For each 3 ml nuclei lysis buffer, 1 ml saturated NaCl (6 M) was added and the mix was vigorously shaken for 15 sec. The supernatant was collected after a 15 min spin at 3000 rpm (4°C) and transferred to a new tube. After an additional 15 min spin at 3000 rpm (4°C), isopropanol was added (1:1) to the supernatant and gently shaken until the DNA precipitated. The precipitated DNA was then collected mechanically using a blunt end glass rod and transferred to a tube containing 400 μl double-distilled H_2O . DNA concentrations were measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). For MS-HRM analysis, 500 ng genomic DNA from each sample was subjected to sodium bisulfite treatment using the EZ-96 DNA Methylation-

Gold™ kit (Zymo Research, Irvine, USA) according to the manufacturer's instructions and eluted in a final volume of 52 µl.

5 *Microarray Analyses*

The microarray based screening for differentially methylated regions (DMRs) was performed as previously described³⁴. Briefly, DNA was extracted from primary tumor and tumor-adjacent normal lung tissue from four LAC patients. After DNA extraction, a methylated DNA immunoprecipitation (MeDIP) was performed in order to enrich the methylated fragments. A detailed description of the MeDIP protocol can be found in³⁴. Two fractions from each sample (MeDIP enriched and input) were subsequently labeled with Cy5 and Cy3 and cohybridized to the NimbleGen Human DNA Methylation 3x720K CpG Island Plus RefSeq Promoter Array (Roche/NimbleGen, Madison, WI, USA). The arrays were processed using NimbleScan software (Roche/NimbleGen, Madison, WI, USA) to generate log2 signal ratios for each probe. The ratios were then averaged within each group (tumor and normal lung) and subsequently processed by the NimbleScan software to generate a relative enrichment score for each group. The enrichment scores for each group were then subtracted to produce a differential methylation score indicating an enrichment or depletion of signal in the tumor group relative to the normal lung group. Hence, negative and positive differential methylation scores indicate potentially hypo- and hypermethylated loci in lung cancer, respectively. A threshold of 2 was applied to the differential methylation score. A large fraction of the probes with differential methylation scores ≥ 2 were located in close proximity and the probes that were located within 5000 bp of each other were therefore grouped into differentially methylated regions (DMRs) with at least two probes targeting each DMR. Eighteen DMRs were then selected for validation based on the differential methylation score and the number of probes that mapped to the region. Previously undescribed and hypermethylated regions were prioritized. All validation experiments were performed using MS-HRM.

30

Methylation-Sensitive High-Resolution Melting (MS-HRM)

Validation of the 18 potential DMRs was performed by MS-HRM analysis^{35,36}. The LightCycler® 480 platform (Roche, Mannheim, Germany) was used for PCR and HRM, and each reaction comprised 1 × MeltDoctor™ HRM Master Mix (Life Technologies,

35

Carlsbad, CA, USA), 3 mM MgCl₂, 500 nM of each primer and 10 ng of bisulfite modified DNA in a final volume of 10 µl. All primers were designed to amplify both methylated and unmethylated DNA as described by Wojdacz et al.³⁷. The methylation status of each DMR was determined by comparing the melting profiles of each sample with a standard dilution series of fully methylated DNA (Universal Methylated Human DNA Standard, Zymo Research, Irvine, CA USA) into unmethylated DNA, which was generated by subjecting DNA extracted from PB to whole genome amplification (WGA) using the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Life Sciences, Piscataway, NJ, USA) according to the manufacturer's instructions. All analyses were performed in duplicates. The technical specifications for each of the 18 assays, including the genomic location of the used primers, PCR cycling and HRM protocol, as well as melting profiles of the standards are included as Supplementary Information.

Statistical analyses and calculation of sensitivity and specificity

Statistical analyses were done using GraphPad Prism version 6 software (GraphPad Software, La Jolla, CA, USA). A Mann-Whitney Test of Ranks was used to assess the statistical significance for each DMR. To perform this test, all samples were ranked based on the determined level of methylation for each DMR; 0-1% methylation was ranked 1, 1-10% methylation was ranked 2, 10-50% methylation was ranked 3 and 50-100% methylation was ranked 4. Two-tailed p -values ≤ 0.05 were considered statistically significant. To evaluate the clinical potential of the candidate biomarkers, we calculated the sensitivity and specificity for each region. The sensitivity was calculated as True Positives / (True Positives + False Negatives) and the specificity was calculated as True Negatives / (True Negatives + False Positives).

Sequences

Assay 1: OTX2

PCR cycling and HRM protocol for the OTX2 MS-HRM assay; 1 cycle of 95°C for 10 minutes. 1 cycle of 95°C for 15 seconds, 40 cycles of 1 minute at 61°C. 95°C for 1 minute, 55°C for 1 minute and a melting phase from 55°C to 95°C with a temperature increase of 0.1°C/sec and 50 fluorescence acquisition points per °C. 95°C for 1 minute. Amplicon length = 92 bp.

Genomic Location Hg38: Chr14: 56809871-56809962

5 TGCCGGGGGTGCCCAGCGCCGAGCCGGAGCCCGACTTGGCAGGTGCTGCTCCGCCTGGTTT
 ||:++|||||::||++:++||:++|||::++|:||||:||||:|:|:++::|||||
 TGTCCGGGGTGTCTTAGCGTCGAGTCGGAGTTCGATTGGTAGGTGTTGTTTCGTTTGGTTT (SEQ ID
 NO: 22)

MS-HRM Primers:

SEQ ID NO: 43: *HOXD10* F: 5' – TAA AGG TCG GTA TGA GTA GAG TTG TT – 3'

SEQ ID NO: 44: *HOXD10* R: 5' – ACC CCG ACA CTC CCT CTC TA – 3'

10

Assay 5: Chr1(q21.1).A

15 PCR cycling and HRM protocol for the Chr1(q21.1).A MS-HRM assay; 1 cycle of 95°C for 10 minutes. 1 cycle of 95°C for 15 seconds, 40 cycles of 1 minute at 56°C. 95°C for 1 minute, 55°C for 1 minute and a melting phase from 55°C to 95°C with a temperature increase of 0.1°C/sec and 50 fluorescence acquisition points per °C. 95°C for 1 minute. Amplicon length = 101 bp.

Genomic Location Hg38: Chr1: 147080162-147080262

20

SEQ ID NO: 5:
 CCGACTCCCGCCTCCAAGCGCGGCTCTTGGCTCGCTGGCGGGCAGCGTCCACAGAGTGTG
 :++|:|:++::|:|++++|:|:||||:|++|+++|:|++|:|:|||||
 TCGATTTTCGTTTTTAAGCGCGGTTTTTGGTTCGTTGGCGGGTAGCGTTTATAGAGTGTG

25

GAACCGCCGAGCCGAGCTCCCGCCCGCTGGCGGGCAGACACTAGCAGGAGAAAGGACA
 |||:++:++||:++||:|:++::++:|++|:|:|:|:|:|:|:|:|:|:|:|:|
 GAATCGTCGTAGTCGTAGTTTTCGTTCGTTGGCGGGTAGATATTAGTAGGAGAAAGGATA

30

CAAGGCCTGCGTGGTGGGAAAGCATGGGAGACCTCGCTTTCCACCGGACGAGAAGGTCT
 :||||:|++|++++|++++|:|++++|:|++|:|:++|++|++++|:|
 TAAGGTTTTCGTTGGTGGGAAAGTATGGGAGATTTCGTTTTTTTATCGGACGAGAAGGTTT (SEQ ID
 NO: 23)

MS-HRM Primers:

35

SEQ ID NO: 45: Chr1(q21.1).A F: 5' – GGT AGC GTT TAT AGA GTG TGG AAT – 3'

SEQ ID NO: 46: Chr1(q21.1).A R: 5' – TTT CCC ACC ACG CAA ACC TTA TA – 3'

Assay 6: HIST1H3G/HIST1H2BI

40 PCR cycling and HRM protocol for the *HIST1H3G/HIST1H2BI* MS-HRM assay; 1 cycle of 95°C for 10 minutes. 1 cycle of 95°C for 15 seconds, 40 cycles of 1 minute at 62°C. 95°C for 1 minute, 55°C for 1 minute and a melting phase from 55°C to 95°C with a

temperature increase of 0.1°C/sec and 50 fluorescence acquisition points per °C. 95°C for 1 minute. Amplicon length = 128 bp.

Genomic Location Hg38: Chr6: 26272252-26272379

5

SEQ ID NO: 6:

TACCTGGAGGCGGTGCTGGAGTACCTGACCGCCAAGATCCTGGAGTTGGCGGGCAACGCG

| | : : | | | | | + + | | | : | | | | | | : : | | | : + + : : | | | | | : : | | | | | | | + + | | : | | + + + +

TATTTGGAGGCGGTGTTGGAGTATTTGATCGTTAAGATTTTGGAGTTGGCGGGTAACGCG

10

GCCCGAGACAAGAAGACCCGCGTCACCCCCGACGCCTGCAGCTCGCCATCCACGACGAG

| : : + + | | : | | | | | : : + + + + | : | : : : : : + + | + + : : | | : | | : | + + : : | | : : | + + | + + | |

GTTCGAGATAAGAAGATTCGCGTTATTTTTCGACGTTTGTAGTTCGTTATTTACGACGAG

15

AAGCTCAACAAGCTGTTGGGCAAAGTTACTATCGCGCAGGGCGGGTGTCTTGCCCAATATT

		:	:	:		:					:				:		++++:			++			::		:::				
--	--	---	---	---	--	---	--	--	--	--	---	--	--	--	---	--	-------	--	--	----	--	--	----	--	-----	--	--	--	--

AAGTTTAATAAGTTGTTGGGTAAAGTTATTATCGCGTAGGGCGGTGTTTTGTTTAATATT (SEQ ID

NO: 24)

20

MS-HRM Primers:

SEQ ID NO: 47: HIST1H3G/HIST1H2BI F: 5' – GAT CGT TAA GAT TTT GGA GTT

GG - 3'

SEQ ID NO: 48: HIST1H3G/HIST1H2BI R: 5' – CGA TAA TAA CTT TAC CCA ACA

ACT T - 3'

25

Assay 7: GHSR

PCR cycling and HRM protocol for the *GHSR* MS-HRM assay; 1 cycle of 95°C for 10 minutes.

1 cycle of 95°C for 15 seconds, 40 cycles of 1 minute at 62°C. 95°C for 1 minute, 55°C for 1 minute and a melting phase from 55°C to 95°C with a temperature increase of 0.1°C/sec and 50 fluorescence acquisition points per °C. 95°C for 1 minute. Amplicon length = 88 bp.

30

Genomic Location Hg38: Chr3: 172448360-172448447

35

SEQ ID NO: 7:

GCTCGTCGCCAGCGAGTCGTTGCCGGGGGAAGCATCCCAGTCCAGGTCGGCCAGTGTGA

| : | ++ | ++ : : : | | ++ | | | ++ | | | : ++ | | | | | | : | | : : : | | | : : | | | ++ | : : | | | | | |

GTTCGTCGTTTAGCGAGTCGTTGTCGGGGGAAGTATTTTAGTTTAGGTCGGTTAGTGTGA

40

GGTTGAACCCCGGCTCTTCGCTGGGCGTCGCGTTCCACATGCTGCCGGCTCAGCTGAACA

| | | | | | | : : : + + | : | : | + + : | | | + + | + + + + | | : : | : | | : | | : + + | : | : | | : | | | : |

GGTTGAATTCGGTTTTTCGTTGGGCGTCGCGTTTTATATGTTGTCGGTTTAGTTGAATA

45

GGCTCTGGGACGTGACTGCGCTGGGAGGCTGGACCGAGCTGGCTCCCGAGGAGGTCCGCT

||:|:|||||++|||:|++:|||||:||||:++||:|:|:++|||||:++:|
 GGTTTTGGGACGTGATTGCGTTGGGAGGTTGGATCGAGTTGGTTTTTCGAGGAGGTTTCGTT (SEQ ID
 NO: 25)

5 MS-HRM Primers:

SEQ ID NO: 49: *GHSR* F: 5' – GTC GGT TAG TGT GAG GTT GAA TT– 3'

SEQ ID NO: 50: *GHSR* R: 5' – CAC GTC CCA AAA CCT ATT CAA CTA– 3'

Assay 8: *SIM1*

10 PCR cycling and HRM protocol for the *SIM1* MS-HRM assay; 1 cycle of 95°C for 10 minutes.

1 cycle of 95°C for 15 seconds, 40 cycles of 1 minute at 54°C. 95°C for 1 minute, 55°C for 1 minute and a melting phase from 55°C to 95°C with a temperature increase of 0.1°C/sec and 50 fluorescence acquisition points per °C. 95°C for 1 minute. Amplicon
 15 length = 95 bp.

Genomic Location Hg38: Chr6: 100465031-100465125

20 SEQ ID NO: 8:
 GCCCCCTCGCCTCCCATTGGCCTGGCGCGGCCAGGGGCGAGGCGCTCATTTGGCCAATAGG
 |:|:|:|:|++:|:|:|:|:|:|:|+++:|:|:|:|++|:|++:|:|:|:|:|:|:|:|:|
 GTTTTTTCGTTTTTTATTGGTTTGGCGCGGTTAGGGGCGAGGCGTTTATTGGTTAATAGG
 25 GCTGAGTGACACGAGTCGGCCCCGAGCCGCCCTCCGCGCGCCCAGGCTCCGGGCTCTGAA
 |:|:|:|:|:|++|:|++|:|+++:|:+++:|:+++++:|:|:|:|++|:|:|:|:|:|:|
 GTTGAGTGATACGAGTCGGTTTCGAGTCGTTTTTCGCGCGTTTAGGTTTCGGGTTTTGAA
 TCTTACTACCCGCGGGACCGCTGGACTCCTAATGAGCTCAGGGTGCCGGGTCGGCGCTGT
 |:|:|:|:|:|+++:|:+++:|:|:|:|:|:|:|:|:|:|:|++|:|++|++:|:|:|
 30 TTTTATTATTCGCGGGATCGTTGGATTTTAAATGAGTTTAGGGTGTCGGGTCGGCGTTGT (SEQ ID
 NO: 26)

MS-HRM Primers:

SEQ ID NO: 51: *SIM1* F: 5' – GTT TAT TGG TTA ATA GGG TTG AGT GAT – 3'

35 SEQ ID NO: 52: *SIM1* R: 5' – ATC CCG CGA ATA ATA AAA TTC AAA ACC – 3'

Assay 9: *OSR1*

PCR cycling and HRM protocol for the *OSR1* MS-HRM assay; 1 cycle of 95°C for 10 minutes. 1 cycle of 95°C for 15 seconds, 40 cycles of 1 minute at 60°C. 95°C for 1
 40 minute, 55°C for 1 minute and a melting phase from 55°C to 95°C with a temperature increase of 0.1°C/sec and 50 fluorescence acquisition points per °C. 95°C for 1 minute. Amplicon length = 102 bp.

increase of 0.1°C/sec and 50 fluorescence acquisition points per °C. 95°C for 1 minute. Amplicon length = 113 bp.

Genomic Location Hg38: Chr7: 143571685-143571797

5

SEQ ID NO: 14:
CCGGGCTCCTCCATAGCGTCGAGGCTGCTCTGGCGGTCACCGCAGTAACACTGGCCACAA
:++|||:::|||++|||::|++|||::++:||||::|::|
TCGGGTTTTTTTATAGCGTCGAGGTTGTTTGGCGGTATCGTAGTAATATTGGTTATAA

10

CAAGCGGTGGAGAACACGCAGCCTTGGGTCTGGAACCCGAATGCGCACGTGACAACCAAC
: ||| ++ || || || || : ++ : || : || || || : || || : ++ || || ++ : ++ || : || : || :
TAAGCGGTGGAGAATACGTAGTTTTGGGTTTGAATTCGAATGCGTACGTGATAATTAAT

15

CGGAGCGGACCACTGTGGAGCGGGCTGCGGGGGGAGCTGGGGAACGCGGGCACCCACAGG
 ++|||++||:::|||||++||:|++|||:|||||++++||:::|:|
 CGGAGCGGATTATTGTGGAGCGGGTTGCGGGGGGAGTTGGGGAACGCGGGTATTTATAGG (SEQ ID
 NO: 32)

20

MS-HRM Primers:

SEQ ID NO: 63: CTAGE15 F: 5' – GGT TAT CGT AGT AAT ATT GGT TAT A – 3'

SEQ ID NO: 64: CTAGE15 R: 5' – CAA CCC GCT CCA CAA TAA TTC – 3'

Assay 15: LOC648987

25

PCR cycling and HRM protocol for the *LOC648987* MS-HRM assay; 1 cycle of 95°C for 10 minutes. 1 cycle of 95°C for 15 seconds, 40 cycles of 1 minute at 60°C. 95°C for 1 minute, 55°C for 1 minute and a melting phase from 55°C to 95°C with a temperature increase of 0.1°C/sec and 50 fluorescence acquisition points per °C. 95°C for 1 minute. Amplicon length = 126 bp.

30

Genomic Location Hg38: Chr5: 43040396-43040521

SEQ ID NO: 15:
TTTAGCAGCACCCCTTTACGGCGCCAAAACGGATATTTGTTTGGCAATACCAGCGCTATCC
|||::|::|::|++|++::|||++| ||||| |||||: ||||::|++:| |:+
TTTAGTAGTATTTTTTACGGCGTTAAAACGGATATTTGTTTGGTAATATTAGCGTTATTC

35

GCTAGGTGCCGGCGCTTGCTAAGTTC AACGCGCCAGTTTCTCGTTTGCAAGGTGGTTAGG
+ : ||||| : ++ | ++ : | : ||||| : | + + + + : ||||| : ++ ||||| : |||||
GTTAGGTGTCGGCGTTTGTTAAGTTT AACGCGTTAGTTTTTCTCGTTTGTAAGGTGGTTAGG

40

GCAGAGCCCTAGCAGACAGTTTTTCCGGTGGCAGCAACGCTCATTTCCCGGAAACGGGTGG
|:||||::|||:|||||:++|||:|:|++:|:||||:++|||++|||
GTAGAGTTTTAGTAGATAGTTTTTCGGTGGTAGTAACGTTTATTTTTTCGGAAACGGGTGG (SEQ ID
NO: 33)

45

MS-HRM Primers:

SEQ ID NO: 65: LOC648987 F: 5' – ACG GAT ATT TGT TTG GTA ATA TTA G – 3'

SEQ ID NO: 66: LOC648987 R: 5' – CTA CCA CCG AAA AAC TAT CTA C – 3'

Assay 16: HIST1H2AJ/HIST1H2BM

- 5 PCR cycling and HRM protocol for the *HIST1H2AJ/HIST1H2BM* MS-HRM assay; 1 cycle of 95°C for 10 minutes. 1 cycle of 95°C for 15 seconds, 40 cycles of 1 minute at 58°C. 95°C for 1 minute, 55°C for 1 minute and a melting phase from 55°C to 95°C with a temperature increase of 0.1°C/sec and 50 fluorescence acquisition points per °C. 95°C for 1 minute. Amplicon length = 156 bp.

10

Genomic Location Hg38: Chr6: 27814577-27814732

SEQ ID NO: 16:

15 TGCCATGTCTGGGCGTGGTAAGCAGGGAGGCCAAAGCTCGCGCCAAGGCCAAGACCCGCTC
| |: : | | | | : | | | | ++ | | | | | : | | | | | : | | | | : | + + + + : : | | | | : | | | | : + + : | :
TGTTATGTTTGGGCGTGGTAAGTAGGGAGGTAAAGTTCGCGTTAAGGTTAAGATTCTGTTT

20

TTCTCGGGCCGGGCTTCAGTTTCCCGTAGGCCGAGTGCATCGCCTGCTCCGCAAAGGCAA
| |: | + + | | : + + | | : | | : | | | | : + + | | | | : + + | | | | : | + + : : | | : | + + : | | | | : | |
TTTTCGGGTCGGGTTTTAGTTTTTCGTAGGTCGAGTGTATCGTTTGTTCGTAAAGGTAA

25

CTATGCGGAGCGGGTCGGTGTGAGAGCGCCGGTGTACCTGGCGGCGGTGCTGGAGTACCT
: | | | | + + | | | + + | | | + + | | : | | | | + + : + + | | | | : : | | | | + + | + + | | : | | | | | | : : |
TTATGCGGAGCGGGTCGGTGTGAGAGCGTCGGTGTATTTGGCGGCGGTGTTGGAGTATTT (SEQ ID
NO: 34)

MS-HRM Primers:

SEQ ID NO: 67: HIST1H2AJ/HIST1H2BM F: 5' – GCG TGG TAA GTA GGG AGG TAA
AG – 3'

30

SEQ ID NO: 68: HIST1H2AJ/HIST1H2BM R: 5' – ACC GAC GCT CCA ACA CC – 3'

Assay 17: HOXA5

- 35 PCR cycling and HRM protocol for the *HOXA5* MS-HRM assay; 1 cycle of 95°C for 10 minutes. 1 cycle of 95°C for 15 seconds, 40 cycles of 1 minute at 54°C. 95°C for 1 minute, 55°C for 1 minute and a melting phase from 55°C to 95°C with a temperature increase of 0.1°C/sec and 50 fluorescence acquisition points per °C. 95°C for 1 minute. Amplicon length = 90 bp.

40

Genomic Location Hg38: Chr7: 27143474-27143563

SEQ ID NO: 17:

CACACATATCAAAAAACAAATGAGCTCTTATTTTGTAAACTCATTTTGGCGGTGCTATCC
: | |: : | | | | : | | | | | : | | | | | | | | : | : | | | | | + + | + + : | | | | : :

TATATATATTAAAAAATAAATGAGTTTTTATTTGTAAATTTATTTTGC GGTCGTTATTT
 AAATGGCCCGGACTACCAGTTGCATAATTATGGAGATCATAGTTCCGTGAGCGAGCAATT
 |||||: :++||: ||: ||||: ||||| ||||| |||||: |||||: ++||| ++||: ||||
 5 AAATGGTTTCGGATTATTAGTTGTATAATTATGGAGATTATAGTTTCGTGAGCGAGTAATT
 CAGGGACTCGGCGAGCATGCACTCCGGCAGGTACGGCTACGGCTACAATGGCATGGATCT
 : |||||: ++|++|: |||: ||: ++|: ||||| ++|: ||++|: ||: |||||: |||||: |
 10 TAGGGATTTCGGCGAGTATGTATTTCCGGTAGGTACGGTTACGGTTATAATGGTATGGATTT (SEQ ID
 NO: 35)

MS-HRM Primers:

SEQ ID NO: 69: HOXA5 F: 5' – TGG TTC GGA TTA TTA GTT GTA TAA T – 3'

SEQ ID NO: 70: HOXA5 R: 5' – TAC CTA CCG AAA TAC ATA CTC – 3'

Assay 18: HIST1H3E

PCR cycling and HRM protocol for the *HIST1H3E* MS-HRM assay; 1 cycle of 95°C for 10 minutes. 1 cycle of 95°C for 15 seconds, 40 cycles of 1 minute at 56°C. 95°C for 1 minute, 55°C for 1 minute and a melting phase from 55°C to 95°C with a temperature increase of 0.1°C/sec and 50 fluorescence acquisition points per °C. 95°C for 1 minute. Amplicon length = 147 bp.

Genomic Location Hg38: Chr6: 26225275-26225421

SEQ ID NO: 18:
 TCCTGCAGCGCCATCACCGCGGAACCTCTGGAAGCGCAGGTCGGTCTTGAAGTCCTGAGCT
 | : : | : | ++ : : | : | : ++ + | | : | : | | | | ++ : | | | ++ | : | | | | : : | | | : |
 TTTTGTAGCGTTATTATCGCGGAATTTTGAAGCGTAGGTTCGGTTTGAAGTTTGTAGTT
 30 ATTTCTCGCACCAGGCGCTGAAACGGCAGCTTCCGGATTAGAAGCTCGGTAGACTTCTGG
 | | | : | ++ : | : | | ++ : | | | | ++ : | : | : ++ | | | | | | : | ++ | | | : | : | | |
 ATTTTTCGTATTAGGCGTTGAAACGGTAGTTTTCGGATTAGAAGTTCGGTAGATTTTGG
 TAGCGACGGATCTCGCGCAGAGCCACGGTGCCAGGGCGGTAGCGATGGGGCTTCTTCACG
 35 | | | ++ | ++ | | : | ++ + : | | | : | ++ | | : | | | ++ | | | ++ | | | : | : | : | ++
 TAGCGACGGATTTTCGCGTAGAGTTACGGTGTTAGGGCGGTAGCGATGGGGTTTTTTTACG (SEQ ID
 NO: 36)

MS-HRM Primers:

SEQ ID NO: 71: HIST1H3E F: 5' – CGC GGA ATT TTG GAA GCG TAG G – 3'

SEQ ID NO: 72: HIST1H3E R: 5' – CGC TAC CGC CCT AAC ACC – 3'

Claims

1. A method of determining non-small cell lung cancer, a predisposition to non-small cell lung cancer, the prognosis of a non-small cell lung cancer, and/or monitoring a non-small cell lung cancer in a subject, said method comprising in a sample from said subject determining the methylation status of at least one gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHRSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).
2. A method for assessing whether a human subject is likely to develop non-small cell lung cancer, said method comprising
 - i. providing a sample from said human subject,
 - ii. determining in said sample the methylation status of at least one gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHRSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).
 - iii. on the basis of said methylation status identifying a human subject that is more likely to develop non-small cell lung cancer.
3. The method according to any of the preceding claims, wherein said method comprises determining the presence of metastatic lung cancer.
4. The method according to any of the preceding claims, wherein said methylation status is determined in the SIM1 gene locus and at least one additional gene locus selected from the group consisting of Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHRSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).
5. The method according to any of the preceding claims, wherein said methylation status is determined in the Chr6(p22.1) gene locus and at least one additional gene locus selected from the group consisting of SIM1, HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHRSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

6. The method according to any of the preceding claims, wherein said methylation status is determined in the HIST1H3G/HIST1H2BI gene locus and at least one additional gene locus selected from the group consisting of SIM1, Chr6(p22.1), HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E,
5 HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).
7. The method according to any of the preceding claims, wherein said methylation status is determined in one or more gene loci selected from the group consisting of SIM1, HIST1H3G/HIST1H2BI, Chr6(p22.1) and/or
10 HOXB3/HOXB4.
8. The method according to any of the preceding claims, wherein the methylation status of at least two gene loci are determined, such as at least three genes, such as at least four gene loci.
9. The method according to any of the preceding claims, wherein said sample
15 comprise lung tissue, such as lung cells and/or genetic material of lung cells.
10. The method according to any of the preceding claims, wherein the methylation status of primary lung tumour cells and/or distant metastases are compared with tumour adjacent normal lung tissue cells.
11. The method according to any of the preceding claims, wherein said methylation
20 status is determined by any method selected from the group consisting of Methylation-Specific PCR (MSP), Whole genome bisulfite sequencing (BS-Seq), HELP assays, ChIP-on-chip assays, Restriction landmark genomic scanning, Methylated DNA immunoprecipitation (MeDIP), Pyrosequencing of bisulfite treated DNA, Molecular break light assays, and Methyl Sensitive
25 Southern Blotting.
12. The method according to any of the preceding claims, wherein methylation status is determined by methylation specific PCR, bisulfite sequencing, COBRA, melting curve analysis, or DNA methylation arrays.
13. The method according to any of the preceding claims, wherein said methylation
30 status is determined by melting curve analysis, such as high resolution melting (HMR) analysis.
14. The method according to any of the preceding claims, wherein said methylation status is determined by a method comprising the steps of

- i) providing a sample, such as a lung tissue sample, from said subject comprising nucleic acid material comprising said gene locus,
 - ii) modifying said nucleic acid using an agent which cleaves nucleic acid sequences in a methylation-dependent manner,
 - 5 iii) amplifying at least one portion of said gene locus using primers, which span or comprise at least one CpG dinucleotide in said gene locus in order to obtain an amplification product, and
 - iv) analyzing said amplification product.
- 10 15. The method according to claim 14, wherein said amplification product is analysed by detecting the presence or absence of amplification product, wherein the presence of amplification product indicates that the target nucleic acid has not been cleaved by said agent, and wherein the absence of amplification product indicates that the target nucleic acid has been cleaved by said agent.
- 15 16. The method according to any of the preceding claims, wherein said methylation status is determined by a method comprising the steps of
- i) providing a sample, such as a lung tissue sample, from said subject comprising nucleic acid material comprising said gene locus,
 - ii) modifying said nucleic acid using an agent which modifies
20 unmethylated cytosine,
 - iii) amplifying at least one portion of said gene locus using primers, which span or comprise at least one CpG dinucleotide in said gene locus in order to obtain an amplification product, and
 - iv) analyzing said amplification product.
- 25 17. The method according to claim 16, wherein said amplification product is analysed for nucleic acid substitutions resulting from conversion of modified cytosine residues, wherein the presence of converted cytosine residues are indicative of unmethylated cytosine residues, and presence of unconverted cytosine residues is indicative of methylated cytosine residues.
- 30 18. The method according to any of the preceding claims, wherein said amplified CpG-containing nucleic acid is analyzed by melting curve analysis

19. The method according to any of the preceding claims, wherein said unmethylated cytosine is modified by bisulfite.
20. The method according to any of the preceding claims, wherein said methylation status is determined by amplifying at least one portion of said gene locus using at least one primer pair selected from the group consisting of SEQ ID NOs: 37-72.
21. The method according to any of the preceding claims, wherein said methylation status is determined by amplifying at least one portion of said at least one gene locus, and wherein the amplified portion is detected using at least one oligonucleotide probe.
22. The method according to any of the preceding claims, wherein said oligonucleotide probe hybridizes to a sequence selected from the group consisting of SEQ ID NO: 1-18 and/or the complement thereof (non-modified strand) or the group consisting of SEQ ID NO: 19-36 and/or the complement thereof (modified strand).
23. The method according to any of the preceding claims, wherein said oligonucleotide probe comprises 10-100 consecutive nucleic acids selected from the group of sequences consisting SEQ ID NO: 1-18 and/or the complement thereof (non-modified strand) or the group consisting of SEQ ID NO: 19-36 and/or the complement thereof (modified strand).
24. A method for categorizing or predicting the clinical outcome of a non-small cell lung cancer of a subject, said method comprising in a sample from said subject determining the methylation status of at least one gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).
25. The method according to claim 24, wherein the presence of methylation is indicative of decreased overall survival, different stage cancer.
26. A method of evaluating the risk for a subject of contracting cancer, said method comprising in a sample from said subject determining the methylation status of a gene locus selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987,

HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

27. A method of treating a non-small cell lung cancer in a human subject, said method comprising the steps of

- 5 i. determining non-small cell lung cancer, a predisposition to non-small cell lung cancer, or the prognosis of a non-small cell lung cancer in a subject by a method as defined in any of the preceding claims,
- ii. selecting human subjects having non-small cell lung cancer, a predisposition to non-small cell lung cancer, or a negative or positive
- 10 prognosis of a non-small cell lung cancer,
- iii. subjecting said subjects identified in step ii. to a suitable treatment for non-small cell lung cancer.

28. The method according to claim 27, wherein said treatment is surgery, chemotherapy and/or radiotherapy.

15 29. A kit for determining non-small cell lung cancer, predisposition to non-small cell lung cancer, or categorizing or predicting the clinical outcome of a non-small cell lung cancer, or monitoring the treatment of a non-small cell lung cancer, said kit comprising

- i. an agent that (a) modifies methylated cytosine residues but not non-
- 20 methylated cytosine residues; or (b) modifies non-methylated cytosine residues but not methylated cytosine residues; or (c) modifies a nucleic acid sequence in a methylation-dependent manner,
- ii. and at least one pair of oligonucleotide primers that specifically hybridizes under amplification conditions to a region of a gene locus selected from
- 25 the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1).

30 30. The kit according to claim 29, wherein said at least one primer pair selected from the group consisting of SEQ ID NO: 37/38 to 71/72, for example at least one primer pair identified as SEQ ID NO: 51 and 52 for SIM1.

31. The kit according to any of claims 27 to 30, wherein said kit comprise at least one oligonucleotide probe comprising 10-100 consecutive nucleic acids selected from the group of sequences consisting SEQ ID NO: 1-18 and/or the

complement thereof (non-modified strand) or the group consisting of SEQ ID NO: 19-36 and/or the complement thereof (modified strand).

- 5 32. The kit according to any of claims 27 to 31, wherein said kit comprise at least one oligonucleotide probe which hybridizes to a sequence selected from the group consisting of SEQ ID NO: 1-18 and/or the complement thereof (non-modified strand) or the group consisting of SEQ ID NO: 19-36 and/or the complement thereof (modified strand).
- 10 33. The kit according to any of claims 27 to 32, said kit further comprising a DNA polymerase.
- 10 34. The kit according to any of claims 27 to 33, wherein said agent is a bisulfite, hydrogen sulfite, and/or disulfite reagent, for example sodium bisulfite.
35. The kit according to any of claims 27 to 33, said kit further comprising a methylation-sensitive restriction enzyme.
- 15 36. Use of oligonucleotide primers comprising a sequence, which is a subsequence of a gene loci selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2, LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1) or the complement thereof for diagnosing non-small cell lung cancer in a method of any of the preceding claims.
- 20 37. The use according to any of the preceding claims, wherein said oligonucleotide primers is selected from the group consisting of SEQ ID NO: 37-72 and/or oligonucleotide primers comprising 10-100 consecutive nucleic acids selected from the group of sequences consisting SEQ ID NO: 1-18 and/or the complement thereof (non-modified strand) or the group consisting of SEQ ID
- 25 NO: 19-36 and/or the complement thereof (modified strand).
38. The use according to any of the preceding claims, wherein said gene locus is selected from the group consisting of SIM1, HOXB3/HOXB4 and OSR1.
39. A method of identifying therapeutically effective agents for treatment of non-small cell lung cancer, said method comprising
- 30 i. providing a non-small cell lung cancer cell line comprising one or more genetic loci selected from the group consisting of SIM1, Chr6(p22.1), HIST1H3G/HIST1H2BI, HOXB3/HOXB4, OSR1, GHSR, OTX2,

LOC648987, HIST1H3E, HIST1H2AJ/HIST1H2BM, HOXD10, HOXD3, HOXA3, HOXA5 and Chr1(q21.1),

- ii. providing one or more potential therapeutic agents,
- 5 iii. treating said non-small cell lung cancer cells by bringing said agents in contact with said non-small cell lung cancer cells,
- iv. determining methylation status of said one or more genetic loci
- v. comparing said methylation status of said treated non-small cell lung cancer cells with the methylation status of said non-small cell lung cancer cells, when untreated, wherein a decreased level of methylation positive
10 alleles is indicative of a therapeutic agent.

1/31

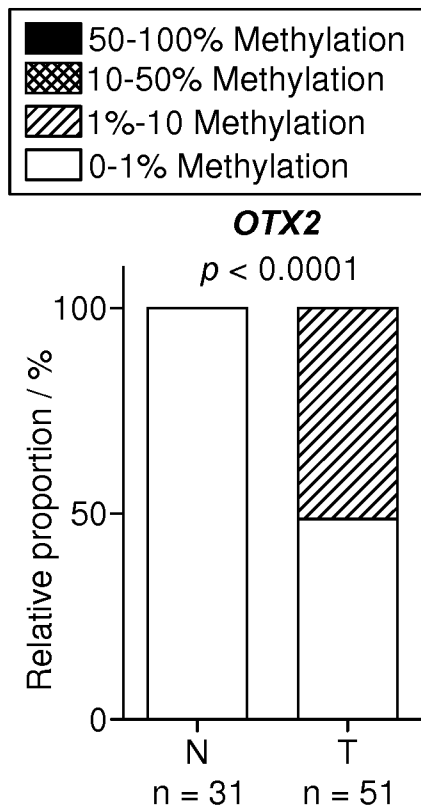


FIG. 1A

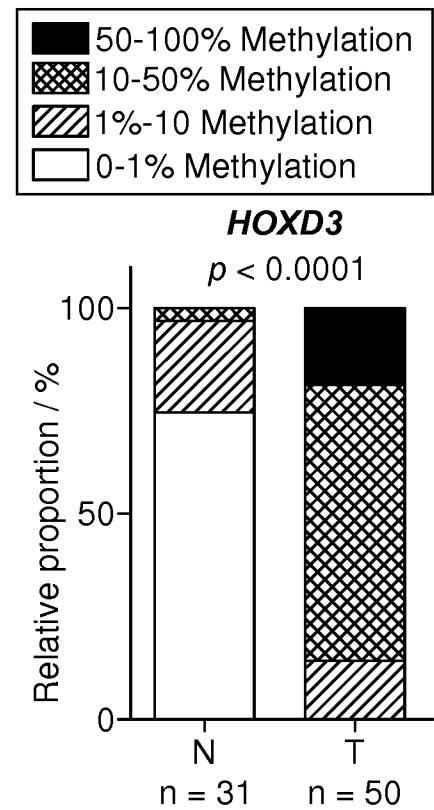


FIG. 1B

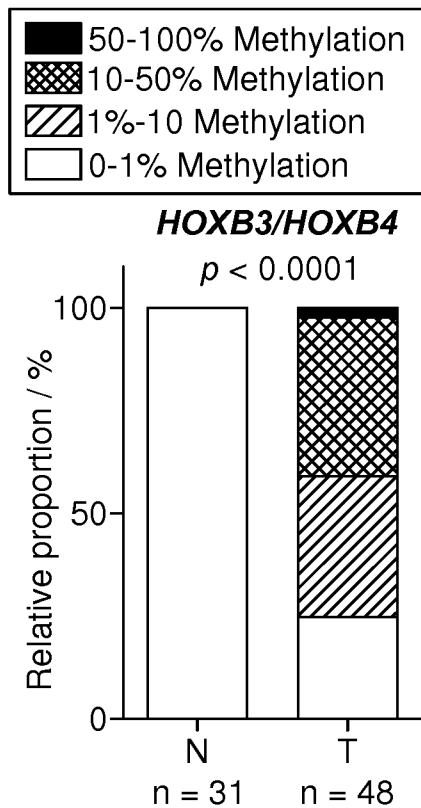


FIG. 1C

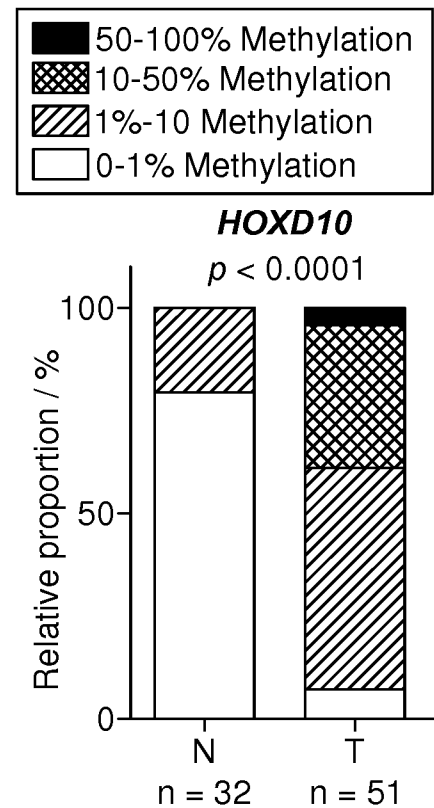


FIG. 1D

2/31

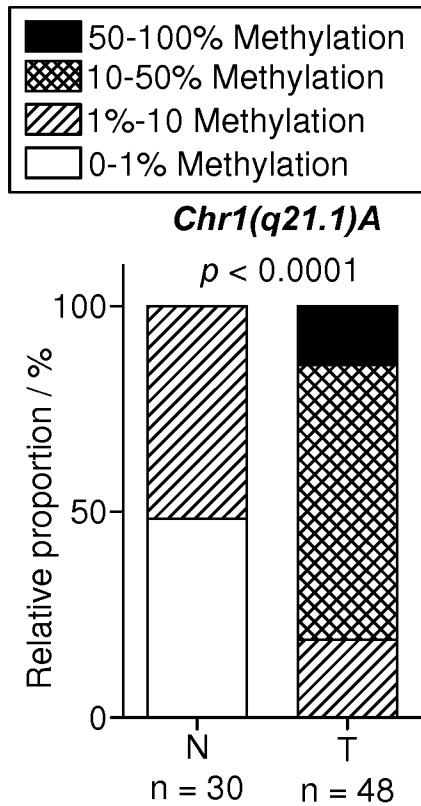


FIG. 1E

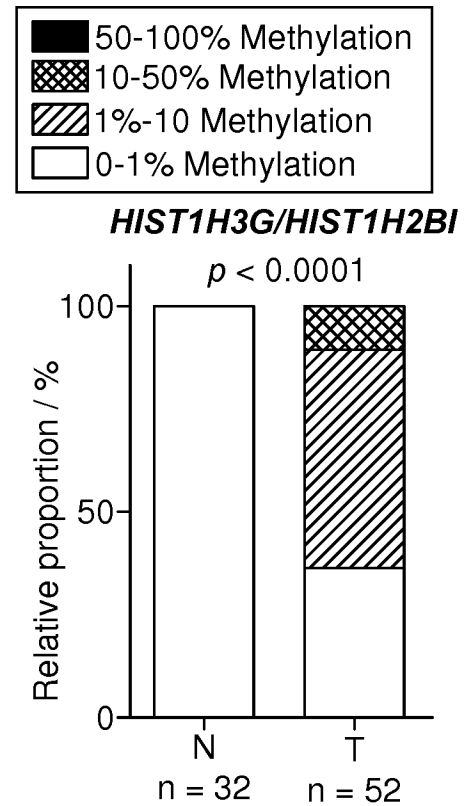


FIG. 1F

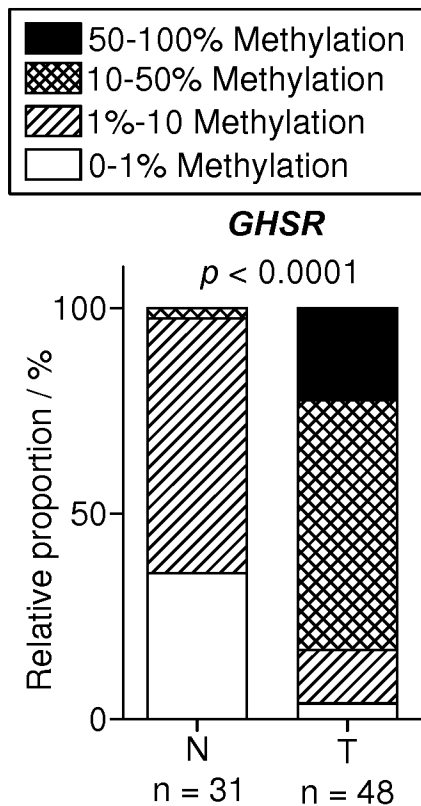


FIG. 1G

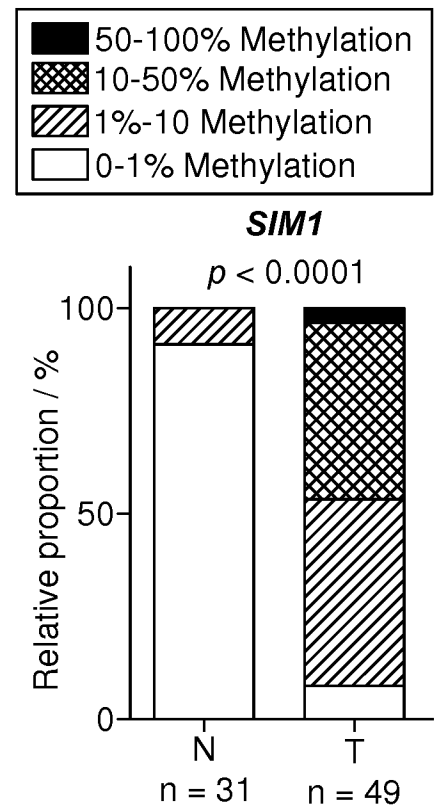


FIG. 1H

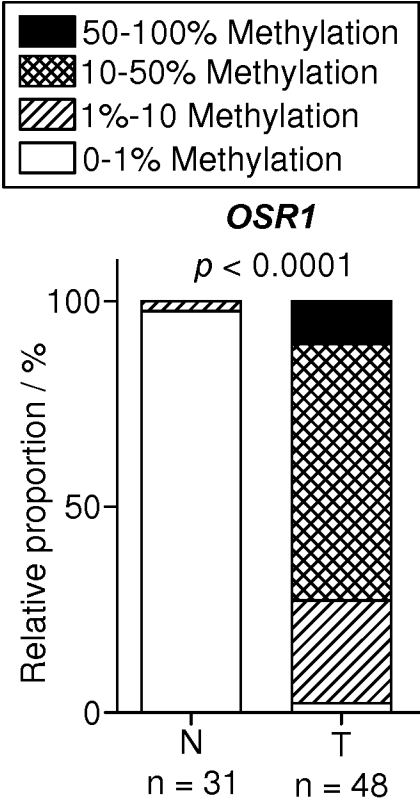


FIG. 1I

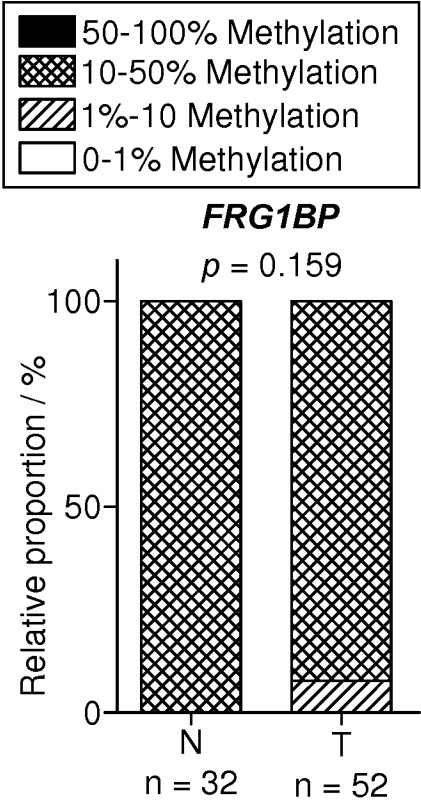


FIG. 1J

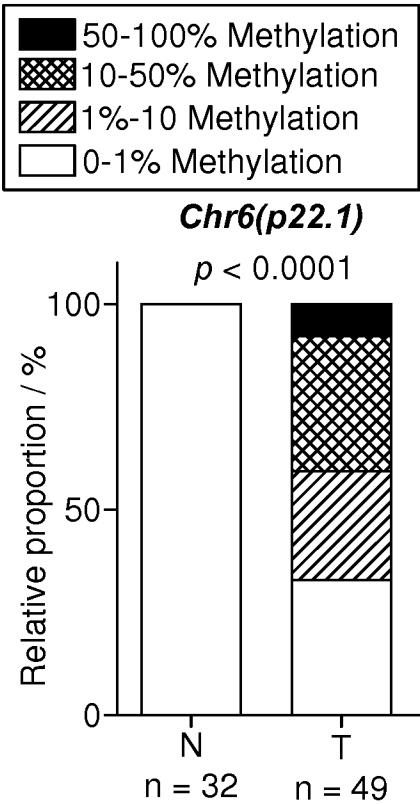


FIG. 1K

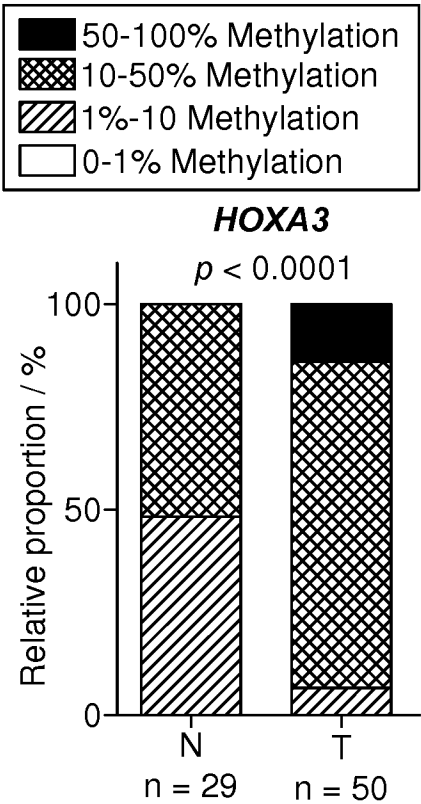
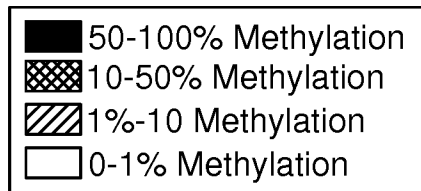
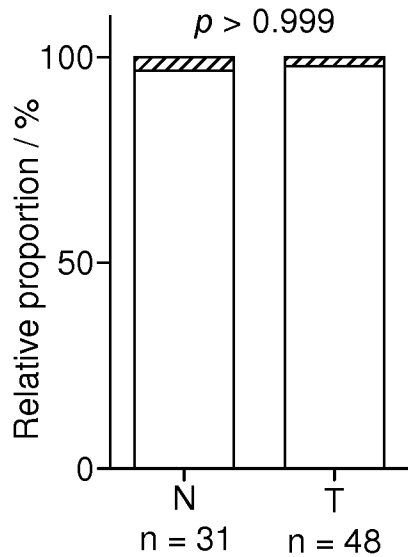
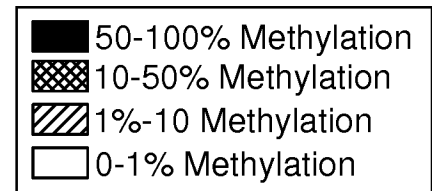
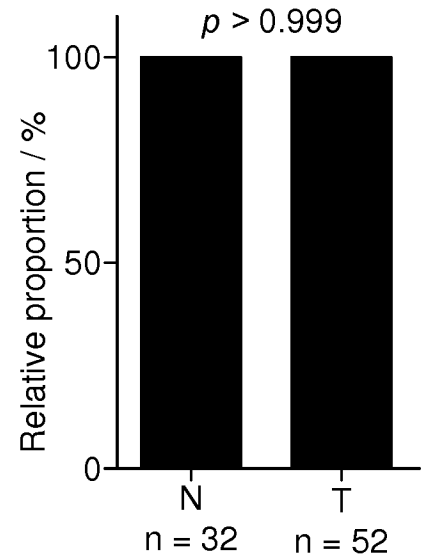
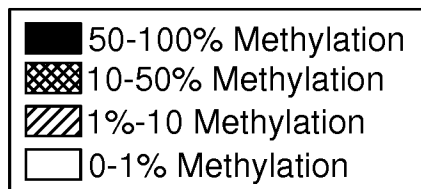
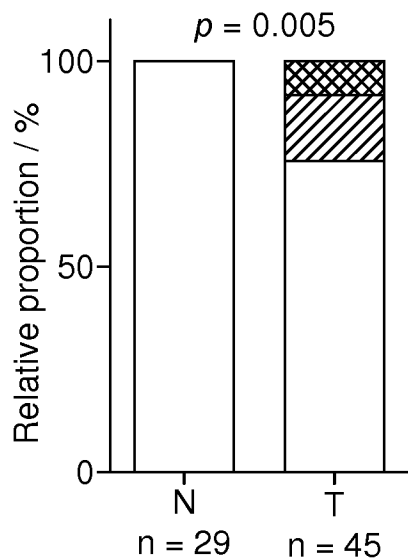
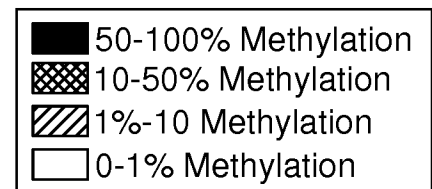
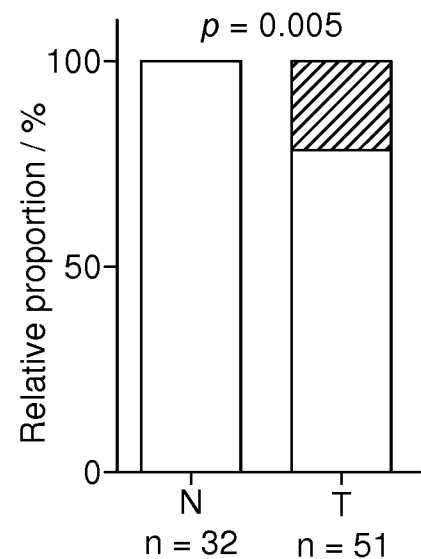


FIG. 1L

4/31

**LY75-CD302****FIG. 1M****CTAGE15****FIG. 1N****LOC648987****FIG. 1O****HIST1H2AJ/HIST1H2BM****FIG. 1P**

5/31

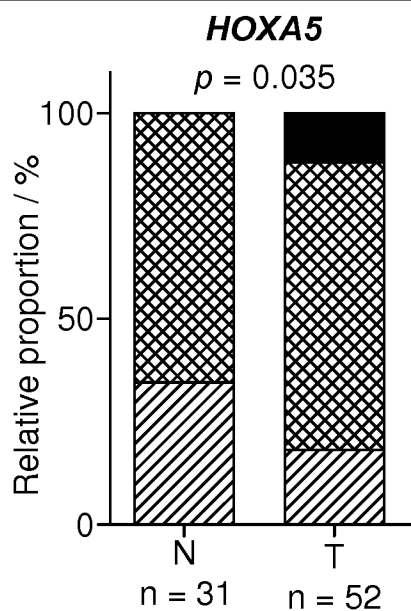
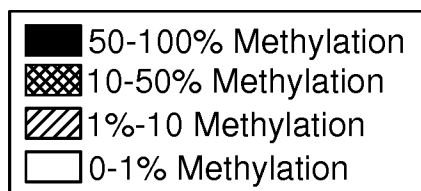


FIG. 1Q

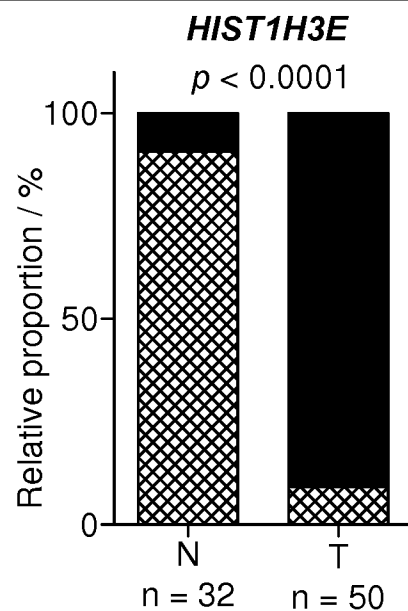
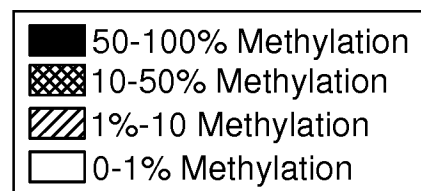


FIG. 1R

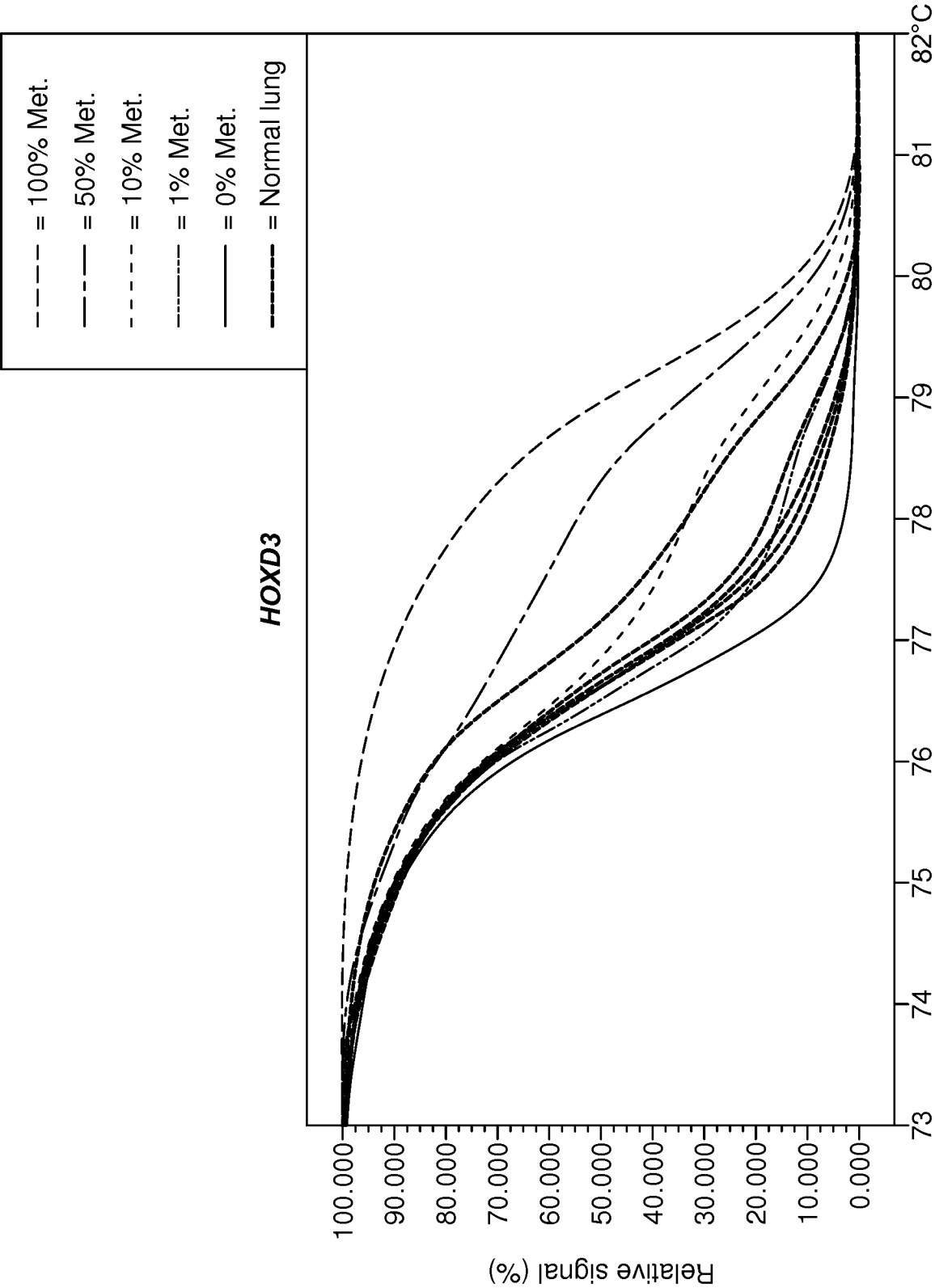


FIG. 2A

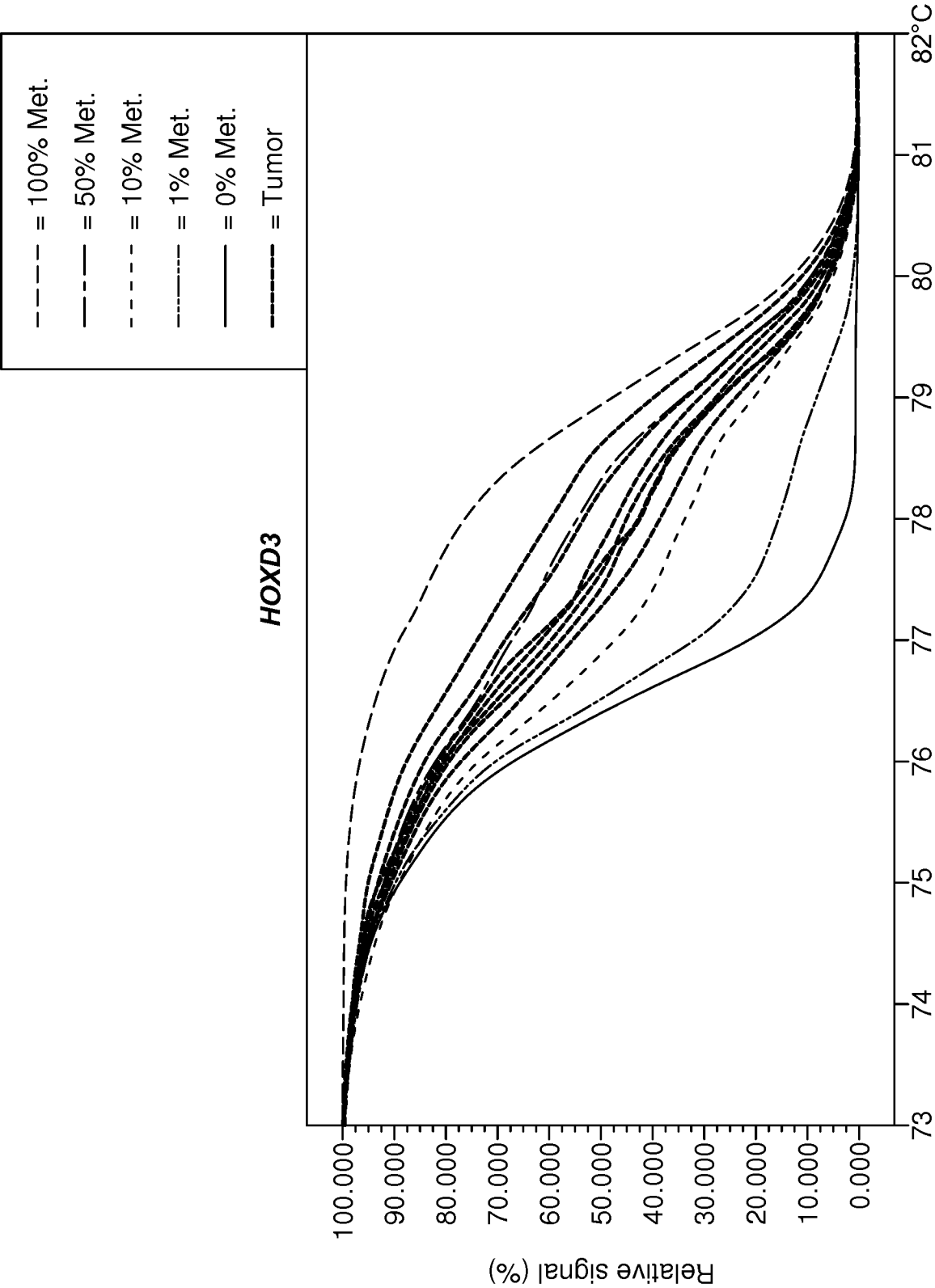


FIG. 2B

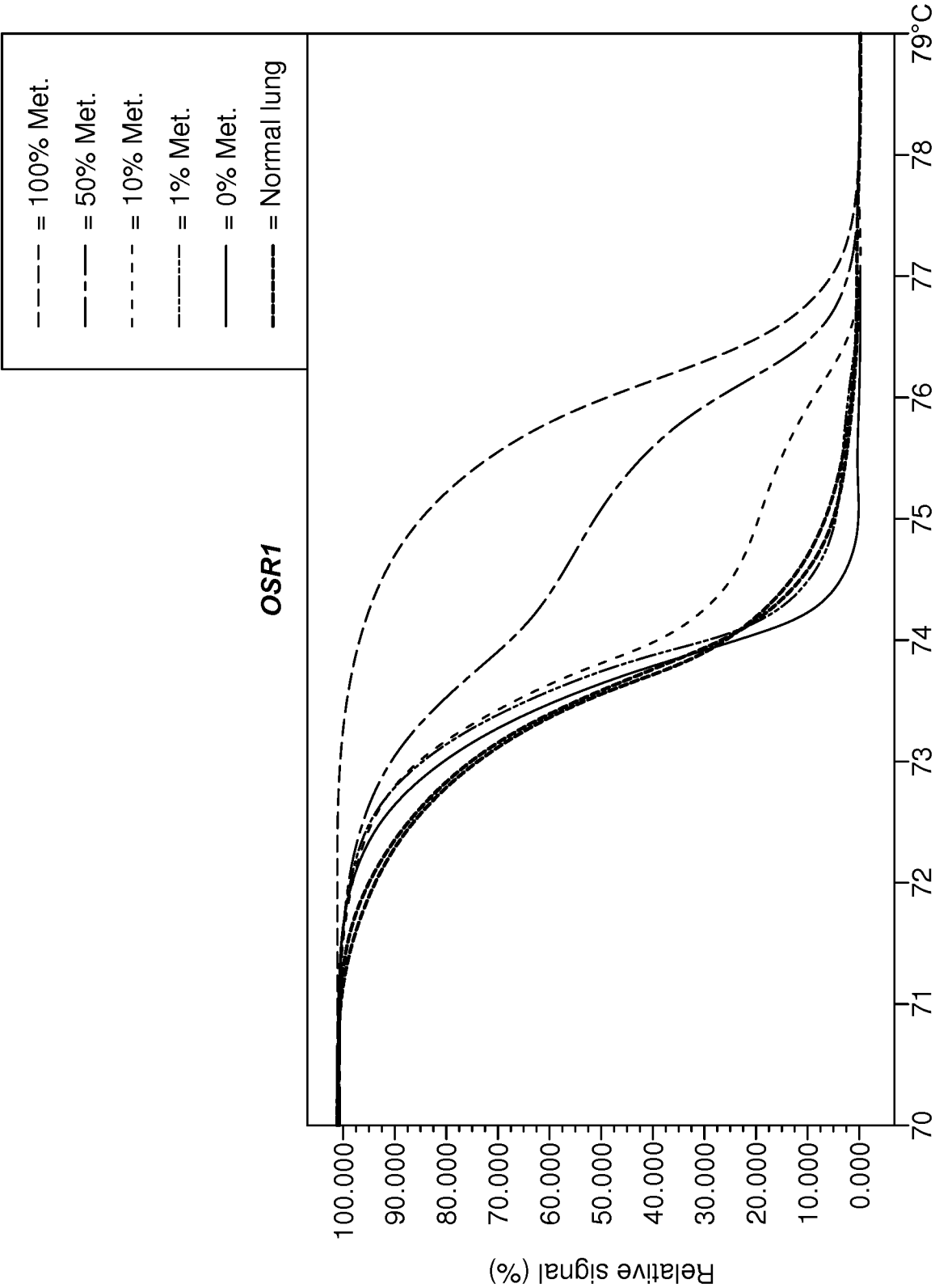


FIG. 2C

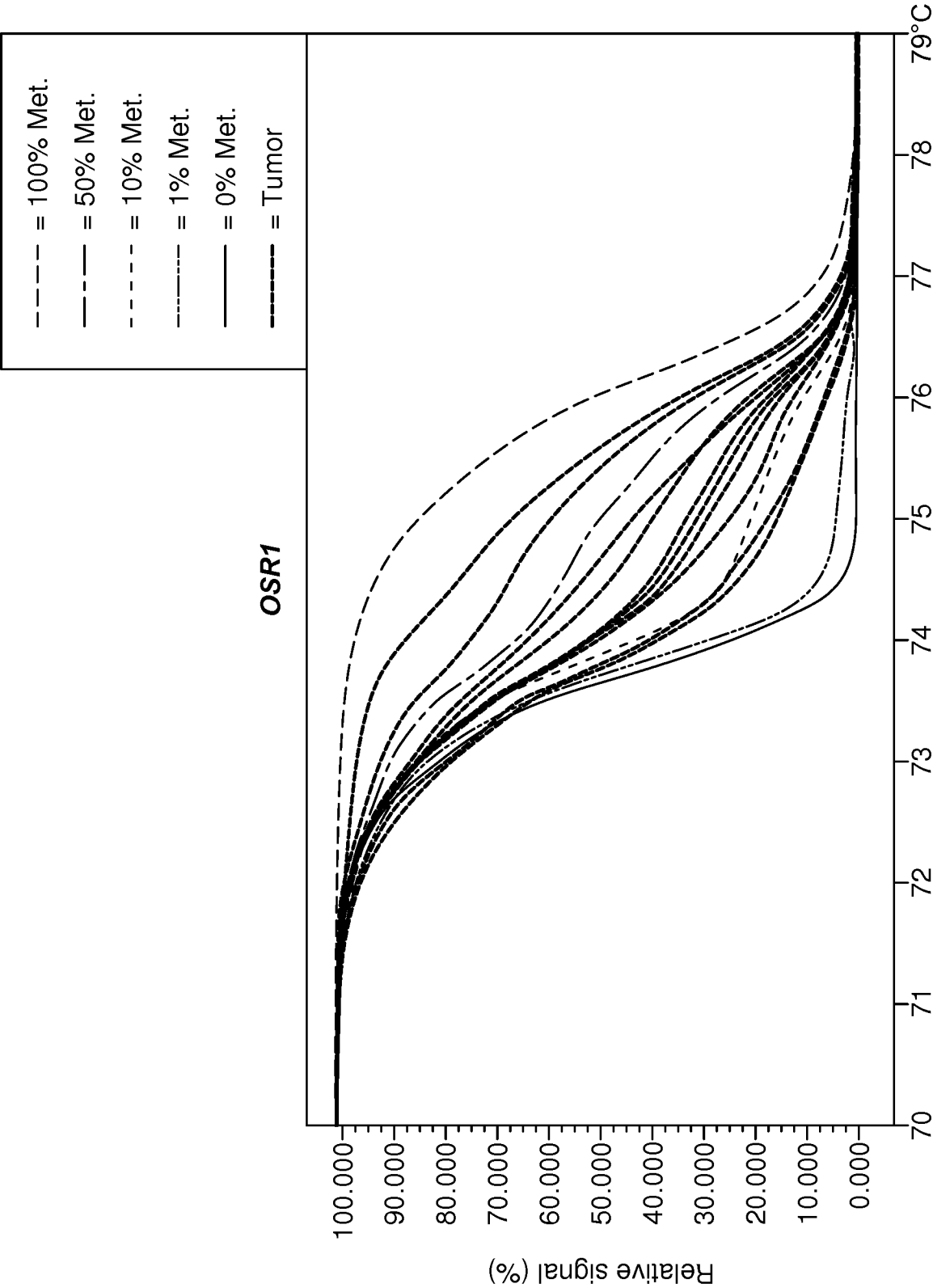


FIG. 2D

10/31

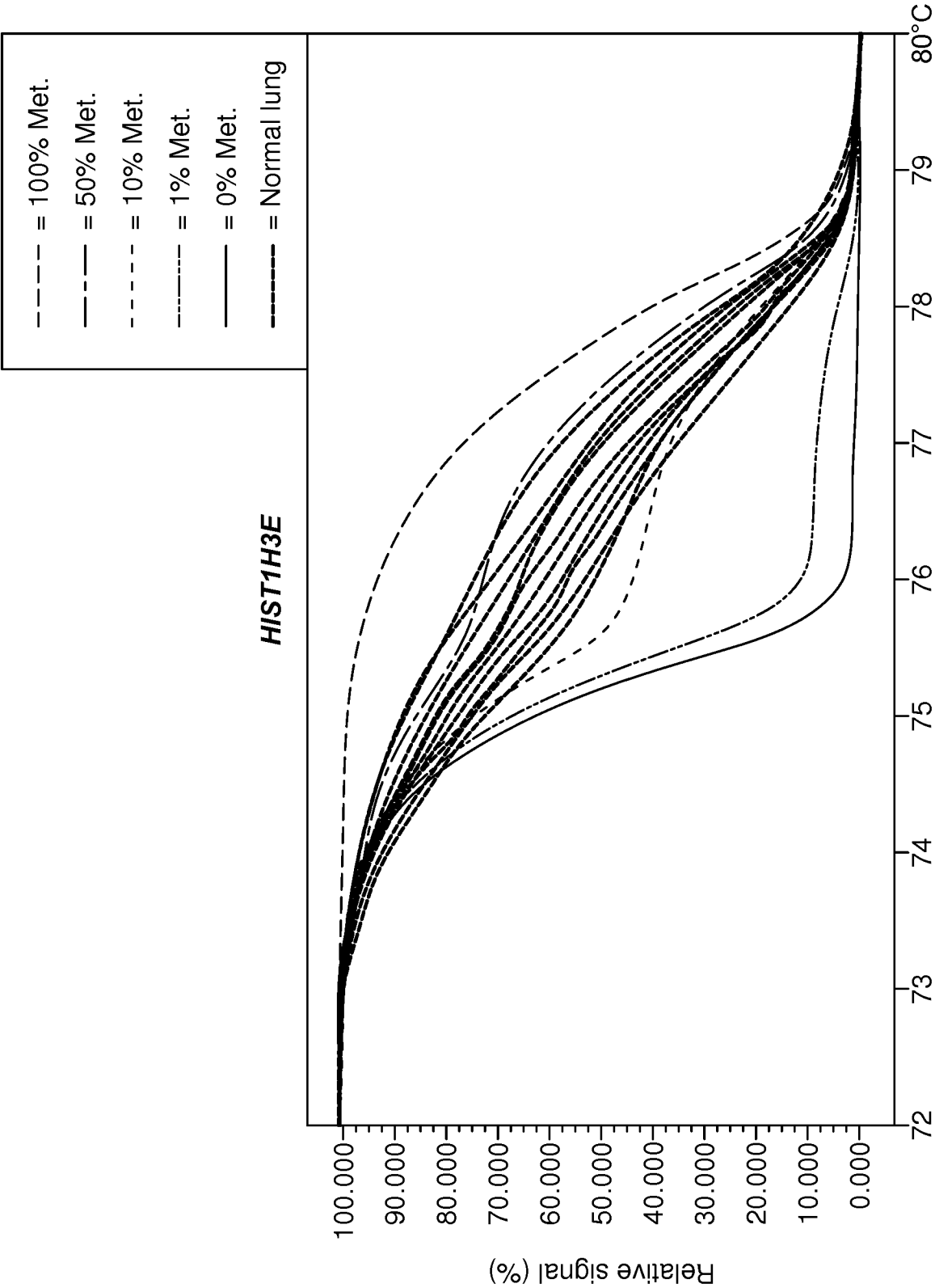


FIG. 2E

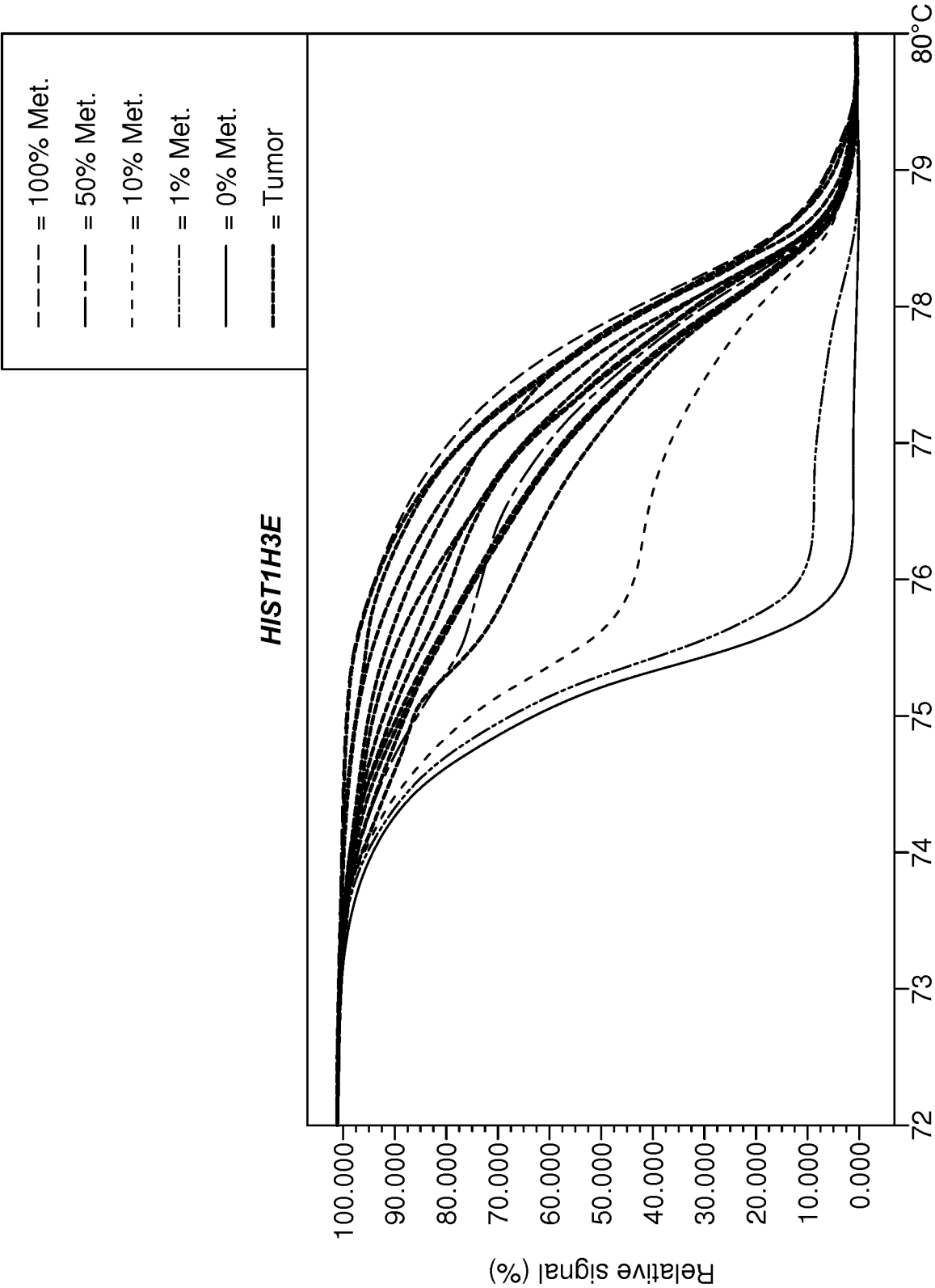


FIG. 2F

12/31

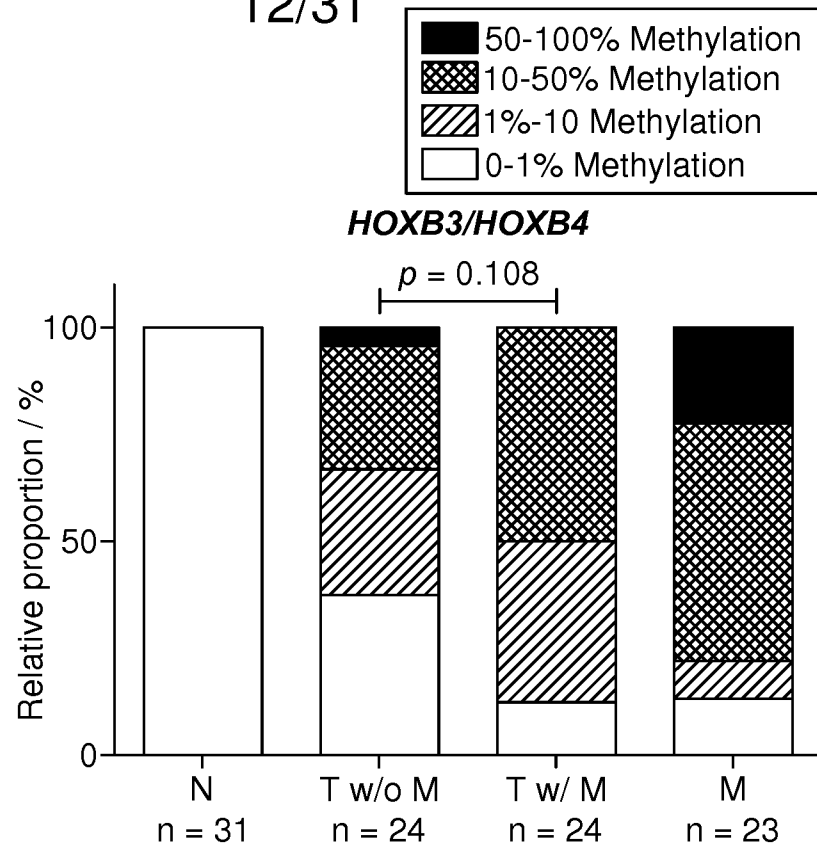


FIG. 3A

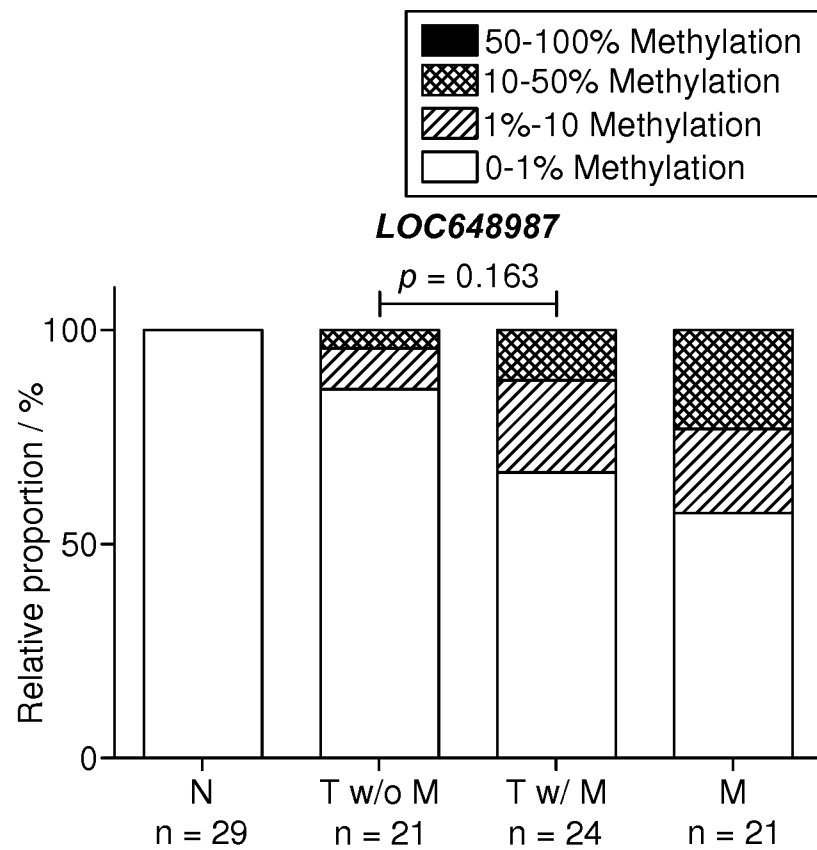


FIG. 3B

13/31

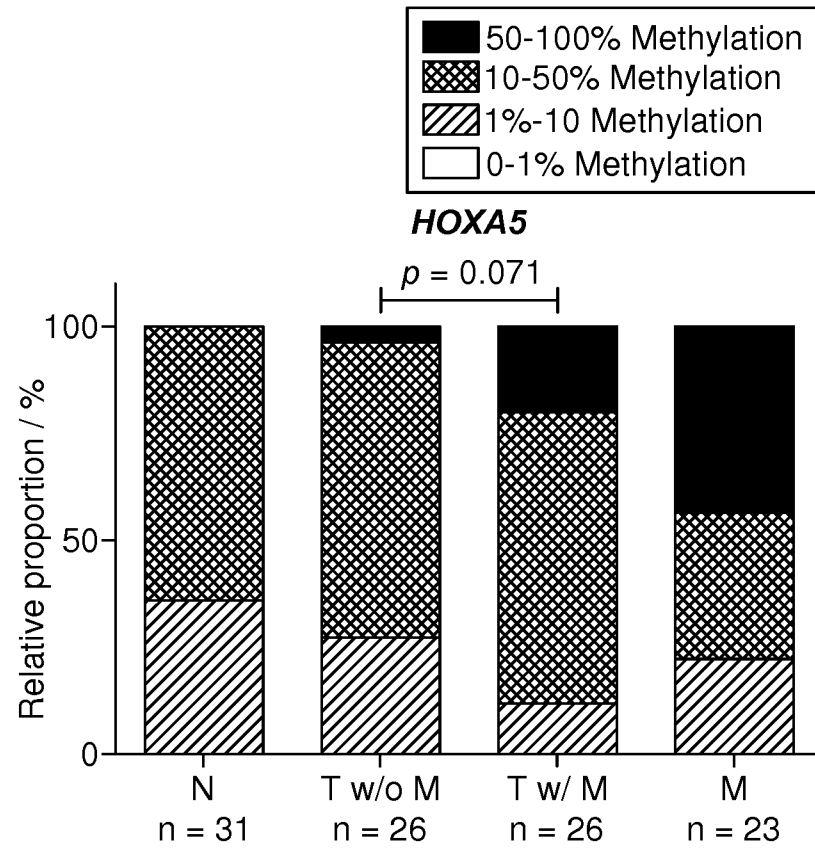


FIG. 3C

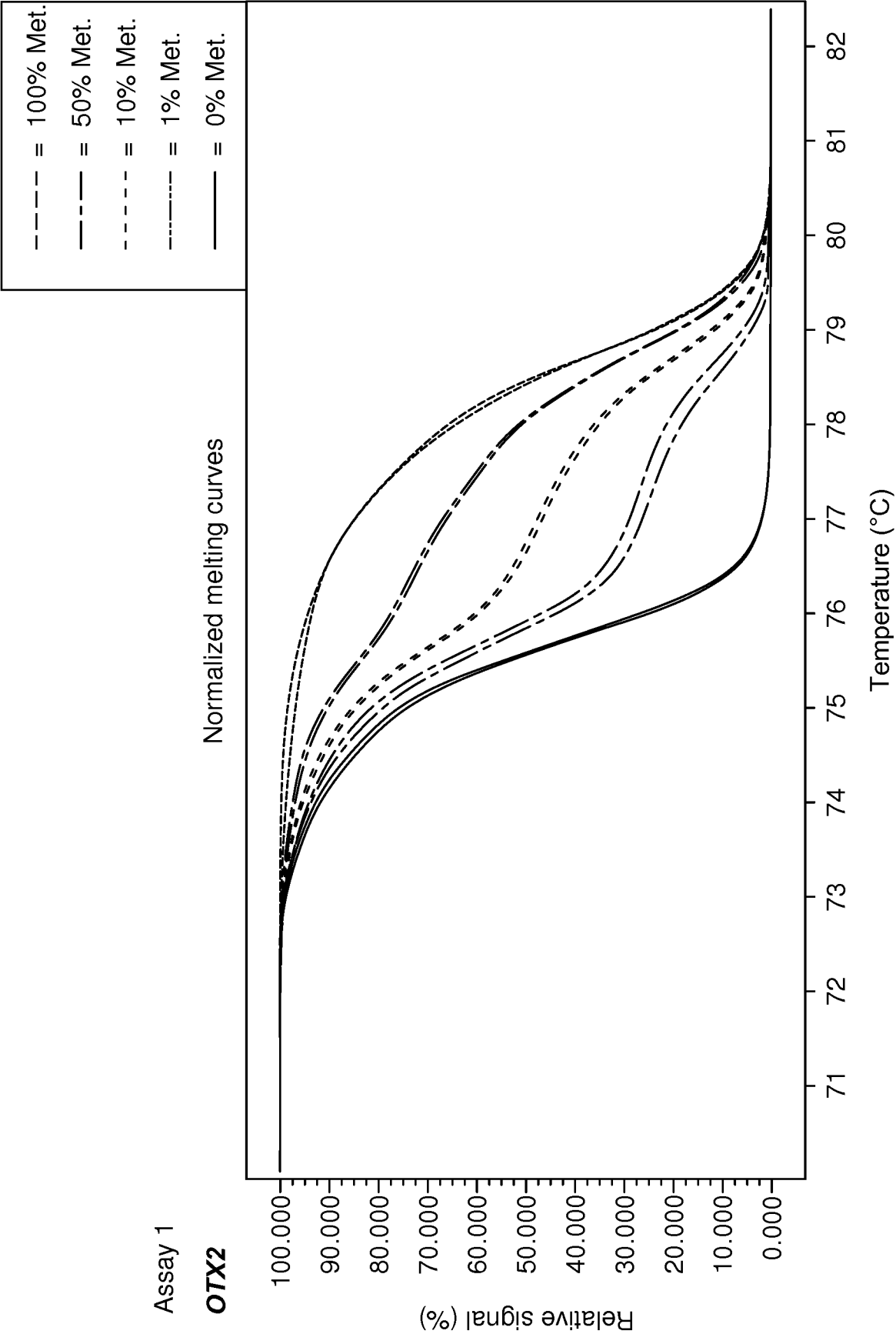


FIG. 4

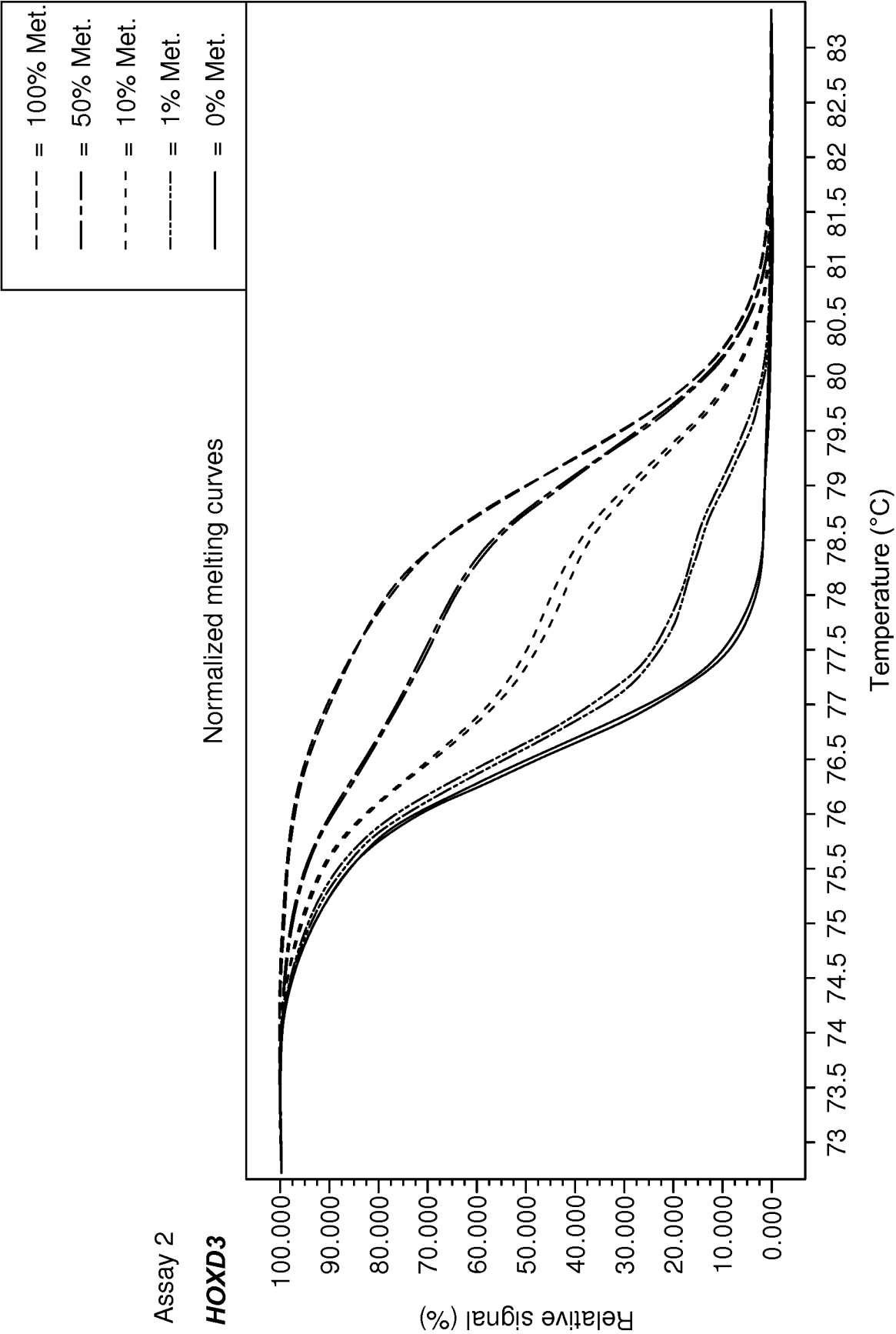


FIG. 4, ctd

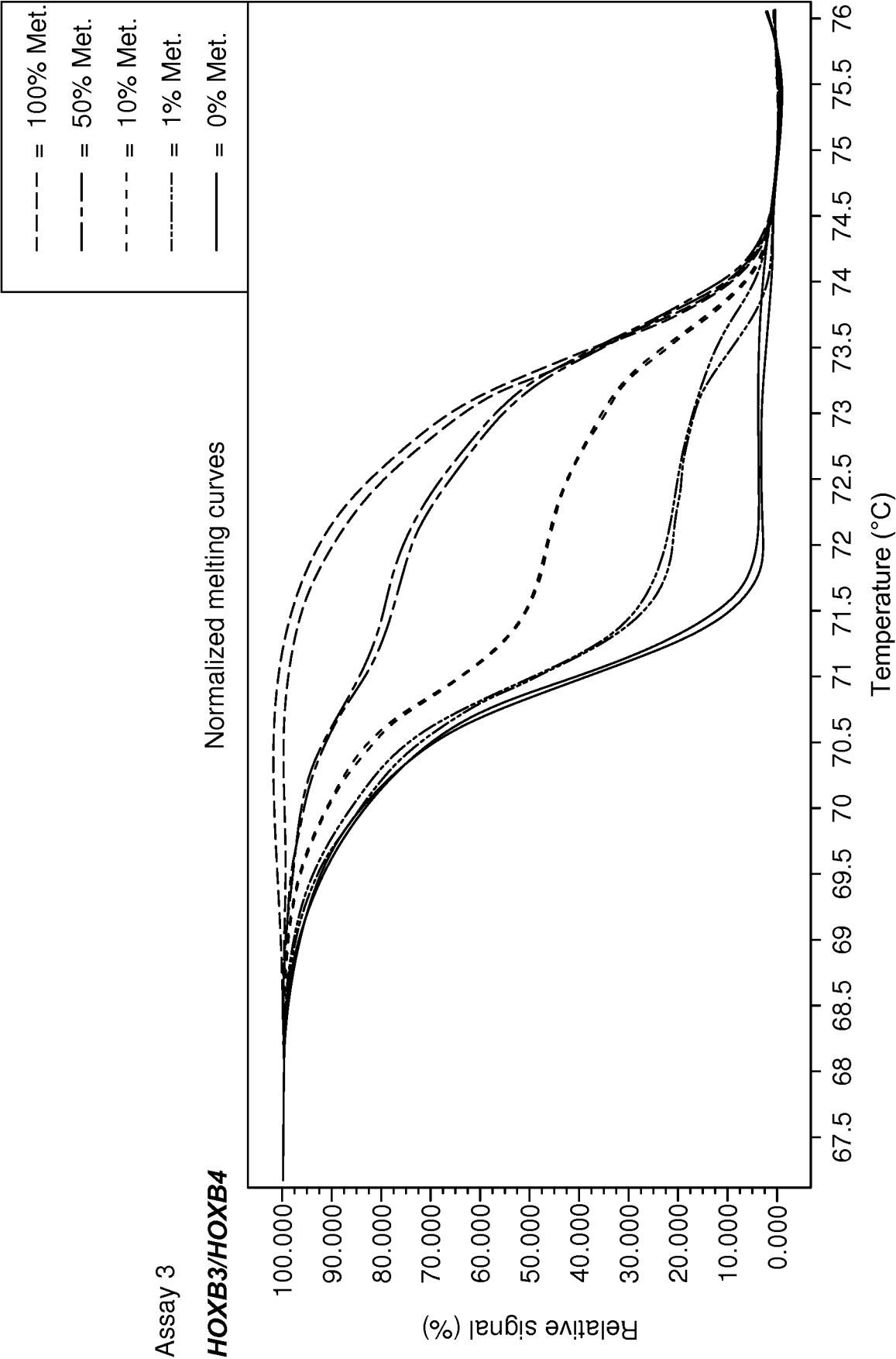


FIG. 4, ctd

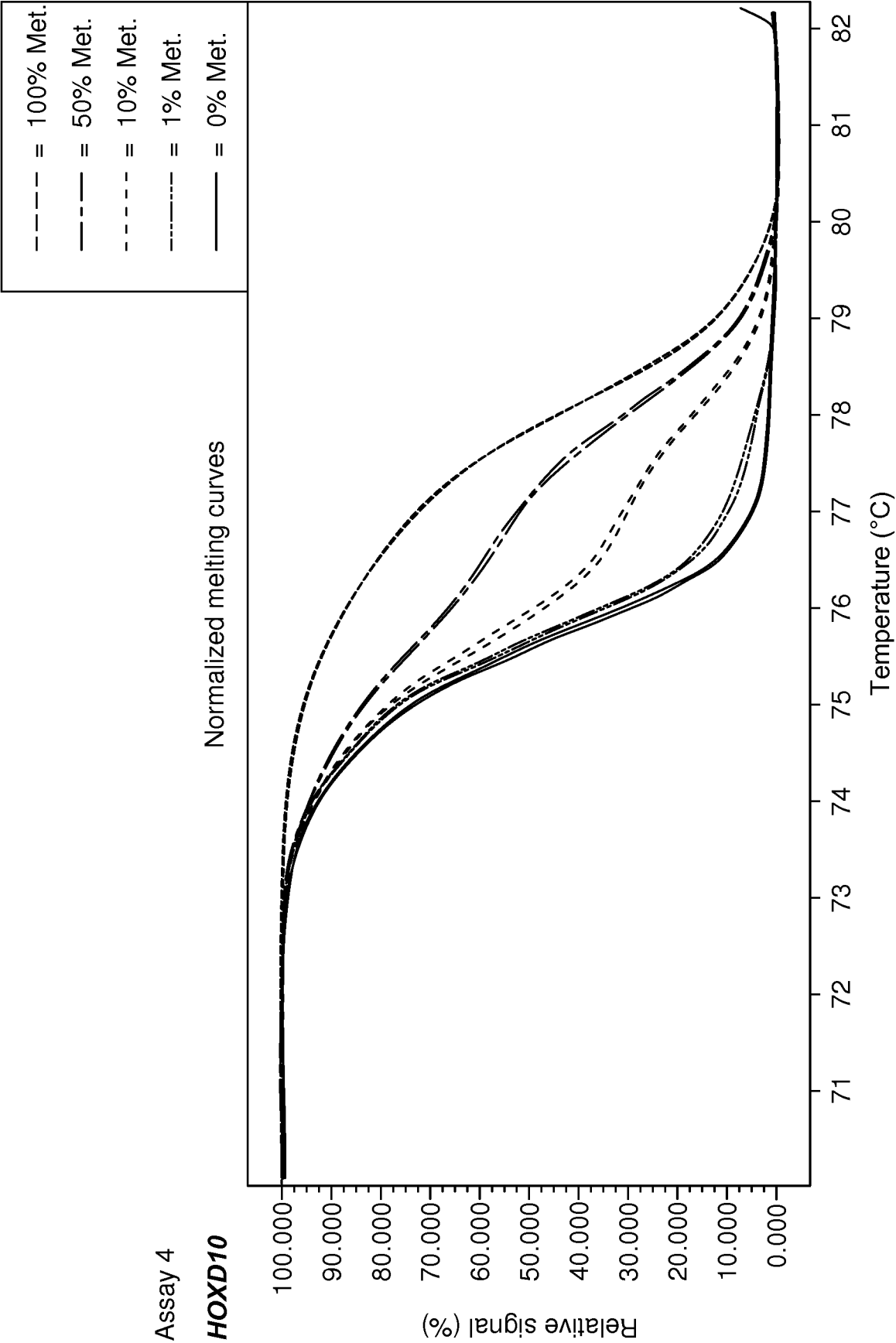


FIG. 4, ctd

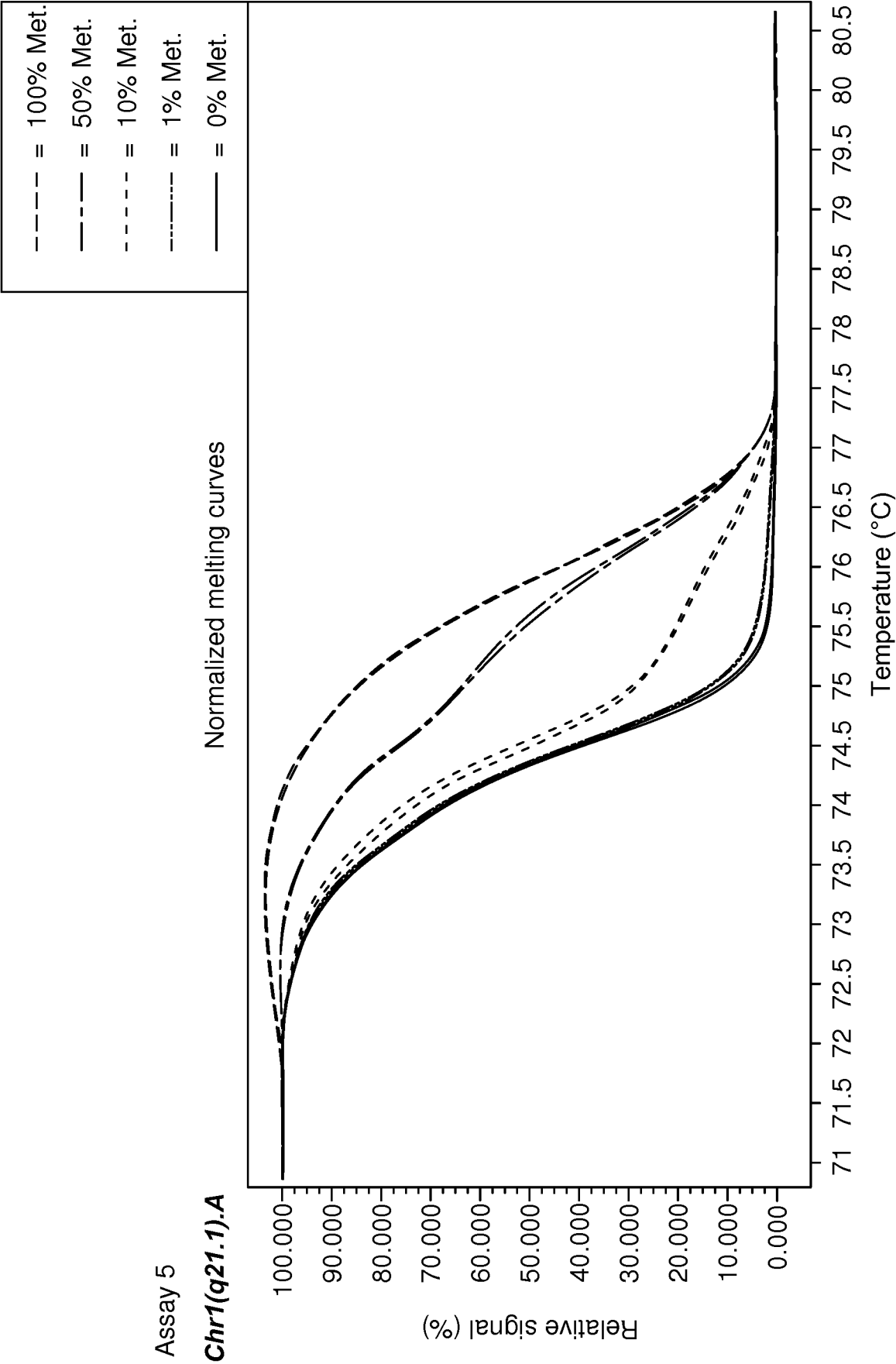


FIG. 4, ctd

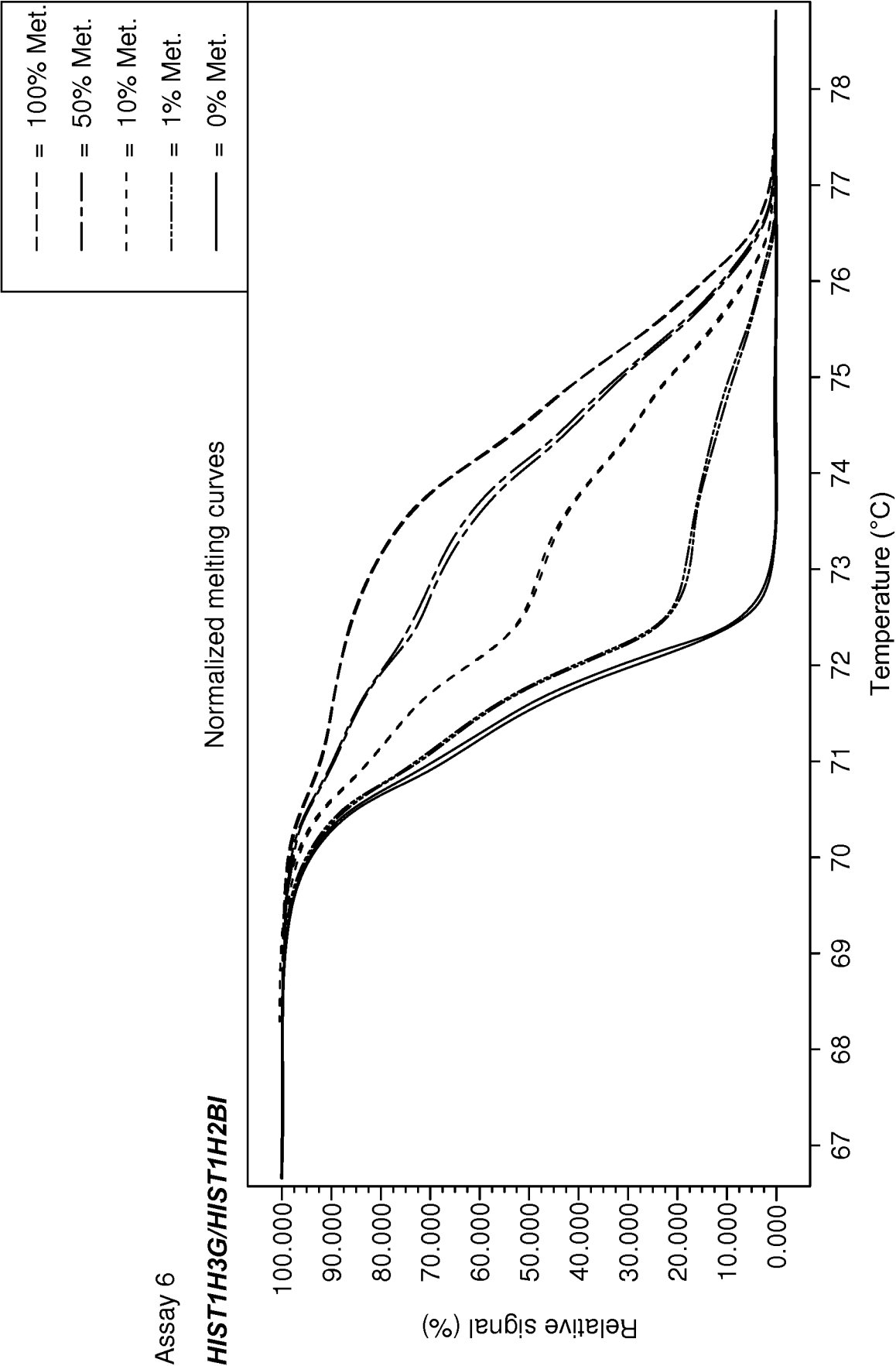


FIG. 4, ctd

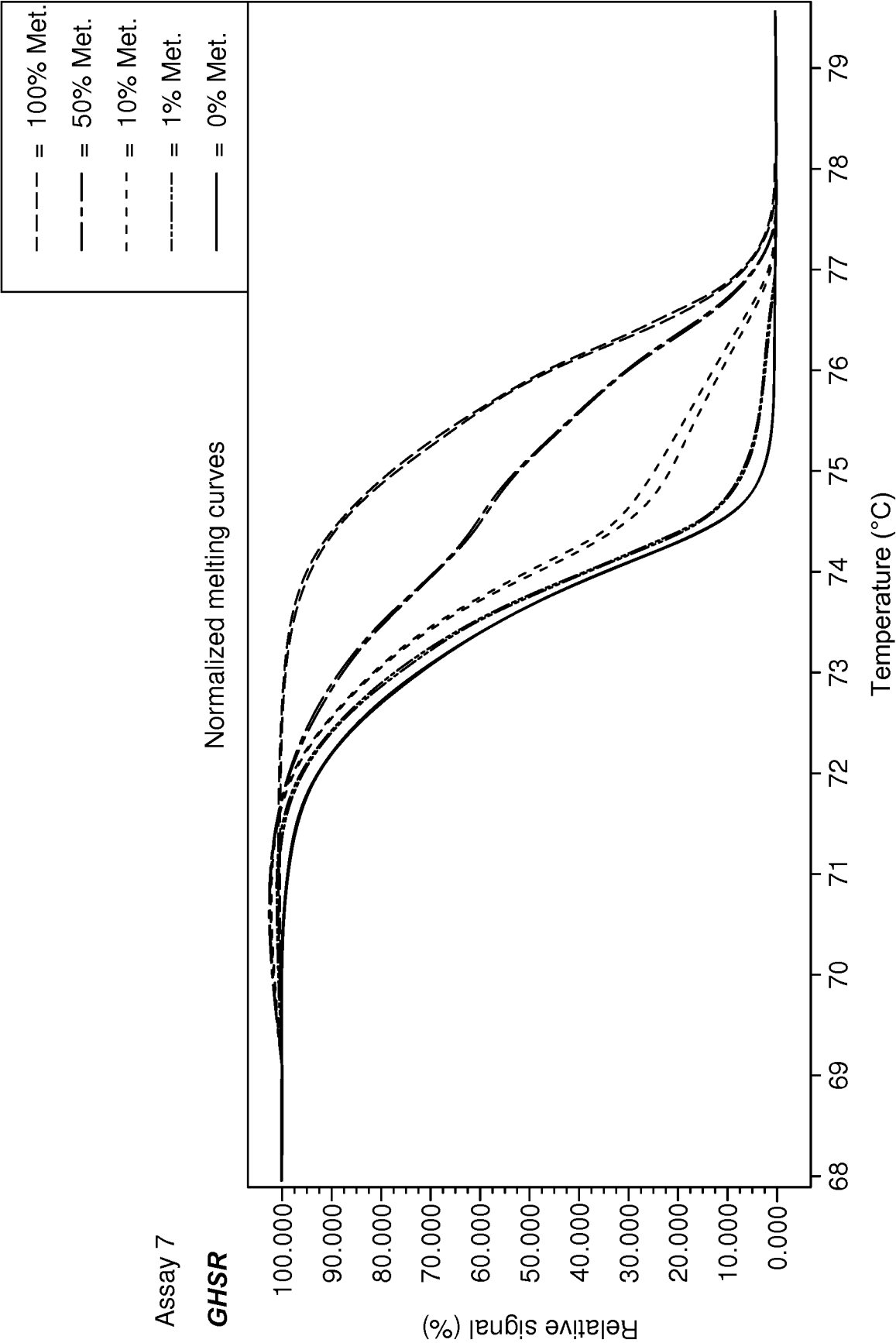


FIG. 4, ctd

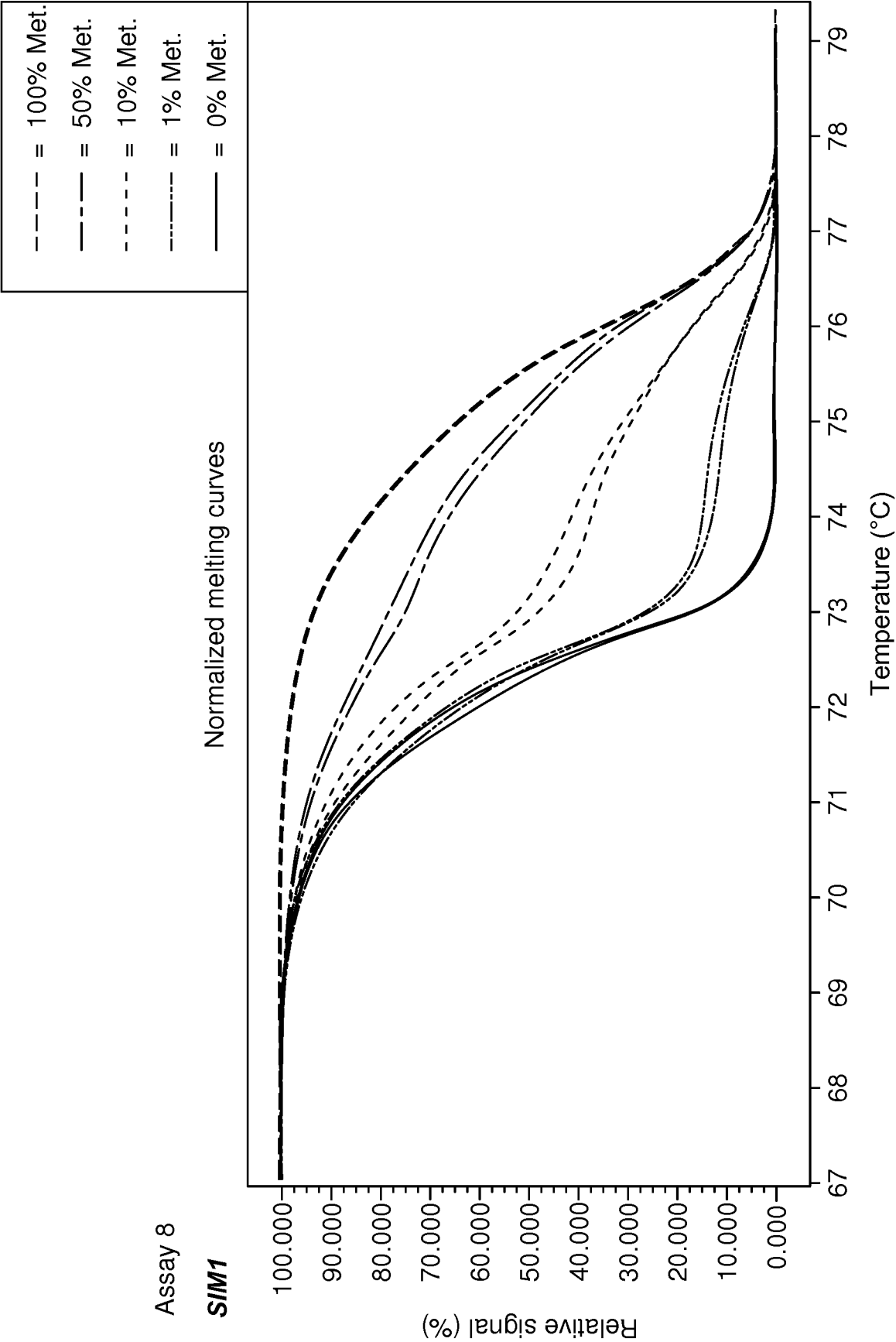


FIG. 4, ctd

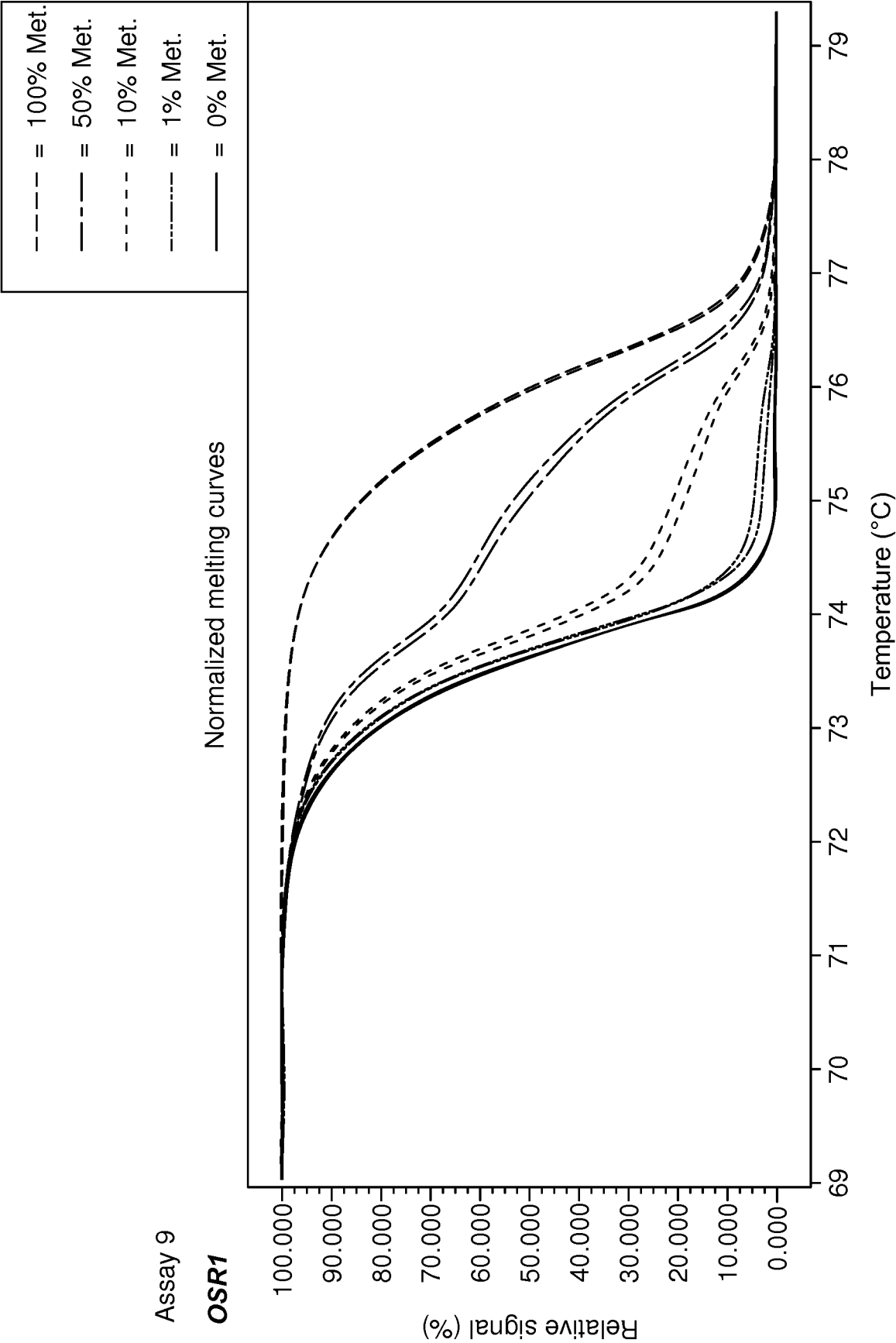


FIG. 4, ctd

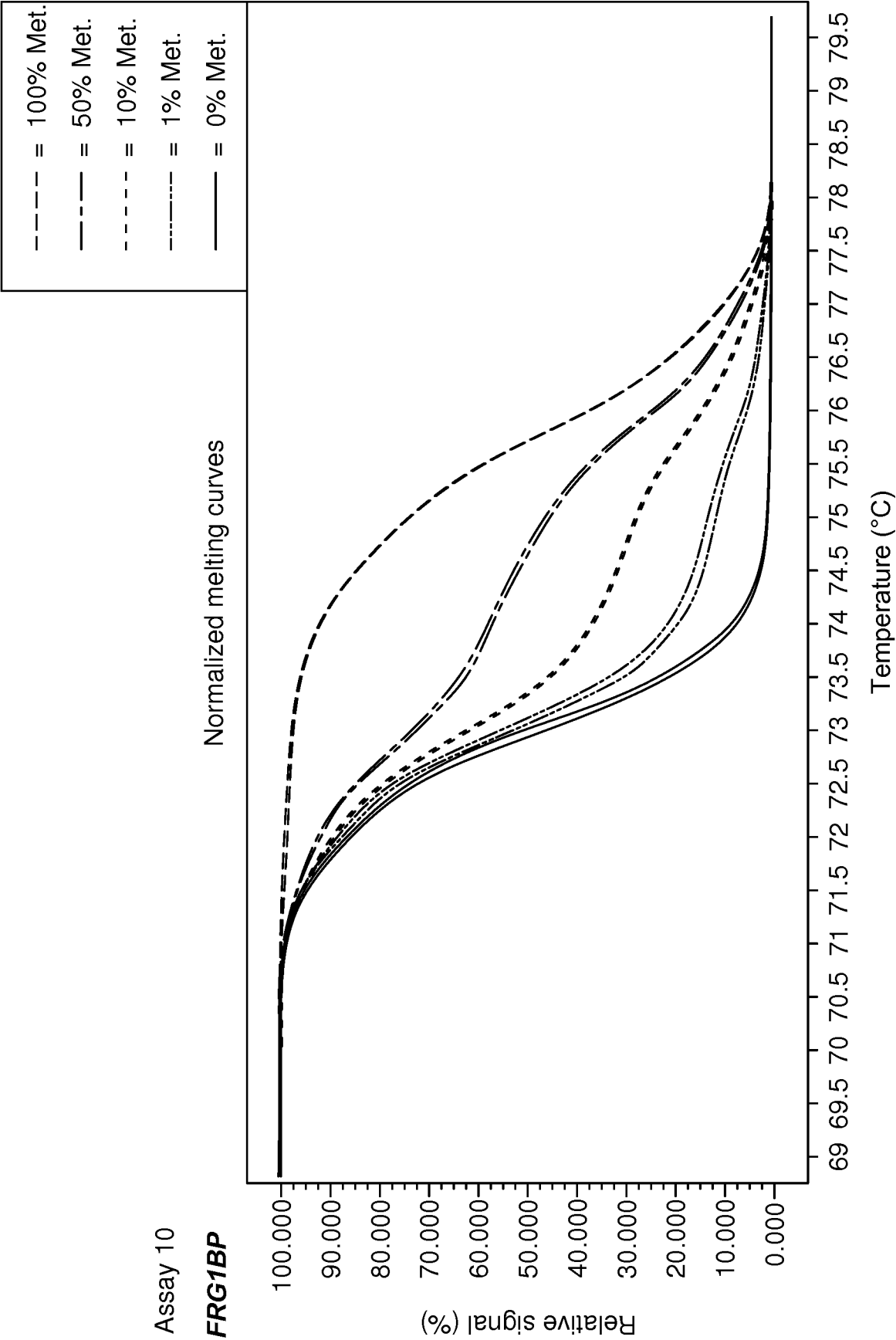


FIG. 4, ctd

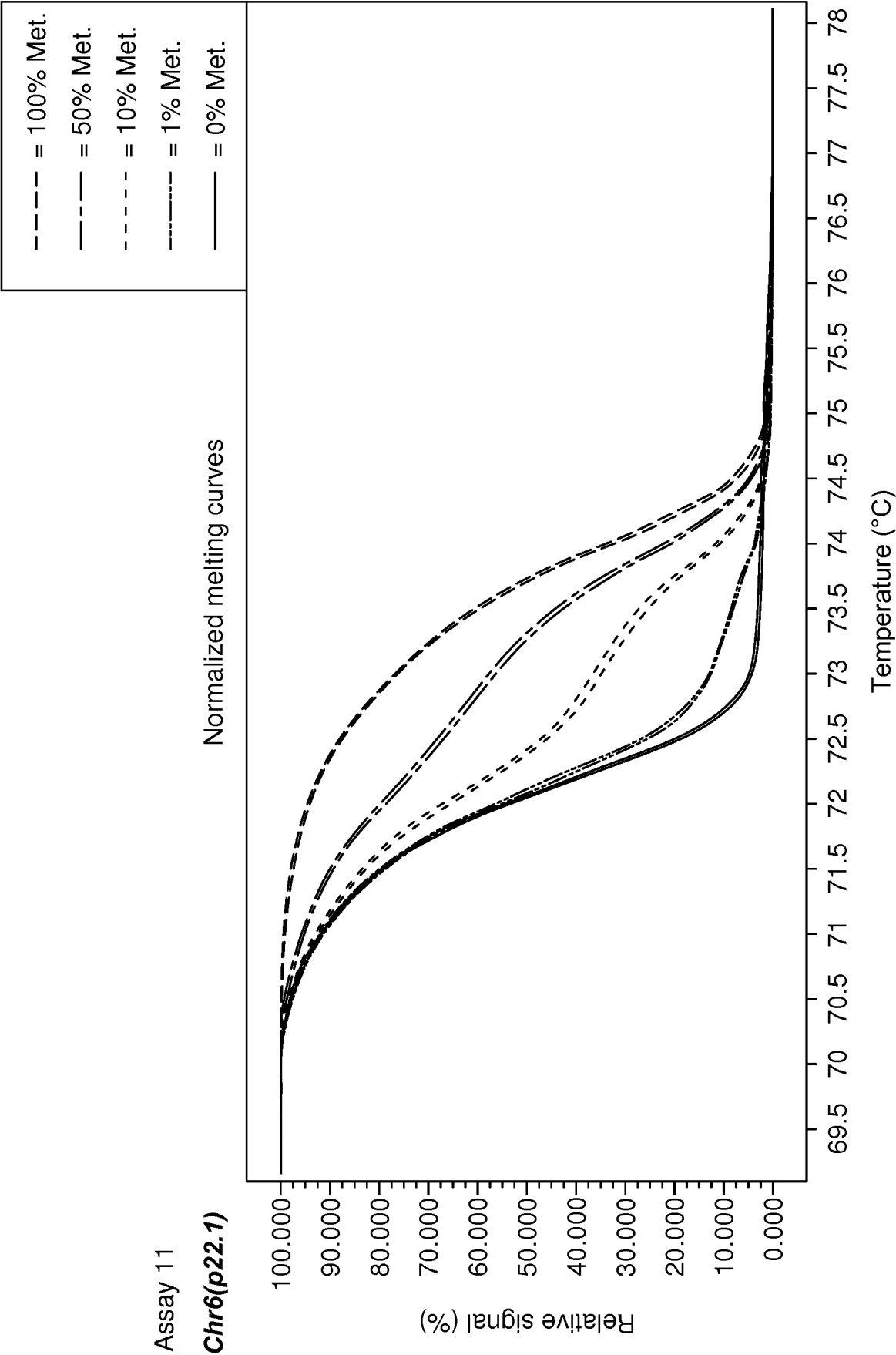


FIG. 4A

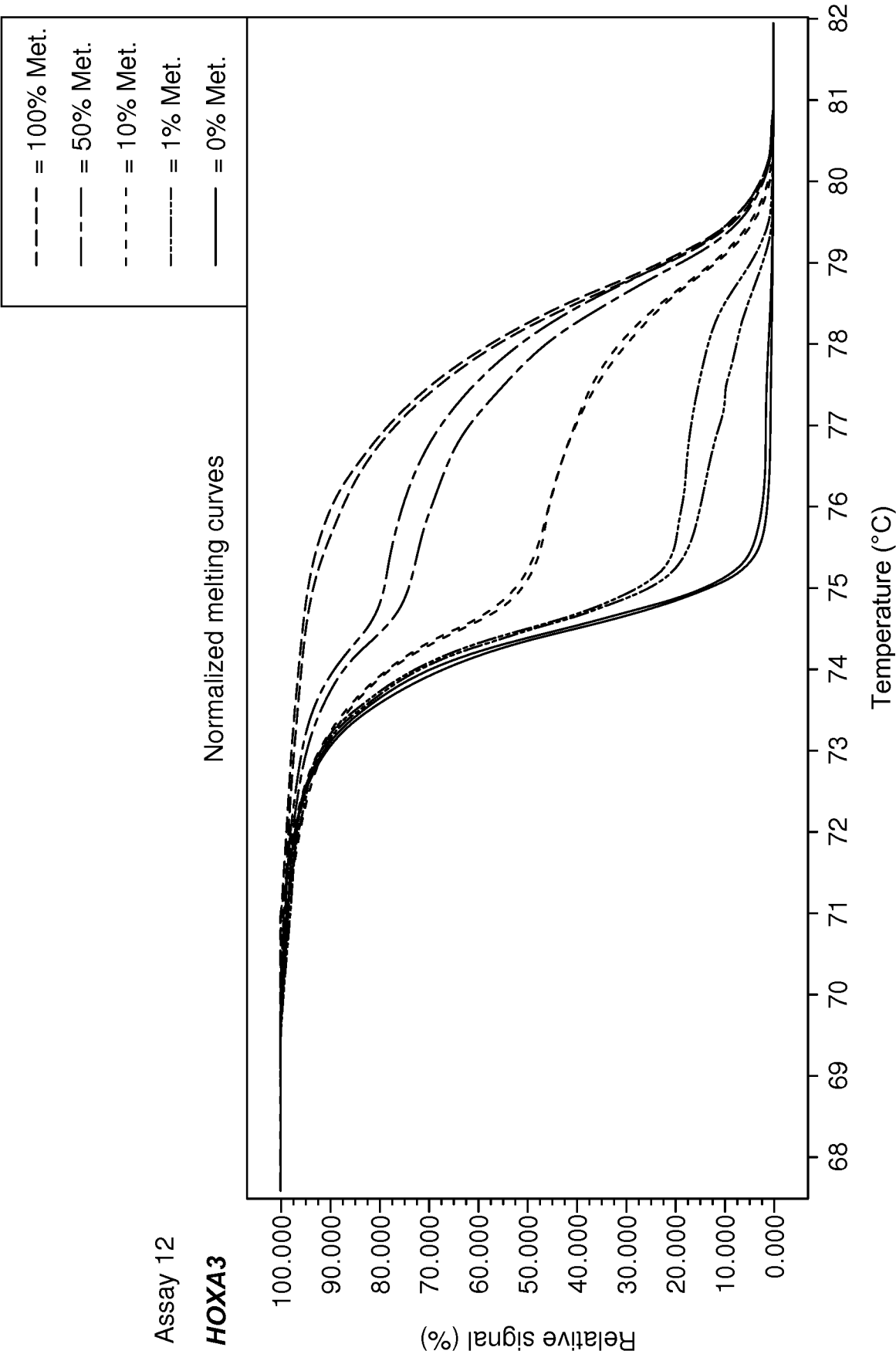


FIG. 4, ctd

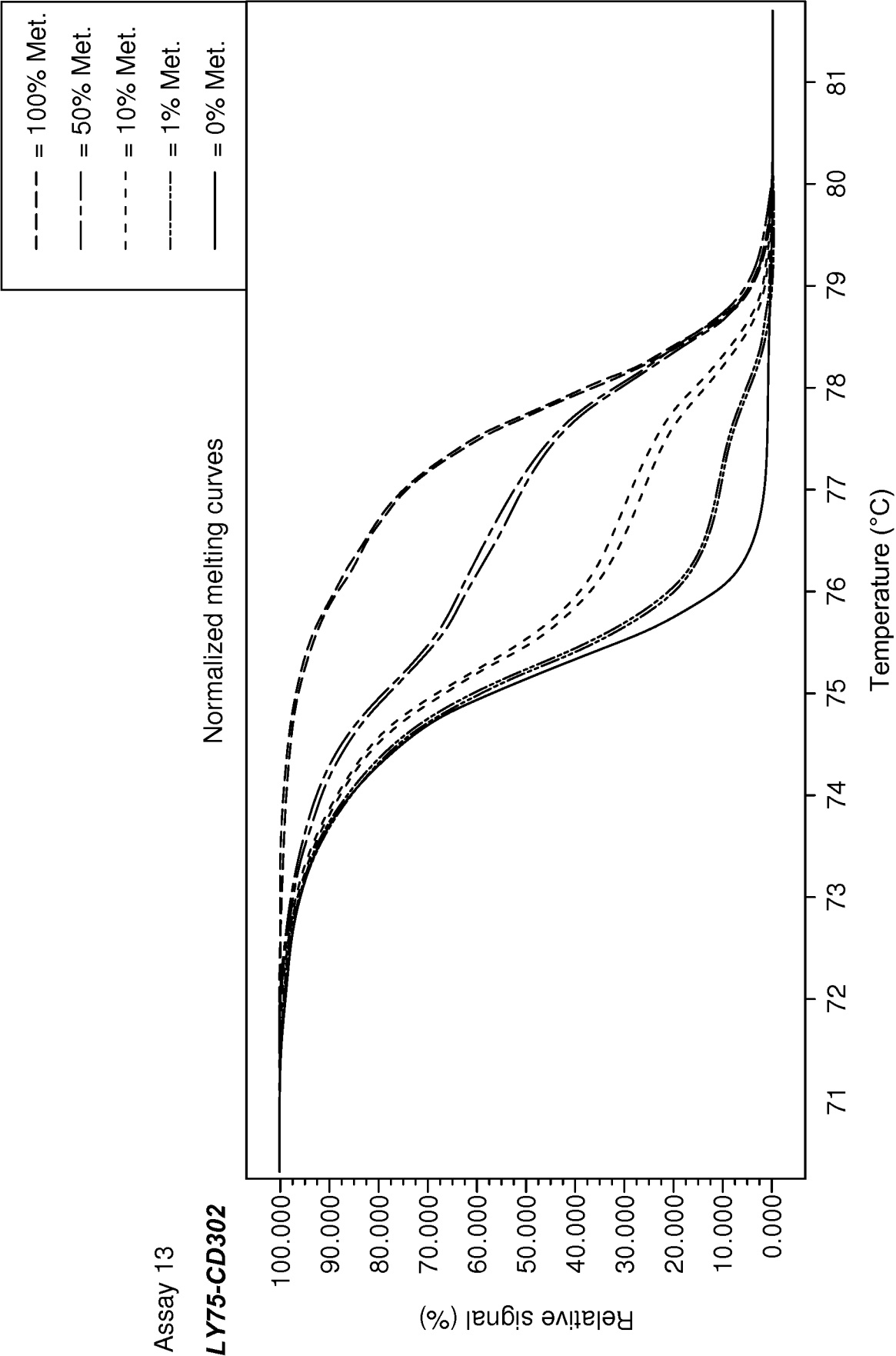


FIG. 4, ctd

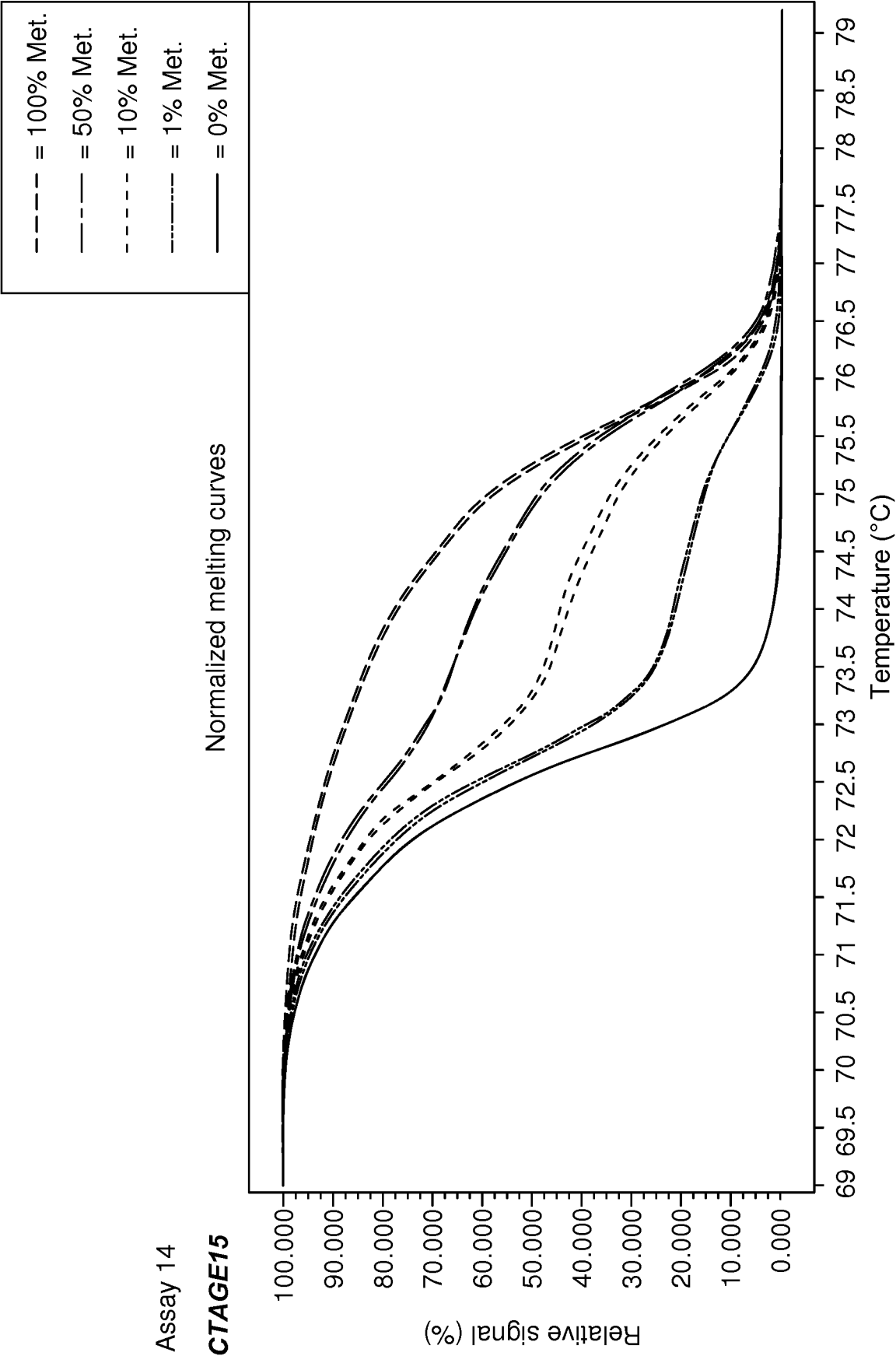


FIG. 4

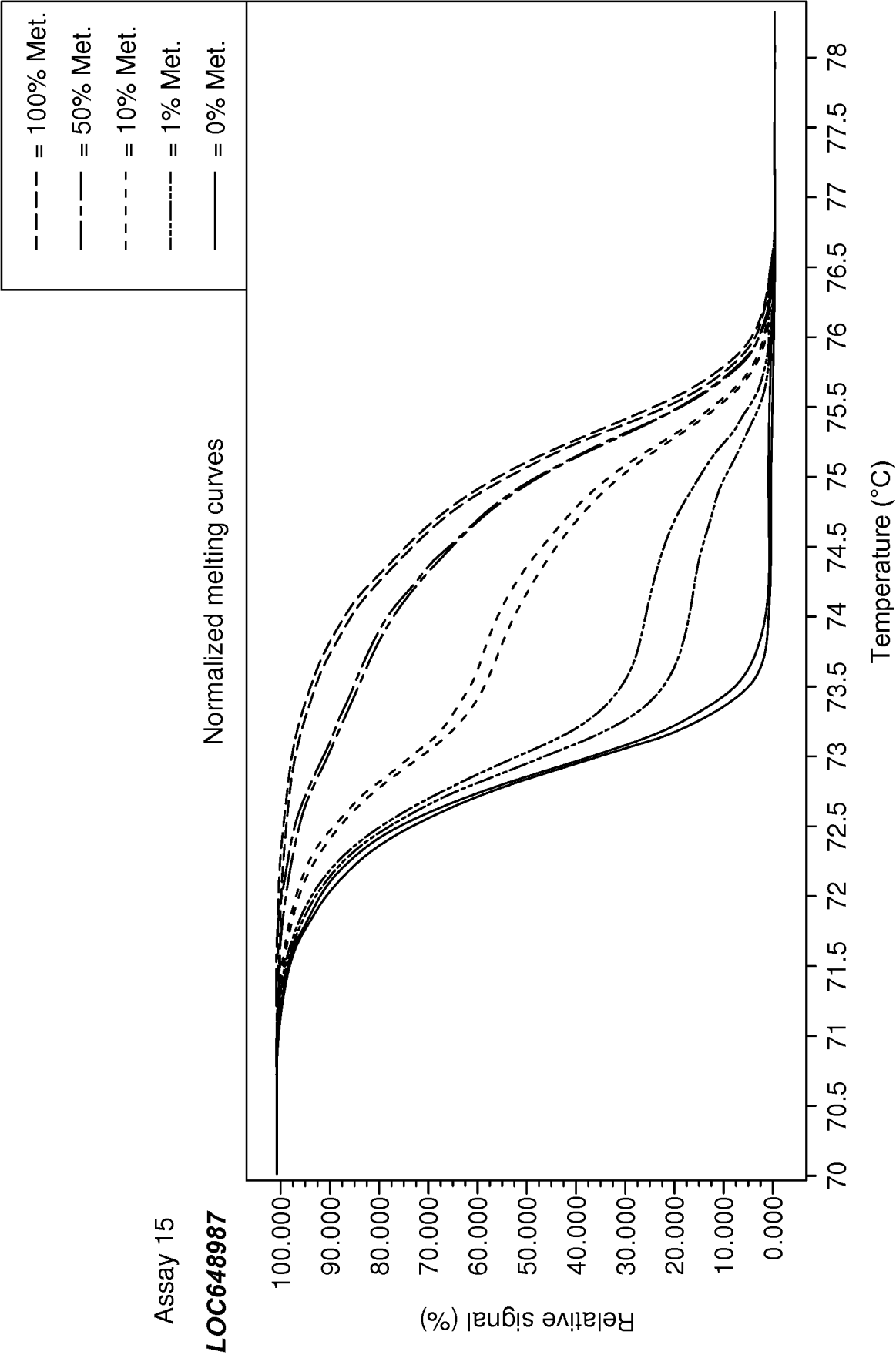


FIG. 4, ctd

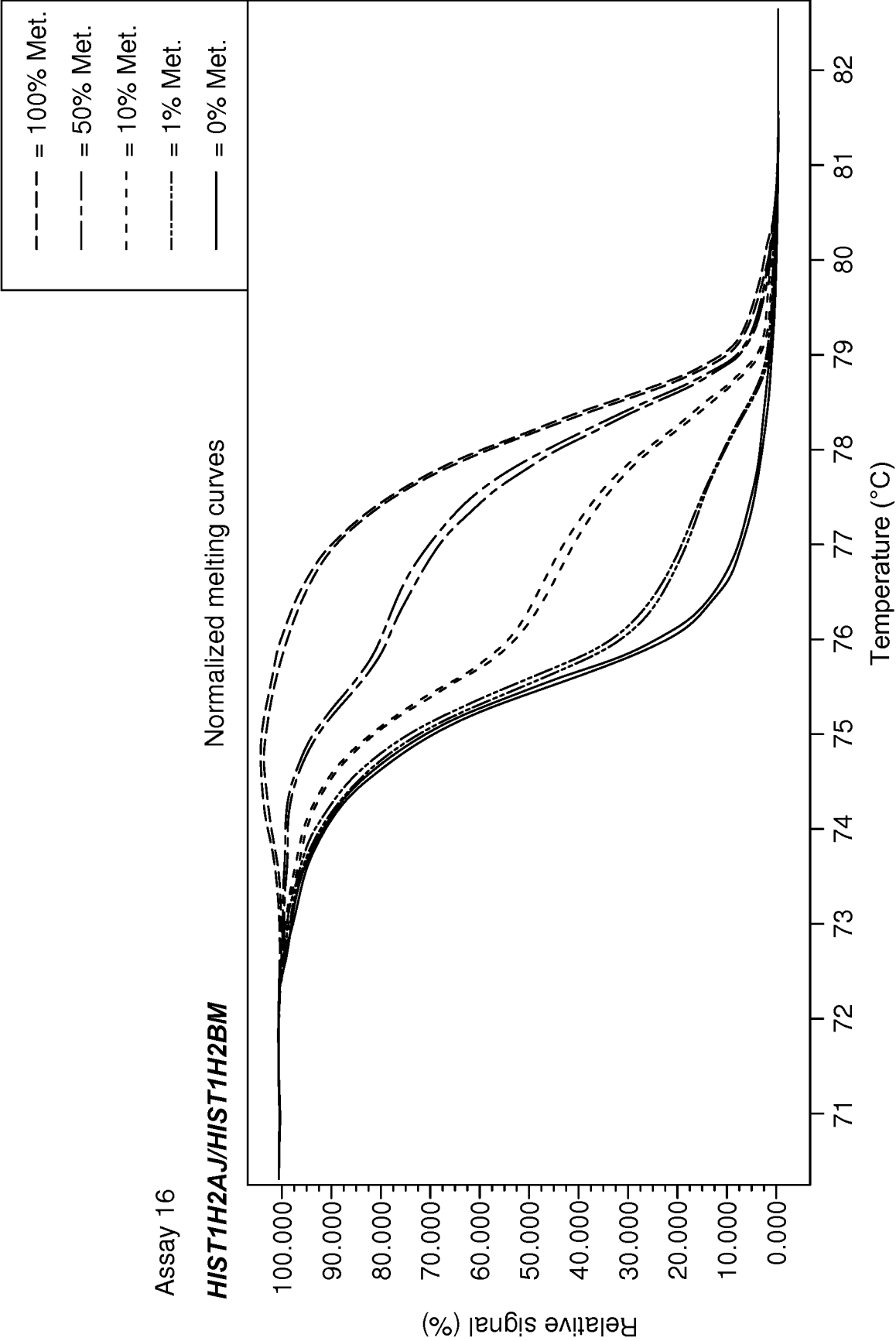


FIG. 4, ctd

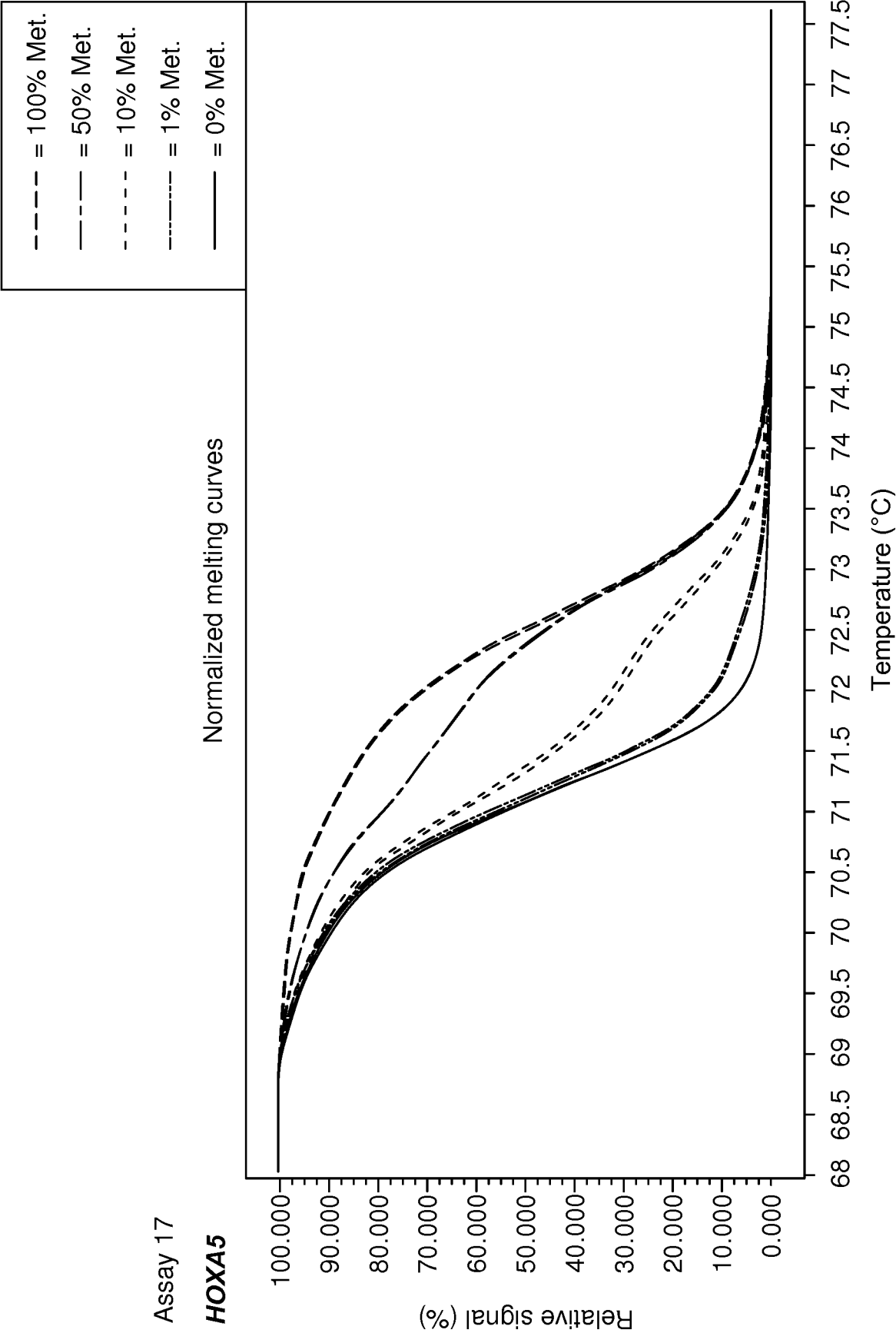


FIG. 4, ctd

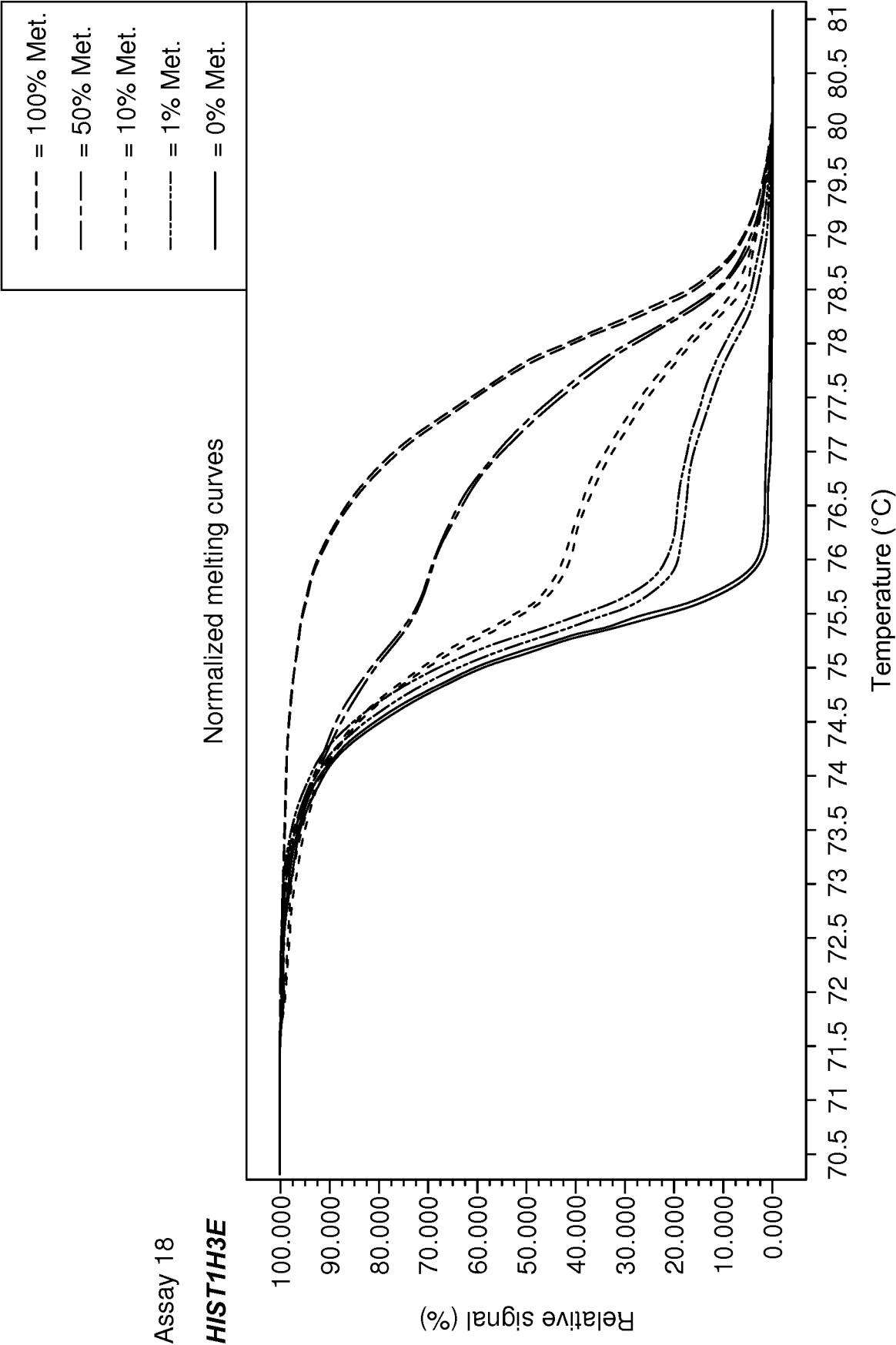


FIG. 4, ctd

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2017/076076

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/6886
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.
X	WO 2012/023648 A1 (NAT CANCER CT [KR]; LEE YEON SU [KR]; HWANG JUNG AH [KR]; PARK JOO BAE) 23 February 2012 (2012-02-23)	24-26, 39
Y	paragraph [0069]; claims 1, 3; figures 1, 11 paragraph [0149]	1-4, 7-23, 27, 28, 30-33, 36-38
X	----- WO 2016/083360 A1 (AIT AUSTRIAN INST TECHNOLOGY [AT]) 2 June 2016 (2016-06-02)	29, 33-35
Y	page 18, line 13 - line 27; claims 3, 4, 12, 14, 15; tables 1, 4; sequences 239, 240 page 32, line 1 - line 3 ----- -/-	1-4, 7-23, 27, 28, 30-33, 36-38



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

10 January 2018

Date of mailing of the international search report

12/03/2018

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

Santagati, Fabio

INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2017/076076

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	RAUCH TIBOR A; WANG ZUNDE; WU XIWEI; KERNSTINE KEMP H; RIGGS ARTHUR D; PFEIFER GERD P: "DNA methylation biomarkers for lung cancer", TUMOR BIOLOGY, vol. 33, no. 2, 1 April 2012 (2012-04-01), - 6 December 2011 (2011-12-06), pages 287-296, XP009502625, DOI: 10.1007/s13277-011-0282-2 tables 2, 3 -----	1-4,7-39
A	JI WOONG SON ET AL: "Genome-wide combination profiling of DNA copy number and methylation for deciphering biomarkers in non-small cell lung cancer patients", CANCER LETTERS, vol. 311, no. 1, 26 June 2011 (2011-06-26), pages 29-37, XP028278511, ISSN: 0304-3835, DOI: 10.1016/J.CANLET.2011.06.021 [retrieved on 2011-06-24] table 2 -----	1-4,7-39
A	CORTESE R ET AL: "Correlative gene expression and DNA methylation profiling in lung development nominate new biomarkers in lung cancer", INTERNATIONAL JOURNAL OF BIOCHEMISTRY AND CELL BIOLOGY, PERGAMON, GB, vol. 40, no. 8, 1 January 2008 (2008-01-01), pages 1494-1508, XP022655992, ISSN: 1357-2725, DOI: 10.1016/J.BIOCEL.2007.11.018 [retrieved on 2007-12-27] table 3 -----	1-4,7-39
X,P	IBEN DAUGAARD ET AL: "Identification and validation of candidate epigenetic biomarkers in lung adenocarcinoma", SCIENTIFIC REPORTS, vol. 6, no. 1, 26 October 2016 (2016-10-26), XP055438917, DOI: 10.1038/srep35807 the whole document -----	1-4,7-39
A	US 2014/271455 A1 (PFEIFER GERD P [US] ET AL) 18 September 2014 (2014-09-18) figure 2; table 6 -----	1-4,7-39

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2017/076076

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

4(completely); 1-3, 7-39(partially)

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 4(completely); 1-3, 7-39(partially)

A method of determining non-small cell lung cancer, a predisposition to non-small cell lung cancer, the prognosis of a non-small cell lung cancer, and/or monitoring a non-small cell lung cancer in a subject, a method for assessing whether a human subject is likely to develop non-small cell lung cancer, a method for categorizing or predicting the clinical outcome of a non-small cell lung cancer of a subject, and a method of evaluating the risk for a subject of contracting cancer, said methods comprising determining the methylation status of at least SIM1 in a sample from said subject. A method of treating a non-small cell lung cancer in a human subject, said method comprising i. determining non-small cell lung cancer, a predisposition to non-small cell lung cancer, or the prognosis of a non-small cell lung cancer in a subject by a method as defined in any of the preceding claims and subjecting said subjects identified to a suitable treatment for non-small cell lung cancer. A kit comprising. an agent that (a) modifies methylated cytosine residues but not non-methylated cytosine residues; or (b) modifies non-methylated cytosine residues but not methylated cytosine residues; or (c) modifies a nucleic acid sequence in a methylation-dependent manner and at least one pair of oligonucleotide primers that specifically hybridizes under amplification conditions to a region of SIM1. Use of oligonucleotide primers comprising a sequence, which is a subsequence of SIM1 for diagnosing non-small cell lung cancer. A method of identifying therapeutically effective agents for treatment of non-small cell lung cancer, said method comprising. providing a non-small cell lung cancer cell line comprising SIM1 providing one or more potential therapeutic agents, treating said non-small cell lung cancer cells by bringing said agents in contact with said non-small cell lung cancer cells, determining methylation status of said one or more genetic loci, comparing said methylation status of said treated non-small cell lung cancer cells with the methylation status of said non-small cell lung cancer cells, when untreated, wherein a decreased level of methylation positive alleles is indicative of a therapeutic agent.

2-15. claims: 1-3, 5-39(all partially)

The same subject-matter as in invention 1, but wherein the gene or locus is

- inv 2: Chr6(p22.1)
- inv 3: HIST1H3G/HIST1H2BI
- inv 4: HOXB3/HOXB4
- inv 5: OSR1
- inv 6: GHSR2

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

inv 7: OTX2
inv 8: LOC648987
inv 9: HIST1H3E
inv 10: HIST1H2AJ/HIST1H2BM
inv 11: HOXD10
inv 12: HOXD3
inv 13: HOXA3
inv 14: HOXA5
inv 15: Chr1(q21.1)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/EP2017/076076

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2012023648 A1	23-02-2012	KR 20120046799 A WO 2012023648 A1	11-05-2012 23-02-2012
WO 2016083360 A1	02-06-2016	EP 3224377 A1 US 2017283887 A1 WO 2016083360 A1	04-10-2017 05-10-2017 02-06-2016
US 2014271455 A1	18-09-2014	NONE	