

Method of early diagnosis of Epstein-Barr virus-related cancer and respective reagents and kits

Field of the invention

5 [0001] The invention relates to methods of diagnosing cancer associated with Epstein-Barr virus infection in subjects. The invention utilizes the determination of the gene load and the gene expression profiling of certain molecular biomarkers related to Epstein-Barr virus related cancers. The invention further relates to reagents and kits for performing methods of the invention.

10

Background of the invention

[0002] The Epstein-Barr virus (EBV) is a lymphotropic human Herpes virus that establishes lifelong latency. In childhood an EBV infection often is asymptomatic. However, in adolescents and adults a primary EBV infection causes the self-limiting lymphoproliferative disorder, infectious mononucleosis (IM). After the primary EBV infection the virus persists in the organism and is lifelong excreted via the saliva. In addition, the virus is latently present in about 1 in 10,000 human periphery B cells. In rare cases, an EBV infection can lead to the so-called X-linked lymphoproliferative syndrome (XLP-syndrome) or Duncan syndrome, which is correlated with a gene defect on the X chromosome.

20

[0003] The EBV is prevalent worldwide and about 95% of the adult population is seropositive for EBV. The B-cells in which the virus latently persists have the capability to develop into immortalized cells in cell culture, whereas the unregulated proliferation of these cells is normally suppressed by a functional cellular immunity *in vivo*. However, in case of immune-suppressed patients, for example patients that underwent organ transplantation or are infected with HIV, the EBV-infected B cells may proliferate uncontrolled and thus cause a B cell lymphoma.

25

[0004] Due to this property to stimulate cell proliferation and immortalize an infected cell, the virus is strongly associated with a number of human malignant disorders, characterized by high titers of anti-EBV antibody profiles and an increased circulating EBV DNA level. Besides the above mentioned B cell lymphomas in immune-suppressed patients, the EBV is very strongly associated with Burkitt lymphoma, nasopharyngeal carcinoma (NPC) and salivary gland cancer and to a lower extent also associated to

30

Hodgkin lymphomas, T-cell lymphomas and gastric adenocarcinomas. However, the specific role of EBV in the etiology of the above malignancies has not been completely elucidated yet.

[0005] Nasopharyngeal carcinoma (NPC) shows a world wide very uneven geographical distribution. NPC is the second most common tumor in Greenland (31 cases per 10⁵ inhabitants per year) and the most common in Southern China (app. 42 cases per 10⁵ inhabitants per year) and South East Asia, including Singapore, Hong Kong, Vietnam, and Indonesia. It is also common in North Africa, while it is comparably rare in other countries (0.6 cases per 10⁵ per year in Sweden). Such a pronounced, uneven distribution of the cancer strongly suggests environmental risk factors besides the well-documented association with Epstein Barr virus (EBV) acting on genetically susceptible populations.

[0006] Exemplary etiology factors may include the life style, such as the consumption of nitrosamines in salt fish and certain herbal medicines, and genetic factors, including inheritable, i.e. somatic, and epigenetic changes.

[0007] In Indonesia, with an ethnically diverse population of 225 million people, NPC is the most common ear-nose-throat (ENT) tumor with high prevalence among native populations and a yearly overall incidence estimated at 6.2 per 100.000 inhabitants. Extremely high incidence was recently documented in native populations living on the island of Sulawesi. In Yogyakarta, central Java, NPC is the most prevalent tumor among man and 4th most prevalent among females, with a male/female ration of 2.4, constituting respectively 22 and 8% of all diagnosed malignancies.

[0008] In all samples of undifferentiated NPC, dysplastic lesions, carcinoma in situ and in metastatic tissue EBV DNA, RNA and protein can be detected. Six EBV genes are known to be expressed in this tumor. One of these proteins expressed in NPC, LMP1, can transform established rodent fibroblasts and human epithelial cells *in vitro* (Wang et al., 1985; Hu et al., 1993) and thus is classified as an oncogene. Clinical and follow-up data from 74 cases of NPC showed that LMP1 positive NPC grow faster and are more expansively than LMP1 negative tumors (Hu et al., 1995). In addition, it was shown that EBV infection was present in all high-grade dysplastic epithelial lesion and invasive NPC but not in any normal epithelia from fetuses or normal adults.

[0009] The Burkitt lymphoma has a high endemic incidence in about seven to nine year old children in equatorial Africa and New Guinea. About 98% of these endemic Burkitt lymphomas show Epstein-Barr viruses in the transformed cells. In the sporadic

form of the Burkitt lymphoma, which is found predominantly in adults, about 25% of the cases are associated with the EBV.

[0010] In all of the EBV associated diseases and cancers discussed above, presence of the viral protein EBNA1 can be demonstrated. EBNA1 has been found to be a DNA binding protein important for the long-term maintenance of the viral genome in the cell nucleus.

[0011] However, the tumorigenesis is thought to be a multi-step process so that even in cases where EBV seems to be an essential factor in the events leading to malignancy, it is clear that the virus-related tumor development and progression is a multistep process also involving secondary cellular events. In this process, both genes (genetic and epigenetic) and environmental factors (including EBV) play a decisive role. In recent years, there is increasing evidence demonstrating that the inactivation of tumor suppressor genes (TSGs) and the activation of oncogenes is a key element in the pathogenesis and development of cancer. These genes can be categorized into cell cycle control, DNA damage repair, apoptosis, tumor invasion and growth factor response genes.

[0012] In recent years, particularly epigenetic changes and their role in tumor development and progression have come into the focus of interest.

[0013] Epigenetic change happens to the modifications of genetic material instead of the genetic material itself. Epigenetic changes are also inheritable via the cell division, and play an important role in many physiological and pathophysiological conditions. Research has shown that epigenetic mechanisms provide an "extra" layer of transcriptional control that regulates how genes are expressed. These mechanisms are critical components in the normal development and growth of cells. Epigenetic abnormalities have been found to be causative factors in cancer, genetic disorders and pediatric syndromes as well as contributing factors in autoimmune diseases and aging.

[0014] Among the epigenetic events such as methylation, deacetylation and chromatin re-modeling, promoter hypermethylation has been the most detailed studied and could be a main mechanism for the inactivation of tumor suppressor genes in cancer. Promoter hypermethylation of known or candidate tumor suppressor genes involved in various fundamental pathways, such as apoptosis, DNA damage repair, cell movement and migration, were reported as an early event in various types of cancers. For example, aberrant methylation of tumor suppressor genes was frequently observed in NPC.

[0015] Cytosine DNA methylation is a covalent modification of DNA, in which a methyl group is transferred from S-adenosylmethionine to the C-5 position of cytosine by a family of cytosine (DNA-5)-methyltransferases. DNA methylation occurs almost exclusively at CpG nucleotides and has an important contributing role in the regulation of gene expression and the silencing of repeat elements in the genome.

[0016] Widespread hypermethylation of CpG islands over the entire genome implied a methylator phenotype in EBV associated malignancies. Besides the silencing of tumor suppressor genes, DNA methylation may also play an important role in the maintenance of specific EBV latency programs in the tumor cells. Methylation of both viral and cellular genes may thus be involved in the cellular transformation.

[0017] Early genetic alterations, such as the chromosome 3p loss of heterozygosity (LOH), have been more frequently identified in the early stages of NPC, even prior to EBV latent infection, and expand clonally thereafter (Chan et al., 2000). A study has shown that patients afflicted by squamous lung cell carcinoma can be diagnosed by the determination of the abnormal methylation of tumor suppressor genes p16 and O6-methylguanine-DNA methyltransferase (MGMT) in sputum 35 months prior to clinical diagnosis.

[0018] Up to now, for early stage NPC patients without involvement of cervical lymph node radiotherapy is still the prime choice of treatment. The survival rate is 50% - 80%. However, more than 70% of NPC patients were found to have the involvement of regional lymph nodes, and 10% of patients show distant metastasis at the time of diagnosis. The early diagnosis is a major factor, as a late diagnosis can adversely affect the effect of treatment.

[0019] Currently, diagnosis of NPC requires a biopsy from the primary tumor site or metastases for the histopathological assessment and demonstration of EBV involvement by *in situ* hybridization for EBER1/2.

[0020] Although NPC patients have higher anti-EBV EA/VCA IgA antibody titers than healthy EBV carriers and elevated antibody titers can be found prior to onset of clinical symptoms in NPC patients with primary or recurrent carcinoma, and this method has been used for risk prediction for the population at high-incidence area, the low detection rate and high rate of false positive results make this method disadvantageous.

[0021] It has been speculated that the circulating EBV DNA value may be useful for prognostic monitoring, but the EBV DNA level in the blood is even in patients with NPC low or negative due to tumor apoptosis or necrosis.

[0022] Importantly, although EBV is essential for the pathogenesis of EBV-related cancers, it is not sufficient, because the initiation of tumors involves multiple other etiological factors with multiple processes and steps. Thus, there still exists a need for a method for the diagnosis of the early stages of EBV-related cancers, in particular NPC, as such a method would greatly help in the diagnosis of these cancers, the choice of subsequent therapy and the evaluation of survival probability. This need is met by the method provided by the present invention.

Summary of the invention

[0023] The present invention provides a method of diagnosing Epstein-Barr virus (EBV) associated cancer by determining the EBV gene load and gene expression level of certain molecular biomarkers of, for example, viral as well as cellular origin that are related to cancers.

[0024] In a first aspect, the invention relates to a method of diagnosing Epstein-Barr virus-related cancer in a subject, comprising collecting a biological sample from said subject, determining in said biological sample the amount of the Epstein-Barr virus gene load by determining the EBNA1 (Epstein-Barr nuclear antigen 1) gene load and the expression level of at least one gene selected from the group consisting of the Epstein-Barr virus gene LMP1 (latent membrane protein 1) and the cellular genes RASSF1 (ras association domain family 1A), CHFR (Checkpoint with fork-head associated and ring finger) and DAPK (death associated protein kinase), and comparing the determined expression levels of said genes in said biological sample with a reference.

[0025] In one embodiment of the method of the invention, the gene load of EBNA1 and the expression levels of at least two genes, or at least three genes selected from the group consisting of the Epstein-Barr virus gene LMP1 and the cellular genes RASSF1, CHFR and DAPK are determined and compared to a reference.

[0026] In one specific embodiment of the invention, the expression levels of LMP1 and RASSF1A are determined. Optionally, the expression levels of CHFR and/or DAPK may also be determined.

[0027] The expression level of the above indicated marker genes can be determined via the epigenetic status of said genes. In this connection the promoter hypermethylation of the tumor suppressor genes RASSF1A, CHFR and DAPK is of particular interest, because a hypermethylated promoter region of these genes leads to a decreased or completely

abolished expression of the gene products and thus to a loss of the tumor suppressor function of these gene products. Moreover, the methylation status of the EBV gene LMP1 is also of major interest, because LMP1 codes for a protein that activates transcription factors and blocks apoptosis and has been associated with tumor development and progression.

[0028] The methylation status of the above genes can, for example, be determined via methylation specific PCR (MSP). In order to expedite the determination step, the MSP can be carried out in form of multiple methylation specific PCR (MMSP).

[0029] In one embodiment of the method of the invention, the primers used to determine the EBV gene load may be designed to take into account the changes of nucleotides after bisulfate treatment (for the conversion of unmethylated cytosines to uraciles) so that the EBV gene load is also determined by means of methylation specific PCR (MSP).

[0030] To determine the gene load of EBNA1 in a biological sample or the DNA isolated and purified from such a sample via PCR, oligonucleotide primers suited for this purpose may be used. Suitable upstream (sense) primers for carrying out the method of the invention may comprise, consists essentially of or consist of the nucleotide sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 3, whereas suitable downstream (antisense) primers may comprise, consists essentially of or consist of the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 4.

[0031] To determine the methylation status of each of LMP1, RASSF1A, CHFR and DAPK (which can occur together or individually) in a biological sample or the DNA, which has been isolated and purified from such a sample, via methylation specific PCR, for each gene one set of primers, i.e. for the methylated or the unmethylated gene, depending on the expected methylation status of the gene in question in EBV-related cancers, may be used. As an illustrative example for such a case, a primer set for the unmethylated variant of LMP1 and/or a primer set for the methylated variant of at least one of RASSF1A, CHFR and DAPK may, for example, be utilized. Alternatively, two sets of oligonucleotide primers may be employed. Both sets of oligonucleotide primers (sense and antisense primer) may be directed to the methylated or unmethylated variant of the target gene. The additional primer for each gene may cover the rare event of a gene variation of the gene to be detected. In still another alternative two sets of oligonucleotide primers, one set of primer for the methylated variant of the gene and the other primer set for the unmethylated

variant, may be employed. Such an approach allows, for example, the quantification of gene methylation in a sample by correlating the detected signals for the unmethylated and the methylated gene variant.

[0032] Exemplary upstream (sense) primers for unmethylated LMP1 may comprise, consist essentially of or consist of the nucleotide sequences set forth in SEQ ID NO: 5 and 7. Exemplary upstream (sense) primers for methylated LMP1 may comprise, consist essentially of or consist of the nucleotide sequence set forth in SEQ ID NO: 23.

[0033] Exemplary downstream (antisense) primers for unmethylated LMP1 may comprise, consist essentially of or consist of the nucleotide sequences set forth in SEQ ID NO: 6 and 8. Exemplary downstream (antisense) primers for methylated LMP1 may comprise, consist essentially of or consist of the nucleotide sequence set forth in SEQ ID NO: 24.

[0034] Exemplary upstream (sense) primers for methylated RASSF1A, CHFR and DAPK may comprise, consist essentially of or consist of the nucleotide sequence set forth in SEQ ID NO: 9 or SEQ ID NO: 11, SEQ ID NO: 13 or SEQ ID NO: 15, and SEQ ID NO: 17 or SEQ ID NO: 19, respectively. Exemplary upstream (sense) primers for unmethylated RASSF1A, CHFR and DAPK may comprise, consist essentially of or consist of the nucleotide sequence set forth in SEQ ID NO: 25, SEQ ID NO: 27 and SEQ ID NO: 29, respectively.

[0035] Exemplary downstream (antisense) primers for methylated RASSF1A, CHFR and DAPK may comprise, consist essentially of or consist of the nucleotide sequence set forth in SEQ ID NO: 10 or SEQ ID NO: 12, SEQ ID NO: 14 or SEQ ID NO: 16, and SEQ ID NO: 18 or SEQ ID NO: 20, respectively. Exemplary downstream (antisense) primers for unmethylated RASSF1A, CHFR and DAPK may comprise, consist essentially of or consist of the nucleotide sequence set forth in SEQ ID NO: 26, SEQ ID NO: 28 and SEQ ID NO: 30, respectively.

[0036] The reference, to which the determined expression levels or the epigenetic status of the selected genes is compared, may be normal (healthy) tissue, such as normal epithelial cells, or any other reference tissue. Thus, said reference tissue can be non-cancerous tissue composed of non-EBV-infected cells, such as non-EBV infected epithelial cells. The reference tissue can either originate from the subject the biological sample is collected from or from any other adequate source, for example, a subject not suffering from an EBV infection or an EBV-related cancer or a non-EBV infected cell line, e.g. a

non-EBV infected epithelial cell line. Alternatively, an EBV-related tumor tissue or cell line, for example the Namalwa BL cell line, may provide a positive control.

[0037] The biological sample can be either assayed directly, or the DNA contained in the sample can be isolated and purified and subsequently subjected to the invented method.

5 In one embodiment the isolated and purified DNA is genomic DNA.

[0038] To ensure the quality of the sample DNA, the method of the invention can also make use of determining the presence or amount of a control gene. This control gene may be any suitable housekeeping gene, for example actin or glyceraldehyde-3-phosphat dehydrogenase (GAPDH).

10 **[0039]** If the control gene is actin and the determination method is PCR, suitable primers may comprise, consist essentially of or consist of the nucleotide sequence set forth in SEQ ID NO: 21 (sense) and SEQ ID NO: 22 (antisense).

[0040] The subject the biological sample is obtained from may be a mammal, for example, but not limited to, a human, a mouse, a rat, a dog or a cat.

15 **[0041]** Illustrative examples of EBV-related cancers that can diagnosed using the method of the present invention include, but are not limited, to B-cell lymphomas, for example in immune-suppressed patients, Burkitt lymphomas, including the endemic and sporadic form, nasopharyngeal carcinomas (NPC), salivary gland cancer, Hodgkin lymphoma, posttransplant lymphomas, AIDS-associated lymphomas, T-cell lymphomas,
20 such as lethal midline granuloma, gastric adenocarcinoma and leiomyosarcoma in immune-suppressed patients, as well as certain benign tumors, such as oral hairy leukoplakia. Although not all of the afore-mentioned EBV-related cancers may express all marker genes disclosed in the present invention, they may be diagnosed by using a suitable subset of the markers. Burkitt lymphomas, for example, do not express LMP1. However,
25 they may still be diagnosed according to the method of the invention by utilizing the other marker genes disclosed herein.

[0042] In another aspect, the present invention relates to oligonucleotide primers comprising, consisting essentially of or consisting of the nucleotide sequence of SEQ ID Nos: 1-30.

30 **[0043]** In yet another aspect, the present invention relates to a method of modifying a DNA molecule, wherein the DNA molecule is denatured, mixed with liquid agarose, pipetted in a chilled liquid to form agarose beads, with the agarose beads containing the

denatured DNA molecules, wherein the resulting agarose beads are subsequently subjected to the following modification and purification steps (cf., paragraph 0119 below).

Brief description of the drawings

5 [0044] The invention will be better understood with reference to the detailed description when considered in conjunction with the non-limiting examples and the drawings, in which:

[0045] Figure 1 demonstrates the feasibility of Multiplex MSP by directly comparing the obtained results with those of single MSP. The biological sample used is the Namalwa cell line, an EBV positive Burkitt lymphoma cell line expressing EBNA1 and LMP1 that contains one copy of EBV genome DNA per cell (and which can be obtained from American Type Culture Collection (ATCC), for example). These cells were subjected to MMSP or single MSP with primers specific for the indicated marker genes. The obtained PCR fragments were then analyzed on an agarose gel. The density of all bands from MMSP are corresponding to one from single MSP, indicating that reaction condition of MMSP is optimal for each gene and maximally reduced the competition among primers.

[0046] Figure 2 shows the results of an agarose gel electrophoresis of a number of samples that demonstrate the sensitivity of MMSP. Namalwa cells were used as a reference by semi-quantitative evaluation for EBV DNA load and by evaluation of the housekeeping gene actin for sample DNA input. The DNA quantity, corresponding to the number of analyzed Namalwa cells, was serially diluted and applied to MMSP. As shown in the picture, the sensitivity of this assay is equivalent to up to 2 cells from 10^5 EBV negative cells. After MMSP amplification, bands of all selected marker genes could be clearly demonstrated.

25 [0047] Figure 3 illustrates results obtained by MMSP from known NPC cell lines. C15 designates a nude mouse bearing an LMP1 expressing NPC. C666-1 is an EBV positive, but LMP1 non-expressing cell line as confirmed by Western Blot and immunostaining analysis. The rest are EBV negative NPC cell lines. EBNA1 and LMP1 could be detected in C15 and EBNA1 in C666-1.

30 [0048] Figure 4 shows MMSP result from biopsy and paraffin NPCs. 5FB, 9FB and 18FB indicate biopsy samples from NPCs, whereas 56 par., 58 par., 60 par., 63 par. indicate paraffin NPC samples. C15 indicates sample from a nude mice bearing an LMP1 expressing NPC and H₂O indicates a negative control.

[0049] Figure 5 shows a comparison of MMSP results between paraffin NPC samples (par.) and brush NPC samples (brush). C15 indicates sample from a nude mice bearing an LMP1 expressing NPC and Namalwa indicates a Namalwa cell sample (EBV positive Burkitt lymphoma cell line expressing EBNA1 and LMP1). H2O indicates a water control.

5

Detailed description of the invention

[0050] One object of the present invention is to provide an early diagnostic and prognostic tool for various EBV-related tumors.

[0051] In order to achieve this object, the invention provides a method of determining the tumor-specific methylation patterns of a sample. The sample may be collected from body fluid that is close to the primary tumor, e.g. sputum from lung cancer patients, urine from bladder or prostate cancer patients, and nasopharyngeal swab and mouth washing fluid from nasopharyngeal carcinoma (NPC) patients. However, detection using a blood sample is also possible even though use of a blood sample may lead to a rather non-specific signal. The method allows the semi-quantitative evaluation of the expression level and pattern of different markers from materials that can be obtained by non-invasive methods and that are simple and fast to collect.

[0052] In an illustrative approach, the molecular basis of NPC was investigated in order to set up a panel of biomarkers, which can be used for the diagnosis and the monitoring of the treatment efficiency of NPCs.

[0053] Pursuing this approach it was surprisingly found that in NPC samples there are several genes including EBV encoding genes and cellular genes that are significantly changed in their expression level in NPC compared to the normal nasopharyngeal epithelium samples. Based on these findings, a panel of up to 6 markers (two genes from EBV, three from cellular and one for internal quality control), which can be used for the early detection of tumor or monitor of treatment efficiency, was set up.

[0054] Thus, in a first aspect, the present invention is directed to a method of diagnosing Epstein-Barr virus-related cancer in a subject, comprising collecting a biological sample from said subject; determining in said biological sample the amount of the Epstein-Barr virus gene load by determining the EBNA1 gene load and the expression level of at least one gene selected from the group consisting of the Epstein-Barr virus gene LMP1 and the cellular genes RASSF1A, CHFR and DAPK, and comparing the determined expression levels of said genes in said biological sample with a reference.

[0055] The method of the invention is typically carried out in a non-invasive manner, for example, by using as a biological sample a bodily fluid, such as mouth washing fluid or a nasopharyngeal swab/brush. However, the biological sample may as well be blood or a tissue sample.

5 **[0056]** The EBV-related cancer may be any malignancy associated with Epstein-Barr virus infection. It may thus include nasopharyngeal carcinoma (NPC), Burkitt lymphoma, salivary gland carcinoma, Hodgkin lymphoma, other B cell lymphomas occurring in immune-suppressed patients, for example after an organ transplantation, an HIV infection or a chemotherapy, T cell lymphomas, such as lethal midline granuloma, gastric
10 adenocarcinomas, leiomyosarcoma in immune-suppressed subjects. Also encompassed are benign tumors associated with EBV, for example oral hairy leukoplakia.

[0057] In one exemplary embodiment, the method of the invention is used for the diagnosis of NPC in a subject.

[0058] The EBV gene load is determined by selection of a suitable marker gene, for
15 example EBNA1, and determining the copy number of said gene per cell, for example by semi-quantitative PCR. To calculate the gene load the obtained result may be compared to a reference, for example a cell line with a defined copy number of EBV genomes per cell. A specific example for such cells is the Namalwa cell line, which is derived from a human Burkitt lymphoma and has a single copy of the EBV genome per cell.

20 **[0059]** EBNA1 is the abbreviation for Epstein-Barr nuclear antigen 1 and is used in the context of the present invention to designate the gene encoding said protein. The gene product of EBNA1 is essential for the virus maintenance and also may be important for growth transformation of human primary B lymphocytes, because it is involved in the replication, segregation, and transcriptional activation of latent Epstein-Barr virus genomes
25 and avoids proteasomal processing and cell-surface presentation. The EBNA1 protein is implicated to influence cellular events by sequestering key regulatory proteins. Accordingly, EBNA1 is a gene which is expressed in essentially every EBV infected cell and as such is suitable to determine the EBV gene load, which is correlated to the clinical course and prognosis of EBV-related cancers as well as their tendency to metastasize and
30 hence may be monitored during therapy.

[0060] In case the EBNA1 gene load in a given sample is determined by any suitable nucleic acid amplification method, for example a PCR technique, the oligonucleotide primers set forth in the present invention may be used. The oligonucleotide primers

disclosed by the present invention to determine the EBNA1 gene load in a sample comprise, consist essentially of or consist of the nucleotide sequences set forth in SEQ ID NOs.: 1-4, with SEQ ID NOs.: 1 and 3 representing upstream, i.e. sense, primers, and SEQ ID NOs.: 2 and 4 representing downstream, i.e. antisense, primers. Moreover, complements, variants and fragments of the nucleotide sequences set forth in SEQ ID NOs.: 1-4 are also contemplated by the present invention.

[0061] The expression levels of the marker genes LMP1, RASSF1A, CHFR and DAPK may be determined by any suitable method known by skilled person. Such methods may include Northern Blot, Western Blot and nucleic acid amplification techniques, for example TMA (transcription mediated amplification) and PCR. In one embodiment of the invention, the epigenetic status, for example the methylation status of the marker genes is determined.

[0062] The methylation status can be determined by use of any suitable method, such as for example, but not limited to, methylation-sensitive DNA restriction enzymes, methylation specific nucleic acid probes and methylation specific polymerase chain reaction (MSP). Particularly useful for carrying out the present invention is the MSP technique. Methylation specific PCR (MSP) is a bisulfite conversion based PCR technique for the study of DNA CpG methylation. For an MSP experiment, oligonucleotide primer pairs specific for methylated DNA (M, M primer) and/or for unmethylated DNA (U) are necessary. To achieve discrimination for methylated and unmethylated DNA, in each primer (or at least one of the pair) sequence, one or more CpG sites are included. First, DNA is modified with sodium bisulfite and purified. The treatment of DNA with bisulfite converts all unmethylated cytosines (C) to uraciles, so that instead of Watson-Crick base-pairing with guanosines (G) they now pair with adenosines (A). The utilized primer pairs are designed to detect the target region with methylated C's by having a G at the respective complementary position (M primer) or the target region with unmethylated C's that have been converted to U's by having an A at the respective complementary position (U primer). Then, PCR reactions are performed using M primer pair and/or U primer pair. Successful amplification from M pair and/or U pair indicates methylation and unmethylation, respectively.

[0063] Besides a single methylation specific PCR reaction, several such reactions with different primers and different targets can be performed simultaneously, as long as the obtained amplified PCR fragments are distinguishable, which in turn can be achieved by

appropriate primer design. Such assay is called multiple methylation specific PCR (MMSP).

[0064] The MMSP technique has the advantage the methylation status of several target genes can be evaluated in a single reaction. Therefore, this approach reduces DNA requirement from samples and primer competition during PCR amplification.

[0065] The EBV protein LMP1 (latent membrane protein 1) is absolutely essential for the immortalization of B lymphocytes by EBV and is expressed in most of the NPCs and in several EBV-associated tumors. It is a direct target of EBNA2 and crucial for the survival and proliferation of latently infected cells. Investigation of deletion mutants has shown that the transmembrane as well as the signaling domain of the LMP1 protein are necessary for its transforming properties. The LMP1 protein has a molecular weight of approximately 62 kDa and is 386 amino acids long, with the N-terminal 186 amino acids forming the transmembrane domain and the C-terminal 200 amino acids forming the cytoplasmic signaling domain. The transmembrane domain consists of six transmembrane regions connected via short loop regions that mediate the multimerization. The multimerization of the LMP1 proteins leads independently from ligand binding to a constitutively activated signaling complex. Although the sequence homology to the members of the tumor necrosis factor receptor (TNF-R) family is scarce, LMP1 shares many functional similarities with these proteins. So does LMP1 interact via its C-terminal activating regions 1-3 (CTAR 1-3) with many of the proteins the TNF-R family proteins interact with, including TRAF 1, 2, 3 and 5, TRADD and RIP. Upon binding of these proteins the canonical and non-canonical NF κ B signaling pathway is activated. Furthermore, the Jak-STAT pathway as well as the MAP kinases p38 JNK and ERK become activated. Via these signaling pathways LMP1 induces the expression of activation and adhesion markers, such as C54, MHC class II, CD 23, CD95, and CD86, and of anti-apoptotic protein, such as Bcl2, BclX and A20. LMP1 is the only viral protein of EBV for which a direct oncogenic potential has been demonstrated. Accordingly, LMP1 expression has been found in several EBV-related malignancies. The level of LMP1 protein expression depends strongly from the methylation status of LMP1. In the unmethylated form found in many EBV-related cancers, the LMP1 protein is constitutively expressed. In contrast, no LMP1 protein expression is found in normal tissues, as those normally do not contain the EBV genome, or non-tumorigenic EBV-infected tissues, wherein LMP1 is typically methylated.

[0066] Accordingly, the methylation status of the LMP1 gene is of high significance in the diagnosis of EBV-associated malignancies.

[0067] The presence or amount of unmethylated LMP1 gene can be determined by methylation specific PCR (MSP) employing oligonucleotide primers comprising, consisting essentially of or consisting of the nucleotide sequence set forth in SEQ ID NO:5 or 7 (upstream, sense) and SEQ ID NO:6 or 8 (downstream, antisense), as well as complements, variants and fragments thereof.

[0068] The RASSF1A (Ras association domain family 1A) protein is a cellular tumor suppressor encoded by RASSF1A, which acts through inhibition of the G1/S phase progression of the cell cycle. The RASSF1A protein has been shown to bind to the Ras-GTP binding protein Nore-1, consistent with its role as a negative regulator of the Ras proto-oncogene (Ortiz-Vega et al. (2002) *Oncogene* **21**(9):1381-1390). From the high frequency with which RASSF1A is inactivated by promoter methylation it has been deduced that it plays a key role in the development of many primary human tumors.

However, the mechanism of RASSF1A action remains yet unknown. RASSF1A also associates with microtubules and it has been shown that this association is essential for RASSF1A to mediate its growth inhibitory effects. Overexpression of RASSF1A promotes the formation of stable microtubules, whereas a dominant-negative fragment of RASSF1A destabilizes microtubule networks. Because RASSF1A localizes to the mitotic spindle and blocks activated Ras-induced genomic instability, it has also been implicated to effect genomic stability. Thus, it is presently assumed that RASSF1A plays a role in the control of microtubule polymerization and potentially in the maintenance of genomic stability (Vos et al. (2004) *Cancer Research* **64**:4244-4250).

[0069] In one embodiment of the present invention, promoter hypermethylation of the RASSF1A gene can be determined via methylation specific PCR utilizing oligonucleotide primers comprising, consisting essentially of or consisting of the nucleotide sequence of SEQ ID NOs.:9 or 11 (upstream, sense) and SEQ ID NO:10 or 12 (downstream antisense), as well as complements, variants and fragments thereof.

[0070] CHFR (Checkpoint with fork-head associated and ring finger) is a cellular tumor suppressor involved in cell cycle control. CHFR is an ubiquitin ligase and the key component in a new mitotic checkpoint that acts at prophase and delays chromosome condensation in response to mitotic stress (Scolnick and Halazonetis (2000) *Nature* **406**(6794):430-5). The cell cycle progression is monitored by checkpoint mechanisms to

ensure the integrity of the genome and the fidelity of sister chromatid separation. Failure of such checkpoint functions results in genomic instability, a condition that predisposes cells to neoplastic transformation and tumor progression. CHFR seems to be required for delaying prophase in human cells. Although CHFR is ubiquitously expressed in normal tissues, it is frequently downregulated in human cancers, mostly owing to hypermethylation of its promoter region. A significant proportion of human esophageal cancer has loss of expression of the CHFR gene (Shibata et al. (2002) *Carcinogenesis* **23**(10): 1695-1700; Toyota et al. (2003) *PNAS* **100**(13):7818-7123; Yu et al. (2005) *Nature Genetics* **37**(4):401-406).

[0071] In one embodiment of the present invention, promoter hypermethylation of the CHFR gene can be determined via methylation specific PCR utilizing oligonucleotide primers comprising, consisting essentially of or consisting of the nucleotide sequence of SEQ ID Nos.:13 or 15 (upstream, sense) and SEQ ID Nos.:14 or 16 (downstream antisense), as well as complements, variants and fragments thereof.

[0072] DAPK (death associated protein kinase) is a cellular tumor suppressor. DAPK is a Ca²⁺/Calmodulin-dependent, cytoskeletal associated protein kinase whose expression is implicated in the sensitivity of cells to apoptotic effects of TNF α and Interferon- γ . Deletion of the Calmodulin-binding domain results in a constitutively active mutant with a more potent cytotoxic effect. In contrast, the catalytically inactive mutant exerts a dominant negative effect, reducing cytotoxicity and protecting cells from Interferon- γ induced cell death. DAPK expression is frequently lost in human carcinomas and B-cell leukemia, and lower levels of expression correlate with high rates of metastasis. A high frequency of death-associated protein kinase (DAPK) promoter hypermethylation has been noted in B-cell malignancies, head and neck cancers, and other solid tumors, and it has thus been used as a tumor marker in molecular detection strategies (Reddy et al. (2003) *Cancer Research* **63**: 7694-7698).

[0073] In one embodiment of the present invention, promoter hypermethylation of the DAPK gene can be determined via methylation specific PCR utilizing oligonucleotide primers comprising, consisting essentially of or consisting of the nucleotide sequence of SEQ ID NO:17 or SEQ ID NO:19 (upstream, sense) and SEQ ID NOs.:18 or 20 (downstream, antisense), as well as complements, variants and fragments thereof.

[0074] Additionally, in another embodiment the methylated form of the LMP1 gene and the unmethylated forms of the RASSF1A, CHFR and/or DAPK genes may be

determined separately or in addition to the above. If the methylated as well as the unmethylated form of a gene is determined both signals can be correlated to determine the extent of gene methylation in a sample. Suitable primers to determine methylated LMP1 comprise, consist essentially of or consist of the nucleotide sequence set forth in SEQ ID NO: 23 (upstream, sense) and SEQ ID NO: 24 (downstream, antisense). For the determination of unmethylated RASSF1A, CHFR and/or DAPK primers that comprise, consist essentially of or consist of the nucleotide sequence set forth in SEQ ID NOs: 25, 27, and 29, respectively (upstream, sense) and SEQ ID NOs: 26, 28, and 30, respectively (downstream, antisense), may be utilized.

10 **[0075]** According to the above, in one embodiment the method of the invention is directed to determining the EBNA1 gene load and the methylation status of the EBV oncogene LMP1.

[0076] In addition, the methylation status of the cellular tumor suppressor gene RASSF1A may be determined.

15 **[0077]** Optionally, the methylation status of the cellular tumor suppressor genes CHFR and DAPK may be also determined.

[0078] In one embodiment the determination of the gene load and the expression level is carried out by PCR, for example MSP or MMSP.

[0079] Prior to the determination step, the DNA may be isolated and purified from the biological sample, and then subjected to the respective determination technique. The isolated and purified DNA may contain both viral and cellular DNA and may be genomic DNA. Suitable methods and protocols to isolate DNA or genomic DNA from a sample are known to the skilled person and kits for performing DNA isolation and purification are commercially available.

20 **[0080]** In order to have a check (internal reference) that the isolated DNA is of quality that is sufficient to carry out the determination of the expression level or the EBNA1 gene load a control gene may be concomitantly determined. Suitable control genes are all housekeeping genes (a housekeeping gene is typically a constitutive gene that is transcribed at a relatively constant level) of a cell, for example actin, ubiquitin or GAPDH.

30 The absence of such a house-keeping gene in a DNA preparation indicates that the DNA preparation is of a quality (purity) that is not suitable for carrying out the diagnostic method described here.

[0081] In case the control gene is actin and the determination method is a PCR technique, suitable oligonucleotide primers for the detection of actin may comprise, consist essentially of or consist of the nucleotide sequences set forth in SEQ ID NO:21 (upstream, sense) and SEQ ID NO:22 (downstream, antisense), as well as complements, variants and fragments thereof.

[0082] For the evaluation of the obtained results, the expression level of the selected marker genes may be compared to a reference. This reference may be normal tissue, i.e. non-cancerous tissue, which is preferably of the same tissue type as the biological sample. The reference may originate from the subject the biological sample was collected from or any other suitable source, such as healthy tissue from another donor or cultured cells. If the expression level of a gene of interest exceeds a certain, for example, predetermined threshold value compared to the reference sample, the expression level of the gene is considered to be altered. The expression level can either be increased or reduced compared to the reference sample, depending on the biological function of the selected gene.

[0083] In this context, the level of LMP1 protein expression, for example, depends strongly from the methylation status of the LMP1 gene. In the unmethylated form found in several EBV-related cancers, the LMP1 protein is constitutively expressed. In contrast, no LMP1 Protein expression is found in normal (healthy) tissues, which are normally EBV genome negative or wherein the LMP1 gene promoter is typically methylated.

[0084] Contrary to the viral oncogene LMP1, the cellular tumor suppressor genes RASSF1A, CHFR and DAPK are usually constitutively expressed in normal (healthy) cells. Upon hypermethylation of the genes, in particular the promoter region, the expression of these genes is decreased or abolished, and the gene products can no longer be found in the affected cells. Accordingly, missing or reduced RASSF1A, CHFR and/or DAPK protein expression is found in many cancers.

[0085] As already stated above, the gene load may be determined by comparison of the obtained results with a reference with a defined EBV gene load. Such a reference may be a cell from a cell line with a defined EBV gene load, such as Namalwa cells.

[0086] The subject the biological sample is collected from may be mammal, in particular a human being.

[0087] In one embodiment, NPC samples are obtained from brush or mouth-washing, being from elapsed tumor cells, and tumor DNAs are modified by chemicals and subsequently amplified by methylation and unmethylation specific PCR through a single

reaction. This design is able to reduce DNA amount from precious samples and primer competition during PCR amplification. In this embodiment, an actin marker can be used as an internal DNA quality control. Namalwa cells can be used as a control for systematic working conditions and a reference for semi-quantity control for EBV DNA load. The “expression” pattern of NPC related genes can then be evaluated for early detection /prognosis. This method is a simple and feasible way to be used in a clinical laboratory.

[0088] The method described here may not only have a value for the diagnosis, but also may for monitoring the treatment of cancer. As such the method may generate a new-criteria for tumor classification from a molecular basis, opening the way to a host of innovative diagnostic, preventive and therapeutic strategies. Accordingly, the results obtained from the method of the invention may be used in developing a therapeutic or preventive strategy for Epstein-Barr virus-associated cancers.

[0089] In another aspect, the present invention is directed to reagents for carrying out the method of the invention. Such reagents allow the determination of the gene load of EBNA1 and/or the expression level of any one of LMP1, RASSF1A, CHFR and DAPK. Exemplary reagents suitable for determining the gene load of EBNA1 and/or the expression level of a gene selected from the group of LMP1, RASSF1A, CHFR and DAPK by means of a nucleic acid amplification technique may be oligonucleotide primers. These primers may comprise, consist essentially of or consist of the nucleotide sequences set forth in SEQ ID Nos. 1-20 and 23-30 as well as complements, variants and fragments thereof.

[0090] In still another aspect, the present invention is directed to a kit for carrying out the method of the invention. Hence, the present invention also features a kit for the diagnosis of EBV-related cancer in a subject, comprising at least one reagent for determining the EBV gene load, for example by determining the EBNA1 gene load, and at least one reagent for determining the expression level of a marker gene, said marker gene being selected from the EBV gene LMP1 and the cellular genes RASSF1A, CHFR and DAPK.

[0091] The kit of the invention may further include reagents for the determination of the expression of a reference gene.

[0092] In one embodiment of the invention, the reagents comprised in the kit are oligonucleotides, preferably oligonucleotide primer pairs. Said oligonucleotides may

comprise, consists essentially of or consist of any one of the nucleotide sequences set forth in SEQ ID NOs.: 1-30 as well as complements, variants and fragments thereof.

[0093] Accordingly, in one embodiment the invented kit comprises a oligonucleotide primer pair for the determination of the EBNA1 gene load and at least one further primer pair for the determination of the expression of a marker gene selected from LMP1, RASSF1A, CHFR and DAPK.

[0094] Additionally, the kit may comprise further reagents, for example buffers, solvents and compounds for detection of the expression level such as nucleic acid polymerase or antibodies for Western Blots.

[0095] As far as not stated otherwise, the terms and expression used herein have the following meaning.

[0096] “Fragment” with respect to nucleic acids relates to 3’ and/or 5’ shortened versions of a given nucleotide sequence. Such fragments include for example nucleic acids which are shortened by one or more nucleotides at their 3’ and/or 5’ end(s) compared to their full length counterpart. It is preferred that these fragments have a sufficient length to retain the functionality of the full length version, i.e. in case of oligonucleotide primers for nucleic acid amplification, such fragments preferably have a length of at least 9, more preferably at least 12, most preferably at least 15 nucleotides.

[0097] The term “variant” with respect to nucleic acids, refers to chemically modified nucleic acids, for example backbone-modified nucleic acids, or mutated nucleic acids, derived from a given nucleotide sequence. The mutants may comprise one or more base substitutions, however it is preferred that the functionality of the original nucleic acid molecule is retained. Thus, it is preferred that a variant shares at least 70, 75, 80, more preferably 90, most preferably 95% nucleotide sequence identity with the original nucleic acid. The base substitution may include any of the naturally occurring standard bases A, G, T, C and U but may also include other bases. The chemical modification may occur at the base, the sugar moiety or the phosphate backbone and may alter the properties of the nucleic acid molecule, e.g. with respect to its stability.

[0098] “Complementary” relates to the property of a nucleic acid molecule to base-pair with another nucleic acid molecule via canonical Watson-Crick base pairs. Accordingly, a “complement” is a nucleic acid that can hybridize with a given nucleotide sequence via Watson-Crick base pairs.

[0099] A “gene” is a part of the genomic DNA, encoding for a polypeptide.

[00100] An “expression product” is a product of gene expression and comprises the transcription product, i.e. an mRNA, as well as the translation product, i.e. a polypeptide.

[00101] “Expression level” as used herein relates to the extent a gene is transcribed. The expression level can be determined in the present invention by any suitable methodology, for example, via Northern Blotting, transcription-mediated amplification or other known techniques to determine the presence or amount of a specific mRNA. The expression level may also be determined by determining the presence of the encoded protein, for example by use of Western Blotting, immunostaining, immunoprecipitation or other known techniques. As disclosed in the present invention, the expression level can also be directly determined by evaluating the epigenetic status of a gene. Promoter hypermethylation indicates that the expression level is low or virtually non-existent, whereas an unmethylated promoter region indicates that the gene of interest is expressed. The degree of methylation in a promoter of a specific gene is correlated with the gene expression level, shown by RT-PCR (mRNA) or Western blotting (protein) or by methylation specific PCR (DNA fragment).

[00102] “Epigenetic status” as used herein relates to modifications of genetic material that are inheritable via the cell division, and play an important role in many physiological and pathophysiological conditions. Epigenetic mechanisms provide an “extra” layer of transcriptional control that regulates how genes are expressed. These mechanisms are critical components in the normal development and growth of cells. Among the epigenetic events such as methylation, deacetylation and chromatin re-modeling, promoter hypermethylation has been the most detailed studied.

[00103] “Methylation” or “DNA methylation” as used herein relates predominantly to the methylation of the 5-position of cytosines in CpG dinucleotides that may be clustered in groups called CpG islands. A methylated gene thus includes 5-methylated cytosines, primarily in the promoter region or in exon 1 of this gene, whereas an unmethylated gene lacks these extra methyl islands at cytosine.

[00104] A “tumor suppressor” is a gene or gene product that is involved in the control of cellular proliferation or apoptosis. In case such a gene is mutated to lose its function or is deleted the cell loses a fundamental control mechanism that would normally help to regulate cell growth, cell proliferation or apoptosis and thus contribute to prevent cellular transformation. Accordingly, loss-of-function mutations of tumor suppressor genes are frequently found in tumors.

[00105] An "oncogene" is a gene that promotes cellular transformation. Such an oncogene may be of viral origin or may originate from a constitutively activated cellular proto-oncogene, for example by an activating mutation. Numerous viral oncogenes and cellular proto-oncogenes have been implicated to play a major role in the development and progression of cancer.

[00106] "Gene load" as used in the present invention, relates to the copy number of a certain gene or genome, in particular the EBV genome, in a cell or cell-containing sample. In other words, the expression "gene load" refers to the amount of a certain gene, for example the EBV gene EBNA1, in a cell or biological sample. This amount can be determined qualitatively, i.e. by determining if it is higher, equal to or lower than that of a reference sample, or quantitatively, e.g. by calculating the copy number of the gene of interest by comparison with a reference with a defined copy number. Determination of the gene load as used herein also includes detecting the presence or absence of a gene such as EBNA1. Over the last years viral DNA load was found to be a good prognosticator for subsequent clinical events (Tan et al. (2006) *BMC Cancer* 6:227).

[00107] The invention illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising", "including", "containing", etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed. Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications and variations are considered to be within the scope of this invention.

[00108] The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[00109] Other features and advantages of the invention will become apparent from the following illustrative examples and the claims. It is to be understood that the examples are of illustrative purpose only and should not be construed to be limiting the scope of the present invention.

5

Examples

[00110] CNE1, CNE2, TW03, C666-1 and HONE1 NPC cell lines were used in this study. Cells were grown at 37°C in IMDM medium with 10% fetal calf serum and antibiotics, except for C666-1, which was cultured according to conditions specified by Cheung et al. (Cheung et al. (1999) *Int. J. Cancer* **83**(1):121-126).

10

[00111] Namalwa cells are an EBV positive lymphoma cell line expressing EBNA1 and LMP1, and containing one copy of EBV genome DNA per cell.

[00112] Xenograft C15 mice (nude mice with a human NPC) were obtained from Dr. P. Busson, Institute Gustave Roussy, Paris.

15

[00113] 133 NPC samples from pathology-verified NPC patients, 35 unmatched NPC and 24 normal nasopharyngeal epithelia were obtained from GMU, GMU and HMU (ethical approval: no.00-312 Stockholm, Sweden and local committee in China) for this study. Among them, 98 cases are paired samples with biopsy/paraffin NPC and brush/swab samples.

20

[00114] NPC brush/swabs were collected under surface anesthesia with 1% cocaine solution. To obtain these samples, cotton/brush topped sticks were inserted into the left and right nasal cavity and moved until they reached the nasopharyngeal wall. The cotton swab was then rotated several times against the posterior and lateral nasopharyngeal walls. Upon exit, the cotton stick with the absorbed material was transferred in a Falcon® tube containing 3 ml saline and immediately processed by centrifugation at 3000x g for 10min. Cell pellets were stored in 400ul TE buffer (10 µM Tris 1 µM EDTA, pH 7.0) at -80°C until further use. Alternatively, the absorbed material was soaked and washed in TE buffer containing 0.5%SDS/50µg/ml proteinase K (Invitrogen, Carlsbad, CA) and the DNA extracted via the phenol/chloroform (CH₃Cl) method.

25

30

[00115] Mouth washing was collected at morning with 20ml of 0.9% NaCl, twice.

[00116] NPC biopsy tissues were snap frozen in liquid nitrogen until further use. Upon thawing the biopsies were homogenized and treated with TE buffer containing 50µg/ml

proteinase K (Invitrogen, Carlsbad, CA) for 3 hours at 56°C. High molecular weight genomic DNA was obtained by conventional phenol/chloroform and ethanol extraction.

[00117] Genomic DNA from the swab cell pellets was also extracted by conventional phenol/chloroform method.

5 **[00118]** DNA from paraffin section was adapted by silicon method. Four sections of a formalin fixed and paraffin embedded biopsy with a thickness of 5 µm were subjected to DNA extraction. To tubes containing the paraffin sections 100 µl 0.5% Tween-20 was added. The tubes were agitated and heated to 90°C for 10 min on a Thermal Cycler and then cooled down to 55°C. The wax remained in solution prior to digestion. 2 µl of 10
10 mg/ml proteinase K were added to the tubes and incubated for 3 hours at 55°C, with gentle shaking every hour. The digested solutions were heated to 99°C for 10 min with 100 µl 5% Chelex-100 and suspended in Tris-EDTA. The suspended solutions were gently shaken and whilst hot centrifuged at 10,500g for 15 min. The tubes were then placed on ice and the wax was allowed to harden to be removed. The tubes were heated to 45°C and 100 µl
15 chloroform were added while gently agitating the tubes. The tubes were centrifuged at 10,500g for 15 min and the upper phase was recovered (~180µl). Ten µl samples were used for the PCR assay. (Servi et al. (2002) *J. Clinical Microbiology* 40(11): 3986–3992).

[00119] To determine the methylation status of the targeted genes, the bisulfite modification procedure was slightly modified, according to the published protocol of
20 Alexander Olek et al. (Olek A et al. (1996) *Nucleic Acids Res.* 24(24):5064-5066). In brief, 500 ng genomic DNA was denatured by incubation in 0.3 M NaOH for 15 minutes at 37°C, then mixed with two volumes of 2% low melting point agarose. The agarose/DNA mixture was pipetted into chilled mineral oil to form agarose beads. Each bead was placed in an individual tube to which we added aliquots of 200 µl 5M bisulfite solution (2.5M sodium
25 metabisulfite, Sigma; 100mM hydroquinone, Sigma; pH 5.0). The reaction mixture was then incubated in darkness for 16 hours at 50°C. Treatments were stopped by equilibration against 1ml of TE buffer followed by a desulfonation step in 500 µl of 0.2 M NaOH. Finally, the beads were washed with 1ml of H₂O, and then used directly in the PCR reactions.

30 **[00120]** The bisulfite treated DNA was subjected to PCR with primers flanking the targeted methylation-specific PCR regions. The sequences of forward and reverse primers are listed in Table 1.

Table 1. Nucleotide sequences of the oligonucleotide primers used in the PCR reaction

Marker gene	Oligonucleotide sequence Forward /Sense Primer	Oligonucleotide sequence Reverse/Antisense Primer	PCR product size bp
EBNA1	AGA GGT TTA GGA GTT TTA GTA GTT AGT TAT (SEQ ID NO:1)	CAC CTT CTT AAT AAT ATT CAA AAT AAT C (SEQ ID NO:2)	129
	AGG GTT AAG ATA TAG AGA TGG TGT T (SEQ ID NO:3)	TAC TCC TAC CCC TCC TAC TCC TAC (SEQ ID NO:4)	130
Unmethylated LMP1	GAG GTA GTA TGG GTA TAG ATT TTT TGA (SEQ ID NO:5)	CAT TTC CTA TTA CAC TTA ACC ACC AC (SEQ ID NO:6)	149
	TGG TTA TGT TAG AGT AAT GTG (SEQ ID NO:7)	TTT CTA CTT CCC CTT TCT ATG (SEQ ID NO:8)	150
Methylated RASFF1A	GGG TTT TGC GAG AGC GCG (SEQ ID NO:9)	GCT AAC AAA CGC GAA CCG (SEQ ID NO:10)	169
	GTT TTG CGA GAG CGC G (SEQ ID NO:11)	GCT AAC AAA CGC GAA CCG (SEQ ID NO:12)	169
Methylated CHFR	GTT TTA ATA TAA TAT GGC GTC GAT C (SEQ ID NO:13)	CTC AAC TAA TCC GCG AAA CG (SEQ ID NO:14)	213
	TAA TTG TAT TCG AAA GGG TTT TTA C (SEQ ID NO:15)	TTA ATC CTA ACC AAA CGA CTT CG (SEQ ID NO:16)	217
Methylated DAPK	GAG GAT AGT CGG ATC GAG TTA AC (SEQ ID NO:17)	CAA CTA AAA AAT AAA TAA AAA ACG CA (SEQ ID NO:18)	227
	CGG TAG GGT TTG GGG TCG (SEQ ID NO:19)	AAA CCT CCC AAC TTC GAT CG (SEQ ID NO:20)	225
BS-Actin	AAG TTA AGT TTT GTT TTT ATT TTT TTT (SEQ ID NO:21)	CAA TAA TCT CCT TCT ACA TCC TAT C (SEQ ID NO:22)	184
Methylated LMP1	AAT TTG TAT AAA GAG GCG CG (SEQ ID NO: 23)	CTT CTA CTT CAT CAC CCG CCG (SEQ ID NO: 24)	149
Unmethylated RASFF1A	GGG GTT TTG TGA GAG TGT G (SEQ ID NO: 25)	CTA ACA AAC ACA AAC CAA ACA (SEQ ID NO: 26)	169
Unmethylated CHFR	TTG TGG TTG TGG GAG GGG GTG (SEQ ID NO: 27)	ATC CCC AAA ACT ACA ACA ACA (SEQ ID NO: 28)	213
Unmethylated DAPK	ATT TTT TAG TTG TGT TTT TGT TGT T (SEQ ID NO: 29)	TCC CAA TTA CTC AAA ACA CTA CCC C (SEQ ID NO: 30)	227

[00121] Twenty-five μ l of bisulfite sequencing PCR reaction contained 20 ng of bisulfite treated DNA as template, 10 pmol of each primer, 100 pmol of each deoxynucleoside triphosphate, 10 $\times$ PCR buffer, and 1 unit of AmpliTaq Gold (Applied Biosystems, Foster City, CA). Cycle conditions were as follows: 95°C for 10 minutes,

followed by 34 cycles at 94°C for 45 seconds, 55°C for 45 seconds, and 72°C for 90 seconds. PCR products were purified using QIA quick Gel Extraction Kit (Qiagen, Chatsworth, CA). The purified PCR product was cloned using the TA Cloning Kit (Invitrogen Corporation, Carlsbad, CA) following the manufacturer's recommendations.

5 Five clones were sequenced using M13 R primer with an ABI3100 DNA sequencer and a Big Dye Terminator v3.0 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA). The completion of bisulfite conversion was confirmed by the presence of substituted thymine for all cytosine residues at non-CpG sites.

[00122] PCR primer sequences, which are specific for the genes EBNA1, unmethylated
10 LMP1, methylated RASSF1A, methylated CHFR, methylated DAPK and Actin, and the expected PCR product sizes are summarized in Table 1. For the PCR reaction, 2 µl of bisulfite-modified DNA were added in a final volume of 25 µl of PCR mixture containing 1x PCR buffer, 1.5 mM MgCl₂, 100 pmol of each deoxynucleotide triphosphate, 100 pmol of each utilized primer and one unit of AmpiTaq Gold (Applied Biosystems, Branchburg,
15 NJ).

[00123] PCR amplification was performed at 95°C for 10 minutes, followed by 34 cycles at 94°C for 30 seconds, the specific annealing temperature 55°C for 45 seconds, and 72°C for 90 seconds. Each reaction included a positive methylation control (Namalwa cell DNA, one EBV genome per cell), a negative methylation control (normal nasopharyngeal
20 epithelial tissue), and a water blank control. MMSP products were analyzed by 2% agarose gel electrophoresis and stained with ethidium bromide. All the experiments were repeated to evaluate the reproducibility of the results and semi-quantitatively evaluated for EBV DNA load with comparison to Namalwa cells as a reference for one copy of EBV genome per cell.

25 **[00124]** The evaluation of the samples yielded the results depicted in Table 2.

30

35

Table 2. Evaluation for the early diagnosis and prognosis of NPC

	Markers	Diagnostic								Predictive					
		EBV tumor						No tumor							
		NPC		NPC & Other											
		Valuable		No value											
EBV	EBNA1	+	+	+	+	+	+	+			+	+	+	+	
	LMP1	+	+	+	+						+				+
Cellular genes	RASSF1A		+			+						+			
	CHFR			+			+					+			
	DAPK				+			+						+	+
Determination [§]		Y	Y	Y	Y	Y	Y	N	N	N	Y	Y	Y	N	N
Degree*		5	5	5	5	4	3	2	0	0	5	4	3	2	

§: Y: NPC, N: no NPC;

*: 5-3 can be diagnosed as NPC, 2: not sure; 0: no NPC

- 5 **[00125]** EBV EBNA1 gene exists in all NPC, but not in normal epithelial cells. Therefore it appears to be essential for diagnosing or predicting an NPC. LMP1 is expressed in more than 60% of NPC. A sample positive for both EBNA1 and LMP1 leads to the diagnosis of an NPC with 100% certainty. To diagnose NPC in a LMP1 negative sample, it requires positive EBNA1 plus positive from either of methylated RASSF1A or
- 10 both methylated CHFR/DAPK. Accordingly, the criterion for the diagnosis of NPC is the detectability of EBNA1 plus either of unmethylated LMP1, methylated RASSF1A or methylated CHFR/DAPK.

- [00126]** The result from 98 matched NPC with biopsy and swab/brush/mouth-washing, 35 unmatched NPC biopsies and 34 normal control samples showed that detection rate of
- 15 cancer is 92 % from swab samples so that the specificity of the MMSP detection method is about 90 % (Table 3). The sensitivity is equivalent to 2 cells from 10^5 cells and the false positive and negative are less than 5%.

Table 3. MMSP results in matched NPC samples

Genes	NPC Biopsy/paraffin section		NPC Swab/brush/mouth-washing		Normal	
	No. cases/total	%	No. cases/total	%	No. cases/total	%
EBNA1	87/91	95.6	76/82	92.7	0	0
LMP1	57/91	62.6	48/82	58.5	0	0
RASSF1A	71/91	78.0	52/82	63.4	1/34	2.9
CHFR	38/58	65.5	25/47	53.2	0	0
DAPK	55/91	60.4	45/82	54.9	2/34	6

[00127] Hence, the present invention provides for the first time a highly specific, highly sensitive, non-invasive method for the early diagnosis of EBV-related cancers, such as nasopharyngeal carcinoma.

Claims:

1. A method of diagnosing Epstein-Barr virus-related cancer in a subject, comprising
 - (a) collecting a biological sample from said subject;
 - 5 (b) determining in said biological sample the amount of the Epstein-Barr virus gene load by determining the EBNA1 gene load and the expression level of at least one gene selected from the group consisting of the Epstein-Barr virus gene LMP1 and the cellular genes RASSF1A, CHFR and DAP; and
 - (c) comparing the determined expression levels of said genes in said biological
10 sample with a reference.
2. The method of claim 1, further comprising isolating and purifying the DNA from said biological sample prior to b).
- 15 3. The method of claim 2, wherein the isolated and purified DNA is genomic DNA.
4. The method according to any one of claims 1-3, wherein the expression level is determined by determining the epigenetic status of at least one gene selected from the group consisting of the Epstein-Barr virus gene LMP1 and the cellular genes RASSF1A,
20 CHFR and DAP.
5. The method according to claim 4, wherein the epigenetic status is the methylation status.
- 25 6. The method according to claim 5, wherein the methylation status is determined by methylation specific PCR (MSP).
7. The method according to claim 6, wherein the methylation specific PCR is multiple methylation specific PCR (MMSP).
- 30 8. The method according to any one of claims 1-7, further comprising determining the gene load of EBNA1 and the expression level of LMP1 and RASSF1A.

9. The method according to claim 8, further comprising determining the expression level of CHFR and/or DAP.
10. The method according to any one of claims 1-9, further comprising determining the expression level of a control gene as a control for the sample DNA quality.
11. The method according to claim 10, wherein a housekeeping gene is used as control gene.
12. The method according to claim 11, wherein actin and/or glyceraldehyde-3-phosphat dehydrogenase (GAPDH) is used as control gene.
13. The method according to any one of claims 1-12, wherein the Epstein-Barr virus-related cancer is selected from the group consisting of B and T cell lymphomas, comprising Burkitt lymphoma, Hodgkin lymphoma, lymphomas associated with immune-suppression, and lethal midline granuloma; carcinomas, comprising nasopharyngeal carcinoma (NPC), salivary gland carcinoma, and gastric adenocarcinoma; and sarcomas, comprising leiomyosarcoma.
14. The method according to any one of claims 1-13, wherein the reference in c) is normal tissue.
15. The method according to any one of claims 1-14, wherein determining the gene load of EBNA1 comprises the use of at least two oligonucleotide primers, wherein the first primer comprises the nucleotide sequence set forth in SEQ ID NO: 1 or SEQ ID NO: 3 or variants, complements and fragments thereof and the second primer comprises the nucleotide sequence set forth in SEQ ID NO: 2 or SEQ ID NO: 4 or variants, complements and fragments thereof.
16. The method according to any one of claims 1-15, wherein determining the expression level of LMP1 comprises the use of at least two oligonucleotide primers, wherein the first primer comprises the nucleotide sequence set forth in SEQ ID NO: 5 or SEQ ID NO: 7 or variants, complements and fragments thereof and the second primer comprises the

nucleotide sequence set forth in SEQ ID NO: 6 or SEQ ID NO: 8 or variants, complements and fragments thereof.

17. The method according to any one of claims 1-16, wherein determining the expression
5 level of RASFF1A comprises the use of at least two oligonucleotide primers, wherein the first primer comprises the nucleotide sequence set forth in SEQ ID NO: 9 or SEQ ID NO: 11 or variants, complements and fragments thereof and the second primer comprises the nucleotide sequence set forth in SEQ ID NO: 10 or SEQ ID NO: 12 or variants, complements and fragments thereof.

10

18. The method according to any one of claims 1-17, wherein determining the expression
level of CHFR comprises the use of at least two oligonucleotide primers, wherein the first
primer comprises the nucleotide sequence set forth in SEQ ID NO: 13 or SEQ ID NO: 15
or variants, complements and fragments thereof and the second primer comprises the
15 nucleotide sequence set forth in SEQ ID NO: 14 or SEQ ID NO: 16 or variants,
complements and fragments thereof.

19. The method according to any one of claims 1-18, wherein determining the expression
level of DAPK comprises the use of at least two oligonucleotide primers, wherein the first
20 primer comprises the nucleotide sequence set forth in SEQ ID NO: 17 or SEQ ID NO: 19
or variants, complements and fragments thereof and the second primer comprises the
nucleotide sequence set forth in SEQ ID NO: 18 or SEQ ID NO: 20 or variants,
complements and fragments thereof.

20. The method according any one of claims 12-19, wherein determining the expression
25 level of actin as a control gene comprises the use of at least two oligonucleotide primers,
wherein the first primer comprises the nucleotide sequence set forth in SEQ ID NO: 21 or
variants, complements and fragments thereof and the second primer comprises the
nucleotide sequence set forth in SEQ ID NO: 22 or variants, complements and fragments
30 thereof.

21. The method according to any one of claims 1-20, wherein the biological sample is a selected from the group consisting of nasopharyngeal swab, mouth washing, and body fluid.
- 5 22. The method according to any one of claims 1-21, wherein the subject is a mammal.
23. The method according to claim 22, wherein the subject is human.
24. Use of the results obtained by the method of any one of claims 1-23 for the
10 development of a preventive or therapeutic strategy for Epstein-Barr virus-related cancer.
25. An oligonucleotide having a nucleotide sequence selected from group of sequences set forth in SEQ ID NO: 1-30.
- 15 26. A kit for diagnosing an Epstein-Barr virus-related cancer in a subject, comprising at least one reagent for determining the EBNA1 gene load and at least one reagent for determining the expression level of at least one gene selected from the group consisting of the Epstein-Barr virus gene LMP1 and the cellular genes RASSF1A, CHFR and DAP.
- 20 27. The kit according to claim 26, further comprising a reagent for determining the gene expression level of a reference gene.
28. The kit according to claim 26 or 27, wherein the reagents for the detection of the EBNA1 gene load, the expression level of at least one gene selected from the group
25 consisting of the Epstein-Barr virus gene LMP1 and the cellular genes RASSF1A, CHFR and DAP and/or the expression level of a reference gene are one or more oligonucleotide primers.
29. The kit according to claim 28, wherein the one or more oligonucleotide primers
30 comprise, consist essentially of or consist of any one of the nucleotide sequences set forth in SEQ ID NOs.:1-30.

30. A method of modifying a DNA molecule, wherein the DNA molecule is denatured, mixed with liquid agarose, pipetted in a chilled liquid to form agarose beads, with the agarose beads containing the denatured DNA molecules, wherein the resulting agarose beads are subsequently subjected to the following modification and purification steps.