



(51) International Patent Classification:

C12Q 1/6886 (2018.01)

(21) International Application Number:

PCT/GB2018/050400

(22) International Filing Date:

14 February 2018 (14.02.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

1702392.0

14 February 2017 (14.02.2017) GB

(71) Applicant: THE UNIVERSITY COURT OF THE
UNIVERSITY OF EDINBURGH [GB/GB]; Old College,
South Bridge, Edinburgh Midlothian EH8 9YL (GB).(72) Inventors: POLLARD, Jeffrey William; Queen's Med-
ical Research Institute, Edinburgh EH16 4TJ (GB).
FRAGKOIANNI, Stamatina; Queen's Medical Re-
search Institute, Edinburgh EH16 4TJ (GB). CASSET-
TA, Luca; Queen's Medical Research Institute, Edinburgh
EH16 4TJ (GB).(74) Agent: HGF LIMITED; Document Handling (London), 1
City Walk, Leeds Yorkshire LS11 9DX (GB).(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).

Published:

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

(54) Title: METHODS FOR CANCER DIAGNOSIS USING A GENE EXPRESSION SIGNATURE

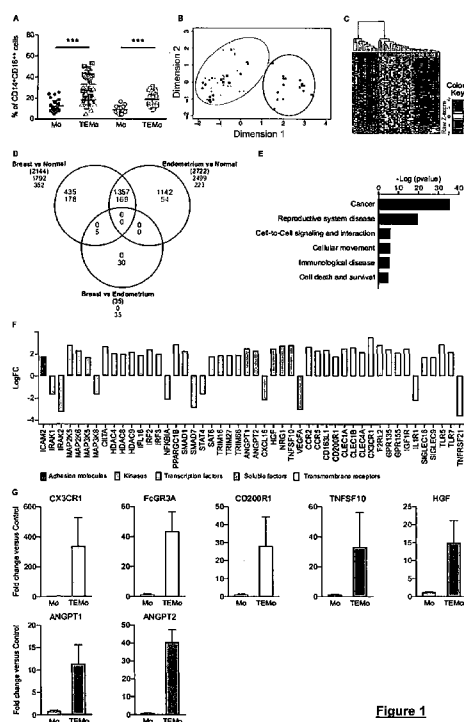


Figure 1

(57) Abstract: The present invention relates to the field of cancer diagnosis and treatment. The present invention provides methods of diagnosing cancer, prognosing cancer, predicting efficacy of treatment for cancer, assessing outcome of treatment for cancer or assessing recurrence of cancer comprising the steps of a) analysing a biological sample obtained from a subject to determine the presence of target molecules representative of expression of at least four biomarkers selected from the group consisting of MCTP1, PIBF1, TMTC2, ANKRD32, CEPT1, ZNF114, CRYBG3, IFI16, PPIF, SCD, RP11-469M7.1, PTP4A1 and NRIP1; and b) comparing the expression levels of the biomarkers determined in (a) with one or more reference values, wherein a difference in the expression of the biomarkers in the sample from the subject compared to the one or more reference values is indicative of cancer presence or absence. The invention further provides methods of treating cancer, and devices and kits for use in the methods of the invention.

METHODS FOR CANCER DIAGNOSIS USING A GENE EXPRESSION SIGNATURE

FIELD OF THE INVENTION

5 The present invention relates to methods for diagnosing cancer, for example breast cancer and endometrial cancer. The invention also relates to methods of assessing the prognosis of a cancer and response to treatment. The invention also relates to methods of treating cancer. Further, the invention concerns kits and assay devices for use in the methods of the invention.

BACKGROUND TO THE INVENTION

10 The tumor microenvironment is a dominant player of tumor progression and growth; cancer cells acquire the ability to “distract and educate” the immune system so that their abnormal proliferation and invasive capacity is not detected, but rather promoted.

It has been previously reported that, in the mouse, bone marrow-derived monocytes migrate to primary^{1,2} and metastatic³ tumour sites where they differentiate into Tumor Associated
15 Macrophages (TAMs) and promote angiogenesis, tumor cell invasion, extravasation, dissemination and overt metastasis^{4,5}.

In contrast, very little is known on the role of human monocytes and macrophages in human cancer. There are at least two circulating monocyte populations in the human blood; classical monocytes (CCR2^{High} CD14⁺⁺CD16⁻) and nonclassical monocytes (CX3CR1^{High}
20 CD14⁺CD16⁺⁺), that share a common progenitor but exhibit different immune functions^{6,7}.

There is growing evidence that neoplastic growths in tissues/organs impacts peripheral and circulating leukocytes in blood, potentially affecting their responses to tissue damage caused by neoplasia.

Recent reports have underlined a positive correlation between circulating monocytes and
25 cancer progression⁸⁻¹¹, thus revealing a significant expansion of the “non-classical” monocyte subset in cancer patients compared to healthy individuals⁹. Moreover total monocytes from a small cohort of renal cell carcinoma patients and a larger colorectal cancer patients have a distinct transcriptional profile as compared to healthy individuals^{12,13}.

Although these data corroborate the hypothesis that cancer is able to systemically perturb
30 immune homeostasis, very little is known on the role of human peripheral monocytes in breast (and endometrial) cancer.

Breast Cancer is the most common cancer in women¹⁴. Early detection of tumours improves survival rates; more than 90% of women diagnosed with early stage breast cancer survive for at least five years¹⁵. Mammographic screening, by enabling early detection, reduces mortality in women 50-74 years of age, whereas efficacy is more limited for younger women 40-49 years old¹⁶, with false positive resulting in overdiagnosis and potentially unnecessary radiation therapy-induced carcinogenesis. At present, alternative early detection screening methods (e.g., MRI, ultrasonography, clinical and self-breast examination) are inadequate to reduce breast cancer mortality. Only invasive tests like mammary biopsies accurately determine if a tissue alteration is neoplastic or not. These data, in combination with the cost-inefficiency of current early screening¹⁷ underlies an urgent medical need for improved early detection of lesions likely to progress into life-threatening malignancy.

With this in mind, the inventors sought to develop a rapid, blood-based, non-invasive diagnostic test that can diagnose breast and other cancers at early stage with high accuracy and specificity. This would allow efficient and frequent screening of patients at risk for developing cancer in order to increase the chances for a rapid pharmacological intervention, and would potentially identify novel myeloid-based targets for therapy.

Significantly, and surprisingly, the inventors have found and show that the presence of many different types of cancer, including breast cancer, can be detected at an early stage using a monocyte-derived gene signature with a high level of accuracy.

SUMMARY OF THE INVENTION

Accordingly, the present invention provides methods of diagnosing and/or prognosing cancer, predicting efficacy of treatment for cancer, assessing outcome of treatment for cancer or assessing recurrence of cancer. The methods comprise the steps of a) analysing a biological sample obtained from a subject to determine the presence of target molecules representative of expression of at least four biomarkers selected from the group consisting of *MCTP1*, *PIBF1*, *TMTC2*, *ANKRD32*, *CEPT1*, *ZNF114*, *CRYBG3*, *IFI16*, *PPIF*, *SCD*, *RP11-469M7.1*, *PTP4A1* and *NRIP1*; and b) comparing the expression levels of the biomarkers determined in (a) with one or more reference values, wherein a difference in the expression of the biomarkers in the sample from the subject compared to the one or more reference values is indicative of cancer presence or absence. Preferably the method is a method of diagnosing cancer.

The invention also provides methods of treating cancer in a subject. The methods comprise the steps of a) analysing a biological sample obtained from a subject to determine the presence of target molecules representative of expression of at least four biomarkers

selected from the group consisting of *MCTP1*, *PIBF1*, *TMTC2*, *ANKRD32*, *CEPT1*, *ZNF114*, *CRYBG3*, *IFI16*, *PPIF*, *SCD*, *RP11-469M7.1*, *PTP4A1* and *NRIP1*; b) comparing the expression levels of the biomarkers determined in (a) with one or more reference values, and in the event that there is a difference in the expression of the biomarkers in the sample
5 from the subject compared to the one or more reference values, identifying the subject as requiring a treatment for cancer or not. Preferably, in the event that the subject is identified as requiring a treatment for cancer, the method further comprises providing the subject with said treatment for cancer.

The invention also provides kits for use in the above methods, the kits comprising binding
10 partners capable of binding to target molecules representative of expression of at least four biomarkers selected from the group consisting of *MCTP1*, *PIBF1*, *TMTC2*, *ANKRD32*, *CEPT1*, *ZNF114*, *CRYBG3*, *IFI16*, *PPIF*, *SCD*, *RP11-469M7.1*, *PTP4A1* and *NRIP1*. Preferably the kits also comprise indicators capable of indicating when said binding occurs.

The invention also provides an assay device for use in the above methods, the device
15 comprising: a) a loading area for receipt of a biological sample; b) binding partners specific for target molecules representative of expression of at least four biomarkers selected from the group consisting of *MCTP1*, *PIBF1*, *TMTC2*, *ANKRD32*, *CEPT1*, *ZNF114*, *CRYBG3*, *IFI16*, *PPIF*, *SCD*, *RP11-469M7.1*, *PTP4A1* and *NRIP1*; and c) detection means to detect the levels of said target molecules present in the sample.

20 DETAILED DESCRIPTION

The methods of the present invention provide simple tests that may be used in diagnosing cancer, prognosing cancer, predicting efficacy of treatment for cancer, assessing outcome of treatment for cancer and assessing recurrence of cancer, and provides methods of treatment using the diagnosis. Such cancers that may be diagnosed include, but are not
25 limited to, breast and endometrial cancer. The kits and devices of the invention are useful for conducting the methods of the invention.

The invention is based upon the inventors' finding that certain biomarkers show differential expression in samples obtained from subjects with cancer as compared to samples obtained from subjects without cancer. The inventors' work in identifying these clinically useful
30 biomarkers has led them to the instant invention, involving new methods of cancer diagnosis and prognosis. The advantages provided in respect of the new methods are notable, and worthy of further comment at this time.

As discussed elsewhere in the specification, the methods of the invention may be put into practice in the form of a blood test. The use of a blood sample to detect cancer, such as

breast cancer and endometrial cancer, provides advantages, such as allowing the use of one simple and minimally-invasive test to detect the possible presence of many different cancer types, making it an efficient and economical screening tool.

Currently there are two main routes by which cancers are detected. Firstly, the subject may present with symptoms that may be suggestive of the subject having cancer, for example a lump or swelling, unexplained bleeding, and/or unexplained weight loss. When such symptoms are identified, there are generally a number of possible causes, and it can be a laborious and time consuming process to rule out each possible cause in turn, meaning a potential delay before cancer can be identified (or not) as the likely cause. In such circumstances the methods of the invention advantageously can provide a simple blood test that can quickly and non-invasively be used to determine whether cancer is the likely cause, leading to earlier diagnosis and treatment, and so a better outcome for the subject.

Secondly, the subject may take part in a routine screening programme that may lead to the detection of the cancer. Such screening programmes are currently directed at detecting specific cancer types, and generally include screening only those considered at highest risk, for example the breast screening by mammography of women aged over 50 in the UK to detect breast cancer. The aims of such screening include detecting the development of cancer at an earlier stage, when treatments are likely to be more successful but the cancer may be generally asymptomatic. The methods disclosed herein provide alternative tests for cancer screening; potential advantages of these alternative tests include the fact that methods involving a simple blood test will be generally less invasive and more convenient for the subject than the screening methods of the prior art, and also the methods disclosed herein are not limited to detecting only one cancer type.

Of course, in some embodiments the methods of the invention may be used in combination with other methods of detecting, diagnosing, prognosing and/or treating cancer, in which case the combination may advantageously increase specificity and sensitivity compared to use of the other methods on their own, and allow the prioritization of the identification, follow-up and treatment of those most likely to have cancer and those most suited to a particular form of treatment.

Thus methods of the invention, such as those employing a blood test, may allow patients with suspected cancer to be identified swiftly, and guide medical staff to commence appropriate treatment promptly. Furthermore, the methods may allow patients without cancer to avoid unnecessary exploratory procedures, such as biopsies.

By measuring human biomarkers in blood, this invention offers to revolutionise, accelerate and improve diagnosis of cancer, guide prompt provision of appropriate treatment and

enhance patient outcomes. This can be of particular benefit in the case of the early diagnosis, or earlier compared to prior art methods of diagnosis, of common cancers such as breast or endometrial cancer.

In order to assist the understanding of the present invention, certain terms used herein will now be further defined, and more generally further details of the invention will be given, in the following paragraphs.

Biological Sample

The biomarker expression levels are analysed in a biological sample obtained from a subject. The biological sample may be a blood sample or a derivative thereof, and preferably the blood sample will be a peripheral blood sample. The biological sample will comprise monocytes, and preferably may be enriched for monocytes or may substantially consist of monocytes. Thus preferably at least 75% of the cells in the biological sample will be monocytes, for example 80%, 85%, 90%, 95%, 96%, 97%, 97.5% 98%, 99%, 99.5% or 99.8% of the cells in the sample will be monocytes. It is particularly preferred that at least 97% of the cells in the sample will be monocytes. Preferably the monocytes in the sample are obtained from peripheral blood. The biological sample may comprise, or be enriched for, a total monocyte population, non-classical monocytes, and/or classical monocytes.

Preferably the expression levels of the biomarkers are selectively detected in monocytes of the biological sample. Therefore, it is particularly preferred that the biological sample in which the levels are detected will be enriched for monocytes or may substantially consist of monocytes. Suitable methods for enriching samples for monocytes are known to those of skill in the art, for example using FACS sorting or commercially available kits like the pan-monocyte extraction kit (from Myltenyi Biotec) and the EasySep™ Human monocyte isolation kit (from STEMCELL™ technologies), which allow the separation of pure monocytes from peripheral blood cells or total blood, using magnetic beads, in less than 1 hour. Such methods for enrichment may be included in the methods of the invention, and associated reagents may be included in the kits and devices for use in the methods of the invention. Alternatively, or additionally, the step of analysing the levels of the biomarkers may specifically target the monocytes for that analysis. For example, the analysis may take place on the magnetic beads to which the monocytes specifically attach, such that even though the biological sample may be of peripheral blood for example, the expression levels analysed substantially correspond only to the levels in the monocytes of the sample. Thus, it may not be necessary to enrich for monocytes when preparing a sample for use in the methods of the invention.

The method may involve obtaining a sample of biological material from the subject, or it may be performed on a pre-obtained sample, e.g. one which has been obtained previously for this or other clinical purposes. Similarly, the biological sample obtained from the subject may be pre-processed before use in methods of the invention, for example to enrich for monocytes, and/or the methods of the invention may include suitable processing steps to enrich for or identify monocytes in the sample, for example through the use of selective magnetic separation systems such as those mentioned above.

In some embodiments the methods of the present invention may make use of a range of biological samples taken from a subject to determine the expression level of a biomarker.

10 **A subject**

In the context of the methods and medical uses of the present invention, a subject may be anyone requiring the diagnosis, prognosis and/or treatment for cancer. Suitably the subject may be a mammal, preferably a primate and further preferably a human subject.

As mentioned elsewhere in the specification, the subject may present with symptoms consistent with cancer. In such a subject, the method of diagnosis may be used to indicate whether or not the subject actually has cancer.

Alternatively, the subject may appear to be asymptomatic. Suitably an asymptomatic subject may be a subject who is believed to be at elevated risk of having cancer. Such an asymptomatic subject may be one who has a family history of early-onset of cancer or who has an increased risk of an age-related cancer.

In some embodiments, the subject may be undergoing routine examination, for example as part of a health check, and a method of diagnosis in accordance with the invention may be used to screen for cancer during that routine examination. Prior art cancer screens used in this way include the cervical smear test, the prostate specific antigen (PSA) test, and mammography; in some embodiments methods of diagnosis of the invention may be used in addition to, or as an alternative to, such prior art cancer screens.

Levels of biomarkers

Methods of the invention involve looking at the expression levels of biomarkers selected from the list consisting of *MCTP1*, *PIBF1*, *TMTC2*, *ANKRD32*, *CEPT1*, *ZNF114*, *CRYBG3*, *IFI16*, *PPIF*, *SCD*, *RP11-469M7.1*, *PTP4A1* and *NRIP1*, i.e. biomarkers corresponding to the genes listed in Table 1. The methods involve looking at the levels of at least four biomarkers in the list, for example four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen of the biomarkers. Preferably the methods involve looking at the expression levels of at least five

of the biomarkers in the list, at least eight of the biomarkers in the list, or all of the biomarkers in the list. The kits and devices of the invention correspondingly provide binding partners for looking at the levels of four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen of the biomarkers of the invention, in accordance with the methods of the invention disclosed herein.

Table 1 Biomarkers differentially expressed in monocytes, including relative expression in subjects having cancer compared to subjects not having cancer

Gene Name	Gene Full Name	Expression compared to subjects not having cancer
MCTP1	multiple C2 and transmembrane domain containing 1	Increased
PIBF1	progesterone immunomodulatory binding factor 1	Increased
TMTC2	transmembrane and tetratricopeptide repeat containing 2	Increased
SLF1	SMC5-SMC6 complex localization factor 1	Increased
CEPT1	choline/ethanolamine phosphotransferase 1	Increased
ZNF114	zinc finger protein 114	Increased
CRYBG3	crystallin beta-gamma domain containing 3	Increased
IFI16	interferon gamma inducible protein 16	Increased
PPIF	peptidylprolyl isomerase F	Decreased
SCD	stearoyl-CoA desaturase	Decreased
RP11-469M7.1	RP11-469M7.1	Decreased
PTP4A1	protein tyrosine phosphatase type IVA, member 1	Decreased
NRIP1	nuclear receptor interacting protein 1	Decreased

10 The biomarkers may be selected from the group consisting of *PIBF1*, *SCD*, *ZNF114*, *CRYBG3*, *PPIF*, *PTP4A1*, *ANKRD32*, and *CEPT1*, for example the methods may comprise determining the expression levels of four, five, six, seven or eight biomarkers selected from *PIBF1*, *SCD*, *ZNF114*, *CRYBG3*, *PPIF*, *PTP4A1*, *ANKRD32*, and *CEPT1*. In preferred

methods the expression levels of *PIBF1*, *SCD*, *ZNF114*, *CRYBG3*, *PPIF*, *PTP4A1*, *ANKRD32*, and *CEPT1* are determined.

In preferred methods the biomarkers may be selected from the group consisting of *PIBF1*, *SCD*, *ZNF114*, *CRYBG3*, and *PPIF*, for example the methods may comprise determining the expression levels of four or five biomarkers selected from *PIBF1*, *SCD*, *ZNF114*, *CRYBG3*, and *PPIF*. In particularly preferred methods, the expression levels of *PIBF1*, *SCD*, *ZNF114*, *CRYBG3*, and *PPIF* are determined.

It will be apparent to the skilled person that the abovementioned combinations of biomarkers represent various minimal marker sets, and additional biomarkers, whether selected from the list of Table 1 or not, can also be included. Alternatively, in some methods, kits and devices of the invention the four, five, six, seven, eight, nine, ten, eleven, twelve, or thirteen biomarkers selected from Table 1 may be the only biomarkers for which the expression levels are assessed. However, the methods, kits and devices may also provide for the assessment of control target molecules in the biological sample, where the assessment of the control target molecules allow for the accuracy of the assessment mechanism to be tested.

The invention involves assessing changes in levels for biomarkers, and in preferred embodiments this change is typically differentially upwards for *MCTP1*, *PIBF1*, *TMTG2*, *ANKRD32*, *CEPT1*, *ZNF114*, *CRTBG3*, and *IFI16*, but differentially downwards for *PPIF*, *SCD*, *RP11-469M7.1*, *PTP4A1* and *NRIP1*, in subjects having cancer.

Throughout, biomarkers in the biological sample(s) from the subject are said to be expressed at different levels, or differentially expressed, where they are significantly up- or down-regulated. Depending on the individual biomarker, cancer may be diagnosed in a biological sample by either an increase or decrease in expression level, optionally scaled in relation to sample mean and sample variance, relative to those of subjects not having cancer or one or more reference values. Clearly, variation in the sensitivity of individual biomarkers, subject and samples mean that different levels of confidence are attached to each biomarker. Biomarkers of the invention are said to be significantly up- or down- regulated when, optionally after scaling of biomarker expression levels in relation to sample mean and sample variance, they exhibit at least a 1.5-fold change, preferably a 2-fold change, compared with subjects not having cancer or one or more reference values (i.e. a $\log_2$ fold change of greater than 0.58 or less than -0.58, preferably greater than +1 or less than -1). Preferably biomarkers will exhibit a 3-fold change or more compared with the reference value. More preferably biomarkers of the invention will exhibit a 4-fold change or more compared with the reference value. That is to say, in the case of increased expression level

(up-regulation relative to reference values), the biomarker level will be more than double that of the reference value. Preferably the biomarker level will be more than 3 times the level of the reference value. More preferably, the biomarker level will be more than 4 times the level of the reference value. Conversely, in the case of decreased expression level (down-regulation relative to reference values), the biomarker level will be less than half that of the reference value. Preferably the biomarker level will be less than one third of the level of the reference value. More preferably, the biomarker level will be less than one quarter of the level of the reference value.

Throughout the term "reference value" may refer to a pre-determined reference value, for instance specifying a confidence interval or threshold value for the diagnosis or prediction of the susceptibility of a subject to cancer, treatment and/or recurrence. Alternatively, the reference value may be derived from the expression level of a corresponding biomarker or biomarkers in a 'control' biological sample, for example a positive (patient diagnosed with cancer and/or not susceptible to treatment and/or with a poor prognosis) or negative (patient not diagnosed with cancer or patient diagnosed with a cancer that proved susceptible to treatment or patient diagnosed with a cancer who had a successful outcome) control. Furthermore, the reference value may be an 'internal' standard or range of internal standards, for example a known concentration of a protein, transcript, label or compound. Alternatively, the reference value may be an internal technical control for the calibration of expression values or to validate the quality of the sample or measurement techniques. This may involve a measurement of one or several transcripts within the sample which are known to be constitutively expressed or expressed at a known level (e.g. an invariant level). Accordingly, it would be routine for the skilled person to apply these known techniques alone or in combination in order to quantify the level of biomarker in a sample relative to standards or other transcripts or proteins or in order to validate the quality of the biological sample, the assay or statistical analysis.

In preferred methods of the invention the reference values correspond to the levels of the biomarkers in samples from subjects not having cancer. Thus the reference values may be representative of corresponding values in subjects not having cancer. Alternatively in methods involving assessing outcome of treatment for cancer or assessing recurrence of cancer, the reference values may correspond to the levels of the biomarkers in samples from subjects who had been diagnosed with cancer and for whom the outcome of treatment for the cancer and/or the recurrence status is known. Thus the reference values may be representative of corresponding values in subjects who have been successfully treated for cancer, in subjects who have been unsuccessfully treated for cancer, and/or in subjects previously successfully treated for whom the cancer has returned. Similarly, in some

methods involving providing a prognosis for a subject and/or predicting a subject's response to (a particular) treatment, the reference values may correspond to the levels of the biomarkers in samples from subjects with a particular known prognosis or response to the treatment.

- 5 Preferably the subjects used to generate the reference values will be "matched" to some extent with those providing the biological sample. For example, if the subject providing the sample is a female suspected of having breast cancer then preferably the subjects providing the reference values will also be female. Similarly, if the subject providing the sample is an adolescent male suspected of having cancer then preferably the subjects providing the reference values will also be adolescent males. Thus the subjects providing the samples to which the reference values correspond may be "matched" according to sex and/or age. Alternatively the subjects providing the samples to which the reference values correspond may comprise a range of ages and/or sexes.

- 15 *MCTP1* (multiple C2 and transmembrane domain containing 1) is located on chromosome 5q15 and encodes a calcium binding protein. The inventors have surprisingly found that this gene is significantly overexpressed in the peripheral blood monocytes of subjects having cancer, compared to the expression in the peripheral blood monocytes of subjects not having cancer. Therefore in methods of the invention in which the expression levels of *MCTP1* are analysed, it is preferred that significant up-regulation of the expression level of *MCTP1* in a subject is associated with the subject having cancer and/or a poorer prognosis. For example, a log₂ fold change of at least 0.6, for example a log₂ fold change of at least 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, or 2.0, in a sample compared to one or more reference values may be indicative of the subject having cancer and/or a poorer prognosis. The skilled person will appreciate that the relative expression levels of *MCTP1* will depend on the reference values used in the comparison; however, in preferred methods the reference values will correspond to the levels of the biomarkers in samples from subjects not having cancer, and a log₂ fold change of at least 1, preferably at least 1.25, in expression levels of *MCTP1* will be indicative of cancer and/or a poorer prognosis.

- 30 *PIBF1* (progesterone immunomodulatory binding factor 1) is located on chromosome 13q. The inventors have surprisingly found that this gene is significantly overexpressed in the peripheral blood monocytes of subjects having cancer, compared to the expression in the peripheral blood monocytes of subjects not having cancer. Therefore in methods of the invention in which the expression levels of *PIBF1* are analysed, it is preferred that significant up-regulation of the expression level of *PIBF1* in a subject is associated with the subject having cancer and/or a poorer prognosis. For example, a log₂ fold change of at least 1, for example a log₂ fold change of at least 1.5, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, or 3.0

in a sample compared to one or more reference values may be indicative of the subject having cancer and/or a poorer prognosis. The skilled person will appreciate that the relative expression levels of *PIBF1* will depend on the reference values used in the comparison; however, in preferred methods the reference values will correspond to the levels of the biomarkers in samples from subjects not having cancer, and a log₂ fold change of at least 1, preferably at least 1.5, 2.0, 2.5, or 3.0, and further preferably at least 2.0, in expression levels of *PIBF1* will be indicative of cancer and/or a poorer prognosis.

TMTC2 (transmembrane and tetratricopeptide repeat containing 2) is located on chromosome 12q21.31 and encodes an integral membrane protein localized to the endoplasmic reticulum. The inventors have surprisingly found that this gene is significantly overexpressed in the peripheral blood monocytes of subjects having cancer, compared to the expression in the peripheral blood monocytes of subjects not having cancer. Therefore in methods of the invention in which the expression levels of *TMTC2* are analysed, it is preferred that significant up-regulation of the expression level of *TMTC2* in a subject is associated with the subject having cancer and/or a poorer prognosis. For example, a log₂ fold change of at least 1, for example a log₂ fold change of at least 1.5, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, or 3.0 in a sample compared to one or more reference values may be indicative of the subject having cancer and/or a poorer prognosis. The skilled person will appreciate that the relative expression levels of *TMTC2* will depend on the reference values used in the comparison; however, in preferred methods the reference values will correspond to the levels of the biomarkers in samples from subjects not having cancer, and a log₂ fold change of at least 1, preferably at least 1.5, 2.0, or 2.5, and further preferably at least 2.0, in expression levels of *TMTC2* will be indicative of cancer and/or a poorer prognosis.

SLF1 (SMC5-SMC6 complex localization factor 1, also referred to as ANKRD32) is located on chromosome 5q15 and encodes a transcription regulator associated with DNA repair. The inventors have surprisingly found that this gene is significantly overexpressed in the peripheral blood monocytes of subjects having cancer, compared to the expression in the peripheral blood monocytes of subjects not having cancer. Therefore in methods of the invention in which the expression levels of *SLF1* are analysed, it is preferred that significant up-regulation of the expression level of *SLF1* in a subject is associated with the subject having cancer and/or a poorer prognosis. For example, a log₂ fold change of at least 1, for example a log₂ fold change of at least 1.5, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, or 3.0 in a sample compared to one or more reference values may be indicative of the subject having cancer and/or a poorer prognosis. The skilled person will appreciate that the relative expression levels of *SLF1* will depend on the reference values used in the comparison; however, in preferred methods the reference values will correspond to the levels of the

biomarkers in samples from subjects not having cancer, and a $\log_2$ fold change of at least 1, preferably at least 1.5, 2.0, or 2.5, and further preferably at least 2.5, in expression levels of *SLF1* will be indicative of cancer and/or a poorer prognosis.

CEPT1 (choline/ethanolamine phosphotransferase 1) is located on chromosome 1q13.3 and encodes a choline/ethanolaminephosphotransferase involved in the synthesis of choline- or ethanolamine- containing phospholipids. The inventors have surprisingly found that this gene is significantly overexpressed in the peripheral blood monocytes of subjects having cancer, compared to the expression in the peripheral blood monocytes of subjects not having cancer. Therefore in methods of the invention in which the expression levels of *CEPT1* are analysed, it is preferred that significant up-regulation of the expression level of *CEPT1* in a subject is associated with the subject having cancer and/or a poorer prognosis. For example, a $\log_2$ fold change of at least 0.6, for example a $\log_2$ fold change of at least 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, or 2.0, in a sample compared to one or more reference values may be indicative of the subject having cancer and/or a poorer prognosis.

The skilled person will appreciate that the relative expression levels of *CEPT1* will depend on the reference values used in the comparison; however, in preferred methods the reference values will correspond to the levels of the biomarkers in samples from subjects not having cancer, and a $\log_2$ fold change of at least 1, preferably at least 1.2, 1.5, or 2.0, and further preferably at least 1.5, in expression levels of *CEPT1* will be indicative of cancer and/or a poorer prognosis.

ZNF114 (zinc finger protein 114) is located on chromosome 19q3.33 and encodes a zinc finger protein. The inventors have surprisingly found that this gene is significantly overexpressed in the peripheral blood monocytes of subjects having cancer, compared to the expression in the peripheral blood monocytes of subjects not having cancer. Therefore in methods of the invention in which the expression levels of *ZNF114* are analysed, it is preferred that significant up-regulation of the expression level of *ZNF114* in a subject is associated with the subject having cancer and/or a poorer prognosis. For example, a $\log_2$ fold change of at least 1, for example a $\log_2$ fold change of at least 1.5, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, or 3.0, in a sample compared to one or more reference values may be indicative of the subject having cancer and/or a poorer prognosis. The skilled person will appreciate that the relative expression levels of *ZNF114* will depend on the reference values used in the comparison; however, in preferred methods the reference values will correspond to the levels of the biomarkers in samples from subjects not having cancer, and a $\log_2$ fold change of at least 1, preferably at least 1.5, 2.0, or 2.5, and further preferably at least 2.5, in expression levels of *ZNF114* will be indicative of cancer and/or a poorer prognosis.

CRYBG3 (crystallin beta-gamma domain containing 3) is located on chromosome 3q11.2. The inventors have surprisingly found that this gene is significantly overexpressed in the peripheral blood monocytes of subjects having cancer, compared to the expression in the peripheral blood monocytes of subjects not having cancer. Therefore in methods of the invention in which the expression levels of *CRYBG3* are analysed, it is preferred that significant up-regulation of the expression level of *CRYBG3* in a subject is associated with the subject having cancer and/or a poorer prognosis. For example, a $\log_2$ fold change of at least 1, for example a $\log_2$ fold change of at least 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, or 2.5, in a sample compared to one or more reference values may be indicative of the subject having cancer and/or a poorer prognosis. The skilled person will appreciate that the relative expression levels of *CRYBG3* will depend on the reference values used in the comparison; however, in preferred methods the reference values will correspond to the levels of the biomarkers in samples from subjects not having cancer, and a $\log_2$ fold change of at least 1, preferably at least 1.5, 2.0, or 2.2, and further preferably at least 2.0, in expression levels of *CRYBG3* will be indicative of cancer and/or a poorer prognosis.

IFI16 (interferon gamma inducible protein 16) is located on chromosome 1q22 and encodes a member of the HIN-200 (hematopoietic interferon-inducible nuclear antigens with 200 amino acid repeats) family of cytokines. The inventors have surprisingly found that this gene is significantly overexpressed in the peripheral blood monocytes of subjects having cancer, compared to the expression in the peripheral blood monocytes of subjects not having cancer. Therefore in methods of the invention in which the expression levels of *IFI16* are analysed, it is preferred that significant up-regulation of the expression level of *IFI16* in a subject is associated with the subject having cancer and/or a poorer prognosis. For example, a $\log_2$ fold change of at least 1, for example a $\log_2$ fold change of at least 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, or 2.0, in a sample compared to one or more reference values may be indicative of the subject having cancer and/or a poorer prognosis. The skilled person will appreciate that the relative expression levels of *IFI16* will depend on the reference values used in the comparison; however, in preferred methods the reference values will correspond to the levels of the biomarkers in samples from subjects not having cancer, and a $\log_2$ fold change of at least 1, preferably at least 1.5 or 1.8, in expression levels of *IFI16* will be indicative of cancer and/or a poorer prognosis.

PPIF (peptidylprolyl isomerase F) is located on chromosome 10q22.3 and encodes a member of the peptidyl-prolyl cis-trans isomerase (PPIase) family. The inventors have surprisingly found that this gene is significantly underexpressed in the peripheral blood monocytes of subjects having cancer, compared to the expression in the peripheral blood

monocytes of subjects not having cancer. Therefore in methods of the invention in which the expression levels of *PPIF* are analysed, it is preferred that significant down-regulation of the expression level of *PPIF* in a subject is associated with the subject having cancer and/or a poorer prognosis. For example, a $\log_2$ fold change of -1 or less, for example a $\log_2$ fold change of less than or equal to -1.5, -2.0, -2.1, -2.2, -2.3, -2.4, -2.5, -2.6, -2.7, -2.8, -2.9, or -3.0, in a sample compared to one or more reference values may be indicative of the subject having cancer and/or a poorer prognosis. The skilled person will appreciate that the relative expression levels of *PPIF* will depend on the reference values used in the comparison; however, in preferred methods the reference values will correspond to the levels of the biomarkers in samples from subjects not having cancer, and a $\log_2$ fold change of -1 or less, preferably -1.5 or less, -2.0 or less, or -2.5 or less, and further preferably -2.5 or less, in expression levels of *PPIF* will be indicative of cancer and/or a poorer prognosis.

SCD (stearoyl-CoA desaturase) is located on chromosome 10q24.31 and encodes an enzyme involved in fatty acid biosynthesis, primarily the synthesis of oleic acid. The inventors have surprisingly found that this gene is significantly underexpressed in the peripheral blood monocytes of subjects having cancer, compared to the expression in the peripheral blood monocytes of subjects not having cancer. Therefore in methods of the invention in which the expression levels of *SCD* are analysed, it is preferred that significant down-regulation of the expression level of *SCD* in a subject is associated with the subject having cancer and/or a poorer prognosis. For example, a $\log_2$ fold change of -1 or less, for example a $\log_2$ fold change of less than or equal to -1.5, -2.0, -2.1, -2.2, -2.3, -2.4, -2.5, -2.6, -2.7, -2.8, -2.9, -3.0, -3.1, -3.2, -3.3, -3.4, or -3.5 in a sample compared to one or more reference values may be indicative of the subject having cancer and/or a poorer prognosis. The skilled person will appreciate that the relative expression levels of *SCD* will depend on the reference values used in the comparison; however, in preferred methods the reference values will correspond to the levels of the biomarkers in samples from subjects not having cancer, and a $\log_2$ fold change of -1 or less, preferably -1.5 or less, -2.0 or less, -2.5 or less, -3.0 or less, or -3.2 or less, and further preferably -2.5 or less, in expression levels of *SCD* will be indicative of cancer and/or a poorer prognosis.

RP11-469M7.1 is located on chromosome 2 and encodes a human long noncoding RNA; the sequence of the gene is provided herein as SEQ ID NO:19. The inventors have surprisingly found that this gene is significantly underexpressed in the peripheral blood monocytes of subjects having cancer, compared to the expression in the peripheral blood monocytes of subjects not having cancer. Therefore in methods of the invention in which the expression levels of *RP11-469M7.1* are analysed, it is preferred that significant down-regulation of the expression level of *RP11-469M7.1* in a subject is associated with the

subject having cancer and/or a poorer prognosis. For example, a $\log_2$ fold change of -1 or less, for example a $\log_2$ fold change of less than or equal to -1.5, -1.6, -1.7, -1.8, -1.9, -2.0, -2.1, -2.2, -2.3, -2.4, -2.5, or -2.6 in a sample compared to one or more reference values may be indicative of the subject having cancer and/or a poorer prognosis. The skilled person will appreciate that the relative expression levels of *RP11-469M7.1* will depend on the reference values used in the comparison; however, in preferred methods the reference values will correspond to the levels of the biomarkers in samples from subjects not having cancer, and a $\log_2$ fold change of -1 or less, preferably -1.5 or less, -2.0 or less, or -2.5 or less, and further preferably -2.0 or less, in expression levels of *RP11-469M7.1* will be indicative of cancer and/or a poorer prognosis.

PTP4A1 (protein tyrosine phosphatase type IVA, member 1) is located on chromosome 6q12 and encodes a member of a small class of prenylated protein tyrosine phosphatases (PTPs). The inventors have surprisingly found that this gene is significantly underexpressed in the peripheral blood monocytes of subjects having cancer, compared to the expression in the peripheral blood monocytes of subjects not having cancer. Therefore in methods of the invention in which the expression levels of *PTP4A1* are analysed, it is preferred that significant down-regulation of the expression level of *PTP4A1* in a subject is associated with the subject having cancer and/or a poorer prognosis. For example, a $\log_2$ fold change of -0.6 or less, for example a $\log_2$ fold change of less than or equal to -1, -1.1, -1.2, -1.3, -1.4, or -1.5, in a sample compared to one or more reference values may be indicative of the subject having cancer and/or a poorer prognosis. The skilled person will appreciate that the relative expression levels of *PTP4A1* will depend on the reference values used in the comparison; however, in preferred methods the reference values will correspond to the levels of the biomarkers in samples from subjects not having cancer, and a $\log_2$ fold change of -1 or less, preferably -1.2 or less, in expression levels of *PTP4A1* will be indicative of cancer and/or a poorer prognosis.

NR1P1 (nuclear receptor interacting protein 1) is located on chromosome 21q and encodes a transcription regulator. The inventors have surprisingly found that this gene is significantly underexpressed in the peripheral blood monocytes of subjects having cancer, compared to the expression in the peripheral blood monocytes of subjects not having cancer. Therefore in methods of the invention in which the expression levels of *NR1P1* are analysed, it is preferred that significant down-regulation of the expression level of *NR1P1* in a subject is associated with the subject having cancer and/or a poorer prognosis. For example, a $\log_2$ fold change of -0.6 or less, for example a $\log_2$ fold change of less than or equal to -1, -1.1, -1.2, -1.3, -1.4, or -1.5, in a sample compared to one or more reference values may be indicative of the subject having cancer and/or a poorer prognosis. The skilled person will

appreciate that the relative expression levels of *NR1P1* will depend on the reference values used in the comparison; however, in preferred methods the reference values will correspond to the levels of the biomarkers in samples from subjects not having cancer, and a log₂ fold change of -1 or less, preferably -1.2 or less, in expression levels of *NR1P1* will be indicative of cancer and/or a poorer prognosis.

The biological samples are analysed to determine the expression levels of the biomarkers. "Gene expression", or more simply "expression" is the process by which information from a gene is used in the synthesis of a functional gene product, such as a protein or non-coding RNA (ncRNA), including ribosomal RNA (rRNA), or transfer RNA (tRNA). As used herein, the term "expression" includes RNA (for example mRNA or non-coding RNA) transcription. Thus suitably the expression level for a biomarker may be determined by looking at the amount of a target molecule selected from the group consisting of the protein expressed from the biomarker and a polynucleotide molecule encoding the biomarker, or a nucleic acid complementary thereto. It is preferred that the target molecule is a nucleic acid molecule, and highly preferred that it is an RNA molecule, for example mRNA or ncRNA, transcribed from the biomarker or a cDNA molecule complementary thereto.

The levels of the target molecules, which are representative of expression of the biomarkers in the biological sample, may be investigated using specific binding partners, polymerase chain reaction (PCR) and/or sequencing techniques. The binding partners may be selected from the group consisting of complementary nucleic acids, aptamers, and antibodies or antibody fragments. Preferably the levels of the biomarkers in the biological sample are investigated using a nucleic acid probe having a sequence which is complementary to the sequence of the relevant mRNA, ncRNA or cDNA against which it is targeted.

Suitable classes of binding partners for any given biomarker will be apparent to the skilled person, and are discussed further below. The expression levels of the biomarkers in the biological sample may be detected by direct assessment of binding between the target molecules and binding partners. The levels of the biomarkers in the biological sample may be detected using a reporter moiety attached to a binding partner. Preferably the reporter moiety is selected from the group consisting of fluorophores; chromogenic substrates; and chromogenic enzymes.

Methods of Diagnosis, Prognosis and Treatment

As explained in the Experimental Results section, the methods of the invention are able to distinguish between samples from individuals with and without cancer. Therefore the term "diagnosing" or "diagnosis" in the context of cancer should be taken as allowing such a

distinction to be made; the term is used to mean both an indication of the presence of cancer and an indication of the initial stages of cancer development. Other physical or biological measurements may be taken, or tests carried out, in conjunction with the measurement of biomarker expression levels as part of the methods of the invention. Preferably the methods of the invention, or at least preferably those that do not involve treatment of the subject, are performed *in vitro* and/or *ex vivo* and/or are not practised on the subject's body. For the avoidance of doubt, it should be noted that the present invention can be used for both initial diagnosis of cancer and for ongoing monitoring of cancer, e.g. indicating the continued presence of cancer despite treatment (response to, or outcome following, treatment) or indicating the presence of cancer after a period of being "cancer free" following treatment (assessing recurrence).

The methods of the invention may be used to diagnose cancer in a subject showing symptoms consistent with such disease. Alternatively, the methods of the invention may be used to diagnose cancer in a subject that appears asymptomatic. Cancer may be asymptomatic, for example, during the early stages of the disease.

The invention also provides methods of prognosing and methods of predicting efficacy of treatment for cancer. Such methods may include methods for predicting the likelihood that a cancer, such as a carcinoma *in situ*, will progress or not, predicting the outcome for the subject in response to therapy or in response to a particular therapy schedule, predicting the likely clinical development of the cancer following therapy or a particular therapy schedule, predicting the response to therapy in a particular sub-class of cancer such as Triple negative or ER positive breast cancer, the likely life expectancy and/or survival of the subject, the likely reduction in symptoms for the subject and/or the likely reduction in the extent of the cancer (including size and number of sites). Preferably, the greater the extent of the differential expression of one or more, or all, of the biomarkers in the method of the invention compared to reference values from subjects not having cancer, the poorer the prognosis for the subject and/or the less likely they are to respond to the cancer treatment.

As used herein the term "cancer" includes: cancer generically; groups or sub-groups of cancers originating from specific organs, tissues and/or cell types; cancer originating from a specific organ, tissue and/or cell type; and cancers of unknown primary origin. For example, a method of the invention may indicate that the subject has cancer, without the site or origin of the cancer being known or indicated, or alternatively a method of the invention may indicate that the subject has a more specific type of cancer, such as breast cancer. Indeed it is a remarkable feature of the present invention that these biomarkers have broad utility for detecting many different types of cancers. The specificity of the cancer diagnosis given may depend, for example, on whether the subject has any symptoms and what those symptoms

are, which may indicate a suspected site for the cancer, and/or further measurements taken or tests carried out in order to indicate a likely origin of the cancer; said measurements or tests may form part of the methods of the invention, or alternatively may be carried out additionally, simultaneously with or separately from the methods of the invention, before or
5 after the methods of the invention. Such measurements or tests that may be part of the methods of the invention, or additional to it, include further blood tests, X-rays, CT scans and endoscopy.

The term "cancer" as used herein may also include pre-cancerous lesions and non-invasive cancers. Therefore it should be understood that methods, kits and assays of the invention
10 may be used as early diagnosis tools and to treat pre-cancerous lesions and non-invasive cancers that could develop into invasive cancer. Examples of such pre-cancerous lesions include carcinoma in situ, such as ductal carcinoma in situ of the breast. Alternatively, some methods, kits and assays of the invention may be considered specific to invasive cancers, and not encompass pre-cancerous lesions.

The cancer detected and/or indicated in the present invention may be breast cancer, endometrial cancer, ovarian cancer, prostate cancer, pancreatic cancer, thyroid cancer, cervical cancer, bladder cancer, blastoma, brain cancer and gliomas, bowel cancer, gastric cancer, head and neck cancer, kidney cancer, liver cancer, lung cancer, mesothelioma, melanoma, oral cancer, pituitary cancer, skin cancer, soft tissue cancer, testicular cancer,
15 20 uterine cancer, heart cancer, and/or eye cancer. Preferably the methods, kits and devices of the invention will be for subjects having, or suspected of having, a solid tumour cancer, for example a carcinoma. Preferably the solid tumour cancer will not be a sarcoma.

Preferably the methods, kits and devices of the invention will not be for subjects having, or suspected of having, a blood cancer, for example preferably the cancer will not be a
25 leukemia, and/or a myeloma. It is particularly preferred that the cancer will not be a myeloid leukemia, for example monocytic leukemia, or a lymphocytic leukemia.

It is preferred in the present invention that the cancer is a hormone-related cancer, for example breast cancer, endometrial cancer, ovarian cancer, prostate cancer, pancreatic cancer, and thyroid cancer. It is particularly preferred that it is an estrogen-dependent
30 cancer, for example a cancer selected from breast cancer, endometrial cancer and ovarian cancer.

The methods of identifying a subject for treatment may further involve providing said treatment to the subject. Alternatively, said methods may not involve any actual treatment of a human or animal body, for example they may be restricted to simply providing a
35 recommendation for treatment. In methods of the invention, for example those methods

involving either a recommendation for treatment, or the actual treatment itself, treatment may involve any of the treatments known in the art for the cancer diagnosed, for example one or more treatments selected from the group consisting of surgery, radiation therapy, chemotherapy, immunotherapy, hormone therapy, and targeted therapy. The therapy may, for example, be used to remove the entire tumour, to debulk the tumour, and/or to ease the cancer symptoms.

Surgery involves removing or destroying tumour tissue and may be open or minimally invasive. It may include, for example, the use of sharp tools to cut the body, cryosurgery, lasers, hyperthermia and/or photodynamic therapy.

Radiation therapy involves the use of high doses of radiation to kill cancer cells and shrink tumours. Treatment using radiation therapy in accordance with the invention includes the use of external beam radiation therapy, where an external source is used to aim radiation at the affected part(s) of the body, internal radiation therapy (brachytherapy), where a solid or liquid radiation source is put into the body, and systemic radiation therapy. Radiation therapies of use in embodiments of the invention include the use of external x-rays or gamma rays, interstitial brachytherapy, intracavitary brachytherapy, episcleral brachytherapy, radioactive iodine, samarium-153-lexidronam (Quadramet) and strontium-89 chloride (Metastron).

Chemotherapy involves the use of chemicals that target the fast dividing cancer cells. It may be used on its own or in combination with other cancer therapies. Chemotherapy drugs of use in embodiments of the invention include one or more of Abraxane (Abraxane), Amsacrine (Amsidine), Azacitidine (Vidaza), Bendamustine, (Levact), Bleomycin, Busulfan (Busilvex, Myleran), Cabazitaxel (Jevtana), Capecitabine (Xeloda), Carboplatin, Carmustine (BiCNU), Chlorambucil (Leukeran), Cisplatin, Cladribine (Leustat, LITAK), Clofarabine (Evoltra), Crisantaspase (Erwinase, asparaginase or L-asparaginase), Cyclophosphamide, Cytarabine, Dacarbazine (DTIC), Dactinomycin (Cosmegen Lyovac), Daunorubicin, Docetaxel (Taxotere), Doxorubicin (Adriamycin), Epirubicin (Pharmorubicin), Eribulin (Halaven), Etoposide (VP-16, Etopophos, Vepesid), Fludarabine (Fludara), Fluorouracil (5FU), Gemcitabine (Gemzar) , Hydroxycarbamide (Hydrea, hydroxyurea), Idarubicin (Zavedos), Ifosfamide (Mitoxana), Irinotecan (Campto), Leucovorin (Folinic acid), Liposomal daunorubicin (DaunoXome), Liposomal doxorubicin (DaunoXome), Melphalan (Alkeran), Mercaptopurine (Puri-Nethol), Mesna (Uromitexan), Methotrexate (Maxtrex), Mitomycin (Mitomycin C Kyowa), Mitotane (Lysodren), Mitoxantrone, Oxaliplatin (Eloxatin), Paclitaxel (Taxol), Pemetrexed (Alimta), Pentostatin (Nipent), Procarbazine, Raltitrexed (Tomudex), Rasburicase (Fasturtec), Streptozocin (Zanosar), Temozolomide (Temodal), Thiotepa,

Tioguanine (Ilanvis), Topotecan (Hycamtin), Trabectedin (Yondelis), Treosulfan, Vinblastine (Velbe), Vincristine (Oncovin), Vindesine (Eldisine), and Vinorelbine (Navelbine).

Immunotherapy includes treatment that help the subject's immune system to target the cancer cells. Immunotherapies of use in embodiments of the invention include monoclonal
 5 antibodies such as those targeting CTLA4 or PD1, adoptive cell transfer which boosts the ability of T cells to fight the cancer, cytokines such as interferons and interleukins, vaccines, and BCG.

Hormone therapy blocks the body's ability to produce hormones, or interferes with how the hormones behave. Hormone therapies of use in embodiments of the invention include
 10 estrogens and anti-estrogens, androgens and anti-androgens, progestins, gonadotropin-releasing hormone (GnRH) analogues and aromatase inhibitors.

Targeted therapy involves selecting drugs that specifically target changes that have occurred during the development of the specific cancer in the subject's body. Examples of targeted
 15 therapies that may be used in embodiments of the invention include small-molecule drugs and monoclonal antibodies. Preferably the targeted therapies in the embodiments of the invention will include one or more from the group consisting of Trastuzumab (Herceptin), ramucirumab (Cyramza), Vismodegib (Erivedge), sonidegib (Odomzo), Atezolizumab (Tecentriq), nivolumab (Opdivo), Bevacizumab (Avastin), Everolimus (Afinitor), tamoxifen (Nolvadex), toremifene (Fareston), fulvestrant (Faslodex), anastrozole (Arimidex),
 20 exemestane (Aromasin), lapatinib (Tykerb), letrozole (Femara), pertuzumab (Perjeta), ado-trastuzumab emtansine (Kadcyla), palbociclib (Ibrance), Cetuximab (Erbix), panitumumab (Vectibix), ziv-aflibercept (Zaltrap), regorafenib (Stivarga), ramucirumab (Cyramza), Lanreotide acetate (Somatuline Depot), pembrolizumab (Keytruda), Imatinib mesylate (Gleevec), sunitinib (Sutent), regorafenib, Denosumab (Xgeva), Alitretinoin (Panretin),
 25 sorafenib (Nexavar), pazopanib (Votrient), temsirolimus (Torisel), axitinib (Inlyta), cabozantinib (Cabometyx), lenvatinib mesylate (Lenvima), crizotinib (Xalkori), erlotinib (Tarceva), gefitinib (Iressa), afatinib dimaleate (Gilotrif), ceritinib (LDK378/Zykadia), osimertinib (Tagrisso), necitumumab (Portrazza), alectinib (Alecensa), Ipilimumab (Yervoy), vemurafenib (Zelboraf), trametinib (Mekinist), dabrafenib (Tafinlar), cobimetinib (Cotellic),
 30 Bortezomib (Velcade), carfilzomib (Kyprolis), panobinostat (Farydak), daratumumab (Darzalex), ixazomib citrate (Ninlaro), elotuzumab (Empliciti), Dinutuximab (Unituxin), olaparib (Lynparza), rucaparib camsylate (Rubraca), enzalutamide (Xtandi), abiraterone acetate (Zytiga), radium 223 dichloride (Xofigo), Cabozantinib (Cometriq), and vandetanib (Caprelsa).

In some embodiments of the methods of treatment provided herein, the cancer is breast cancer and the treatment is one or more selected from the group consisting of surgery, radiation therapy, chemotherapy, hormonal therapy, immunotherapy, and targeted therapy. Preferably the chemotherapy involves treatment with one or more drugs selected from the group consisting of Capecitabine (Xeloda), Carboplatin (Paraplatin), Cisplatin (Platinol), Cyclophosphamide (Neosar), Docetaxel (Docefrez, Taxotere), Doxorubicin (Adriamycin), Pegylated liposomal doxorubicin (Doxil), Epirubicin (Ellence), Fluorouracil (5-FU, Adrucil), Gemcitabine (Gemzar), Methotrexate (multiple brand names), Paclitaxel (Taxol), Protein-bound paclitaxel (Abraxane), Vinorelbine (Navelbine), Eribulin (Halaven), mitoxantrone (Mitozantrone or Onkotrone), mitomycin C, Ixabepilone (Ixempra) and megestrol (Megace). Preferably the hormonal therapy involves treatment with one or more treatments selected from the group consisting of Tamoxifen, aromatase inhibitors (AIs) such as Anastrozole (Arimidex) and Exemestane (Aromasin), Letrozole (Femara), Fulvestrant (Faslodex), ovarian suppression or ablation such as using goserelin (Zoladex), megestrol acetate (Megace) and high-dose estradiol. Preferably the targeted therapy and/or immunotherapy involves treatment with one or more selected from the group consisting of palbociclib (Ibrance), Everolimus (Afinitor), Trastuzumab, Pertuzumab (Perjeta), Ado-trastuzumab emtansine or T-DM1(Kadcyla), Lapatinib (Tykerb), Bisphosphonates, and Denosumab (Xgeva).

In some embodiments of the methods of treatment provided herein, the cancer is endometrial cancer and the treatment is one or more selected from the group consisting of surgery, radiotherapy, chemotherapy and hormone therapy. Preferably the surgery includes a hysterectomy. Preferably the radiotherapy comprises brachytherapy and/or external radiotherapy. Preferably the chemotherapy involves treatment with one or more drugs selected from the group consisting of Carboplatin (Paraplatin), Cisplatin (Platinol), Cyclophosphamide (Neosar), Doxorubicin (Adriamycin), Paclitaxel (Taxol), Protein-bound paclitaxel (Abraxane). Preferably the hormonal therapy involves treatment with one or more treatments selected from the group consisting of progesterone such as medroxyprogesterone acetate (Provera) and megestrol (Megace), Tamoxifen, and Letrozole (Femara).

Binding Partners

In certain embodiments of the invention, expression levels of the biomarkers in a biological sample may be investigated using binding partners which bind or hybridize specifically to a target molecule for the biomarkers, or a fragment thereof. In relation to the present invention the term 'binding partners' may include any ligands, which are capable of binding specifically to the relevant biomarker and/or nucleotide or peptide variants thereof with high affinity.

Said ligands include, but are not limited to nucleic acids (DNA or RNA), proteins, peptides, antibodies, synthetic affinity probes, carbohydrates, lipids, artificial molecules or small organic molecules such as drugs. In certain embodiments the binding partners may be selected from the group comprising: complementary nucleic acids; aptamers; antibodies or antibody fragments. In the case of detecting mRNAs and ncRNAs, nucleic acids represent highly suitable binding partners.

In the context of the present invention, a binding partner specific to a biomarker should be taken as requiring that the binding partner should be capable of binding to at least one target molecule for such biomarker in a manner that can be distinguished from non-specific binding to molecules that are not target molecules for biomarkers. A suitable distinction may, for example, be based on distinguishable differences in the magnitude of such binding.

In preferred embodiments of the methods or devices of the invention, the target molecule for the biomarker is a nucleic acid, preferably an mRNA or ncRNA molecule, and the binding partner is selected from the group consisting of complementary nucleic acids and aptamers.

Suitably the binding partner is a nucleic acid molecule (typically DNA, but it can be RNA) having a sequence which is complementary to the sequence of the relevant mRNA, ncRNA or cDNA against which is targeted. Such a nucleic acid is often referred to as a 'probe' (or a reporter or an oligo) and the complementary sequence to which it binds is often referred to as the 'target'. Probe-target hybridization is usually detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labeled targets to determine relative abundance of nucleic acid sequences in the target.

Probes can be from 25 to 1000 nucleotides in length. However, lengths of 30 to 100 nucleotides are preferred, and probes of around 50 nucleotides in length are commonly used with great success in complete transcriptome analysis.

While the determination of suitable probes can be difficult, e.g. in very complex arrays, there are many commercial sources of complete transcriptome arrays available, and it is routine to develop bespoke arrays to detect any given set of specific mRNAs using publically available sequence information. Commercial sources of microarrays for transcriptome analysis include Illumina and Affymetrix.

HGNC Symbol	Probe Sequence (Illumina Human HT 12 V4 probes)	Probe Name	SEQ ID NO.	Gene ID
CRYBG3	GCCCAGGGTGAAATATCTGACAGTCACTGGAAGTCTAGCAGACACACAGGGC	ILMN_2190684	1	ENSG00000080200
PIBF1	GTCCACTACGAGGTACTTCAAAGCCAGTAATGGTGTCAGATACCATG	ILMN_2227473	2	ENSG00000083535
PIBF1	CCAGACTCAACAGCCTTACAGGTATCTCATTTGAATCAGTGCCTCAGAGAG	ILMN_1758111	3	ENSG00000083535
SCD	TAGGAATGTCCACCATGAAC TTGATACGTCCGTGTGCCAGATGCTGTC	ILMN_1689329	4	ENSG00000099194
PPIF	CTGAA TCCCAGCTCCTCCCTTAGTGACCCCCAAGTCTGTTCCCTCAGCT	ILMN_1809607	5	ENSG000000108179
PTP4A1	TTACGTTTGGGTAACCAAGCCTAAGTGGAATCTTGGAAAGGAAAGTAAGGG	ILMN_2154654	6	ENSG000000112245
PTP4A1	CTTGGGTTTGGATACCCCTTAGTGGGATGATGTAAATAGAGGCTAGCTACC	ILMN_1760575	7	ENSG000000112245
SLF1	GGGTACACACACTGAGGCCCTTGATGATAACAATTGGAATGATGTGTCGGTC	ILMN_1805985	8	ENSG000000133302
SLF1	TGCGTACCTATAAACTTGTAGCTCCTGACTCTTAGGGATGATTTTGAGG	ILMN_2214278	9	ENSG000000133302
CEPT1	GAGCCAGATCCGCAGTGGTGGAGAGTTCTAATGTTGACTGTTTGCAGGCC	ILMN_1676588	10	ENSG000000134255
CEPT1	GCAC TGCCTGGGGCAGGAGGAGAGTAGGGGAAACCAAGGAAGAACAGACA	ILMN_1674405	11	ENSG000000134255
IFI16	CCTTTCACACTTGTGGCTGATGTGAATGCTGACCGAAACATGGAGATCCC	ILMN_1710937	12	ENSG000000163565
MCTP1	GGCCATTGTAGCCCTCTGTGTGTTACAGGCCATCCTGTACTGCATTCCGC	ILMN_1654685	13	ENSG000000175471
MCTP1	AGAAGAGCTCCTCTCTGCCGGGCACCGCCTGCCCTGGAGCAGCTGCTGGAG	ILMN_1738097	14	ENSG000000175471
ZNF114	GGGAGCTCAAGGGTTAGCATGAGTGAGAACATCTTCCCTGAACTCTCGT	ILMN_1746831	15	ENSG000000178150
ZNF114	ACGGAAATAAAAAATGTAGGAAAGACCTGCCGGCCGCGGTGGCTCATGCC	ILMN_2164290	16	ENSG000000178150
TMTC2	GCATTAGCCTGCATGTTTCTAGGCTTCTCTGCCAAGTGCTTTTCTCAGTTG	ILMN_1805645	17	ENSG000000179104
NRIP1	GTGATTACCTGCTGCATGAAAAGTGATGGGGACCCCTGTGCATCTGTGC	ILMN_1718629	18	ENSG000000180530

Table 2: Probe sequences and accession numbers for biomarkers differentially expressed in monocytes.

In one embodiment the probe sequences will comprise sequences selected from those listed in Table 2 [SEQ ID NOs 1 to 18]. However, nucleotide probe sequences may be designed to any sequence region of the biomarker transcripts (accession numbers listed in Table 2) or a variant thereof. Nucleotide probe sequences, for example, may include, but are not limited to those listed in Table 2. The person skilled in the art will appreciate that equally effective probes can be designed to different regions of the transcript than those targeted by the probes listed in Table 2, and that the effectiveness of the particular probes chosen will vary, amongst other things, according to the platform used to measure transcript abundance and the hybridization conditions employed. It will therefore be appreciated that probes targeting different regions of the transcript may also be used in accordance with the present invention.

In other suitable embodiments of the invention, the target molecule for the biomarker may be a protein, and the binding partner is selected from the group consisting of antibodies, antibody fragments and aptamers.

Polynucleotides encoding any of the specific binding partners of target molecules for biomarkers of the invention recited above may be isolated and/or purified nucleic acid molecules and may be RNA or DNA molecules.

Throughout, the term "polynucleotide" as used herein refers to a deoxyribonucleotide or ribonucleotide polymer in single- or double-stranded form, or sense or anti-sense, and encompasses analogues of naturally occurring nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides. Such polynucleotides may be derived from Homo sapiens, or may be synthetic or may be derived from any other organism.

Commonly, polypeptide sequences and polynucleotides used as binding partners in the present invention may be isolated or purified. By "purified" is meant that they are substantially free from other cellular components or material, or culture medium. "Isolated" means that they may also be free of naturally occurring sequences which flank the native sequence, for example in the case of nucleic acid molecule, isolated may mean that it is free of 5' and 3' regulatory sequences.

In a preferred embodiment the nucleic acid is mRNA or ncRNA. There are numerous suitable techniques known in the art for the quantitative measurement of RNA transcript levels in a given biological sample. These techniques include but are not limited to; "Northern" RNA blotting, Real Time Polymerase Chain Reaction (RT-PCR), Quantitative Polymerase Chain Reaction (qPCR), digital PCR (dPCR), multiplex PCR, Reverse Transcription Quantitative Polymerase Chain Reaction (RT-qPCR), branched DNA signal amplification or by high-throughput analysis such as hybridization microarray, Next Generation Sequencing (NGS) or by direct mRNA quantification, for example by "Nanopore"

sequencing. Alternatively, "tag based" technologies may be used, which include but are not limited to Serial Analysis of Gene Expression (SAGE). Suitable techniques also include nCounter™ systems of NanoString technologies™, zip coding, and targeted hybridization and sequencing. Commonly, the levels of biomarker mRNA transcript in a given biological sample may be determined by hybridization to specific complementary nucleotide probes on a hybridization microarray or "chip", by Bead Array Microarray technology or by RNA-Seq where sequence data is matched to a reference genome or reference sequences.

In a preferred embodiment, where the nucleic acid is RNA, the present invention provides methods wherein the levels of biomarker transcript(s) will be determined by PCR. Preferably mRNA and ncRNA transcript abundance will be determined by qPCR, dPCR or multiplex PCR. More preferably, transcript abundance will be determined by multiplex-PCR. Nucleotide primer sequences may be designed to any sequence region of the biomarker transcripts (accession numbers listed in Table 2) or a variant thereof. The person skilled in the art will appreciate that equally effective primers can be designed to different regions of the transcript or cDNA of biomarkers listed in Table 2, and that the effectiveness of the particular primers chosen will vary, amongst other things, according to the platform used to measure transcript abundance, the biological sample and the hybridization conditions employed. It will therefore be appreciated that primers targeting different regions of the transcript may also be used in accordance with the present invention. However, the person skilled in the art will recognise that in designing appropriate primer sequences to detect biomarker expression, it is required that the primer sequences be capable of binding selectively and specifically to the cDNA sequences of biomarkers corresponding to the nucleotide accession numbers listed in Table 2 or fragments or variants thereof.

Many different techniques known in the art are suitable for detecting binding of the target molecule sequence and for high-throughput screening and analysis of protein interactions. According to the present invention, appropriate techniques include (either independently or in combination), but are not limited to; co-immunoprecipitation, bimolecular fluorescence complementation (BiFC), dual expression recombinase based (DERB) single vector system, affinity electrophoresis, pull-down assays, label transfer, yeast two-hybrid screens, phage display, in vivo crosslinking, tandem affinity purification (TAP), ChIP assays, chemical cross-linking followed by high mass MALDI mass spectrometry, strep-protein interaction experiment (SPINE), quantitative immunoprecipitation combined with knock-down (QUICK), proximity ligation assay (PLA), bio-layer interferometry, dual polarisation interferometry (DPI), static light scattering (SLS), dynamic light scattering (DLS), surface plasmon resonance (SPR), fluorescence correlation spectroscopy, fluorescence resonance energy transfer (FRET), isothermal titration calorimetry (ITC), microscale thermophoresis (MST),

chromatin immunoprecipitation assay, electrophoretic mobility shift assay, pull-down assay, microplate capture and detection assay, reporter assay, RNase protection assay, FISH/ISH co-localization, microarrays, microsphere arrays or silicon nanowire (SiNW)-based detection. Where biomarker protein levels are to be quantified, preferably the interactions between the binding partner and biomarker protein will be analysed using antibodies with a fluorescent reporter attached.

In certain embodiments of the invention, the expression level of a particular biomarker may be detected by direct assessment of binding of the target molecule to its binding partner. Suitable examples of such methods in accordance with this embodiment of the invention may utilise techniques such as electro-impedance spectroscopy (EIS) to directly assess binding of binding partners (e.g. antibodies) to target molecules (e.g. biomarker proteins).

In certain embodiments of the present invention the binding partner may be an antibody, or antibody fragment, and the detection of the target molecules utilises an immunological method. In certain embodiments of the methods or devices, the immunological method may be an enzyme-linked immunosorbent assay (ELISA) or utilise a lateral flow device.

A method of the invention may further comprise quantification of the amount of the target molecules indicative of expression of the biomarkers that is present in the patient sample. Suitable methods of the invention, in which the amount of the target molecule present has been quantified, and the volume of the patient sample is known, may further comprise determination of the concentration of the target molecules present in the patient sample which may be used as the basis of a qualitative assessment of the patient's condition, which may, in turn, be used to suggest a suitable course of treatment for the patient.

Reporter moieties

In preferred embodiments of the present invention the expression levels of the protein in a biological sample may be determined. In some instances, it may be possible to directly determine expression, e.g. as with GFP or by enzymatic action of the protein of interest (POI) to generate a detectable optical signal. However, in some instances it may be chosen to determine physical expression, e.g. by antibody probing, and rely on separate test to verify that physical expression is accompanied by the required function.

In preferred embodiments of the invention, the expression levels of a particular biomarker will be detectable in a biological sample by a high-throughput screening method, for example, relying on detection of an optical signal, for instance using reporter moieties. For this purpose, it may be necessary for the specific binding partner to incorporate a tag, or be labelled with a removable tag, which permits detection of expression. Such a tag may be,

for example, a fluorescence reporter molecule translationally-fused to the protein of interest (POI), e.g. Green Fluorescent Protein (GFP), Yellow Fluorescent Protein (YFP), Red Fluorescent Protein (RFP), Cyan Fluorescent Protein (CFP) or mCherry. Such a tag may provide a suitable marker for visualisation of biomarker expression since its expression can be simply and directly assayed by fluorescence measurement in vitro or on an array. Alternatively, it may be an enzyme which can be used to generate an optical signal. Tags used for detection of expression may also be antigen peptide tags. Similarly, reporter moieties may be selected from the group consisting of fluorophores; chromogenic substrates; and chromogenic enzymes. Other kinds of label may be used to mark a nucleic acid binding partner including organic dye molecules, radiolabels and spin labels which may be small molecules.

Preferably, the levels of a biomarker or several biomarkers will be quantified by measuring the specific hybridization of a complementary nucleotide probe to the target molecule for the biomarker of interest under high-stringency or very high-stringency conditions.

Preferably, probe-target molecule hybridization will be detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labelled probes to determine relative abundance of biomarker nucleic acid sequences in the sample. Alternatively, levels of biomarker mRNA or ncRNA transcript abundance can be determined directly by RNA sequencing or nanopore sequencing technologies.

The methods or devices of the invention may make use of target molecules selected from the group consisting of: the biomarker protein; and nucleic acid encoding the biomarker protein.

Nucleotides and Hybridization Conditions

Throughout, the term "polynucleotide" as used herein refers to a deoxyribonucleotide or ribonucleotide polymer in single- or double-stranded form, or sense or anti-sense, and encompasses analogues of naturally occurring nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides.

Exemplary probe sequences are provided in Table 2, although it will be appreciated that minor variations in these sequences may work. The person skilled in the art would regard it as routine to design nucleotide probe sequences may be designed to any sequence region of the biomarker transcripts (accession numbers listed in Table 2) or a variant thereof. This is also the case with nucleotide primers used where detection of expression levels is determined by PCR-based technology. Nucleotide probe sequences, for example, may include, but are not limited to those listed in Table 2. The person skilled in the art will

appreciate that equally effective (and in some cases more beneficial) probes can be designed to different regions of the transcript than those targeted by the probes listed in Table 2, and that the effectiveness of the particular probes chosen will vary, amongst other things, according to the platform used to measure transcript abundance and the hybridization conditions employed. It will therefore be appreciated that probes targeting different regions of the transcript may also be used in accordance with the present invention.

Of course the person skilled in the art will recognise that in designing appropriate probe sequences to detect biomarker expression, it is required that the probe sequences be capable of binding selectively and specifically to the transcripts or cDNA sequences of biomarkers corresponding to the nucleotide accession numbers listed in Table 2 or fragments or variants thereof. The probe sequence will therefore be hybridizable to that nucleotide sequence, preferably under stringent conditions, more preferably very high stringency conditions. The term "stringent conditions" may be understood to describe a set of conditions for hybridization and washing and a variety of stringent hybridization conditions will be familiar to the skilled reader. Hybridization of a nucleic acid molecule occurs when two complementary nucleic acid molecules undergo an amount of hydrogen bonding to each other known as Watson-Crick base pairing. The stringency of hybridization can vary according to the environmental (i.e. chemical/physical/biological) conditions surrounding the nucleic acids, temperature, the nature of the hybridization method, and the composition and length of the nucleic acid molecules used. Calculations regarding hybridization conditions required for attaining particular degrees of stringency are discussed in Sambrook et al. (2001, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY); and Tijssen (1993, *Laboratory Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes Part I*, Chapter 2, Elsevier, NY). The T_m is the temperature at which 50% of a given strand of a nucleic acid molecule is hybridized to its complementary strand.

In any of the references herein to hybridization conditions, the following are exemplary and not limiting:

Very High Stringency (allows sequences that share at least 90% identity to hybridize)

Hybridization: 5x SSC at 65°C for 16 hours

Wash twice: 2x SSC at room temperature (RT) for 15 minutes each

Wash twice: 0.5x SSC at 65°C for 20 minutes each

High Stringency (allows sequences that share at least 80% identity to hybridize)

Hybridization: 5x-6x SSC at 65°C-70°C for 16-20 hours

Wash twice: 2x SSC at RT for 5-20 minutes each

Wash twice: 1x SSC at 55°C-70°C for 30 minutes each

5

Low Stringency (allows sequences that share at least 50% identity to hybridize)

Hybridization: 6x SSC at RT to 55°C for 16-20 hours

Wash at least twice: 2x-3x SSC at RT to 55°C for 20-30 minutes each.

Diagnostic Devices and Kits

- 10 The invention also provides an assay device for use in the above methods, the device comprising: a) a loading area for receipt of a biological sample; b) binding partners specific for target molecules representative of expression of at least four biomarkers selected from the group consisting of *MCTP1*, *PIBF1*, *TMTC2*, *ANKRD32*, *CEPT1*, *ZNF114*, *CRYBG3*, *IFI16*, *PPIF*, *SCD*, *RP11-469M7.1*, *PTP4A1* and *NRIP1*; and c) detection means to detect
- 15 the levels of said target molecules present in the sample.

Suitably the device comprises specific binding partners to the target molecules of the biomarkers being amplified. A variety of suitable PCR amplification-based technologies are well known in the art.

- The binding partners are preferably nucleic acid primers adapted to bind specifically to the
- 20 mRNA, ncRNA, or cDNA transcripts of the biomarkers, as discussed above.

The detection means suitably comprises means to detect a signal from a reporter moiety, e.g. a reporter moiety as discussed above.

The device is adapted to detect and quantify the levels of said biomarkers present in the biological sample.

- 25 The invention provides kits for use in the above methods, the kits comprising binding partners capable of binding to target molecules representative of expression of at least four biomarkers selected from the group consisting of *MCTP1*, *PIBF1*, *TMTC2*, *ANKRD32*, *CEPT1*, *ZNF114*, *CRYBG3*, *IFI16*, *PPIF*, *SCD*, *RP11-469M7.1*, *PTP4A1* and *NRIP1*. Preferably the kits further comprise indicators capable of indicating when said binding
- 30 occurs.

Preferably the kits and devices comprise binding partners capable of binding to target molecules representative of expression of four, five, six, seven, eight, nine, ten, eleven, twelve or all of the biomarkers.

PCR applications are routine in the art and the skilled person will be able to select appropriate polymerases, buffers, reporter moieties and reaction conditions.

The binding partners are preferably nucleic acid primers adapted to bind specifically to the mRNA, ncRNA, or cDNA transcripts of biomarkers, as discussed above. The nucleic acid primers may be provided in a lyophilized or reconstituted form, or may be provided as a set of nucleotide sequences. In one embodiment, the primers are provided in a microplate format, where each primer set occupies a well (or multiple wells, as in the case of replicates) in the microplate. The microplate may further comprise primers sufficient for the detection of one or more housekeeping genes as a positive control. The kit may further comprise reagents and instructions sufficient for the amplification of expression products from the biomarkers.

As well as the binding partners for the target molecules of the biomarkers, the devices and kits may further comprise binding partners capable of binding to target molecules representative of expression of additional genes. For example, such genes may be “housekeeping genes”, which can act as a positive control and/or to normalize expression across samples, and/or such genes may give an indication of the concentration of the monocyte population within the biological sample.

Preferably said devices and kits provide binding partners capable of binding to target molecules representative of expression of less than 50, 40, 30, 20, 15, or 10 genes, including the biomarkers and any housekeeping or other control genes.

BRIEF DESCRIPTION OF THE FIGURES

The invention will now be described in detail with reference to the accompanying experimental results and figures, in which:

Figure 1 shows that monocytes sub-populations and transcriptional profiles are influenced by the presence of cancer:

A) Relative distribution of non-classical monocytes (CD14+CD16++) in healthy controls (Mo) and breast and endometrial cancer patients (TEMo) determined by flow cytometry; Left part, Cohort 1 (Mo, dots; N = 31, TEMo breast cancer, full squares, N= 22; TEMo endometrial cancer, empty triangles, N=12); right part, Cohort 2 Breast cancer and controls only (Mo, empty dots; N = 18, TEMo, empty squares; N= 33); ***= p<0.0001; (Mean ± SEM)

B) Multidimensional Scaling Plot (MDS) performed on Mo samples from healthy patients versus TEMo from cancer patients. Dots corresponds to the number of healthy subjects (Mo, dots, N = 22) or patients (TEMo breast cancer, full squares; TEMo endometrial cancer, triangles N = 37);

5 C) Hierarchical clustering of top 100 most significant genes between Mo (black) and TEMo (grey). Samples are clustered using complete linkage and Pearson correlation. Light grey colour in the Heatmap indicates upregulation, and dark grey colour indicates down regulation.

10 D) Venn diagram of commonly upregulated and downregulated between Breast TEMo vs Mo (left circle), endometrial TEMo vs Mo (right circle), Breast TEMo vs Endometrial TEMo (bottom circle). Numbers on top of each pair indicate upregulated genes, numbers on the bottom of each pair indicate downmodulated genes. Total number of genes is provided in brackets.

15 E) Functional analysis of DEGs in TEMo compared to Mo using Ingenuity Pathway Analysis (IPA). List of top disease predicted to be involved in TEMo ranked by the negative log of the *P* value of the enrichment score.

F) Bar plot of selected genes whose expression is differentially regulated in TEMo ($\text{Log}_2\text{FC} > 1.5$, $\text{FDR} < 0.05$) shaded and sorted (in the order shown in the key) by type.

20 G) Real-time RT-PCR validation of RNA-seq results. The expression of CX3CR1, FcGR3A (CD16), CD200R1, TNFSF10, HGF, ANGPT1 and ANGPT2 mRNA was quantified by SYBR Green real time RT-PCR in Mo and TEMo. The expression of each gene was normalized to GAPDH using the $\Delta\Delta\text{Ct}$ method; depicted are results obtained from 3-5 patients independent from the RNA seq cohort (Mean $\pm$ SD of triplicate wells)

25 **Figure 2** shows the gating strategy for human monocytes in healthy control women and female breast and endometrial cancer patients. A-E; monocyte gating strategy based on physical (A,B,C) and fluorescence parameters (D-F). G-I: classical (bottom gate) and non-classical monocytes (upper gate) separation in healthy controls (G) and cancer patients (I).

30 **Figure 3** shows the sorting strategy for the isolation of human monocytes in normal and cancer patients. A-E; monocyte gating strategy based on physical (A,B,C) and fluorescence parameters (D-F). Validation of gating strategy was performed by backgating (G) and nuclei coloration (Giemsa, staining, H).

Figure 4 shows additional results from the transcriptome and flow cytometry analysis on TEMo.

A) Volcano plot showing the gene expression profiles of monocytes from cancer compared to healthy samples. Log₂ Fold change is plotted against the x-axis and the p-value for statistical difference on the y-axis. Points shown in dark grey are genes that are not significantly differentially expressed. Points shown in light grey whose values are less than 0 are genes significantly decreased in monocytes from cancer compared to healthy patients. Points shown in light grey whose values are higher than 0 are genes significantly increased in monocytes from cancer compared to healthy patients.

B) Top canonical pathways predicted to be involved in TEMo using IPA pathways, ranked by the negative log of the *P* value of the enrichment score.

C) Gene Set Enrichment Pathway (GSEA) analysis of TEMo; light grey bars represent downregulated pathways, dark grey bars upregulated pathways; values are expressed as Normalized Enrichment Score (NES).

Figure 5 shows that breast cancer patients show altered levels of chemokines in blood and an altered monocytic phenotype.

A) Quantification of CX3CL1 and CCL2 in the sera of control (CTR, N=15) and Breast cancer patients (Cancer, N = 45) by ELISA and Legendplex Array, ***= $p < 0.0001$ (Mean $\pm$ SEM);

B) The expression of CX3CR1, CCR2, CD86, CD95, CD64, CD16 and CD80 was assessed by flow cytometry on classical and non-classical monocyte subpopulations as described in Fig 2. The expression of all markers was compared with their relative Fluorescent Minus One (FMO) control and the net GEO mean was calculated. Shown is the ratio of Geometric Mean of Non-Classical vs Classical monocytes (left panel) and GEO mean (central and right panel). Mo; N= 10, TEMo; N = 31. *** = $p < 0.0001$, ** = $p < 0.001$, * $p < 0.01$ (Mean $\pm$ SEM).

C) Intracellular flow cytometric staining for IL-1 β ; values are expressed as Net Geo mean (Geo mean stain – Geo mean Isotype control) before (black circles = untreated samples) and after treatment (black squares = LPS treated samples). Results from both classical and non-classical populations are showed for Mo and TEMo. N=4 * = $p < 0.01$.

D) Amplitude of response to LPS in Mo and TEMo. Values are expressed as Net Geo mean (Geo mean LPS treated – Geo mean untreated) in Mo (Black bar) and TEMo (Grey bar). N=4 * = $p < 0.01$.

Figure 6 shows performance of the derived 13-gene signature and random classifiers validation.

A) Dot plot of the performance of top 20 genes as ranked by the X^2 statistic. X-axis shows the number of genes used and y-axis their performance. 13-genes gave the best performance.

B) Background distribution based on the performance of the random signatures. Signatures were tested using a leave-one-out cross validation (LOOCV). Solid vertical line represents the performance of our 13-gene signature, dotted vertical line represents the performance of the mode classifier.

C) Background distribution based on classification performance of the 13-gene signature on permuted data (samples).

10 **Figure 7** shows the development and validation of a monocyte-derived signature for cancer diagnosis.

A) Hierarchical clustering of the TEMo signature comprised of 13 genes, between Mo (black) and TEMo (grey) in the internal validation set. Samples are clustered using complete linkage and Pearson correlation;

15 B) ROC curves of RF- X^2 classifier in the training and independent validation set;

C) Classification results matrix of RF- X^2 classifier in the internal set using 5-times 10-fold cross validation (n=59) and in the independent validation set (19 samples);

D) Classification results matrix of RF- X^2 classifier in the negative validation set from patients with Lyme's disease (n = 41) and chronic periodontitis (n = 10) and their controls.

20 **Figure 8** shows the validation of a monocyte-derived 8 gene signature for cancer diagnosis.

A) ROC curves of RF- X^2 classifier for the 13-genes and 8-genes; cross validation in the internal validation dataset;

B) Classification results matrix of RF- X^2 classifier in the internal validation dataset for the 13-gene signature and 8-gene signature.

25 **Figure 9** shows the validation of a monocyte-derived 5 gene signature for cancer diagnosis.

A) ROC curves of RF- X^2 classifier for the 13-genes and 5-genes; cross validation in the internal validation dataset;

B) Classification results matrix of RF- X^2 classifier in the internal validation dataset for the 13-gene signature and 5-gene signature.

30 **Figure 10** shows that the TEMo gene predictor can detect the presence of early stage breast cancer.

(A) MDS plot on all breast cancer samples (internal and independent) using the 13-gene predictor highlighting the DCIS samples in green.

(B) Volcano plot of differentially expressed genes between normal samples compared to samples coming from DCIS patients highlighting the number of genes up-regulated and down-regulated.

Below is the polynucleotide sequence of RP11-469M7.1, one of the biomarkers in accordance with the invention.

[SEQ ID NO: 19] Homo sapiens long non-coding RNA RP11-469M7.1 gene, lincRNA.

```

10      1 tttttttttt ttttttttga gacggagtct cgctctgtca ccaggctgga gtgcagtggc
      61 ggcgactttg ctcactgcaa gttccgcctc ccgggttcac gccattctcc tgcctcagcc
     121 tcccagagtag ctgggactac aggcgcccgc caccatgccc gactaatttt tttgtatttt
     181 tagtagagac gaggtttcac cggattagcg aggatggtct caatctcctg accttgtgat
     241 ccaccgcct cagcctccca aagtgtctggg attatgggct tgagccaccg cgcccagcta
15      301 gacttttttc taatagagtc ccccataaat tttattacca tagatacgct gtgtatgtac
     361 tgtttctgtg ctttgtacat gaaaagagta aaatgttttt gtttttgttt tggagacagt
     421 ctcactctgt tacctgggct ggagtgcaat ggcatgatct cagctcactg caatctctgc
     481 ctctctgggt caagcgattc tcttgcttca gcctcctgag tagctgggat tacaggcggc
     541 caccaccacc cctggctaata ttttgtattt ttagtagagg tttcaccacg ttgatcaggc
20      601 tgggtctcaa ctctgacct cgtggtctgc ctgccttggc ctcccaaagt gctgggatta
     661 caggcgtgag ccaccatgcc caaccaattt ttttttttta aatccttggg agtgattgac
     721 ccccatgag aatgcatgct ctaatatattt ttaaaaggga agaagaccta gtgatagacc
     781 tttgcattgg caggcatgtc tttatgttta tacagtgaag tagtattttt tctgtcaggg
     841 tcaatagaaa gtagacgaaa ttcattttca gccaatcttt caccagtgtg gctttgttcc
25      901 ttacccacc acaggcaaca gacaaagcat ttctgtctgt ctttcaaggt cttaacagat
     961 tctttcttga ttataaatgt attactaagt agaatcattt gggttctttc tataaaaatc
    1021 agtgaanaat ctctagatga atgaattaaa gttgtaggca taacactgat aaacctctgc
    1081 tctcactactg aagtaggctg cttgcaggaa ctgacaaacta ttggttggct ttaaagttaa
    1141 tgtagatgcc aaagttttag tgtacagtgt tacttaaatt accaaattac ctttgtacaa
30      1201 atattcctca gatgctgtct acagggtgcca tataaaatag gttgataagt atttgcaaat
     1261 tctggtaaat tgccttgcta tgattttttca tgctcagtat tagtctctca gtatcaatac
     1321 gaagtttttc tattccccag cctactaagg cctacagggtt aaaaaccag gaataaaaag
     1381 gttttcagtg gacttgattt aagtggaaat tggatgatgt gacaagtacc tttgtctttg
     1441 gagtaatgag tttataattg gcttttggct aaaaactctt aagtggatta atatctgaac
35      1501 aatcttagac ataattatat gaaactggat cagaaaggct ttctgactcc tttctcttca
     1561 cttttttctc ccagcttaac tagaggcaat taacagtcat ttcaaacta attattttta
     1621 ttttccagtg tacttatcat cctttcatct ttatgctagg agtatagtct cagttcttaa
     1681 tgagttgatg ggatgaagat atagatagat agatagatag atagatattt
     1741 tttttttttt tttttttttt tttttttttt gagacagagt ctgctctgtg cgcccaggcg
40      1801 ggagtgcagt ggcacaatct cggctcactg caagctccgc cttggggggt cacaccattc
     1861 tcctgtctca gcctcctagc tgggattaca ggcgccacc aaccaagccc ggctaaggta
     1921 tttttaaatg tactcaatca ctgtcatttt agtaaccttg aatcagaacc tcagttgtta
     1981 aaggcttaac tgcttgtggc aactactata gtgctctagt ggggacagtg gttaaaatct
     2041 gcagaatccc ttgctttgtc atcttatcca gaaggactga tagcttggtt aactgtggta
45      2101 tgattagaat gttggtgggt gtgtagtgca tttgggggtg tacaacgtta atcaaatata
     2161 attgagctgt tctgtttcca gtcagtgtaa ccatttaaaa ataccttggc aaaacaagct
     2221 aggagctagt tcatgctaata attcttaagg acaaaaatag ttacagctt tttttttttt
     2281 aatctgaaat cctgaaggct atgcttaaaa gttagagata aacctgttgg aaataagtga
     2341 ttcattttata gactgaagcc tctatgactt caaaaagata ctcaacagtc tctggcattt
50      2401 gaagaacaaa atattttctc tgtaaataca cctcatttcc attctagtta ggagcaatgg
     2461 cgcccaggac ggcacacaga atggagaaaa ctggatagct ggtaactcat ttagctcttg
     2521 gcactctaaa aaacctcaaa tacagccatc caagccagta attctattgc tgcgttattt

```

```

2581 ctgtgttttaa ctgtgaaact tgcttcttgt ctgtaccctt gaaatggaat aaaatttcac
2641 gagactcctt gttaatgtag agaaaa

```

EXPERIMENTAL RESULTS

5 1. Materials and Methods

Patient and control samples

All study protocols were approved by the IRB of the Albert Einstein Medical College (Bronx, NY, USA) and by The University of Edinburgh (Edinburgh, UK) ethics committees as appropriate. Informed consent was obtained from all human subjects included in this study.

10 For control samples mononuclear cells were isolated from peripheral blood obtained from female healthy individuals through the New York blood center, USA or Cambridge Biosciences, UK. In some case blood was also donated from volunteers in the Bronx, NY who were age and weight matched to the Bronx cancer cohort. Peripheral blood (20ml) from breast and endometrial cancer patients attending the Montefiore Medical Center, Bronx, NY, USA and breast cancer patients from NHS, Edinburgh, Scotland. Breast cancer tissue (0.1-1 grams) and endometrial cancer tissue (0.1-1 grams) was obtained from Montefiore Medical Center, NY, USA and from NHS, Edinburgh, Scotland (breast only). Pathologically the breast cancer patients consisted of invasive breast cancers. The endometrial cancer patients consisted in Type I (endocarcinoma) and Type II (UPSC) cancers (see Table 3 for detailed clinical information). Normal Breast tissue from mammoplasty reduction surgeries (25-50 grams) was obtained from the Human Tissue Procurement Facility (HTPF), Ohio State University, USA; normal/benign endometrial tissue (1-2 grams) was obtained after surgery for conditions unrelated to cancer from Montefiore Medical Center, NY, USA; breast tissue (0.5-1 grams) from patients with benign conditions was obtained from NHS, Edinburgh, Scotland. The exclusion criteria for cancer patients at baseline included systemic metastatic disease, any inflammatory disorder, active infection or immunocompromised status not related to cancer. All the patients recruited were chemotherapy and radiotherapy naive before collection. Blood collection was performed before biopsy or surgery.

Isolation of human blood monocytes

30 All control and cancer blood samples were collected and processed by the same person according to site, HZ in the Bronx and LC Edinburgh. They were processed as attained and not batched together according to sample type. All the blood samples were collected in Venous Blood Collection Tubes containing EDTA and stored immediately at 4C after

collection. Blood was centrifuged at 700 RCF for 10 min at 4°C in a swinging bucket rotor to separate cells from plasma, that was then ultracentrifuged in conical tubes for 10 min at 16,000 x g at 4°C in a fixed angled rotor, immediately aliquoted and stored at -80°C. After red blood cell lysis (RBC lysis buffer, Biolegend) cells were centrifuged 500 RCF 5 min at 4°C, counted and stained for FACS analysis; the remaining cells were frozen in 10% v/v DMSO, 90% v/v Fetal bovine serum solution for subsequent cell sorting and RNA extraction.

Table 3 Clinical information of patients and controls.

Variables	Cohort1	Cohort1 CTR	Cohort2	Cohort2 CTR
No. of patients	60	46	65	52
Age, Years, median (range)	65(42-81)	48(38-62)	60(23-86)	45(27-72)
Gender, male/female	0/60, 60/60	0/46, 46/46	0/65, 65/65	0/52, 52/52
Type of cancer, Breast, Endometrium	40/60, 20/60	n/a	65/65, 0/65	n/a
Breast Cancer Grade (invasive)	1 (16/40) 2 (11/40) 3(13/40)	n/a	1 (18/65) 2 (25/65) 3 (22/65)	n/a
Breast Cancer ER+	32/40	n/a	65/65	n/a
Breast Cancer PR+	28/40	n/a	7/65	n/a
Breast Cancer HER2+	8/40	n/a	1/65	n/a
Breast Cancer ER- /PR-/HER2-	8/40	n/a	0/65	n/a
Endometrial Cancer Grade (invasive)	1 (4/20) 2(5/20) 3(11/20)	n/a	n/a	n/a
Endometrial Cancer Type 1	10/20	n/a	n/a	n/a
Endometrial Cancer Type 2	10/20	n/a	n/a	n/a

FACS Sorting and analysis

PBMCs or total blood cells were counted and resuspended in PBS 1% w/v Bovine Serum Albumin (BSA, Sigma-Aldrich); blocking of Fc receptors was performed by incubating samples with 10% v/v human serum (Sigma Aldrich) for 1h on ice. For FACS analysis 5x10⁵ cells were stained in a final volume of 100 μ L using the following antibodies: CD45-PE-Texas Red, CD3-, CD56-, CD19-BV711, CD11b-BV605, CD14-BV510, CD16-EF450, CX3CR1-FITC, HLA-DR-BV650, CD64-APC-CY7, CD80-PECY7, CCR2-PECY7, CD86-APC, CD95-PECY7 (Biolegend). For monocyte sorting cells were stained and antibody concentration was scaled up based on cell number; cells were stained with the following antibodies: CD45-AlexaFluor 700, CD3-, CD56-, CD19- PECY5, CD14-FITC, CD11b-PECy7, CD16- PE-Texas Red.

FACS analysis was performed using a 6-laser Fortessa flow cytometer (BD); FACS sorting was performed using FACS Arian and FACS Fusion sorters (BD). Cell sorting was performed at 4°C in 1.5 ml RNase and DNase free tubes (Simport, Canada) pre-filled with 750 μ L of PBS 0.1% w/v BSA; at the end of each isolation a sorting purity check was performed. A minimum of 5000 events in the monocyte gate was acquired for FACS analysis. Results were analyzed with Flowjo (Treestar) or DIVA software (BD).

RNA sequencing and Bioinformatic Analysis

Immediately after sorting all the samples were centrifuged at 450 RCF for 10 min at 4°C. The cell pellet was resuspended in 350 μ L of RLT lysis buffer and RNA extracted with RNeasy Microkit (Qiagen) according to manufacturer's instructions. RNA quantity was determined by QUBIT (Invitrogen); total RNA integrity was assessed by Agilent Bioanalyzer and the RNA Integrity Number (RIN) was calculated; samples that had a RIN > 7 were selected for RNA amplification and sequencing. RNA was amplified with Ovation RNAseq Amplification kit v2 (Nugen) according to manufacturer's instructions; amplified RNA was sent to Albert Einstein Genomic Facility (<https://www.einstein.yu.edu/departments/genetics/resources/genomics-core.aspx>) or BGI (Philadelphia; http://www.genomics.cn/en/navigation/show_navigation?nid=271) where library preparation, fragmentation and paired-end multiplex sequencing were performed (HiSeq 2000 and 2005, Illumina). All samples were processed and randomly assigned to lanes without knowledge of clinical identity to avoid bias and batch effects.

Sequencing alignment and Quantification

FastQ files of 2x100bp paired-end reads from normal monocytes and TEMo were quality controlled using FASTQC ¹⁸. Reads were aligned to the GENCODE Human reference genome Release 19 (GRCh37.p13) using STAR aligner ¹⁹(version 2.3) and quantified using HTSeq ²⁰.

Statistical analysis for differentially expressed genes

Quantified genes were pre-filtered and tested for differential expression using R Bioconductor (version 3.2.3) ²¹. Genes with logCPM > 1 in at least N samples (N number of the fewest replicates of a phenotype) were retained. Samples were corrected for batch effects using Combat function of the Surrogate Variable Analysis (SVA) package ²² (version 3.18). Limma statistical software was used (version 3.26.7) ²³ to identify significantly differentially expressed genes with controlled False Positive Rate at 5% (FDR <0.05).

Upregulated genes were selected with a minimum log₂ fold change of 1.5 for up and downregulated genes with a minimum log₂ fold change -1.5. In microarray datasets, the raw data were log₂ transformed, quantile normalized and filtered for probes without annotation using Limma ²³. Duplicated probes were collapsed to the average expression of each gene. Significant genes were used for gene ontology (GO) analysis using DAVID ²⁴ (Database for Annotation, Visualization and Integrated Discovery) database. GSEA (Gene Set Enrichment Analysis) and IPA (Ingenuity Pathway Analysis) were used to identify enriched pathways.

Publically available datasets

An RNA-seq study on Lyme disease from Bouguet et al ²⁵ was analyzed through the GEO database (GSE63085) to use as a negative control for validation of the diagnostic signature. For this purpose we selected only samples from patients diagnosed with Lyme disease that haven't received any treatment (n=28) and healthy patients (n=13). FPKM normalized values were normalized using voom from the statistical package *Limma* ²³ in Bioconductor. Additionally, an RNA-seq study on isolated human monocytes from Chronic periodontitis was downloaded from GEO (GSE61490)⁴⁵ to act as additional negative control for the TEMo signature. Raw data were downloaded, aligned to the GENCODE Human reference genome Release 19 (GRCh37.p13) using STAR aligner ¹⁹(version 2.3) and quantified using HTSeq ²⁰. Reads were normalised using the cpm function from the statistical package Limma ²³.

Construction and validation of diagnostic monocyte-derived classifier

A Random forest (RF) model was trained as implemented ²⁶ in the *caret* package in Bioconductor ²⁷. Variable *ntree* was kept constant at 1000 trees and variable *mtry* was kept constant at $\sqrt{p}$, where p is the total number of genes. Feature selection based on the Chi-square (X^2) score was used to assess the independence of each gene in respect to the cancer class. Genes were ranked in descending order based on their X^2 score. X^2 scores were calculated using the *chi.square* function as implemented in the *FSelector* package in Bioconductor ²⁸. The model was trained on a training set of 59 samples (Bronx, USA sample). To evaluate the performance of the classifier on the training set we used 5 times repeated 10-fold cross-validation. The model was evaluated using the following metrics, Accuracy, sensitivity, specificity and area under the curve (AUC). Overall accuracy, sensitivity and specificity are calculated using the cross-validated predictions. ROC curves were drawn using the *ROCR* package in Bioconductor²⁹. Performance of the optimal classifier was evaluated on an independent cohort (Edinburgh, UK) of 19 samples (5 healthy donors and 14 cancer donors) that haven't been used for training or gene selection. To determine the accuracy rates of the classifiers and gene signatures that can be obtained by chance, performance of randomly extracted gene signatures and performance of permutations expression values in each sample was calculated. This process was performed for 1000 LOOCV classification by random forest using the R package *SigCheck* ³⁰.

Quantitative PCR.

Cells were lysed and RNA extracted with RNAeasy Microkit (Qiagen) according to manufacturer's instructions. Typically, 0.1 ug of total RNA was reverse transcribed using Super Script Vilo kit (Invitrogen) and the cDNA generated was used for semi quantitative PCR on a 7900 Real Time cycler (Applied Biosystem) as per manufacturer's instructions. Target gene expression was normalized to the expression of the housekeeping gene GAPDH. Relative gene expression was calculated using the standard $2^{-\Delta\Delta CT}$ method. Primers were designed using Primer Bank ³¹. The full list of primers used can be found in Table 4.

Table 4. qPCR primer list

Target		Sequence (5'→3')	Seq ID No.
GAPDH	FWD	GGAGCGAGATCCCTCCAAAAT	20
	REV	GGCTGTTGTCATACTTCTCATGG	21

Target		Sequence (5'→3')	Seq ID No.
CX3CR1	FWD	ACTTTGAGTACGATGATTTGGCT	22
	REV	GGTAAATGTCGGTGACACTCTT	23
HGF	FWD	GCTATCGGGGTAAAGACCTACA	24
	REV	CGTAGCGTACCTCTGGATTGC	25
TNFSF10	FWD	TGCGTGCTGATCGTGATCTTC	26
	REV	GCTCGTTGGTAAAGTACACGTA	27
ANGPT1	FWD	AGAACCTTCAAGGCTTGGTTAC	28
	REV	GGTGGTAGCTCTGTTTAATTGCT	29
ANGPT2	FWD	CTCGAATACGATGACTCGGTG	30
	REV	TCATTAGCCACTGAGTGTTGTTT	31
CD200R1	FWD	CAGAGGCATAGTGGTAACACCT	32
	REV	GTGCCATTGCCCCAGTATTCT	33
CD16	FWD	GTACAGGGTGCTCGAGAAGG	34
	REV	AACCACTGTGTGGAATTGTCC	35

Legendplex array and CX3CL1 ELISA on human plasma.

CCL2 levels were assessed in plasma from 15 healthy donors and 42 breast cancer patients using Legendplex bead-based immunoassays (Biolegend) according to manufacturer's protocol. Data was collected using the C4 Accuri (BD).

- 5 ELISA for human CX3CL1 was performed using a human CX3CL1 Quantikine ELISA kit (R&D Systems) as per manufacturer's instructions.

LPS stimulation of monocytes and intracellular staining.

- 10 Whole blood was withdrawn from healthy individuals and cancer patients by venipuncture and collected in Venous Blood Collection Tubes containing EDTA. 50 ul of blood was incubated for 6 hours in the presence or absence of gram-negative bacterial LPS (1ug/ml) for 6 hours in a 37C water bath with brefeldin A (eBiosciences) at 1:1000 dilution.

For staining of surface antigens, CD45-PE-Texas Red, CD3-, CD56-, CD19-BV711, CD11b-BV605, CD14-BV510, CD16-EF450, CX3CR1-FITC, HLA-DR-BV650 (Biolegend) antibodies were added 1 hour before termination of stimulation and incubated at 37 °C. 1 ml of 1X Lyse/Fix buffer (BD Biosciences) was added to each tube with gentle vortexing and incubated at 37 °C in a water bath for 10 min. After this fixation/lysis step, blood was washed twice with 2 ml cold PBS followed by incubation with 500 ul 1X permeabilization buffer IV (BD Biosciences) for 20 min at room temperature. Post permeabilization, PE-IL-1 β or PE-Isotype control antibodies (Biolegend) were added and samples incubated at room temperature in the dark for 30 min. The cells were washed with two mL cold PBS, followed by centrifugation and resuspension in 200 ul PBS and acquired on a BD LSR Fortessa flow cytometer.

Statistics

Statistical significance was calculated by Student's t test when comparing two groups or by one-way or two-way ANOVA when comparing three or more groups. A p value < 0.05 was considered as statistically significant.

2. Expansion of the non-classical monocytic population and distinct monocyte transcriptional profile in cancer patients

Multicolor cytofluorimetric analysis on monocytes from two independent cohorts of patients (Cohort 1 and cohort 2) was performed and the percentage of classical to non-classical monocytes were calculated. Cohort 1 consisted of endometrial and breast cancer patients, while Cohort 2 contained only breast cancer patients (Table 3, Fig 1A and Fig 2). Cancer patients exhibit a significant expansion of non-classical monocytes as compared to non-cancer controls in both analyzed cohorts, indicating that cancer affects monocyte ratios in the blood. In cohort 1, there were no significant differences between endometrial and breast cancer patients in terms of expansion of non-classical monocytes (Fig1A).

In order to understand whether cancer influences the transcriptional signatures of TEMo, these were isolated from patients with breast and endometrial cancer and from appropriate controls by FACS (Fig 3), and the RNA was isolated and subjected to paired-end multiplex RNA-sequencing analysis. Sequences were obtained and aligned to the genome and subjected to Multidimensional Scaling (MDS) Analysis and hierarchical clustering. MDS segregated the transcriptomic profile of TEMo from breast and endometrial cancer patients, indicating the populations to be transcriptionally very different (Fig 1B, C). Differential expression analysis of the population identified 2169 genes modulated in a significant way

(1946 upregulated and 223 downregulated; Log₂FC 1.5, FDR < 0.05) in TEMo compared to Mo (Fig 4A).

3. Distinct gene expression profiles of cancer and normal monocytes

We compared the transcriptional profiles of TEMo from endometrial and breast cancer compared to normal controls (Fig 1D). Only a minority of genes were differentially expressed between breast and endometrial monocytes, indicating that the presence of cancer modified the transcriptomic profile of circulating monocytes in a similar way at least in two distinct cancer types of the female reproductive system. Furthermore in preliminary experiments the differences between monocyte sub-sets were much smaller than those between monocytes isolated from cancer patients and normals. Thus to provide an easy blood diagnostic test, we isolated and profiled the total monocyte population.

Differentially expressed genes (DEGs) between normal and cancer monocytes were analyzed using Ingenuity Pathway Analysis Software (IPA) in order to identify enriched signaling pathways and gene lists; the core analysis reported “cancer” as the most significant disease enriched in TEMo, followed by “reproductive system disease”, “cell to cell signaling”, “cellular movement” and “immunological disease” (Fig 1E). Canonical pathways analysis identified immune-related genes family enriched in the TEMos, including Pattern recognition receptors, TREM1 signalling, Tissue Factors in cancer, IL-1 signaling and G coupled receptors signaling (Fig 4B). GSEA (Gene Enrichment Analysis) identified 9 gene sets with a significant negative enrichment score, e.g., “Regulation of immune response”, “cell to cell communication” and “IL1R pathway” (Fig 4C).

In order to analyze the TEMo transcriptomic profile in detail, we focused on DEGs encoding transmembrane receptors, soluble factors, transcription factors and enzymes (Fig. 1F). Consistent with the previous analysis TEMos exhibited an increased expression of transcripts encoding the chemokine receptors CCR2, CCR5 and CX3CR1, important players in monocyte recruitment at the tissue level ³², TLR5 and TLR7, mainly involved in innate immune system activation and CD200R1, a key regulator of adaptive immune response. Moreover, this analysis revealed upregulation of the C-type lectins Clec1a, 1b and 4a, and the Immunoglobulin-like lectins Siglec 9 and 16, whose main function is immune system regulation ^{3,33,34}. FcγR3 (CD16), an Fc receptor mainly expressed by non-classical monocytes ^{6,35}, was significantly upregulated in TEMo as compared to control, as well as CD163L1, a receptor recently reported to be expressed on tissue macrophages ³⁶. We also identified genes encoding soluble factors; the proangiogenic genes Angiopoietin 1 and 2 together with Hepatocyte Growth factor were significantly unregulated in TEMo; conversely VEGF-A expression was downregulated in TEMo compared to controls. Interestingly the

death ligand TNFSF10 (TRAIL), known to induce apoptosis via death receptors DR4 and DR5 54, was also upregulated in TEMo,

To validate the gene-expression profile of TEMos, we performed semi-quantitative PCR (qPCR) on selected genes identified by the previous analysis (Fig. 1G); qPCR results analysis confirmed significant increased expression of the non-classical monocyte receptors CD16 (FcGR3A) and CX3CR1, the immunosuppressive receptor CD200R1 and the death ligand TRAIL. The proangiogenic factors HGF and Angiopoietin 1 and 2 were also confirmed to be upregulated in TEMo compared to control; conversely CCR2 expression did not show significant difference (data not shown).

Together, these results indicate that TEMo from breast and endometrial cancer patients are influenced by presence of cancer, and their transcriptional profile is significantly different from homeostatic Mo of non-cancer individuals.

4. TEMo show phenotypical and functional alterations compared to normal Mo

Tumors or other cells “educated” by them could release in the blood circulation soluble factors responsible for TEMo transcriptional regulation. Classical and non-classical monocytes respond to different chemokines, mainly CCL2 and CX3CL1 respectively. We collected serum from healthy donors and cancer patients and we evaluated the levels of CCL2 and CX3CL1 in the blood by ELISA/Legendplex. We observed a significant reduction of CCL2 in cancer patients’ sera compared to control; on the contrary CX3CL1 was significantly higher than control (Fig 5A).

To further analyze the monocytes, we performed 16-colour multicolor FACS analysis of classical and non-classical monocytes in normal controls and cancer patients; TEMo exhibit an altered non-classical/classical ratio of expression of CX3CR1 and CD86 (Fig 5B); no differences were observed for CCR2 and for CD64, CD80, CD16 (Fig 5B). Interestingly CD95 expression was significantly downregulated in both classical and non-classical TEMo compared to normal with no change in the relative expression ratio between the two subsets. Thus gene expression changes were in both populations of monocytes and the dramatic transcriptional differences of TEMOs from normal were not just due to changes in proportion of different monocytic populations but alterations in transcriptomes of both populations.

Human monocytes exhibit pro-inflammatory features in a variety of disease contexts⁴⁰; in order to evaluate the pro-inflammatory function of TEMo we incubated the blood from normal donors and cancer patients with or without Lipopolysaccharide (LPS) for 6 hours and we evaluated the level of expression of the pro-inflammatory cytokine IL-1Beta before and after LPS stimulation (Fig 5C). As previously reported, monocytes from healthy donors responded

to LPS stimulation by producing significantly high levels of IL-1Beta. In contrast TEMo failed to produce a significant response to LPS as compared to healthy Mo, indicating that presence of cancer impairs production of the pro-inflammatory factor IL-1Beta (Fig 5C, D).

Together, these data indicate that cancer affects TEMo not only at transcriptional level, but also at phenotypic and functional level. Thus we suggest the term *Tumor-Educated Monocytes* (TEMo).

5. TEMo gene signature can be used to detect cancer

The analyses performed thus far indicated that a blood-based diagnostic signature might be contained within the TEMos transcriptome that could distinguish cancer patients from healthy individuals. We employed supervised predictive analysis using Random forest (RF) and Chi-square feature selection on the Bronx, NY training cohort (n=59 samples, 22 healthy donors and 37 cancer samples) using 5 times 10-fold cross validation. For the RF classifier with Chi-square feature selection (X^2 -RF) we selected the 13 highest ranked genes (Fig S6) in the training cohort (n=59) that maintained the clustering between Mo and TEMo (Fig 7A and Table 5). We achieved accuracy of 94%, sensitivity of 98%, specificity 86% (Figure 7B) and area under curve (AUC) of 97.8% (Figure 7C). Further evaluation of the performance of the classifier on an independent cohort of patients (n=19) that was not used during training or during feature selection, achieved perfect accuracy of 100%, sensitivity 100%, specificity 100% (Figure 7C) and area under the curve 100% (Figure 7B).

Table 5 Details of the identified 13 predictive genes

Gene Name	Chromosome	Log ₂ FC	q-value (FDR)	Type	Gene ontology Process
MCTP1	5q15	1.57	4.44E-07	Calcium Binding Protein	calcium ion binding; calcium-mediated signaling;
PIBF1	13q22.1	3.1	1.76E-08	-	interleukin-4 receptor binding; immune system process; negative regulation of IL-12; negative regulation of NK-cell;
TMTC2	12q21.31	2.8	1.50E-07	Endoplasmic Reticulum Protein	protein binding; calcium ion homeostasis;
SLF1/ANKRD32	5q15	2.99	1.80E-08	Transcription regulator	DNA repair; positive regulation of protein complex assembly; response to DNA damage stimulus;
CEPT1	1p13.3	1.73	2.51E-07	Enzyme	CDP-choline pathway; lipid metabolic process; phospholipid biosynthetic process; phospholipid metabolic process;
ZNF114	19q.3.33	3.06	1.15E-08	Zinc finger protein	regulation of transcription, DNA-dependent;
CRYBG3	3q11.2	2.37	8.25E-06	-	transcription, DNA-dependent carbohydrate binding
IFI16	1q22	1.85	6.05E-05	Transcription regulator	activation of innate immune response; autophagy; B cell receptor signaling pathway; cellular response to interferon-beta; regulation of transcription;
PPIF	10q22.3	-3.04	3.54E-07	Enzyme	apoptotic mitochondrial changes; apoptotic process; necroptosis; programmed cell death;
SCD	10q24.31	-3.5	8.82E-10	Enzyme	fatty acid biosynthetic process; cellular lipid metabolic process; fatty acid metabolic process
RP11-469M7.1	-	-2.55	1.87-06	-	-
PTP4A1	6q12	-1.38	1.31E-03	Phosphatase	cell cycle; dephosphorylation; multicellular organismal development;
NRIP1	21q11.2	-1.36	2.28E-05	Transcription regulator	androgen receptor signaling pathway; cellular response to estradiol stimulus; lipid storage; ovarian follicle rupture; ovulation; positive regulation of transcription

To further underline the uniqueness and specificity of our model to discriminate between cancer and healthy samples we extracted 1000 randomly selected gene signatures comprising 13 genes and tested their performance using a leave-one-out cross validation (LOOCV). Our 13-gene showed the highest accuracy compared to random signatures that yielded a mean accuracy of 80% (SD $\pm$ 4%, $p < 0.001$) (Figure 6B). Furthermore, random classifiers as determined by 1000 rounds of random sample permutations, had no predictive power (mean accuracy: 61%, SD: 6%, $p < 0.001$, Fig 6C).

To establish whether X^2 -RF classifier was specific to cancer, we sought to compare data with a dataset from patients with infectious diseases but not cancer. We compared against a dataset (RNA seq) from peripheral blood mononuclear cells (PBMCs) isolated from individuals with Lyme disease³⁷ and an RNA-seq dataset from isolated human circulating monocytes from healthy individuals and chronic periodontitis patients. This Lyme disease dataset comprises 29 Lyme disease patients and 13 matched controls and the chronic periodontitis dataset comprises 5 chronic periodontitis patients and 5 healthy controls. Out of the 13 gene predictors selected from the X^2 -RF classifier, 12 were also found in the Lyme dataset. X^2 -RF classifier successfully classified all Lyme diseased patients as normal, and all

normal patients as normal. Similarly, the X^2 -RF classifier successfully classified all chronic periodontitis patients as normal and all normal subjects as healthy, thus indicating specificity of the cancer classification (Fig 7D).

To test whether all 13 genes would be required to provide a signature useful for distinguishing cancer patients, we compared results using the 13 genes to those obtained using only 8 of the 13 genes, namely CRYBG3, PIBF1, SCD, PPIF, PTP4A1, ANKRD32, CEPT1, and ZNF114. For these 8 genes we achieved accuracy of 93%, sensitivity of 97%, specificity 86% (Figure 8B) and area under curve (AUC) of 96% (Figure 8A). Further evaluation of the performance of the classifier on an independent cohort of patients (n=19) that was not used during training or during feature selection, achieved perfect accuracy of 100%. The 8 gene X^2 -RF classifier successfully classified all Lyme diseased patients as normal, and all normal patients as normal thus indicating specificity of the 8 gene cancer signature. Therefore, a selection of 8 of the 13 genes also provides a useful gene signature.

Similarly, we also compared results using the 13 genes to those obtained using only 5 of the 13 genes, namely CRYBG3, PIBF1, SCD, PPIF and ZNF114. For these 5 genes we achieved accuracy of 93%, sensitivity of 97%, specificity 86% (Figure 9B) and area under curve (AUC) of 96% (Figure 9A). Further evaluation of the performance of the classifier on an independent cohort of patients (n=19) that was not used during training or during feature selection, achieved perfect accuracy of 100%. Similar results were also achieved using a selection of 4 of the 13 genes in the 13 gene signature. Therefore a selection of 4 or 5 of the 13 genes also provides a useful gene signature.

6. TEMo signature detects breast cancer at an early stage

Within the 19 samples of the independent validation set, 4 patients have been diagnosed with DCIS. Several studies have carried out gene expression profiling of DCIS, however the diagnosis and monitoring of DCIS patients remains very challenging⁴⁶⁻⁴⁸. A careful look at the results of the classification revealed that all samples from patients with DCIS were correctly classified as having cancer suggesting that TEMos can identify the presence of pre-malignant disease even if the cancer cells are non-invasive and still confined within the basement membrane. In order to further investigate the transcriptional profiles of monocytes in DCIS patients, a MDS plot was drawn using only the 13-gene predictor on samples coming from invasive breast cancer (n = 52, 27 healthy, 21 invasive breast cancer and 4 DCIS)

The MDS plot clearly separated the invasive breast cancer patients from healthy individuals, with all the DCIS samples included in the breast cancer cluster. Differential expression

analysis between TEMos coming from DCIS patients compared to healthy individuals showed 2382 up-regulated genes and 506 down-regulated genes with Log₂FC greater or less than 1.5/-1.5, FDR ≤ 0.05. It is clear that TEMos from DCIS patients are different from monocytes coming from healthy individuals.

- 5 These data indicate that circulating TEMos are influenced by the formation and growth of tumours from an early stage and thus may provide a method for early detection of pre-cancer states.

7. Discussion

10 The inventors have shown herein that the transcriptional profile of circulating blood monocytes is severely altered by the presence of breast and endometrial cancer; these data are consistent with previous reports on renal carcinoma and colorectal cancer patients showing alteration of gene expression in circulating monocytes^{12,13} and confirm the hypothesis that cancer can induce a systemic alteration of the immune system.

15 The transcriptional changes are associated with the expansion of the non-classical monocytic population CD14+CD16++, a result that is consistent with previous studies that reported altered frequencies of non-classical monocytes in cancer^{8,9,38,39}.

Non-classical monocytes are important players during infection and inflammation as they are rapidly recruited through a CX3CL1 dependent mechanism by the injured or infected tissue to resolve the inflammation⁴⁰.

20 Interestingly we detected higher levels of CX3CL1 but not CCL2 in the sera of breast cancer patients compared to healthy controls, suggesting that non-classical TEMo could be potentially activated or recruited when cancer is present; this correlates with the significantly higher CX3CR1 expression ratio observed between classical and non classical populations in TEMo compared to control.

25 Moreover TEMo showed a reduced expression of CD86 (B7-2), key molecule whose function is to provide survival and activation stimuli to T cells⁴¹ and reduced expression of CD95 (Fas), reported to be associated with pro-inflammatory cytokines production in monocytes and macrophages⁴².

30 Non-classical TEMo showed reduced ability to respond to a pro-inflammatory stimuli like LPS compared to normal monocytes suggesting a potential cancer-induced de-activation of these cells, like suggested by others^{43,44}.

Significant transcriptional changes could be also detected during the early stage of breast cancer progression; TEMo from DCIS patients show already a significant level of

transcriptional deregulation compared to monocytes from healthy individuals and a significant shift in monocytic sub population distribution, suggesting the use of these changes for early detection of breast cancer.

Supervised machine learning identified a 13-gene signature in TEMo that is able to detect the presence of breast and endometrial cancer with higher specificity and precision than mammography, suggesting that TEMo's gene alteration has diagnostic value for early detection of breast and endometrial cancer.

REFERENCES

1 Augier, S. et al. Inflammatory blood monocytes contribute to tumor development and represent a privileged target to improve host immunosurveillance. *Journal of immunology* 185, 7165-7173, doi:10.4049/jimmunol.0902583 (2010).

2 Lin, E. Y. & Pollard, J. W. Role of infiltrated leucocytes in tumour growth and spread. *British journal of cancer* 90, 2053-2058, doi:10.1038/sj.bjc.6601705 (2004).

3 Qian, B. Z. et al. CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature* 475, 222-225, doi:10.1038/nature10138 (2011).

4 Yeo, E. J. et al. Myeloid WNT7b Mediates the Angiogenic Switch and Metastasis in Breast Cancer. *Cancer research* 74, 2962-2973, doi:10.1158/0008-5472.CAN-13-2421 (2014).

5 Lin, E. Y. et al. Macrophages regulate the angiogenic switch in a mouse model of breast cancer. *Cancer research* 66, 11238-11246, doi:10.1158/0008-5472.CAN-06-1278 (2006).

6 Ancuta, P. et al. Transcriptional profiling reveals developmental relationship and distinct biological functions of CD16+ and CD16- monocyte subsets. *BMC Genomics* 10, 403, doi:10.1186/1471-2164-10-403 (2009).

7 Ziegler-Heitbrock, L. The CD14+ CD16+ blood monocytes: their role in infection and inflammation. *J Leukoc Biol* 81, 584-592, doi:10.1189/jlb.0806510 (2007).

8 Subimerb, C. et al. Circulating CD14(+) CD16(+) monocyte levels predict tissue invasive character of cholangiocarcinoma. *Clin Exp Immunol* 161, 471-479, doi:10.1111/j.1365-2249.2010.04200.x (2010).

9 Feng, A. L. et al. CD16+ monocytes in breast cancer patients: expanded by monocyte chemoattractant protein-1 and may be useful for early diagnosis. *Clin Exp Immunol* 164, 57-65, doi:10.1111/j.1365-2249.2011.04321.x (2011).

- 10 Mytar, B. et al. Tumor cell-induced deactivation of human monocytes. *Journal of leukocyte biology* 74, 1094-1101, doi:10.1189/jlb.0403140 (2003).
- 11 Sanford, D. E. et al. Inflammatory monocyte mobilization decreases patient survival in pancreatic cancer: a role for targeting the CCL2/CCR2 axis. *Clinical cancer research : an official journal of the American Association for Cancer Research* 19, 3404-3415, doi:10.1158/1078-0432.CCR-13-0525 (2013).
- 12 Chittezath, M. et al. Molecular profiling reveals a tumor-promoting phenotype of monocytes and macrophages in human cancer progression. *Immunity* 41, 815-829, doi:10.1016/j.immuni.2014.09.014 (2014).
- 10 13 Hamm, A. et al. Tumour-educated circulating monocytes are powerful candidate biomarkers for diagnosis and disease follow-up of colorectal cancer. *Gut*, doi:10.1136/gutjnl-2014-308988 (2015).
- 14 http://www.breastcancer.org/symptoms/understand_bc/statistics
- 15 Breast cancer incidence statistics. <http://www.cancerresearchuk.org/cancer-info/cancerstats/types/breast/incidence/uk-breast-cancer-incidence-statistics>.
- 16 Lauby-Secretan, B., Loomis, D. & Straif, K. Breast-Cancer Screening--Viewpoint of the IARC Working Group. *N Engl J Med* 373, 1479, doi:10.1056/NEJMc1508733 (2015).
- 17 Pharoah, P. D., Sewell, B., Fitzsimmons, D., Bennett, H. S. & Pashayan, N. Cost effectiveness of the NHS breast screening programme: life table model. *BMJ* 346, f2618 (2013).
- 18 Andrew S. et al. FastQC: a quality control tool for high throughput sequence data. 2010.
- 19 Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 25 20 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166-169, doi:10.1093/bioinformatics/btu638 (2015).
- 21 Team R. R; a language and environment for statistical computing. (2014)
- 22 Leek JT et al. SVA: surrogate variable analysis. R package.
- 30 23 Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43, e47, doi:10.1093/nar/gkv007 (2015).

- 24 Dennis, G. et al. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4, P3 (2003).
- 25 Bouquet, J. et al. Longitudinal Transcriptome Analysis Reveals a Sustained Differential Gene Expression Signature in Patients Treated for Acute Lyme Disease. *MBio* 7, doi:10.1128/mBio.00100-16 (2015).
- 26 Breiman L et al. Random Forests , *Machine Learning*, Vol. 45 5-32 (2001).
- 27 Kuhn M : CARET: classification and regression ,*Astrophysics Source Code library*, (2015).
- 28 Romanski P. F selector : selecting attributes. Vienna: R foundation for statistical computing, (2009).
- 29 Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCr: visualizing classifier performance in R. *Bioinformatics* 21, 3940-3941, doi:10.1093/bioinformatics/bti623 (2005).
- 30 Stark R et al. SiGCheck; check a gene signature's prognostic performance against random signatures, known signatures and permuted data/metadata. R package version 2.5.0. (2016).
- 31 Wang, X., Spandidos, A., Wang, H. & Seed, B. PrimerBank: a PCR primer database for quantitative gene expression analysis, 2012 update. *Nucleic Acids Res* 40, D1144-1149, doi:10.1093/nar/gkr1013 (2012).
- 32 Weber, C. et al. Differential chemokine receptor expression and function in human monocyte subpopulations. *J Leukoc Biol* 67, 699-704 (2000).
- 33 Zhang, J. Q., Nicoll, G., Jones, C. & Crocker, P. R. Siglec-9, a novel sialic acid binding member of the immunoglobulin superfamily expressed broadly on human blood leukocytes. *J Biol Chem* 275, 22121-22126, doi:10.1074/jbc.M002788200 (2000).
- 34 Macauley, M. S., Crocker, P. R. & Paulson, J. C. Siglec-mediated regulation of immune cell function in disease. *Nat Rev Immunol* 14, 653-666, doi:10.1038/nri3737 (2014).
- 35 Mukherjee, R. et al. Non-Classical monocytes display inflammatory features: Validation in Sepsis and Systemic Lupus Erythematosus. *Sci Rep* 5, 13886, doi:10.1038/srep13886 (2015).
- 36 González-Domínguez, É. et al. CD163L1 and CLEC5A discriminate subsets of human resident and inflammatory macrophages in vivo. *J Leukoc Biol* 98, 453-466, doi:10.1189/jlb.3HI1114-531R (2015).

- 37 Bouquet, J. et al. Longitudinal Transcriptome Analysis Reveals a Sustained Differential Gene Expression Signature in Patients Treated for Acute Lyme Disease. *MBio* 7, doi:10.1128/mBio.00100-16 (2015).
- 38 Sponaas, A. M. et al. The proportion of CD16(+)CD14(dim) monocytes increases
5 with tumor cell load in bone marrow of patients with multiple myeloma. *Immun Inflamm Dis* 3, 94-102, doi:10.1002/iid3.53 (2015).
- 39 Maffei, R. et al. The monocytic population in chronic lymphocytic leukemia shows altered composition and deregulation of genes involved in phagocytosis and inflammation. *Haematologica* 98, 1115-1123, doi:10.3324/haematol.2012.073080 (2013).
- 10 40 Mukherjee, R. et al. Non-Classical monocytes display inflammatory features: Validation in Sepsis and Systemic Lupus Erythematosus. *Sci Rep* 5, 13886, doi:10.1038/srep13886 (2015).
- 41 Sharpe, A. H. & Freeman, G. J. The B7-CD28 superfamily. *Nat Rev Immunol* 2, 116-126, doi:10.1038/nri727 (2002).
- 15 42 Park, D. R. et al. Fas (CD95) induces proinflammatory cytokine responses by human monocytes and monocyte-derived macrophages. *J Immunol* 170, 6209-6216 (2003).
- 43 del Fresno, C. et al. Tumor cells deactivate human monocytes by up-regulating IL-1 receptor associated kinase-M expression via CD44 and TLR4. *J Immunol* 174, 3032-3040 (2005).
- 20 44 Mytar, B. et al. Tumor cell-induced deactivation of human monocytes. *J Leukoc Biol* 74, 1094-1101, doi:10.1189/jlb.0403140 (2003).
45. GSE61490 - <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4767619/>
46. Hannemann, J et al. Classification of ductal carcinoma in situ by gene expression profiling. *Breast Cancer Research* 8:R61 (2006)
- 25 47. Adeyinka, A. et al. Analysis of Gene Expression in Ductal Carcinoma in Situ of the Breast. *Clin Can Res* 8:12, 3788-3795 (2002)
48. Seth A, et al. Gene expression profiling of ductal carcinomas *in situ* and invasive breast tumors. *Anticancer Res* 23, 2043 – 2051 (2003)

CLAIMS

1. A method comprising:

a) analysing a biological sample obtained from a subject to determine the presence of target molecules representative of expression of at least four biomarkers selected from the group consisting of MCTP1, PIBF1, TMTC2, ANKRD32, CEPT1, ZNF114, CRYBG3, IFI16, PPIF, SCD, RP11-469M7.1, PTP4A1 and NRIP1; and

b) comparing the expression levels of the biomarkers determined in (a) with one or more reference values, wherein a difference in the expression of the biomarkers in the sample from the subject compared to the one or more reference values is indicative of cancer presence or absence.

2. A method according to claim 1 wherein said method is a method of diagnosing cancer, of prognosing cancer, of predicting efficacy of treatment for cancer, of assessing outcome of treatment for cancer, and/or of assessing recurrence of cancer, preferably wherein said method is a method of diagnosing cancer.

3. A method according to claim 1 or claim 2 wherein the biomarkers are selected from the group consisting of PIBF1, SCD, ZNF114, CRYBG3, PPIF, PTP4A1, ANKRD32, and CEPT1.

4. A method according to any of claims 1 to 3 comprising determining the expression levels of four, five, six, or seven biomarkers selected from PIBF1, SCD, ZNF114, CRYBG3, PPIF, PTP4A1, ANKRD32, and CEPT1.

5. A method according to any preceding claim wherein the expression levels of PIBF1, SCD, ZNF114, CRYBG3, PPIF, PTP4A1, ANKRD32, and CEPT1 are determined.

6. A method according to any of claims 1 to 4 wherein the biomarkers are selected from the group consisting of PIBF1, SCD, ZNF114, CRYBG3, and PPIF.

7. A method according to claim 6 wherein the expression levels of PIBF1, SCD, ZNF114, CRYBG3, and PPIF are determined.

8. A method according to any preceding claim wherein the biological sample is a blood sample or a derivative thereof, and preferably wherein the blood sample is a peripheral blood sample.

9. A method according to any preceding claim wherein the expression levels of the biomarkers are selectively detected in monocytes of the biological sample.

10. A method according to any preceding claim wherein the biological sample is enriched for monocytes or substantially consists of monocytes
11. A method according to any preceding claim wherein the subject is human.
12. A method according to any preceding claim wherein the subject has, or is suspected of having, breast cancer, endometrial cancer, ovarian cancer, prostate cancer, pancreatic cancer, thyroid cancer, cervical cancer, bladder cancer, blastoma, brain cancer and gliomas, bowel cancer, gastric cancer, head and neck cancer, kidney cancer, liver cancer, lung cancer, mesothelioma, melanoma, oral cancer, pituitary cancer, skin cancer, soft tissue cancer, testicular cancer, uterine cancer, heart cancer, and/or eye cancer.
- 10 13. A method according to any preceding claim wherein the subject has, or is suspected of having, a solid tumour cancer, for example a carcinoma, or a precancerous lesion, for example a carcinoma in situ.
14. A method according to any preceding claim wherein the subject has, or is suspected of having, a hormone-related cancer, for example breast cancer, endometrial cancer, 15 ovarian cancer, prostate cancer, pancreatic cancer, and thyroid cancer.
15. A method according to claim 14 wherein the subject has, or is suspected of having, an estrogen-dependent cancer, for example a cancer selected from breast cancer, endometrial cancer and ovarian cancer.
16. A method according to any preceding claim wherein the target molecule for one or 20 more of the biomarkers is nucleic acid, preferably an mRNA or ncRNA molecule.
17. A method according to claim 16 wherein the levels of target molecules for one or more of the biomarkers are determined by PCR, preferably quantitative PCR, digital PCR or multiplex PCR.
18. A method according to any of claims 1 to 15 wherein the target molecule for one or 25 more of the biomarkers is protein, and preferably the analysis of the presence of the biomarkers comprises using a binding partner selected from the group consisting of antibodies, antibody fragments and aptamers.
19. A method according to any preceding claim wherein the reference values correspond to the levels of the biomarkers in samples from subjects not having cancer.
- 30 20. A method comprising:
- a) analysing a biological sample obtained from a subject to determine the presence of target molecules representative of expression of at least four biomarkers selected from

the group consisting of MCTP1, PIBF1, TMTC2, ANKRD32, CEPT1, ZNF114, CRYBG3, IFI16, PPIF, SCD, RP11-469M7.1, PTP4A1 and NRIP1; and

b) comparing the expression levels of the biomarkers determined in (a) with one or more reference values, and in the event that there is a difference in the expression of the biomarkers in the sample from the subject compared to the one or more reference values, identifying the subject as requiring a treatment for cancer or not.

21. A method according to claim 19 further comprising providing a subject, identified as requiring treatment, with said treatment for cancer.

22. A method according to claim 20 or claim 21 wherein the biomarkers are selected from the group consisting of PIBF1, SCD, ZNF114, CRYBG3, PPIF, PTP4A1, ANKRD32, and CEPT1.

23. A method according to any of claims 20 to 22 comprising determining the expression levels of four, five, six, or seven biomarkers selected from PIBF1, SCD, ZNF114, CRYBG3, PPIF, PTP4A1, ANKRD32, and CEPT1.

24. A method according to any of claims 20 to 23 wherein the expression levels of PIBF1, SCD, ZNF114, CRYBG3, PPIF, PTP4A1, ANKRD32, and CEPT1 are determined.

25. A method according to any of claims 20 to 24 wherein the biomarkers are selected from the group consisting of PIBF1, SCD, ZNF114, CRYBG3, and PPIF.

26. A method according to claim 25 wherein the expression levels of PIBF1, SCD, ZNF114, CRYBG3, and PPIF are determined.

27. A method according to any of claims 20 to 26 wherein the biological sample is a blood sample or a derivative thereof, and preferably wherein the blood sample is a peripheral blood sample.

28. A method according to any of claims 20 to 27 wherein the expression levels of the biomarkers are selectively detected in monocytes of the biological sample.

29. A method according to any of claims 20 to 28 wherein the biological sample is enriched for monocytes or substantially consists of monocytes

30. A method according to any of claims 20 to 29 wherein the subject is human.

31. A method according to any of claims 20 to 30 wherein the subject has, or is suspected of having, breast cancer, endometrial cancer, ovarian cancer, prostate cancer, pancreatic cancer, thyroid cancer, cervical cancer, bladder cancer, blastoma, brain cancer and gliomas, bowel cancer, gastric cancer, head and neck cancer, kidney cancer, liver

cancer, lung cancer, mesothelioma, melanoma, oral cancer, pituitary cancer, skin cancer, soft tissue cancer, testicular cancer, uterine cancer, heart cancer, and/or eye cancer.

32. A method according to any of claims 20 to 31 wherein the subject has, or is suspected of having, a solid tumour cancer, for example a carcinoma, or a precancerous lesion, for example a carcinoma in situ.

33. A method according to any of claims 20 to 32 wherein the subject has, or is suspected of having, a hormone-related cancer, for example breast cancer, endometrial cancer, ovarian cancer, prostate cancer, pancreatic cancer, and thyroid cancer.

34. A method according to claim 33 wherein the subject has, or is suspected of having, an estrogen-dependent cancer, for example a cancer selected from breast cancer, endometrial cancer and ovarian cancer.

35. A method according to any of claims 20 to 34 wherein said treatment comprises one or more treatments selected from the group consisting of surgery, radiation therapy, chemotherapy, immunotherapy, hormone therapy, and targeted therapy, preferably wherein said cancer is breast cancer.

36. A method according to any of claims 20 to 35 wherein said cancer is endometrial cancer and the treatment is one or more selected from the group consisting of surgery, radiotherapy, chemotherapy and hormone therapy.

37. A method according to any of claims 20 to 36 wherein the target molecule for one or more of the biomarkers is nucleic acid, preferably an mRNA or ncRNA molecule.

38. A method according to claim 37 wherein the levels of target molecules for one or more of the biomarkers are determined by PCR, preferably quantitative PCR, digital PCR or multiplex PCR.

39. A method according to any of claims 20 to 38 wherein the target molecule for one or more of the biomarkers is protein, and preferably the analysis of the presence of the biomarkers comprises using a binding partner selected from the group consisting of antibodies, antibody fragments and aptamers.

40. A method according to any of claims 20 to 39 wherein the reference values correspond to the levels of the biomarkers in samples from subjects not having cancer or from subjects who have not responded to said treatment.

41. A kit comprising binding partners capable of binding to target molecules representative of expression of at least four biomarkers selected from the group consisting of

MCTP1, PIBF1, TMTC2, ANKRD32, CEPT1, ZNF114, CRYBG3, IFI16, PPIF, SCD, RP11-469M7.1, PTP4A1 and NRIP1.

42. A kit according to claim 41 for use in a method according to any of claims 1 to 40.

43. A kit according to claim 41 or 42 further comprising indicators capable of indicating when said binding occurs.

44. A kit according to any of claims 41 to 43 wherein the biomarkers are selected from the group consisting of PIBF1, SCD, ZNF114, CRYBG3, PPIF, PTP4A1, ANKRD32, and CEPT1.

45. A kit according to any of claims 41 to 44 wherein the biomarkers comprise four, five, six, or seven biomarkers selected from PIBF1, SCD, ZNF114, CRYBG3, PPIF, PTP4A1, ANKRD32, and CEPT1.

46. A kit according to any of claims 41 to 45 wherein the expression levels of PIBF1, SCD, ZNF114, CRYBG3, PPIF, PTP4A1, ANKRD32, and CEPT1 are determined.

47. A kit according to any of claims 41 to 45 wherein the biomarkers are selected from the group consisting of PIBF1, SCD, ZNF114, CRYBG3, and PPIF.

48. A kit according to claim 47 wherein the expression levels of PIBF1, SCD, ZNF114, CRYBG3, and PPIF are determined.

49. A kit according to any of claims 41 to 48 wherein the target molecule for one or more of the biomarkers is nucleic acid, preferably an mRNA or ncRNA molecule.

50. A kit according to claim 49 wherein one or more of the binding partners are selected from the group consisting of complementary nucleic acids and aptamers, and preferably wherein one or more of the binding partners are nucleic acid primers adapted to bind specifically to the mRNA, ncRNA, or cDNA transcripts of the biomarkers.

51. A kit according to any of claims 41 to 50 wherein the target molecule for one or more of the biomarkers is a protein, and the binding partner is selected from the group consisting of antibodies, antibody fragments and aptamers.

52. An assay device, the device comprising:

a) a loading area for receipt of a biological sample;

b) binding partners specific for target molecules representative of expression of at least four biomarkers selected from the group consisting of MCTP1, PIBF1, TMTC2, ANKRD32, CEPT1, ZNF114, CRYBG3, IFI16, PPIF, SCD, RP11-469M7.1, PTP4A1 and NRIP1; and

c) detection means to detect the levels of said target molecules present in the sample.

53. An assay device according to claim 52 for use in a method according to any of claims 1 to 40.

Figure 1

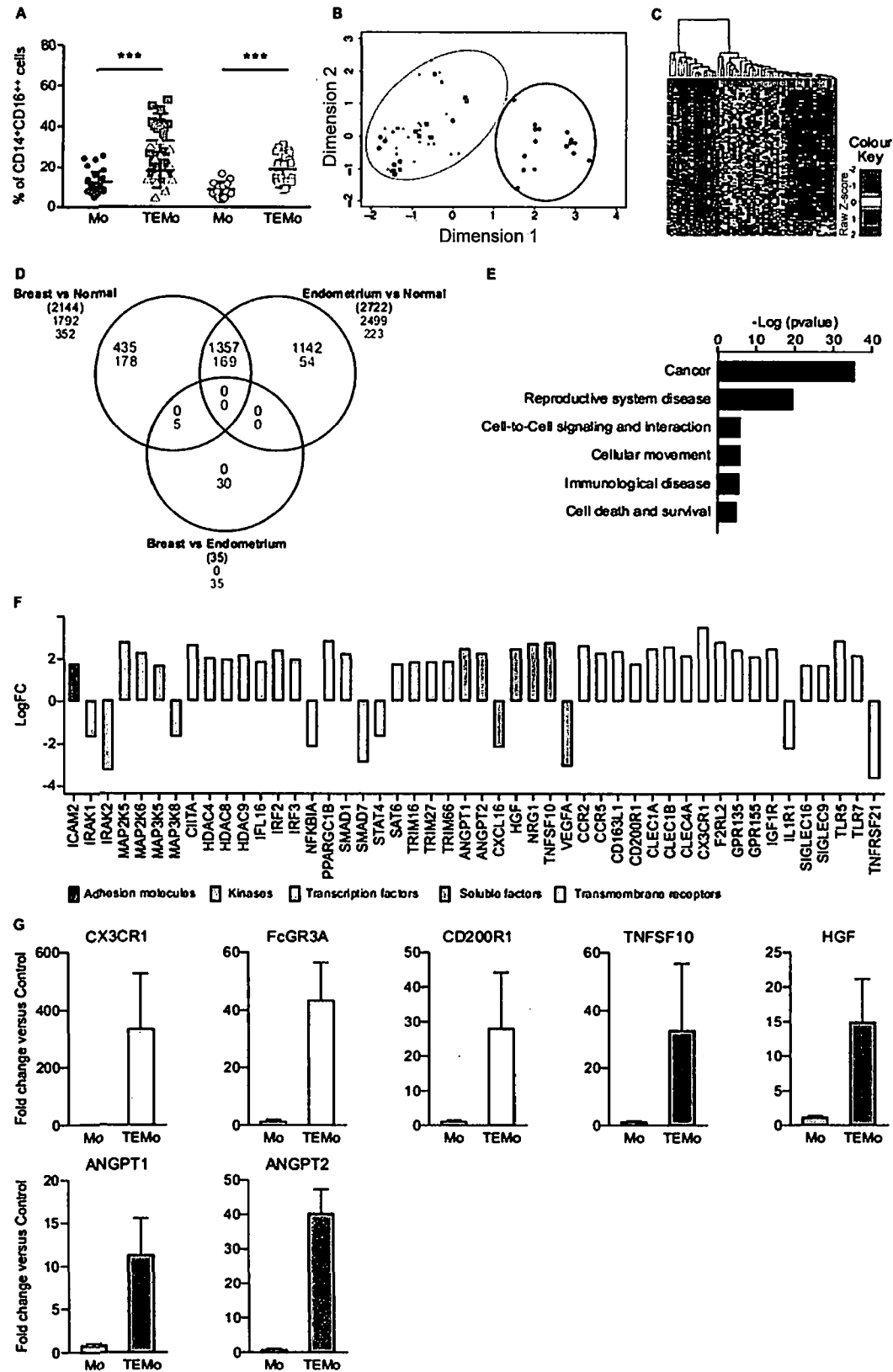


Figure 2

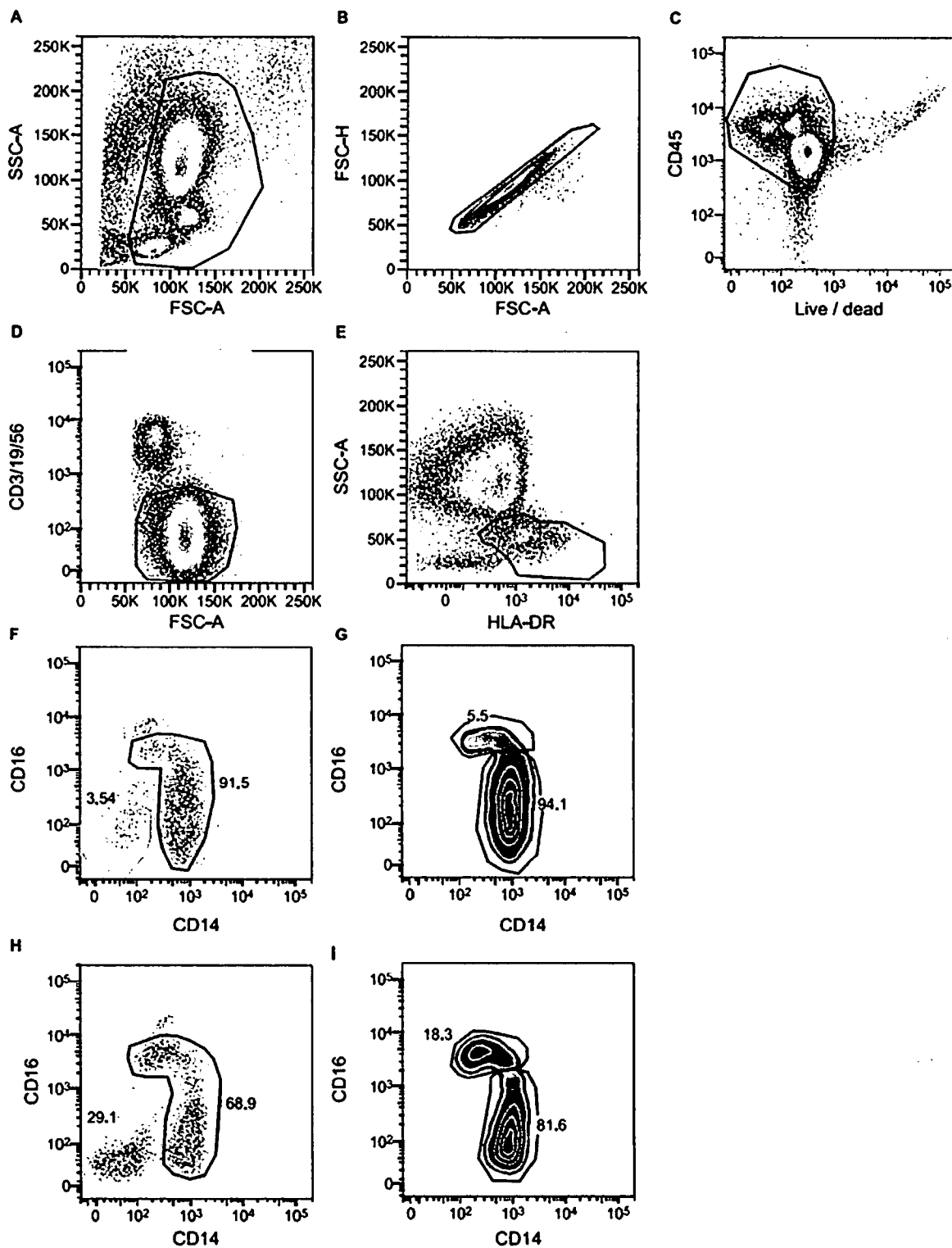


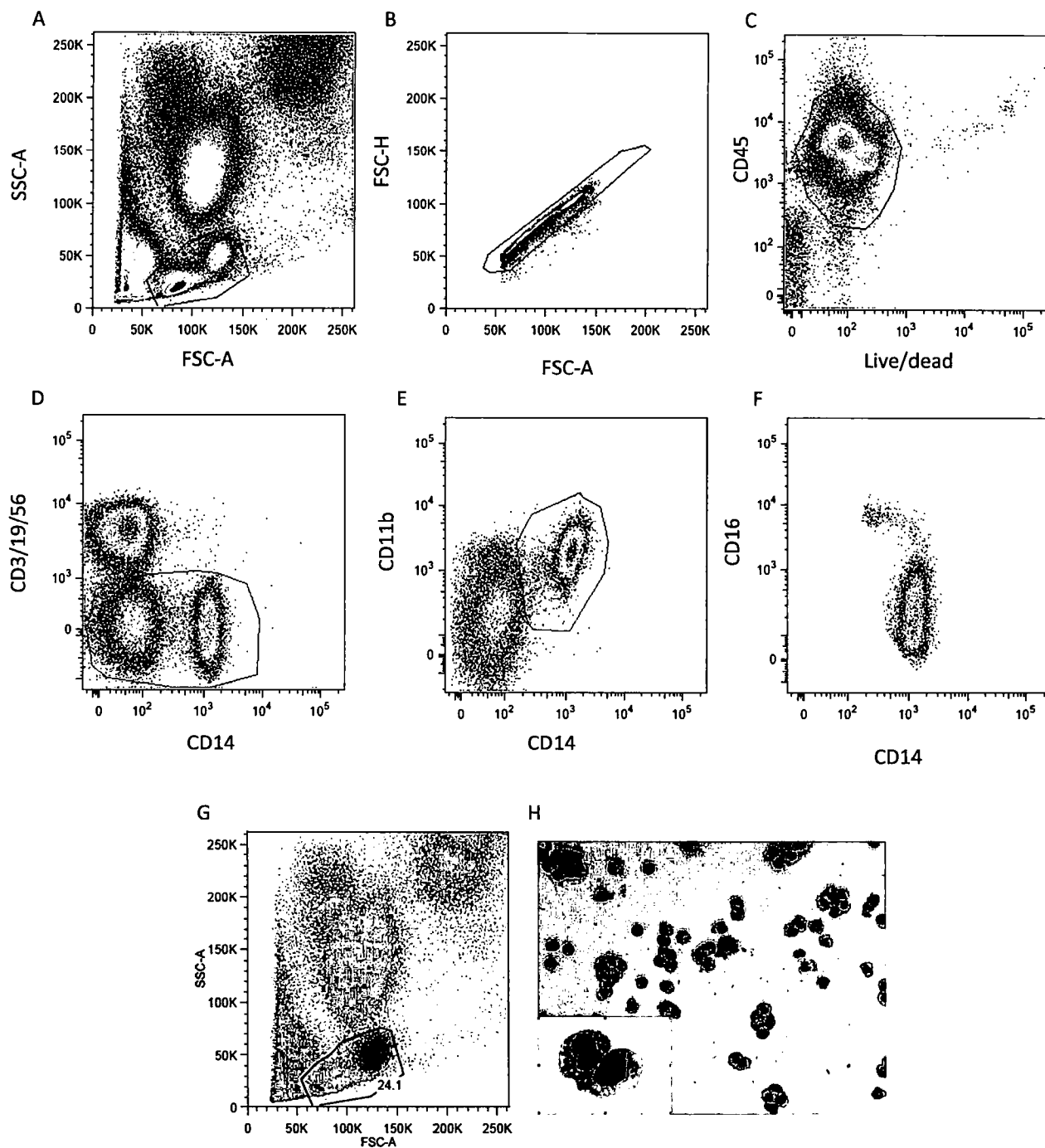
Figure 3

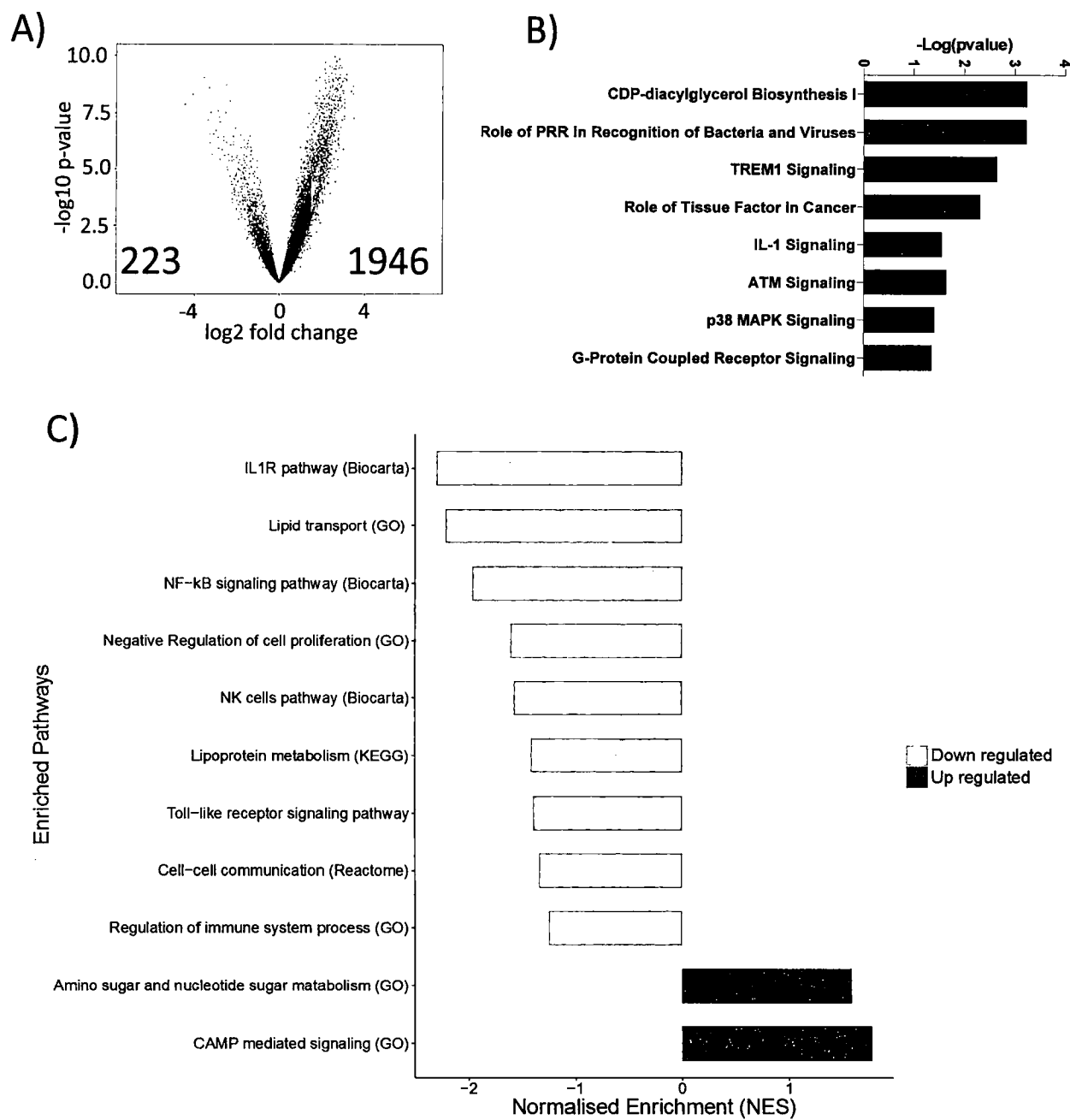
Figure 4

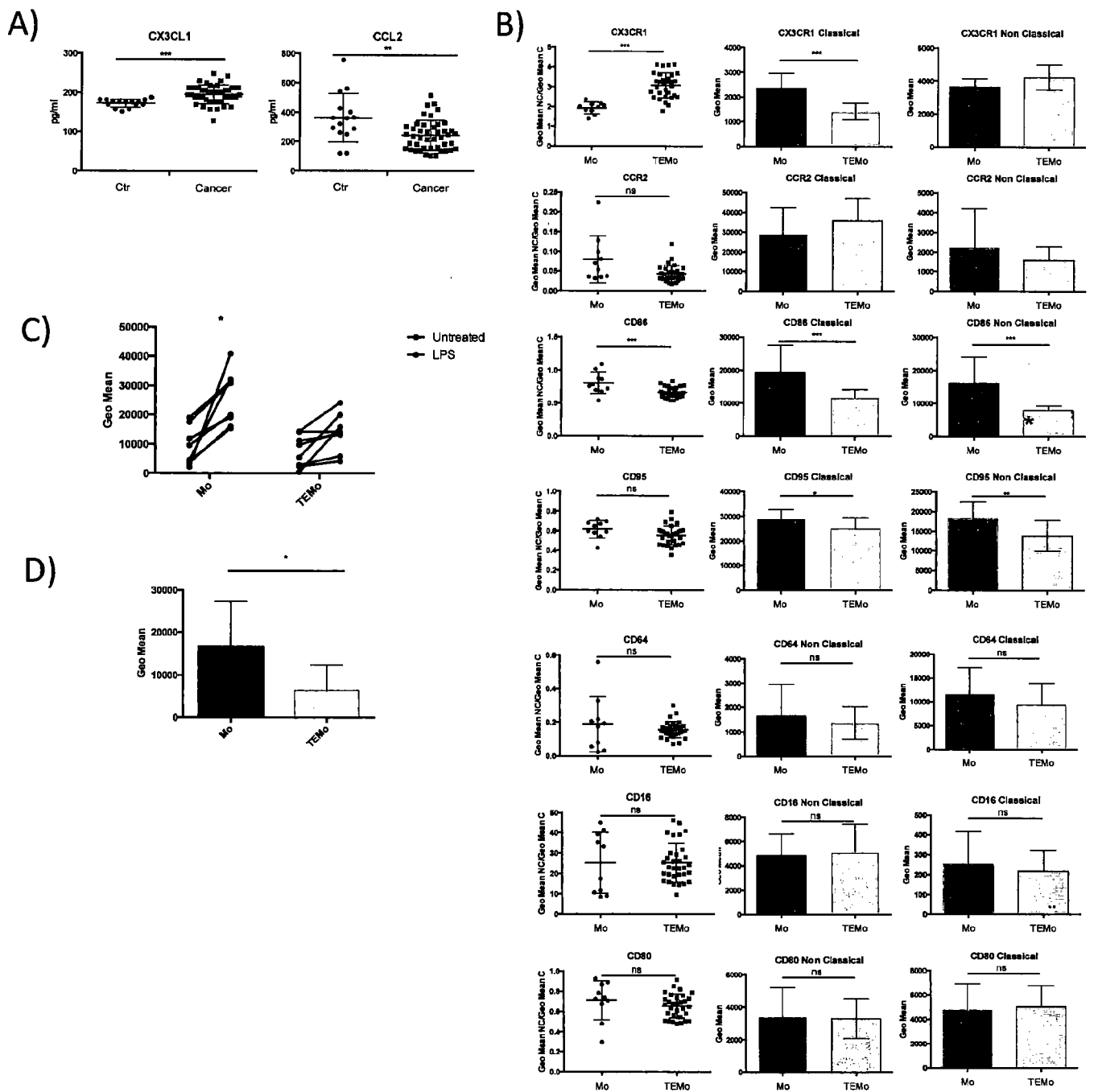
Figure 5

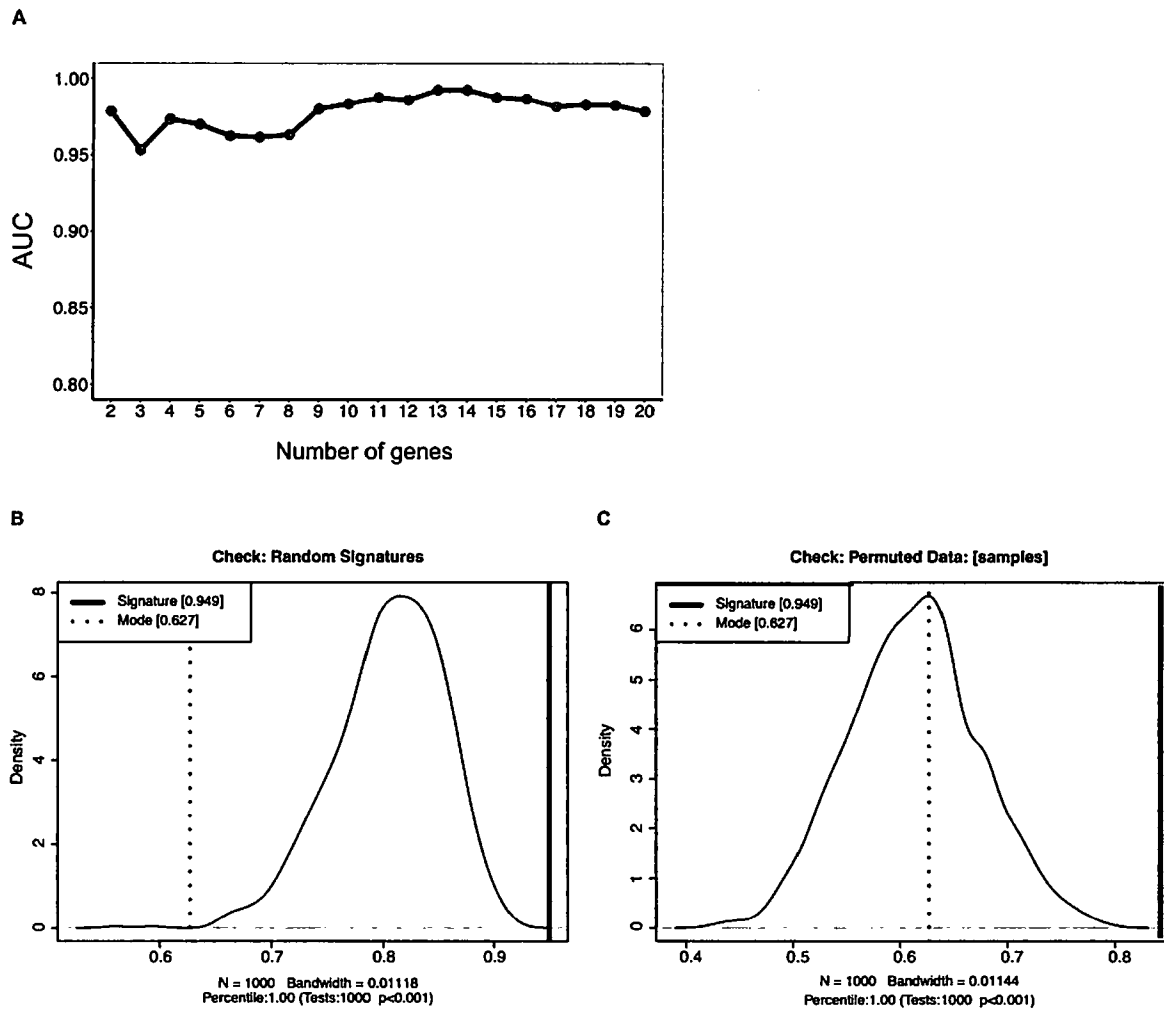
Figure 6

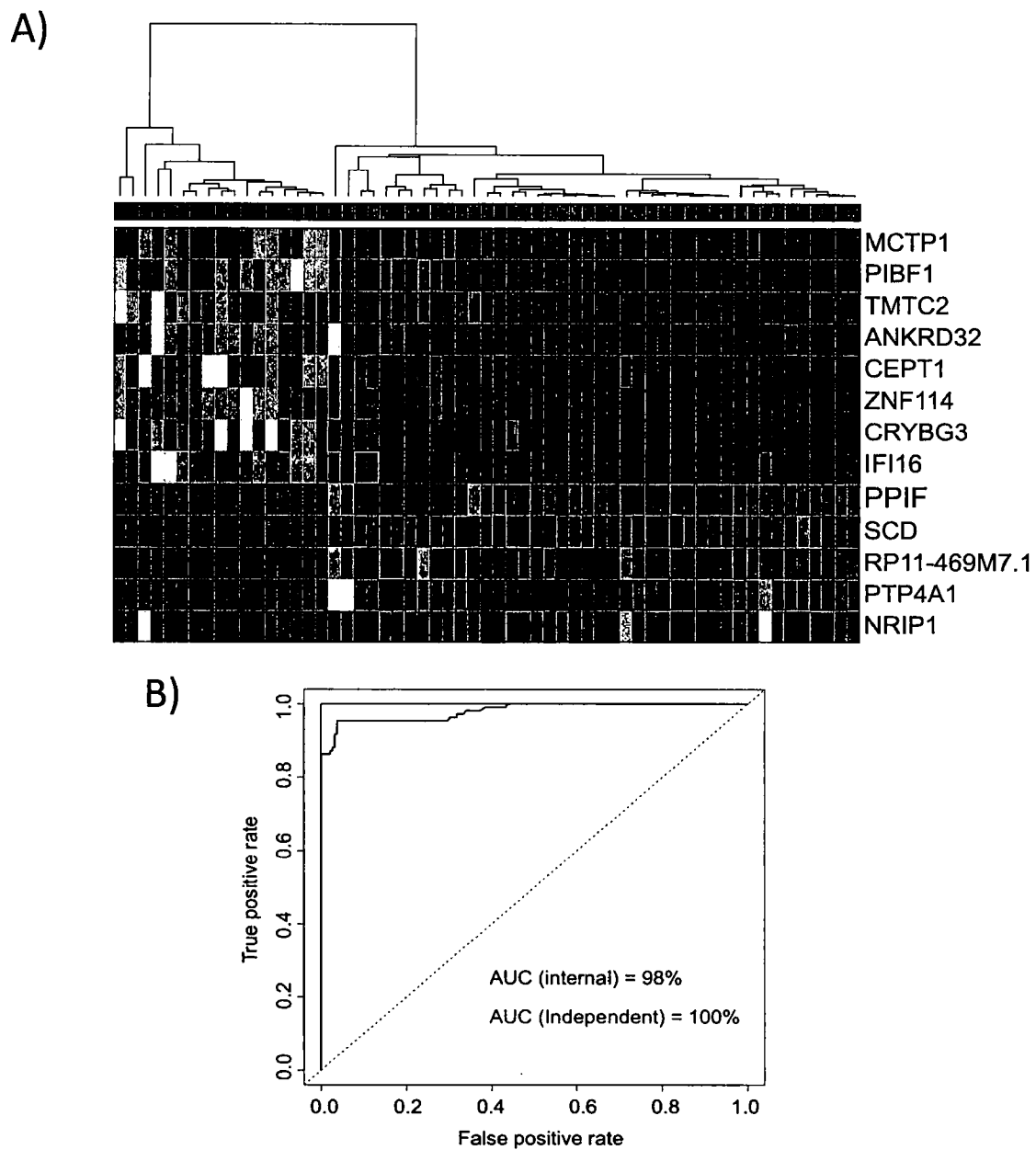
Figure 7

Figure 7 contd

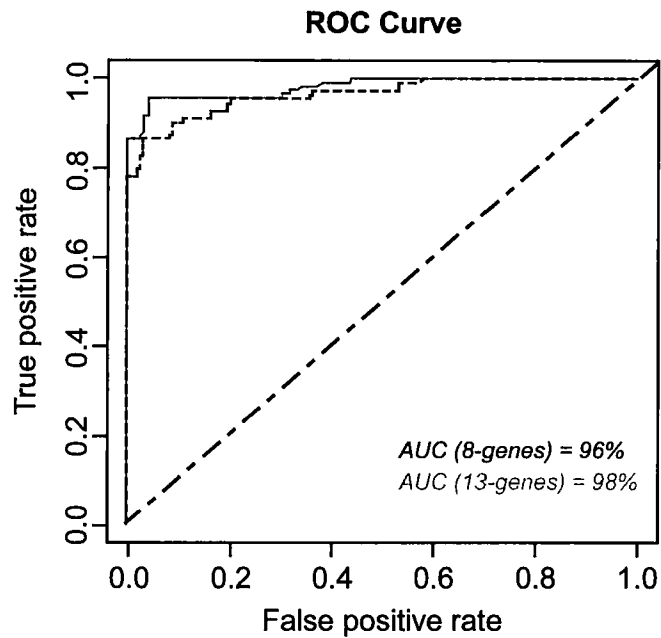
C)

Random Forest & chi-square	Training Validation (22 normal, 37 cancer)	External validation (5 normal, 14 cancer)
Accuracy	94%	100%
Sensitivity	98%	100%
Specificity	86%	100%
Positive predictive Value (PPV)	92%	100%
Negative Predictive Value (NPV)	97%	100%
Area under the curve (AUC)	98%	100%

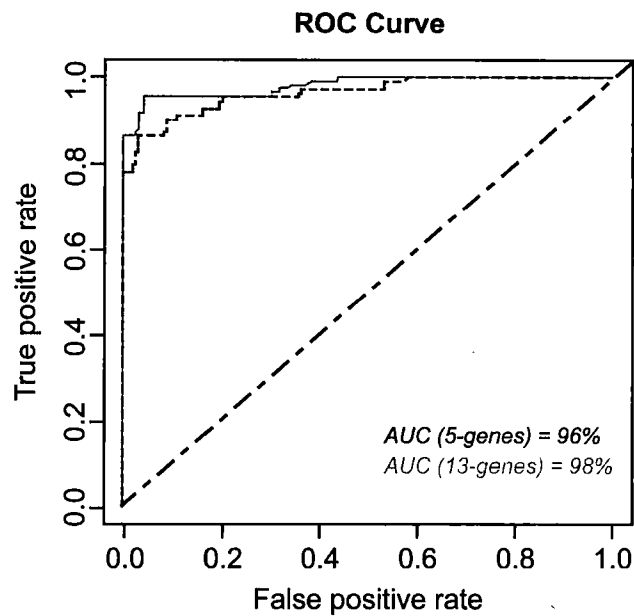
D)

Random Forest & chi-square	Negative validation / Lyme disease (13 normal, 28 Lyme disease)	Negative validation / Chronic periodontitis (5 normal, 5 disease)
Sensitivity	0%	0%
Specificity	100%	100%
Positive Predictive Value (PPV)	0%	0%
Negative Predictive Value (NPV)	100%	100%

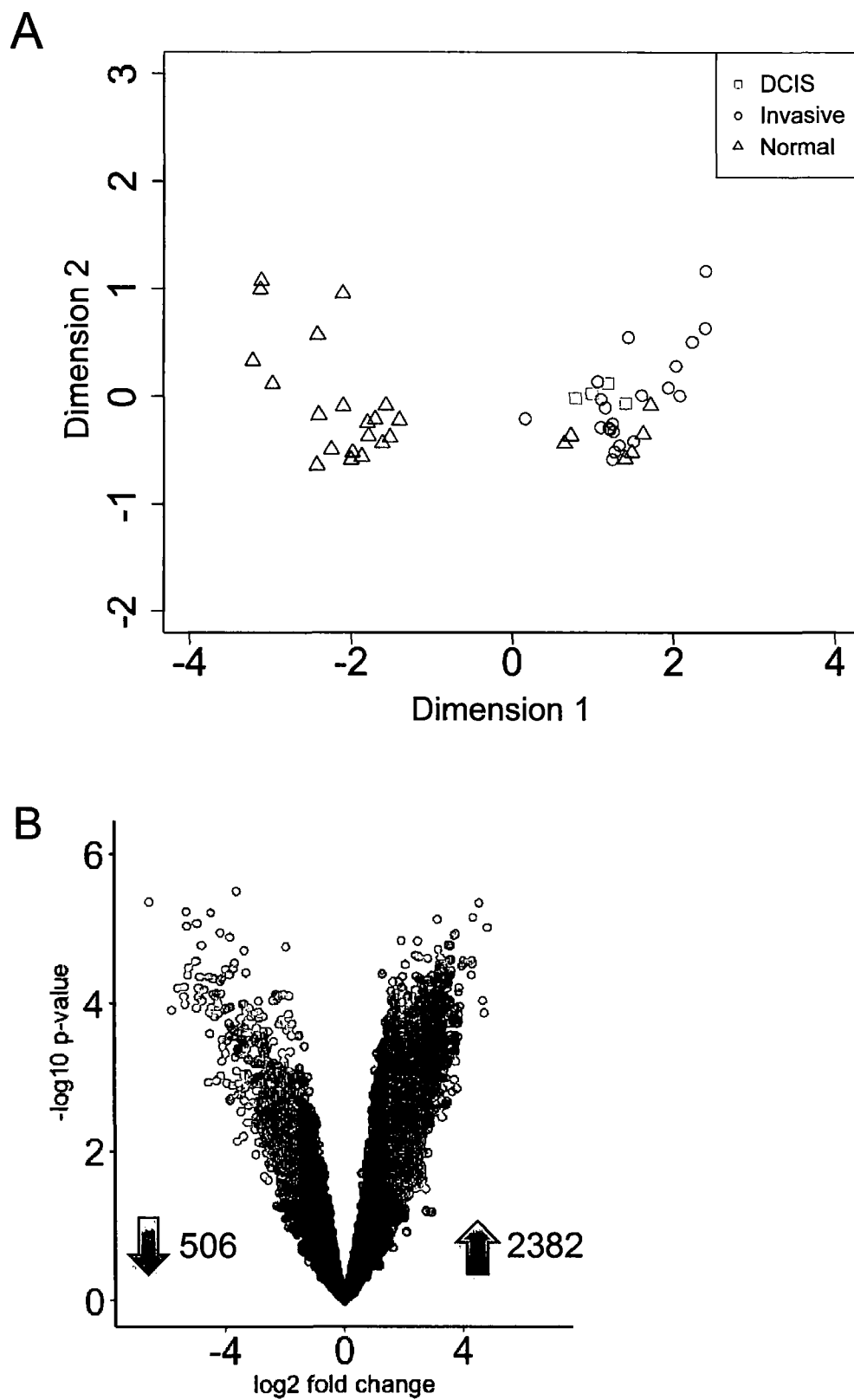
5

Figure 8**A****B**

Random Forest & chi-square	13-gene signature	8-gene signature
Accuracy	94%	93%
Sensitivity	98%	97%
Specificity	86%	86%
Positive predictive Value (PPV)	92%	92%
Negative Predictive Value (NPV)	96%	95%
Area under the curve (AUC)	98%	96%

Figure 9**A****B**

Random Forest & chi-square	13-gene signature	5-gene signature
Accuracy	94%	93%
Sensitivity	98%	97%
Specificity	86%	86%
Positive predictive Value (PPV)	92%	92%
Negative Predictive Value (NPV)	96%	95%
Area under the curve (AUC)	98%	96%

Figure 10

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2018/050400

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 02/44734 A1 (CISTEM BIOTECHNOLOGIES GMBH [AT]; SZEKERES BARTHO JULIA [HU]; NAGY ESZ) 6 June 2002 (2002-06-06) examples 3,6	1-8, 11-27, 30-40
X	----- Illumina: "HumanHT-12 v4 BeadChip Product Information", 1 January 2010 (2010-01-01), pages 1-2, XP055178072, Retrieved from the Internet: URL:http://www.illumina.com/documents/products/product_information_sheets/product_info_humanht-12.pdf [retrieved on 2015-03-20] the whole document ----- -/-	42-53



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

17 April 2018

Date of mailing of the international search report

26/04/2018

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

Gabriels, Jan

INTERNATIONAL SEARCH REPORT

International application No.

PCT/GB2018/050400

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed:
 - ☒ in the form of an Annex C/ST.25 text file.
 - ☐ on paper or in the form of an image file.
 - b. ☐ furnished together with the international application under PCT Rule 13~~ter~~.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
 - c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
 - ☐ in the form of an Annex C/ST.25 text file (Rule 13~~ter~~.1(a)).
 - ☐ on paper or in the form of an image file (Rule 13~~ter~~.1(b) and Administrative Instructions, Section 713).
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2018/050400

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2011/009637 A2 (GEADIC BIOTEC AIE [ES]; ABAL POSADA MIGUEL [ES]; DOLL ANDREAS [ES]; GI) 27 January 2011 (2011-01-27) claim 1 -----	1
A	WO 2013/110817 A1 (VIB VZW [BE]; LIFE SCIENCES RES PARTNERS VZW [BE]) 1 August 2013 (2013-08-01) claims 1-8 -----	1
A	WO 03/004989 A2 (MILLENIUM PHARMACEUTICALS INC [US]; LILLIE JAMES [US]; GANNAVARAPU MAN) 16 January 2003 (2003-01-16) PPIF and SCD; page 13, line 1 - line 4; claim 1 -----	1
A	WO 03/073911 A2 (UNIV GEORGETOWN [US]) 12 September 2003 (2003-09-12) Breast cancer specific gene IFI16 under-expressed in breast cancer; table 5 & DATABASE Geneseq [Online] 11 June 2007 (2007-06-11), "Breast cancer specific gene IFI16 under-expressed in breast cancer.", retrieved from EBI accession no. GSN:ACF79937 Database accession no. ACF79937 sequence -----	1
A	US 2006/088876 A1 (BAUER A R [US]) 27 April 2006 (2006-04-27) claim 1 -----	1,9,10, 28,29

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2018/050400

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0244734	A1	06-06-2002	AT 357662 T 15-04-2007
		AT 410753 B 25-07-2003	
		AU 2490102 A 11-06-2002	
		AU 2002224901 B2 09-11-2006	
		CA 2431837 A1 06-06-2002	
		DE 60127442 T2 29-11-2007	
		EP 1337850 A1 27-08-2003	
		ES 2283461 T3 01-11-2007	
		JP 4025645 B2 26-12-2007	
		JP 2004514909 A 20-05-2004	
		JP 2006230407 A 07-09-2006	
		US 2004142394 A1 22-07-2004	
		WO 0244734 A1 06-06-2002	
WO 2011009637	A2	27-01-2011	AU 2010275699 A1 02-02-2012
		CA 2767914 A1 27-01-2011	
		CN 102549169 A 04-07-2012	
		CN 105132529 A 09-12-2015	
		DK 2456889 T3 01-12-2014	
		EP 2456889 A2 30-05-2012	
		EP 2848700 A1 18-03-2015	
		ES 2523683 T3 28-11-2014	
		ES 2654340 T3 13-02-2018	
		HR P20141062 T1 02-01-2015	
		JP 5934093 B2 15-06-2016	
		JP 2013500001 A 07-01-2013	
		KR 20120047334 A 11-05-2012	
		PT 2456889 E 21-11-2014	
		RU 2012106432 A 27-08-2013	
		US 2012122726 A1 17-05-2012	
		US 2015337392 A1 26-11-2015	
		WO 2011009637 A2 27-01-2011	
WO 2013110817	A1	01-08-2013	AU 2013213563 A1 18-09-2014
		CA 2862559 A1 01-08-2013	
		EP 2807274 A1 03-12-2014	
		NZ 629555 A 24-12-2015	
		US 2015017159 A1 15-01-2015	
		WO 2013110817 A1 01-08-2013	
WO 03004989	A2	16-01-2003	AU 2002322280 A1 21-01-2003
		US 2003124128 A1 03-07-2003	
		US 2008311573 A1 18-12-2008	
		US 2010105051 A1 29-04-2010	
		US 2011092388 A1 21-04-2011	
		WO 03004989 A2 16-01-2003	
WO 03073911	A2	12-09-2003	AU 2003217774 A1 16-09-2003
		US 2004005644 A1 08-01-2004	
		WO 03073911 A2 12-09-2003	
US 2006088876	A1	27-04-2006	US 2006088876 A1 27-04-2006
		WO 2007053785 A2 10-05-2007	