Title: METHODS AND MEANS FOR SUBTYPING INVASIVE LOBULAR BREAST CANCER

Abstract: The invention relates to methods and means of typing a sample from a breast cancer patient, preferably an invasive lobular breast cancer patient. The invention further relates to methods of assigning treatment to a patient suffering from invasive lobular breast cancer, and to methods of treating a patient suffering from invasive lobular breast cancer.
Breast cancer is a heterogeneous disease and has traditionally been subdivided into distinct histological subtypes based on cell morphology. About 60-75% of breast cancers are invasive ductal carcinomas (IDC) (Guiu et al., 2014. Critical Rev Oncol/Hematol 92: 235-257). The next most common subtype is invasive lobular carcinoma (ILC), representing 5-15% of all breast cancers (Guiu et al., 2014. Critical Rev Oncol/Hematol 92: 235-257; Pestalozzi et al., 2008. J Clin Oncol 26: 3006-3014). ILC can be subdivided into five more specific histological subtypes (Iorfida et al., 2012. Breast Cancer Res Treatment 133: 713-723). ILCs are typically estrogen receptor (ER) and/or progesterone (PR) positive and exhibit frequent loss of the cellular adhesion molecule E-cadherin (CDH1) expression (Guiu et al., 2014. Critical Rev Oncol/Hematol 92: 235-257). A subset of ILCs is HER2 positive. ILCs have very similar survival to IDCs at both five and 10 years, but despite this similar survival, the clinical course is distinct: ILCs are three times more likely to metastasize to the peritoneum, gastrointestinal tract, and ovaries and are more frequently bilateral (Arpino et al., 2004. BCR 6: R149-156).

Gene expression-based molecular subtypes have been used as a reference to describe breast cancers (Perou et al., 2000. Nature 406: 747-752; Sorlie et al., 2003. PNAS USA 100: 8418-8423). Such subtypes are relatively well reflected in the immunohistochemistry (IHC)-based diagnosis used in the clinic (de Ronde et al., 2010. Breast Can Res Treatment 119: 119-126). However, they were mainly defined based on IDCs. Some molecular studies have been performed on ILC, using comparative genomic hybridization (Weigelt B, et al., 2010. J Pathol 220: 45-57) or gene expression profiling (Bertucci et al., 2008. Oncogene 27: 5359-5372), and more recently targeted DNA sequencing in advanced disease (Ross et al., 2013. Clin Cane Res: 19: 2668-2676). Two recent studies extensively

Treatment decisions made by oncologists for breast cancer are mainly based on results obtained in large trials, in which ILCs are only a minor subgroup. It is, therefore, not always the case that the conclusions from "breast cancer" trials also apply to ILC. For instance, ILCs were found to be less responsive to neo-adjuvant chemotherapy than IDCs, but this can possibly be explained by differences in molecular characteristics, particularly hormonal status and HER2 expression (Lips et al., 2012. Breast Cane Res Treatment 136: 35-43). In general, evidence that ILCs have a similar or worse clinical outcome as compared to IDCs are not robust (Guiu et al., 2014. Critical Rev Oncol/Hematol 92: 235-257), but the distinct pattern of distant metastases indicates that ILCs have unique biology.

There is thus a need for typing ILCs to identify subtypes that are more aggressive and life-threatening, and which should be treated with aggressive treatments. In addition, knowledge about subtypes may help to identify subtypes that are more sensitive to a drug or treatment regimen than others.

The invention therefore provides a method of typing a sample from a breast cancer patient, the method comprising providing a sample comprising RNA expression products from breast cancer cells from the patient, determining a level of RNA expression for at least two genes that are selected from Table 1 in said sample, comparing said determined level of RNA expression of the at least two genes to the level of expression of the genes in a reference sample; and typing said sample based on the comparison of the determined levels of RNA expression.

A sample comprising RNA expression products from a breast cancer cell of a patient is preferably provided after the removal of all or part of a breast cancer sample from the patient during surgery biopsy. For
example, a sample comprising RNA may be obtained from a needle biopsy sample or from a tissue sample comprising breast cancer cells that was previously removed by surgery. The surgical step of removing a relevant tissue sample, in this case a breast cancer sample, from an individual is not part of a method according to the invention.

A sample from a breast cancer patient comprising RNA expression products from a tumor of the patient can be obtained in numerous ways, as is known to a skilled person. For example, the sample can be freshly prepared from cells or a tissue sample at the moment of harvesting, or it can be prepared from samples that are stored at -70°C until processed for sample preparation. Alternatively, tissues or biopsies can be stored under conditions that preserve the quality of the protein or RNA. Examples of these preservative conditions are fixation using e.g. formaline and paraffin embedding, RNase inhibitors such as RNAsin® (Pharmingen) or RNasecure® (Ambion), aqueous solutions such as RNALater® (Assuragen; US06204375), Hepes-Glutamic acid buffer mediated Organic solvent Protection Effect (HOPE; DE 10021390), and RCL2 (Alphelys; WO04083369), and non-aqueous solutions such as Universal Molecular Fixative (Sakura Finetek USA Inc.; US7138226).

RNA may be isolated from a breast tissue sample comprising breast cancer cells by any technique known in the art, including but not limited to Trizol (Invitrogen; Carlsbad, California), RNAqueous® (Applied Biosystems/Ambion, Austin, Tx), Qiazol® (Qiagen, Hilden, Germany), Agilent Total RNA Isolation Lits (Agilent; Santa Clara, California), RNA-Bee® (Tel-Test. Friendswood, Texas), and Maxwell™ 16 Total RNA Purification Kit (Promega; Madison, Wisconsin). A preferred RNA isolation procedure involves the use of Qiazol® (Qiagen, Hilden, Germany). RNA can be extracted from a whole sample or from a portion of a sample generated by, for example section or laser dissection.
The level of RNA expression of a gene of Table 1 can be determined by any method known in the art. Methods to determine RNA levels of genes are known to a skilled person and include, but are not limited to, Northern blotting, quantitative Polymerase chain reaction (qPCR), also termed real time PCR (rtPCR), microarray analysis and next generation RNA sequencing. The term qPCR refers to a method that allows amplification of relatively short (usually 100 to 1000 basepairs) of DNA sequences. In order to measure messenger RNA (mRNA), the method is extended using reverse transcriptase to convert mRNA into complementary DNA (cDNA) which is then amplified by PCR. The amount of product that is amplified can be quantified using, for example, TaqMan® (Applied Biosystems, Foster City, CA, USA), Molecular Beacons, Scorpions® and SYBR® Green (Molecular Probes). Quantitative Nucleic acid sequence based amplification (qNASBA) can be used as an alternative for qPCR.

A preferred method for determining a level of RNA expression is microarray analysis. For microarray analysis, a hybridization mixture is prepared by extracting and labelling of RNA. The extracted RNA is preferably converted into a labelled sample comprising either complementary DNA (cDNA) or cRNA using a reverse-transcriptase enzyme and labelled nucleotides. A preferred labelling introduces fluorescently-labelled nucleotides such as, but not limited to, cyanine-3-CTP or cyanine-5-CTP. Examples of labelling methods are known in the art and include Low RNA Input Fluorescent Labelling Kit (Agilent Technologies), MessageAmp Kit (Ambion) and Microarray Labelling Kit (Stratagene).

A labelled sample preferably comprises two dyes that are used in a so-called two-colour array. For this, the sample is split in two or more parts, and one of the parts is labelled with a first fluorescent dye, while a second part is labelled with a second fluorescent dye. The labelled first part and the labelled second part are independently hybridized to a microarray.
The duplicate hybridizations with the same samples allow compensating for dye bias.

More preferably, a sample from a breast cancer patient is labelled with a first fluorescent dye, while a reference sample, for example a sample from a breast cancer pool or a sample from a relevant cell line or mixture of cell lines, is labelled with a second fluorescent dye. The labelled sample and the labelled reference are co-hybridized to a microarray. Even more preferred, a sample is labelled with a fluorescent dye and hybridized to a microarray without a reference sample.

The labelled sample can be hybridized against the probe molecules that are spotted on the array. A molecule in the labelled sample will bind to its appropriate complementary target sequence on the array. Before hybridization, the arrays are preferably incubated at high temperature with solutions of saline-sodium buffer (SSC), Sodium Dodecyl Sulfate (SDS) and bovine serum albumin (BSA) to reduce background due to nonspecific binding.

The arrays are preferably washed after hybridization to remove labelled sample that did not hybridize on the array, and to increase stringency of the experiment by reducing cross hybridization of the labelled sample to a partial complementary probe sequence on the array. An increased stringency will substantially reduce non-specific hybridization of the sample, while specific hybridization of the sample is not substantially reduced. Stringent conditions include, for example, washing steps for five minutes at room temperature 0.1x Sodium chloride-Sodium Citrate buffer (SSC)/0.005% Triton X-102. More stringent conditions include washing steps at elevated temperatures, such as 37 degrees Celsius, 45 degrees Celsius, or 65 degrees Celsius, either or not combined with a reduction in ionic strength of the buffer to 0,05x SSC or 0,01x SSC as is known to a skilled person.

Image acquisition and data analysis can subsequently be performed to produce an image of the surface of the hybridised array. For
this, the slide can be dried and placed into a laser scanner to determine the amount of labelled sample that is bound to a target spot. Laser excitation yields an emission with characteristic spectra that is indicative of the labelled sample that is hybridized to a probe molecule. In addition, the amount of labelled sample can be quantified.

The level of expression, preferably mRNA expression levels of genes depicted in Table 1, is preferably compared to levels of expression of the same genes in a reference sample. A reference sample is preferably an RNA sample isolated from a tissue of a healthy individual, preferably comprising breast cells. A preferred reference sample comprises a RNA sample from a relevant cell line or mixture of cell lines. The RNA from a cell line or cell line mixture can be produced in-house or obtained from a commercial source such as, for example, Stratagene Human Reference RNA. A further preferred reference sample comprises RNA isolated and pooled from normal adjacent tissue from cancer patients, preferably breast cancer patients.

A more preferred reference sample comprises an RNA sample from an individual suffering from breast cancer, more preferred from multiple individuals suffering from breast cancer. It is preferred that said multiple samples are pooled from more than 10 individuals, more preferred more than 20 individuals, more preferred more than 30 individuals, more preferred more than 40 individuals, most preferred more than 50 individuals. A most preferred reference sample comprises a pooled RNA sample that is isolated from tissue comprising breast cancer cells from multiple individuals suffering from breast cancer.

As an alternative, a static reference can be generated which enables performing single channel hybridizations for this test. A preferred static reference is calculated by measuring the median background-subtracted level of expression (rMeanSignal) of a gene across 5 hybridizations of a reference sample, preferably obtained from pooled breast
cancer samples, on a microarray. The level of expression may be normalized as is known a skilled person. Subsequently, log-ratios for each gene/probe hybridization is generated relative to the value of the static reference.

Typing of a sample can be performed in various ways. In one method, a coefficient is determined that is a measure of a similarity or dissimilarity of a sample with said reference sample. A number of different coefficients can be used for determining a correlation between the RNA expression level in an RNA sample from an individual and a reference sample. Preferred methods are parametric methods which assume a normal distribution of the data.

The result of a comparison of the determined expression levels with the expression levels of the same genes in at least one reference sample is preferably displayed or outputted to a user interface device, a computer readable storage medium, or a local or remote computer system. The storage medium may include, but is not limited to, a floppy disk, an optical disk, a compact disk read-only memory (CD-ROM), a compact disk rewritable (CD-RW), a memory stick, and a magneto-optical disk.

The expression data are preferably normalized. Normalization refers to a method for adjusting or correcting a systematic error in the measurements of detected label. Systemic bias results in variation by inter-array differences in overall performance, which can be due to for example inconsistencies in array fabrication, staining and scanning, and variation between labeled RNA samples, which can be due for example to variations in purity. Systemic bias can be introduced during the handling of the sample in a microarray experiment.

To reduce systemic bias, the determined RNA levels are preferably corrected for background non-specific hybridization and normalized using, for example, Feature Extraction software (Agilent Technologies). Other methods that are or will be known to a person of ordinary skill in the art, such as a dye swap experiment (Martin-Magniette
et al., Bioinformatics 21:1995-2000 (2005)) can also be applied to normalize differences introduced by dye bias. Normalization of the expression levels results in normalized expression values.

Conventional methods for normalization of array data include global analysis, which is based on the assumption that the majority of genetic markers on an array are not differentially expressed between samples [Yang et al., Nucl Acids Res 30: 15 (2002)]. Alternatively, the array may comprise specific probes that are used for normalization. These probes preferably detect RNA products from housekeeping genes such as glyceraldehyde-3-phosphate dehydrogenase and 18S rRNA levels, of which the RNA level is thought to be constant in a given cell and independent from the developmental stage or prognosis of said cell.

Therefore, a preferred method according to the invention further comprises normalizing the determined RNA levels of said at least two genes listed in Table 1 in said sample.

Said normalization preferably comprises median centering, in which the "centers" of the array data are brought to the same level under the assumption that the majority of genes are not changed between conditions. Said normalization preferably comprises Lowess (LOcally WEighted Scatterplot Smoothing) local regression normalization to correct for both print-tip and intensity-dependent bias.

In a preferred embodiment, genes are selected of which the RNA expression levels are largely constant between individual tissue samples comprising cancer cells from one individual, and between tissue samples comprising cancer cells from different individuals. It will be clear to a skilled artisan that the RNA levels of said set of normalization genes preferably allow normalization over the whole range of RNA levels. An example of a set of normalization genes is provided in WO 2008/039071, which is hereby incorporated by reference.
The sample from a breast cancer patient preferably comprises invasive lobular breast cancer cells.

Breast carcinomas are divided into 2 main types, ductal and lobular, based on their phenotypic appearance. Invasive ductal carcinoma (IDC) is the most common type of breast cancer, which originates in a duct. Invasive lobular carcinoma (ILC) accounts for about 10% of all breast cancers and originates in a lobule. Methods and means for differentiating between IDC and ILC are known in the art. For example, classical ILC is a low-grade tumor with little or no nuclear atypia and a low mitotic rate. ILCs are characterized by cytologically uniform cells with round nuclei and inconspicuous nucleoli, as well as discohesive architecture with a linear or non-linear growth pattern (Martinez and Azzopardi, 1979.Histopathology 3: 467-488; Dixon et al., 1982. Histopathology 6: 149-161).

In addition, E-cadherin is differentially expressed in lobular and ductal breast cancers. Membrane expression of E-cadherin is present in invasive and in situ ductal carcinomas, while almost all lobular carcinomas showed complete loss of expression of E-cadherin (Acs et al., 2001. Am J Clin Pathol 115: 85-98).

Therefore, routine pathology, including immuno-histochemical staining for E-cadherin, enables to determine whether a sample from a breast cancer patient comprises invasive lobular breast cancer cells, or invasive.

A method for typing breast cancer cells according to the invention preferably further comprises determining the presence of a mutation in the GATA3 gene and/or in the ERBB2 gene.

The term "GATA3", as used herein, refers to a gene on human chromosome 10 that is also known as GATABinding Protein 3, and GATA-Binding Factor 3. GATA3 encodes a transcription factor. Database references for GATA3 include Entrez Gene Number 2625 and Ensembl Number ENSG00000 107485. The gene encodes four alternatively spliced
niRNA molecules, having Reference Sequence (RefSeq) numbers
NM_00 1002295.1, NM_002051.2, XM_005252442.1 and XM_005252443.2.

The term "ERBB2", as used herein, refers to a gene on human chromosome 17 that is also termed proto-oncogene C-ErbB-2,
neuroblastoma/glioblastoma derived oncogene homolog 2, and proto-oncogene Neu. ERBB2 encodes a receptor tyrosine kinase. Database references for ERBB2 include Entrez Gene Number 2064 and Ensembl Number ENSG00000141736. The gene encodes ten alternatively spliced mRNA molecules, having Reference Sequence (RefSeq) numbers
NM_00 1005862.2, NM_00 1289936.1, NM_00 1289937.1, NM_00 1289938.1,
NM_004448.3., XM_005257 140.1, XM_00672 1765.1, XM_00672 1766.1,
XM_00672 1767.1, and XM_00672 1768.1.

The presence of a mutation in the GATA3 and/or in the ERBB2 gene is preferably determined by comparing a sequence of the GATA3 gene and/or the ERBB2 gene that is obtained from a breast cancer cell with a reference sequence. A reference sequence of GATA3 and/or ERBB2 corresponds to the wild type nucleotide sequence of GATA3 and/or ERBB2. As a model for the wild type sequences, reference is made to the indicated RefSeq numbers. However, a reference can also be derived from a non-affected individual such as, for example, a non-affected close relative.

A nucleotide sequence of GATA3 and/or ERBB2 that is mutated or altered compared to a nucleotide sequence of GATA3 and/or ERBB2 in a reference sample preferably refers to an altered nucleotide sequence that results in an alteration of the amino acid sequence of the encoded protein.

A nucleotide sequence of GATA3 and/or ERBB2 can be determined by any method known in the art, including but not limited to sequence analysis of a genomic region encoding GATA3 and/or ERBB2; sequence analysis of a mRNA product or a derivative of a mRNA product such as a cDNA product, of GATA3 and/or ERBB2.
Sequence determination can be performed by any method known in the art, including but not limited to dideoxy sequencing, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, and sequencing by hybridization, including hybridization with sequence-specific oligonucleotides and hybridization to oligonucleotide arrays. The nucleotide sequence of GATA3 and/or ERBB2 can also be determined by application of mutation analysis methods such as single stranded conformation polymorphism, DNA heteroduplex analysis, denaturing gradient gel electrophoresis and thermal gradient gel electrophoresis.

Sequence analyses can be performed either by direct sequence analysis of a relevant nucleic acid molecule comprising the GATA3 gene and/or ERBB2 gene, or a mRNA product thereof, or by indirect sequencing of a relevant nucleic acid after amplification of all or any part of the GATA3 and/or ERBB2 gene or a mRNA product thereof. Alternative direct or indirect methods comprise hybridization protection assay, allele-specific amplification, ligase-mediated detection, primer extension, and restriction fragment length analysis.

In an alternative embodiment, the presence of a mutation in GATA3 and/or ERBB2 can be determined by analysis of the encoded protein by, for example, protein sequence determination, two dimensional gel electrophoresis, multidimensional protein identification technology, ELISA, liquid chromatography-mass spectrometry (LC-MS), matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF), or the use of antibodies that interact with either a non-mutated wild type form of GATA3 or ERBB2, or with a mutated variant form of GATA3 or ERBB2.

A method for typing breast cancer cells according to the invention preferably further comprises determining a level of expression of PDL1, PD1, and/or CTLA4.

The term "PD1", as used herein, refers to a gene on human chromosome 2 that is also termed CD279, Programmed Cell Death 1.
PDCD1 and Systemic Lupus Erythematosus Susceptibility 2. PD1 encodes a cell surface membrane protein. Database references for PDL1 include Entrez Gene Number 5133 and Ensembl Number ENSG00000188389. The gene encodes three alternatively spliced mRNA molecules, having Reference Sequence (RefSeq) numbers NM_005018.2, XM_006712573.1, and XM_006724986.1.

The term "PDL1", as used herein, refers to a gene on human chromosome 9 that is also termed CD274, Programmed Cell Death 1 ligand 1 and B7 homolog 1. PDL1 encodes a type I transmembrane protein. Database references for PDL1 include Entrez Gene Number 29126 and Ensembl Number ENSG00000120217. The gene encodes three alternatively spliced mRNA molecules, having Reference Sequence (RefSeq) numbers NM_001267706.1, NM_014143.3, and XM_006716759.1.

The term "CTLA4", as used herein, refers to a gene on human chromosome 2 that is also termed Cytotoxic T-Lymphocyte-Associated Protein 4, CD2, CELIAC3, GSE, and CD152. CTLA4 encodes a member of the immunoglobulin superfamily. Database references for CTLA4 include Entrez Gene Number 1493 and Ensembl Number ENSG00000163599. The gene encodes two alternatively spliced mRNA molecules, having Reference Sequence (RefSeq) numbers NM_001037631.2 and NM_005214.4.

Methods to determine a level of expression of PDL1, PD1, and/or CTLA4 include two dimensional gel electrophoresis, multidimensional protein identification technology, ELISA, liquid chromatography-mass spectrometry (LC-MS), matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF), and the in situ use of antibodies that specifically interact with PDL1, PD1, and/or CTLA4. Said antibodies are preferably directly or indirectly labeled with a detectable label, for example a fluorescent or radioactive label. The term "indirect labeling" is known to the skilled person and includes the use of labeled secondary antibodies that react with the antibodies that specifically interact with
PDL1, PD1, and/or CTLA4, and/or the use of biotin-streptavidin interactions.

The level of expression of expression of PDL1, PD1, and/or CTLA4 is preferably compared to a level of expression in a reference, which preferably is sample isolated from a tissue of a healthy individual, preferably comprising breast cells. A more preferred reference comprises protein expression products from an individual suffering from breast cancer, more preferred from multiple individuals suffering from breast cancer. It is preferred that said multiple samples are pooled from more than 10 individuals, more preferred more than 20 individuals, more preferred more than 30 individuals, more preferred more than 40 individuals, most preferred more than 50 individuals. As an alternative, a static reference can be generated which enables performing single channel determinations such as ELISA for this test. The level of expression may be normalized as is known a skilled person.

In a preferred method of the invention, said at least two genes that are selected from Table 1 comprise GTPase, IMAP Family Member 6 (GIMAP6), also termed Immune Associated Nucleotide 2, Immune Associated Nucleotide 6, Entrez Gene: 474344, and Purinergic Receptor P2Y, G-Protein Coupled, 14 (P2RY14), also termed BPR105, GPR105, or KIAA0001, Entrez Gene: 9934. It was found that an increase of expression of GIMAP6 and P2RY14 in a sample comprising expression products of a breast cancer cell, when compared to an average level of expression of GIMAP6 and P2RY14 in ILCs, is indicative of an Immune Related (IR) subtype.

Said at least two genes that are selected from Table 1 preferably comprise GIMAP6, P2RY14 and CD79B (CD79b Molecule, Immunoglobulin-Associated Beta, Ig-beta, IGB, B-Cell-Specific Glycoprotein B29, Entrez Gene: 974); more preferably GIMAP6, P2RY14, CD79B and GIMAP5 (GTPase, IMAP Family Member 5, IROD, IAN4L1, IMAP3, Immune
Associated Nucleotide 4 Like 1, Entrez Gene: 55340), more preferably GIMAP6, P2RY14, CD79B, GIMAP5 and GIMAP8 (GTPase, IMAP Family Member 8, Immune-Associated Nucleotide-Binding Protein 9, Human Immune Associated Nucleotide 6, Entrez Gene: 155038), more preferably GIMAP6, P2RY14, CD79B, GIMAP5, GIMAP8 and LOC 100290 115 (Entrez Gene: 100290115), more preferably GIMAP6, P2RY14, CD79B, GIMAP5, GIMAP8, LOC 100290 115 and IL33 (Interleukin 33, Nuclear Factor For High Endothelial Venules (NF-HEV), C9orf26, DVS27-Related Protein), more preferably GIMAP6, P2RY14, CD79B, GIMAP5, GIMAP8, LOC 100290 115, IL33 and SEL1L3 (Sel-1 Suppressor Of Lin-12-Like 3, Suppressor Of Lin-12-Like Protein 3, KIAA0746 Proteinl, Entrez Gene: 23231), more preferably GIMAP6, P2RY14, CD79B, GIMAP5, GIMAP8, LOC 100290 115, IL33, SEL1L3 and SLC9A9 (Solute Carrier Family 9, Subfamily A (NHE9, Cation Proton Antiporter 9), Member 9, Putative Protein Product Of NblaOO 118, NHE9, Sodium/Hydrogen Exchanger 9, Entrez Gene: 285195).

Said at least two genes preferably include B4GALNT4 (Beta-1,4-N-Acetyl-Galactosaminyl Transferase 4, betaGT4, Beta-1,4-N-Acetylgalactosaminyltransferase IV, Entrez Gene: 338707), more preferably B4GALNT4 and TTLL12 (Tubulin Tyrosine Ligase-Like Family, Member 12, dJ526I14.2, KIAA0153, Entrez Gene: 23170), more preferably B4GALNT4, TTLL12 and NKAIN1 (Na+/K+ Transporting ATPase Interacting 1, FAM77C, Family With Sequence Similarity 77, Member C, Na(+) / K(+) -Transporting ATPase Subunit Beta-1-Interacting Protein 1, Entrez Gene: 79570), more preferably B4GALNT4, TTLL12, NKAIN1 and MMP11 (Matrix Metallopeptidase 11, stromelysin 3, Entrez Gene: 4320). A decrease of expression of B4GALNT4, TTLL12, NKAIN1 and MMP11 in a sample comprising expression products of a breast cancer cell, when compared to an average level of expression in ILC, is indicative of an Immune Related (IR) subtype.
Said breast cancer cell preferably is an invasive lobular breast cancer cell.

In a further preferred method according to the invention, a level of RNA expression of at least five genes from Table 1 is determined.

In a further preferred method according to the invention, a level of RNA expression of all genes from Table 1 is determined.

A preferred method according to the invention further comprises determining a metastasizing potential of the sample from the patient. Methods of determining a metastasizing potential of a sample comprising breast cancer cells are known in the art and include, for example, a 70 gene Amsterdam profile (MammaPrint ®; (van ’t Veer et al., 2002. Nature 415: 530) and other multigene expression tests such as a 21 gene signature (Oncotype DX®, Paik et al., 2004. New Engl J Med 351: 2817) and EndoPredict (Filipits et al., 2011. Clinical Cancer Research 17: 6012). A preferred method of determining a metastasizing potential is the 70 gene Amsterdam profile.

Using the methods of the invention, the presence of mutations in GATA3, loss of part or all of the long arm of chromosome 1, and enhanced expression of PDL1, PD1 and CTLA4, when compared to a reference, were found to be characteristic for an Immune Related (IR) subtype of invasive lobular breast cancer cells. In contrast, the presence of mutations in ERBB2, gain of part or all of the long arm of chromosome 8 and loss of part or all of the long arm of chromosome 11, and enhanced expression of progesterone receptor, estrogen receptor, GATA3, and fibronectin 1 (FN1), when compared to a reference, were found to be characteristic for an Hormone Response (HR) subtype of invasive lobular breast cancer cells.

Methods for determining a level of expression of progesterone receptor (PR), estrogen receptor (ER) and fibronectin 1 (FN1) are known in the art, including IHC, determination of mRNA levels, and the use of gene profiles that are indicative of the activity of PR and/or ER, for example as is
described in US patent application 13/546,755, which is incorporated herein by reference.

The analyses indicated herein above resulted in the identification of an Immune Related (IR) subtype for about 2/3 of all invasive lobular breast cancer cells, and an Hormone Response (HR) subtype for about 1/3 of all invasive lobular breast cancer cells.

Based on this classification, the method according to the invention may further comprise the determination of a strategy for treatment of the patient.

The invention therefore provides a method of assigning treatment to a patient suffering from breast cancer, comprising (a) typing a sample comprising RNA expression products from breast cancer cells from the patient according to a method of the invention, (b) classifying said sample as Immune Related (IR) or Hormone Response (HR), and (c) assigning anti-Estrogen Receptor directed therapy to a patient of which the sample is classified as HR.

Enhanced expression of ER in cancer cells of the HR subtype may be indicative for applying anti-estrogen receptor-directed therapy (anti-ER), for example comprising a selective estrogen receptor modulator (SERM) such as tamoxifen to said breast cancer patient. Tamoxifen and tamoxifen derivatives such as toremifene, are known antagonistic compounds of the estrogen receptor. Tamoxifen may be administered at 20 to 200 mg/kg per day, for example as Tamoxifen Citrate Tablets USP for oral administration. Toremifene similarly can be administered as toremifene citrate at 10 to 800 mg/d orally.

Further preferred SERMs include fulvestrant \((7\alpha,17\beta)-7\{(4,4,5,5,5-pentafluoropentyl)sulfinyl\}nonyl\)estra-1,3,5(10)-triene-3,17-diol, which is an estrogen receptor antagonist with no agonist effects, which works by down-regulating the estrogen receptor. It is administered as a once-monthly injection at 500 mg; raloxifene \((6\text{-hydroxy-2-(4-
hydroxyphenyl)-benzothiophen-3-yl]-[4-[2-(1-piperidyl)ethoxy]phenyl]-methanone). It is an estrogen receptor antagonist in breast cells, including breast cancer cells. It can be orally administered at 60-240 mg/kg/day; and lasofoxifene ((5R,6S)-6-phenyl-5-[4-(2-pyrrolidin-1-ylethoxy)phenyl]-5,6,7,8-tetrahydronaphthalen-2-ol). It is an estrogen receptor antagonist in breast cells, including breast cancer cells. It can be orally administered at 0.001 mg/kg - 1.0 mg/kg/day.

A further preferred anti-ER directed therapy comprises the administration of an aromatase inhibitor. These non-steroidal inhibitors inhibit the synthesis of estrogen via reversible competition for the aromatase enzyme. Preferred aromatase inhibitors include anastrozole (2,2'-[5-(1H-1,2,4-triazol-1-ylmethyl)-1,3-phenylene]bis(2-methylpropanenitrile) and exemestane (6-Methylideneandrosta-1,4-diene-3,17-dione). Anastrozole can be orally administered at 1.0 - 10 mg/day. Exemestane can be orally administered at 25-50 mg/day.

The invention therefore provides a method of assigning treatment to a patient suffering from breast cancer, comprising (a) typing a sample comprising RNA expression products from breast cancer cells from the patient according to a method of the invention, (b) classifying said sample as Immune Related (IR) or Hormone Response (HR), and (c) assigning therapy targeting PD1, PDL1 and/or CTLA4 and/or assigning therapy comprising at least one DNA damaging agent to a patient of which the sample is classified as IR.

The programmed cell death 1 (PD1) receptor is an inhibitory receptor expressed by T cells following prolonged exposure to antigens. Its primary ligand, PDL1 is often expressed within the tumor microenvironment, including IR invasive lobular breast cancer cells. The PDL1 receptor has a second ligand, PDL2 (also known as B7-DC or CD273), that is preferentially expressed by antigen-presenting cells. PDL1 negatively regulates the effector phase of T-cell responses after ligation of PDL1.
expressed within the tumor. Manipulation of the immune checkpoints, which include PD1, PDL1 and CTLA4, might induce T cell responses against tumors. Hence, immune checkpoint inhibitors can restore T cell responses and thus impede the evasion of IR invasive lobular breast cancer cells.

Similarly, activation of CTLA4 on cytotoxic T lymphocytes (CTLs) result in down regulation of the immune system. Antibodies that turn off this inhibitory mechanism allow CTLs to continue to destroy cancer cells.

Preferred therapy targeting PD1, PDL1 and/or CTLA4 comprises antibodies to PD1, PDL1 and/or CTLA4.

Known antibodies that react with PD1 include nivolumab (BMS-936558; Bristol-Myers Squibb, Princeton, NJ), pembrolizumab (MK-3475; lambrolizumab; Merck & Co., Kenilworth, NJ), pidilizumab (CureTech Ltd., Yavne, Israel) and AMP224 and AMP514 (Amplimmune Inc., Gaithersburg, MD).

Known antibodies that react with PDL1 include BMS-936559 (previously MDX-1105; Bristol-Myers Squibb, Princeton, NJ), MSB0010718C (EMD-Serono; Merck KGaA, Darmstadt, Germany), MED 14736 (AstraZeneca, London, UK), and MPDL 3280A (Roche, Basel, Switzerland).

Known antibodies that react with CTLA4 include ipilimumab (MDX-010 and MDX-101; Bristol-Myers Squibb, Princeton, NJ), tremelimumab (CP-675,206; Pfizer, New York, NY)

Antibodies to PD1, PDL1 and/or CTLA4 are preferably administered to an individual suffering from an IR invasive lobular tumor by parenteral injection and/or infusion, including intramuscular, intrapleural, intravenous, and subcutaneous injection and/or infusion. A typical treatment schedule or dosing regimen comprises parenteral administration, preferably intramuscular injection, of one dosage unit. The term "one dosage unit", as is used herein, refers to an effective amount of
the antibody or antibodies, meaning an amount that produces an effect on the cancer to be treated.

A preferred dosage unit of antibodies to PD1, PDL1 and/or CTLA4 is between 0.1 and 20 mg/kg, preferably between 0.5 and 10 mg/kg. Said dosage unit preferably is applied daily, more preferred every second day, more preferred twice a week, more preferred once a week, more preferred every 2 weeks, more preferred every 3 weeks, more preferred once a month.

Antibodies to PD1, PDL1 and/or CTLA4 are preferably administered together with immune-stimulants. Said immune-stimulants may comprise recombinant, synthetic and natural preparations. Preferred immune-stimulants interleukins such as IL-2, IL-7, and/or IL-12, and interferons, but may also include imiquimod (3-(2-methylpropyl)-3,5,8-triazatricyclo[7.4.0.02,6]trideca-l(9),2(6),4,7,10,12-hexaen-7-amine), synthetic cytosine phosphate-guanosine (CpG), glucans, and cellular membrane fractions from bacteria such as Mycobacterium spp., Corynebacterium parvum, C. granulosum, Bordetella pertussis and Neisseria meningitides, and/or the isolated membrane-bound product N-acetyl muramyl-L-alanyl-D-isoglutamine.

Therapy comprising at least one DNA damaging agent comprises an agent that induces damage to the genomic DNA of a cancerous cell. Said genomic DNA damage includes base modifications, single strand breaks and, preferably, crosslinks, such as intrastrand and interstrand cross-links. A preferred DNA damaging agent is selected from an alkylating agent such as nitrogen mustard, e.g. cyclophosphamide, mechlorethamine or mustine, uramustine and/or uracil mustard, melphalan, chlorambucil, ifosfamide; nitrosourea, including carmustine, lomustine, streptozocin; an alkyl sulfonate such as busulfan, an ethylenime such as N,N,N’-triethylenetriphosphoramid (thiotepa) and analogues thereof, a hydrazine/triazine such as dacarbazine, altretamine, mitozolomide, temozolomide, altretamine, procarbazine, dacarbazine and temozolomide; an
intercalating agent such as a platinum-based compound like cisplatin, carboplatin, nedaplatin, oxaliplatin and satraplatin; a topoisomerase I inhibitor such as irinotecan, topotecan, camptothecin and lamellarin D, anthracyclines such as doxorubicin, daunorubicin, epirubicin and idarubicin; mitomycin-C, dactinomycin, bleomycin, adriamycin, mithramycin, and poly ADP ribose polymerase (PARP)-inhibitors such as 3-aminobenzamide, AZD-2281, AGO14699, ABT-888, and BMN-673.

A preferred DNA damaging agent comprises bleomycin, cisplatin and/or a topoisomerase 1 inhibitor selected from irinotecan (Pfizer) and topotecan (GlaxoSmithKline), or active metabolites thereof such as SN-38 \((+)-(4S)-4,11\text{-dihydroxy-1H-pyranol}[3',4':6,7]\text{-indolizino}[1,2-b]quinohne-3,14(4H,12H)-dione\), which is the active metabolite of the topoisomerase-I inhibitor irinotecan.

Bleomycin is preferably administered at 0.1-1 USP units/kg, preferably at 0.2-0.5 USP units/kg, whereby USP specifies a potency of 1.5-2 USP units per mg weight. Bleomycin is preferably administered daily, for a period of at least 4 days, more preferred every second day, more preferred twice a week, more preferred once a week. Bleomycin is preferably administered by parenteral injection and/or infusion, including intrapleural, intravenous, intramuscular and subcutaneous injection and/or infusion.

Cisplatin is preferably administered at 10-100 mg/m\(^2\), preferably at 50-70 mg/m\(^2\). Cisplatin is preferably administered once a week, more preferred once every two weeks, more preferred once every four weeks. Cisplatin is preferably administered by parenteral injection and/or infusion, preferably intravenous injection.

Irinotecan is preferably administered at 25-500 mg/m\(^2\), preferably at 100-300 mg/m\(^2\). SN-38 is preferably administered from about 2 mg/m\(^2\) to about 150 mg/m\(^2\), preferably between 50-100 mg/m\(^2\). Irinotecan and SN-38 are preferably administered once a week, more preferred once every two weeks, more preferred once every three weeks. Irinotecan and SN-38 are
preferably administered by parenteral injection and/or infusion, preferably intravenous infusion.

Additional therapy that may be assigned to a patient of which a cancer sample is classified as IR, either alone or in combination with therapy targeting PDL, PDL1 and/or CTLA4 and/or therapy comprising at least one DNA damaging agent, comprises a cyclin dependent kinase (CDK) inhibitor, preferably AT-7519 (4-[(2,6-dichlorobenzene)amido]-N-(piperidin-4-yl)-IH-pyrazole-3-carboxamide; Astex, Cambridge, UK), an inhibitor of nicotinamide phosphoribosyltransferase, preferably FK866 (E)-N-(4-(1-benzoylpiperidin-4-yl)butyl)-3-(pyridin-3-yl)acrylamide; Apoxis, Lausanne, Switzerland) and/or an inhibitor of kinesin spindle protein, preferably SB-715992 (N-(3-Aminopropyl)-N-[(IR)-l-[7-chloro-3,4-dihydro-4-oxo-3-(phenylmethyl)-2-quinazolinyl]-2-methylpropyl]-4-methylbenzamide; Cytolinetics, South San Francisco, USA).

The invention therefore provides a method of treating a patient suffering from breast cancer, comprising (a) typing a sample comprising RNA expression products from breast cancer cells from the patient according to a method of the invention, (b) classifying said sample as Immune Related (IR) or Hormone Response (HR), and (c) treating a patient of which the sample is classified as HR with anti-Estrogen Receptor directed therapy.

The invention further provides a method of treating a patient suffering from breast cancer, comprising (a) typing a sample comprising RNA expression products from breast cancer cells from the patient according to the method according to the invention; (b) classifying said sample as Immune Related (IR) or Hormone Response (HR); (c) treating a patient of which the sample is classified as IR with therapy targeting PDL, PDL1 and/or CTLA4 and/or with therapy comprising at least one DNA damaging agent.
For the purpose of clarity and a concise description, features are described herein as part of the same or separate aspects and preferred embodiments thereof, however, it will be appreciated that the scope of the invention may include embodiments having combinations of all or some of the features described.

The invention will now be illustrated by the following examples, which are provided by way of illustration and not of limitation and it will be understood that many variations in the methods described and the amounts indicated can be made without departing from the spirit of the invention and the scope of the appended claims.
Figure legends

Figure 1 Gene expression clustering reveals two ILC subtypes. We defined two robust clusters of ILC samples by consensus clustering on the genome-wide gene expression data: Immune Responsive (IR) and Hormone Receptor signaling (HR). (A) Gene expression of top 250 upregulated and top 250 down-regulated genes in one subtype versus the other. (B) Candidate somatic variants are indicated (truncating mutations in dark and missense mutations in light), while white indicates the absence of variant. PI3K is blue when any of the PI3K pathway genes is mutated. Samples with a high somatic mutation rate (>=10) are indicated at the bottom.

Figure 2 Differential drug response
The boxplots show the drug response (In IC50) of the cell lines in the IR and HR subtypes for the six drugs with FDR<0.25. The red dotted line is the maximum screening concentration. The first three drugs in the figure are DNA-damaging agents.

Figure 3 Survival analysis of the IR and HR subtypes
Kaplan-Meier plot of the stratification of the cohort based on the IR and HR subtypes.

Figure 4 Classification performance of ranked genes
Accuracy of nearest centroid classifier from LOOCV is shown on Y-axis. The number of ranked genes, starting with selected is shown on X-axis.

Figure 5 Classification performance of random subsets of the 300 genes.
Examples

Example 1

Materials and methods

Study design

All patients with an ILC (based on pathology report) treated in the NKI-AVL since 1980 were extracted from the hospital database. We excluded all patients for which no fresh frozen (FF) tissue was available in the NKI-AVL tissue bank. We selected consecutive tumors without neo-adjuvant treatment and with a preference for those also without adjuvant hormonal therapy. All patients diagnosed with ILC (based on pathology report) treated in the Addenbrookes Hospital Cambridge UK since 1997 and with available FF material were included in this study. Clinical data were extracted from the Addenbrookes Hospital Cambridge database. In some cases, we also sourced FF tissue from adjacent matched normal tissues. Subsequently, we collected matched formalin fixed paraffin embedded (FFPE) tissue blocks for tissue microarray (TMA) construction. The NKI-AVL and Cambridge medical ethical committees approved the study and the use of anonymized archival tissue in this study.

Histopathology

Tumor samples were revised for tumor percentage of the FF material, histological grade and ILC sub-classification (Iorfida et al., 2012. Breast Cane Res Treatment 133: 713-723). TMAs were stained for ER, PR and HER2. Samples were defined to be ER-positive or PR-positive when 10% or more of the tumor cells showed positive staining of ER or PR respectively based on immunohistochemistry. Samples were defined HER2 positive for intensity >=2.

Survival analysis

We considered only breast cancer specific survival, due to the presence of competing events and (distant) recurrence free survival. For
patient stratification in Kaplan-Meier plots, we used Kaplan-Meier estimator and calculated p-values with the logrank test. Cox proportional hazards regression model was stratified by biobank and, unless otherwise specified, fitted with commonly used clinical variables: tumor size, grade, number of positive lymph nodes, treatment (hormonal, radiotherapy and/or adjuvant chemotherapy) and age at diagnosis.

Microarray Hybridization

Samples with RNA Integrity number (RIN) above 5 (2100 Bioanalyzer, Agilent Technologies) were selected for further analysis. RNA was amplified, labeled and hybridized to the Agendia custom-designed whole genome microarrays (Agilent Technologies) and raw fluorescence intensities were quantified using Feature Extraction software (Agilent Technologies) according to the manufacturer's protocols. Quality of the microarray process was monitored by an internal Agendia quality control (QC) model using >250 control probes related to background issues, general array signal intensity array uniformity and control genes (positive and negative). The microarray expression dataset was imported into R/Bioconductor software (R version 3.02, www.bioconductor.org) for pre-processing (Gentleman et al., 2004. Genome Biol 5: R80).

Gene expression normalisation

Feature signal intensities were processed and extracted according to the hmmA Bioconductor R package with background subtraction using an offset of 10. All probe intensities < 1 were set as missing values. These missing values were imputed by 10-nearest neighbor imputation (R-package impute) prior to analysis that cannot deal with them. The log2 transformed probe intensities were quantile normalized (Bolstad et al., 2003. Bioinformatics 19: 185-193) using limma. A principal component analysis showed a batch effect for biobank, and an additional batch of samples was identified that were cut at the same time (identifiers RL1110-RL1130). Both batch-effects were adjusted for using ComBat (Johnson et al.,
2007. Biostatistics 8: 118-127). Genes with multiple probes were summarized by the first principal component of a correlating subset (all probes with correlation to any other probe >0.5), if such a subset existed or by the most variable probe if no such subset existed. After summarizing by first principal component, signs were adjusted and scaling was carried out to match variance with the most variable probe of gene. Some genes (Wang et al., 2007. Genome Res 17: 1665-1674) showed a discordant signal over multiple probes, so were not summarized and thus kept as separate probes.

Gene expression clustering

We applied several different clustering algorithms on the top 1000 genes with highest median absolute deviance: hierarchical clustering with Pearson distance and ward D1, single, average and complete linkage, as well as non-negative matrix factorization (NMF). The ward D1, average and NMF methods gave stable clustering results as assessed by consensus clustering. When choosing two clusters, all three methods found largely the same two clusters. To define subtypes, we first performed consensus clustering with average linkage, two clusters, and 90% feature resampling. Then, the consensus matrix was hierarchically clustered with complete linkage and Euclidean distance. Finally, the resulting tree was cut at a quarter of maximum height, defining two big clusters. Samples not falling into one of these two clusters were not assigned to any cluster (n=42). NMF was done with the R package NMF, consensus clustering with the ConsensusClusterPlus package. To assign cell lines to clusters, we normalized together the raw gene expression data of cell lines and tumor samples. Then, we applied the same clustering approach described above, but cut the tree at maximum height such that all cell lines where assigned to a cluster. All tumor samples assigned to a cluster were assigned to the same cluster in both clustering results with and without cell lines.
**Gene optimization**

To determine the genes required for classification of ILC tumors in one of the two subtypes, we obtained a measure of feature importance from a random forest classifier (Breiman, 2001. Machine Learning 45: 5-32). We used 10 000 a bootstrapped trees with 100 randomly selected features each. Feature importance was measured by the mean decrease in Gini index upon resampling of that feature. Differential expression between the two subtypes was tested with a Mann-Whitney U test, p-values were corrected according to the Benjamini-Hochberg method for controlling the false discovery rate (FDR).

**Validation dataset**

We used the ILC samples of METABRIC (Curtis et al., 2012. Nature 486: 346-352) as a validation set for the gene expression subtypes. 76 samples are in common between RATHER and METABRIC, and were removed from the validation set, resulting in a set of 103 samples. We mapped probes to genes with the ReMOAT annotation (Barbosa-Morais et al., 2010. Nucl Acids Res 38: el741).

**Drug sensitivity**

Drug sensitivity was assessed on the Sanger cell line panel (internal version 17). We used the cell lines common in our ILC cell line panel and in the Sanger cell line panel. We used cell line AU565 instead of SK-BR-3, which is derived from the same patient. Among the 262 drugs, we focused our assessment on 88 agents that had measurement in at least three cell lines per subtype. With this dataset, we performed a two-sided t-test between the AUC of the dose-response curves of the cell lines in the two subtypes, correcting for multiple testing with the Benjamini-Hochberg method.

**Results**

**Genomic profiling of ILCs**
To explore the biology of invasive lobular carcinomas (ILCs), we performed comprehensive molecular profiling of 144 untreated tissue samples from primary ILC tumors with 6.8 years median clinical follow-up.

We first used gene expression data to explore the biological processes at play in ILC. Extensive clustering stability analysis based on a variety of clustering approaches identified two robust expression subtypes (Figure 1A). Based on a gene sub-sampling analysis, we found that we could stably classify 71% (102/144) of samples into one of two subtypes (63 and 39 samples). Among these 102 samples, 89 were characterized by mutation analysis and gene expression data and are represented in Figure 1.

A first subtype, referred to as immune responsive (IR), shows upregulation of genes characteristic of cytokine/chemokine signaling and normal-like breast cancer. GATA3 mutations were found to be enriched in the IR subtype, even though not significantly so (Figure 1B). The IR subtype shows up-regulation of a range of lymphoid signaling molecules at the mRNA level, such as TGFB2, IL11RA, TNFRSF17, CCL15, CCL14, CCR2, CD27, XCL2, IFNAR2, CD40LG, suggesting alterations in either the make-up or functional activity of immune cells within these tumors. Interestingly, negative regulators of the immune response PDCD1 (PD1), CD274 (PDL1) and CTLA4 are expressed at higher mRNA levels in the IR subtype (Data not shown). Further expression analyses suggested that the immune infiltrate in the IR subtype is enriched in T cells.

A second subtype, referred to as hormone receptor (HR) signaling subtype, shows up-regulation of estrogen receptor (ER) and progesterone receptor (PR). The HR subtype is characterized by ERBB2 mutations (Figure 1B). Collectively, these findings support elevated levels of hormone receptor signaling in this subtype.

The two ILC subtypes are validated in an independent dataset
To validate the existence of hormone receptor signaling and immune responsive ILC subtypes, we investigated the ILC samples of the METABRIC consortium (Curtis et al., 2012. Nature 486: 346-352). Using the same clustering approach de novo on the METABRIC gene expression data, we also identified two robust subtypes (data not shown).

**Modeling therapeutic response in cell line models**

To identify candidate therapeutic options, we profiled a set of 15 ILC cell lines as *in vitro* models. We used their gene expression profiles to map them to the IR (10 cell lines) and HR (5 cell lines). We then used the response data for 88 drugs on a subset of these cell lines, as provided by the Wellcome Trust Sanger Institute (Garnett et al., 2012. Nature 483: 570-575), to test for differential drug sensitivity between the subtypes. We retained six drugs at an FDR<0.25 (Figure 2). We found that cell lines of the IR subtype are more sensitive to three different DNA-damaging agents: Bleomycin, Cisplatin and the topoisomerase 1 inhibitor SN-38. This suggests that the IR subtype may be more responsive to chemotherapy. Conversely, we did not find any drugs showing a significantly larger effect in the HR subtype.

**Survival analysis**

Kaplan-Meier plot of the stratification of the cohort based on the IR and HR subtypes are shown in Figure 3. There is no significant difference in survival.

**Gene optimization**

A total of 300 genes were identified that were differentially expressed between IR and HR ILCs (see Table 1). In this Table, +1 indicates that the gene is upregulated in IR, while -1 indicates that the gene is downregulated in IR, when compared to a level of expression in HR, or an average level of expression in ILCs. To assess classification performance of a set of genes we trained a nearest centroid classifier. Classification
performance was measured by accuracy in leave-one-out cross validation (LOOCV).

As can be seen in Figure 4, maximum classification performance is reached with the top 9 ranked genes, while the top 2 ranked genes (GIMAP6 and P2RY14) already show good performance.

Using 100 random subsets for each number of genes indicated on the X-axis, it is evident that 2 randomly selected genes already allow discrimination between the two subsets (Figure 5). A good performance is obtained with 5 randomly selected genes from the group of genes indicated in Table 1. Performance keeps increasing up to 18 randomly selected genes.
<table>
<thead>
<tr>
<th>Seq ID No</th>
<th>Gene</th>
<th>Sequence</th>
<th>Gini</th>
<th>Sign</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G1MAP6</td>
<td>CCATCTTCCTAGACACGACAGCAGACATCTGAGAAGGCTTCAGCATTTCCTTCTGCTAACAGA</td>
<td>0.3728383</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>P2RY14</td>
<td>TTTTTCTGGAAAAACAGACGAGGATTTTACCTTCTGAGACATGCACTAGCTGTTAAGTTCTGACTAT</td>
<td>0.3140863</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>CD79B</td>
<td>CGCTGAAGGATGGTATCATCATGATTCAGCCTGCTATCATCCTTCTTTCATCTGACCTG</td>
<td>0.3070698</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>G1MAP5</td>
<td>TCTTTCTGTTGATAGAAGGATTTTACTTCTGGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2810331</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>G1MAP8</td>
<td>TCTTTCTGTTGATAGAAGGATTTTACTTCTGGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2735827</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>LOC100290115</td>
<td>TTCTTCTGGGTAGATACAGGATAAGATAGATACAGGGTAAAATGTTGAGCACTTTGTACA</td>
<td>0.2561993</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>IL33</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2527823</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>SEL1L3</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2527823</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>SLC9A9</td>
<td>TTCTTCTGTTGATAGAAGGATTTTACTTCTGGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2440941</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>C1C2</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2387606</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>ICAM2</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2387606</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>IL11RA</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2387606</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>G1MAP7</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2387606</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>LOC652159</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2296262</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>MEL1</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2174771</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>MAL</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2174771</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>LOC652106</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2097388</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>LOC642838</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2032299</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>POU2AF1</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2032299</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>IGHM1</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2032299</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>LOC100132941</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2032299</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>AM PD1</td>
<td>AAATGAAATACCTGAGAGCCAGAGCCAGAGCCAGTCCAGAGACATGGCATACGGTTACTGACTTAT</td>
<td>0.2032299</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Gene symbol</td>
<td>Genomic Location</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-------------</td>
<td>------------------</td>
<td>---------</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>MZB1</td>
<td>GCACATGTTTGGCATCTACTTGGGGCGATGGAGGACCCAGATCTATGAAAGCACCACCAAC</td>
<td>0.2031787</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>TGFBR2</td>
<td>CAGCTGATACCCTCCCCCATAGCTGTTGGGCCCATAGCTTGAAGTTGCTTCTTGAAG</td>
<td>0.2019837</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>ENST0000039868</td>
<td>GCACACAGCTCGAGACTCTAGCAGAGGGCCACATTCCACCCAGGAATGATGCTGATTCTGCTCAG</td>
<td>0.1992365</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>IGKV1.27</td>
<td>TGATCTATGCTGATTCCCGTGGGCGGTCCTATTCTGAGTGGGACGGTATGATGCTTGCAT</td>
<td>0.1961856</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>TCTCCTTGTGCAGCTTAGTTCAGTGCATGGTTGCTATGTAGCTGTTGGCAATAGGAGGGA</td>
<td>0.1925129</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>SPNS2</td>
<td>GCCTGCTAGTGTGAAATCTCAAGGCTACTTGGCTTGCAGCAGATATGCTGTCCTTGTAAT</td>
<td>0.191244</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>LOC1000000098</td>
<td>GCCACGGGATTTGAGGAAATGTTTGTACATCTCTACCTGCTAGGCTGAGTTTCCTGAT</td>
<td>0.190084</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>GIMAP1</td>
<td>AGGACTGATCATTCCGAGAAGCCCTGAGATCTGAGGAGAAGCCGAAAATTCATTTCCCTTCTGCTGATC</td>
<td>0.1861181</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>FLJ</td>
<td>GGGCCGTTTTAATGCTGACTTATTGGAGAAAGCGAAGCACGAGTGTAACATTTACGAGG</td>
<td>0.185451</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>B4GALNT4</td>
<td>TACAAGTACGGACTTTGACCGGGTTGGAGGAATGACACGGAGTGGGGATTTCCTGAGT</td>
<td>0.1843703</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>LOC10000000415</td>
<td>ACGCAGTCTACATCGGAGAAGCCCTGAGATCTGAGGAGAAGCCGAAAATTCATTTCCCTTCTGCTGATC</td>
<td>0.1814927</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>KIAA0125</td>
<td>CCACTTTTAAGATGGCTACCTTGGAGACCATATGAGTGGCTAGTCTGTGGTACATTTACGAGG</td>
<td>0.1789083</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>IGKV1.16</td>
<td>GACGAGTCCCGCTACTTGGCTGAGGAGAAGCCGAAAATTCATTTCCCTTCTGCTGATC</td>
<td>0.1762107</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>IGI</td>
<td>TTGGGTTGAGTGAATACCCACCTTCCCTCGGACACAAATTAAATATTCACGACTTGTATGAT</td>
<td>0.1736023</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>LRM P</td>
<td>AGGTTCTTGACAGCTACGACATGCTAGGCTAGCCTGCTGAGGAGAATGGAATGACATTTACGAGG</td>
<td>0.1727218</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>IGKV1D.33</td>
<td>GCAGAGCTGAGGACGGTTGAGAAATGACGATTACCTGTGCAACAGTGTAATCATTTACGAGG</td>
<td>0.169694</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>SVEP1</td>
<td>TTGGTCTGAGGACGGTTGAGAAATGACGATTACCTGTGCAACAGTGTAATCATTTACGAGG</td>
<td>0.1695933</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>IRF4</td>
<td>'TTTTGATTATTTTTTACCGAGGAAATGGAATGACATTTACGAGGAAATGGAATGACATTTACGAGG</td>
<td>0.1685261</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>IGKV3.7</td>
<td>AGGACTGAGACCTTGGACCAGTGACCTGAGGAGAATGGAATGACATTTACGAGGAAATGGAATGACATTTACGAGG</td>
<td>0.1651686</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>LOC10000000723</td>
<td>GCAGAGCTGAGGACGGTTGAGAAATGACGATTACCTGTGCAACAGTGTAATCATTTACGAGG</td>
<td>0.1640216</td>
<td></td>
</tr>
<tr>
<td>43</td>
<td>IGL</td>
<td>CCCAAGGCTCAGACGGCCTTCTCCCGTGACACTGAACACCCCTCCTAAATATTTACGAGGAAATGGAATGACATTTACGAGG</td>
<td>0.1629227</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>IGK C</td>
<td>CCACAGGACAGCTTGGACCAGTGACCTGAGGAGAATGGAATGACATTTACGAGGAAATGGAATGACATTTACGAGG</td>
<td>0.1622682</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>CAPS10</td>
<td>CCGTCAATGTTCTTGGCTACCTGAGGAAATGGAATGACATTTACGAGGAAATGGAATGACATTTACGAGG</td>
<td>0.1585399</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>IGKV1D.27</td>
<td>TACCTGAGCCTGAGGAAATGGAATGACATTTACGAGGAAATGGAATGACATTTACGAGGAAATGGAATGACATTTACGAGG</td>
<td>0.1575712</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>IGKV4.1</td>
<td>GCAGAGCTGAGGACGGTTGAGAAATGACGATTACCTGTGCAACAGTGTAATCATTTACGAGGAAATGGAATGACATTTACGAGG</td>
<td>0.1569229</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>ITM2A</td>
<td>CTTAGGCTGCTGAGGAAATGGAATGACATTTACGAGGAAATGGAATGACATTTACGAGGAAATGGAATGACATTTACGAGG</td>
<td>0.1565255</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>LOC10000000862</td>
<td>CAGGAGGAGGGCCTGGAGACGGCTGAGGAAATGGAATGACATTTACGAGGAAATGGAATGACATTTACGAGGAAATGGAATGACATTTACGAGG</td>
<td>0.1551169</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>TTLL12</td>
<td>GAGGGCAGGGCTAGGTTAGGGGCTGCTGAGGAAATGGAATGACATTTACGAGGAAATGGAATGACATTTACGAGGAAATGGAATGACATTTACGAGG</td>
<td>0.1541169</td>
<td></td>
</tr>
</tbody>
</table>
79  INMT  GAGAGAGCTAGTGTTAAGCTGTTGGACAGGCTGTTGCTGATAGAAACCTATATTCACCTGC  0.1176619
80  ACAP1  GTTCCACTTCTCCAAGTTCCGGCTCTGACCTGACTCATGGAGACCAAGACTAGTGAAAG  0.1175861
81  KLRB1  TCAAACCTTTGGAATACAGCTAGCTAAGCTGTTGCTCACCACAAGACCTACGCTGTGCTTTA  0.1171576
82  TANC2  AGGTTTCTAGTACGAGTAAGAGAGTTCTGCTACGATATATTCAAATGCTAG  0.1169085
83  PTGDS  CAAAGCAACCTGCCCACCTGAGGGTCTTCTCAGCAAAATAACCTCAGGAGCAAGTACAGT  0.1159179
84  IGLV2.14  CATCACTGTCCTCAGCAGCCAGAGGAGGCTGTTAATTACTGAGCTCATATACAAGCAG  0.1159175
85  IGHV3.66  TCTCAGTATATTATGCTGTTGAGCAGTATCTAGCAAGTCCCGGATTCA  0.1138371
86  TFPI  GCCCTTGCGTTATCTGAGTATCTGATTTCTTAGGATTGGTATGCAATTTCGCTCTATGT  0.1130304
87  FCRL1  CAAAGGTCCACCTACCTCAACTACACCTACCCAGCAGGTACGCTATATATGAAAAT  0.1128008
88  PGMS  CTTAACAGCCACACACCTCAGGCTGGAGACAAGCGGGAGGTGTGAGTGAATTGTAATGTTT  0.1110997
89  IGLL1  TCCAGGCCAAAGCTACCTGTTGGTTCTCTGATGACTATTTATCCGGGAAATCTTG  0.1104732
90  DARDC  TCCCTCCACGTAGAATCTTCAAGCTCAGTTGAGGTGTGAAATGGGATATTCCCTAT  0.108999
91  CCLI4  ACTATATCAAGGACATGAGGAGAAGCTAGGAGCCAGAAGGGGGTGAGGCACCAGCT  0.1085745
92  LOC400456  TTATCTCAGTATCTGATTCAGTATCCACGAGAAAGCCTAGTCAACAGCTTATTTATCCGAAATGCGCAT  0.1075516
93  C7  ACAGGCTGCGTGCTGGTTTCTTGAATAGGTGTAACCTTCTCCTGAGGCTCTGGTTTTTTTTAAAA  0.1047268
94  CR616309  GGCCTACTTCAAGTATAACAGAGGAGGAAATGCGTTGGAATAAGCATGGCACAAGG  0.1019357
95  NKA1N1  CCAAGATTCTTCTTTTCAGTGACGAGGGAGGAGGTGCTCTCTGGTTTCTGGCA  0.1010148
96  TNXB  AGTTTCTGGTGCTTCGAGGAAATGGAGAAGGAGAAATGCGTTGGAATAAGCATGGCACAAGG  0.1000135
97  MMP11  GCACAAAAAGGTCTCCACGTGAATATGGAGGAGGTTTCTCTGATGAGGGAGAGGGCTCGGAGGCC  0.0999101
98  IGLV6.57  AACTCTCGCTCCCTCCACATCTCCTGCTAGGAGTGGCAAGGGAGGCTCTCTGGTTTCTGGCA  0.0990432
99  SHE  TGCCCTTCTTCTGGGAAAGGGGTGTTTTGAATGCTCTAATAACACTATCAGTCAACAGCTCCCT  0.0970628
100  IGHV3.73  CCAAGAGTAGTATTCCAAGAAGACGGGCTGCTCTGCAATAGAAACGGCTTAAAACCGAGACA  0.0941202
101  ZNF366  CCAACGGGCGCCTGAGCTAAGGGCTGAGCCTTCTCATATAACCCCAAAGGACTAGCTGGTAGATTT  0.0935525
102  MALL  CTTGGAAAGGCTCTTATGTGCTCCTCTTTTT1GGTGCTCTGCTAGGATGACTCAT  0.0930558
103  PLCG2  GAGTGCAAGCAAGCAGAGTTTTTACTCTAGAAGGCTGGGTGTTGTGTAAGGCTATTGTTGT  0.0929387
104  IGHV1.2  CTCCTAGCTGAAAGGCTGCTCGAGATACACCTTCCACCAGCTCTTAAATTGACTCAGT  0.0924283
105  PPP2R2B  AGTCTTCTCTGGTGCGTGTTAGTAAATTTGTTTCTAGGTTGCTGGCGTATAAATGAGGGT  0.0920309
106  IGHV30R16.7  CCGTATCTTGCAATAGACAGCTTGAGCTGGAGACCCGGGTTGTTTATCTGCTGGAATA  0.0915109
<table>
<thead>
<tr>
<th></th>
<th>Gene</th>
<th>Sequence</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>107</td>
<td>IGKV2.D.29</td>
<td>TTGTATTTGCTCCAGCAAGCCAGCCGAGCTCCACAAGCTGCTGTATGATGAGTTTCC</td>
<td>0.0909582</td>
</tr>
<tr>
<td>108</td>
<td>IGKV2.18</td>
<td>GGGGACTCTCCACGAGCTGTATTATAGGGTTTCCAAATCATCTTCTTTTGGGGTCCAGAC</td>
<td>0.0895279</td>
</tr>
<tr>
<td>109</td>
<td>BC144530</td>
<td>TATTTGAACTTGATCTCAAGGAGACAGGAGGACTTGCTCTATCTGAGTTT</td>
<td>0.0892731</td>
</tr>
<tr>
<td>110</td>
<td>ACSL5</td>
<td>CTGCCTTACTCTACCAAGGTAACAGTAGTTGTTTTCCTACTTGTAATGTAAATGTCTA</td>
<td>0.088697</td>
</tr>
<tr>
<td>111</td>
<td>TXNDC5</td>
<td>GGGCAGACACTTGAATCTGAACGCTGAGACCTTTGTGTTGGCCGAAGAGGAAGCAGCTTTAG</td>
<td>0.0879397</td>
</tr>
<tr>
<td>112</td>
<td>BC045756</td>
<td>TGGGTGTTT<em>N</em>G*M**TTGGTTCTGTTTACAATCTCTGATCTGACAAAGATGGAG</td>
<td>0.0873954</td>
</tr>
<tr>
<td>113</td>
<td>C10orf105</td>
<td>CCAAGGACAAGCTCTTTGAGATCTTCTGACAAATGGAATACACTGACAGC*TTAAAAATTA</td>
<td>0.0871425</td>
</tr>
<tr>
<td>114</td>
<td>ZDHHC18</td>
<td>AGATATTGCCAACCTTCCTCTGGCCCTCAAATGGAATGGGCTAAGCCTTTGAAAG</td>
<td>0.086863</td>
</tr>
<tr>
<td>115</td>
<td>FAM180B</td>
<td>GGTGGCCGGAGAGAGGAGGCTCTTCTGTCTCTTCTTCTAATGTTCTGTT</td>
<td>0.0867138</td>
</tr>
<tr>
<td>116</td>
<td>VAM5</td>
<td>CTGCAGACGAGTTCCTGACCAACTCTCTAGATAGATGCTCAACCTACACAGCAAGCTACCA</td>
<td>0.0860009</td>
</tr>
<tr>
<td>117</td>
<td>CPNE5</td>
<td>TGGGTTCTGTGCCCCGCTCTGAGACAGTCCTGTGTGAATTTGGCTTTAAACTGAAAT</td>
<td>0.0855679</td>
</tr>
<tr>
<td>118</td>
<td>IGHG1</td>
<td>GGCGAGCACTCTCTCAGGCACCGAAGATGCTAAGAAGACACACACATCT</td>
<td>0.0833098</td>
</tr>
<tr>
<td>119</td>
<td>TAR1P</td>
<td>CTGGCACAAGGAAACACAGATGTATCGTCAGACAGTAGAATAAAACAGGATGTTCAAA</td>
<td>0.0823656</td>
</tr>
<tr>
<td>120</td>
<td>MUSTN1</td>
<td>AGAAGCTGTCCTTTAAGACGCAAGCCCGCAAGCCCAAGGATGCTCTGCAGTAAAGGTG</td>
<td>0.0797104</td>
</tr>
<tr>
<td>121</td>
<td>TNFRSF17</td>
<td>GATCTCTTATAGGATGACTGATA<em>TTTT1GAC1TGCGAT1AGACG</em>TTTTGTCTCTAAGCT</td>
<td>0.0791632</td>
</tr>
<tr>
<td>122</td>
<td>CFHR3</td>
<td>TCAAGAAAGCTCTGAGGCTACATTGCAACAAAAATGGTGGTCACACGACACACACCC</td>
<td>0.0788775</td>
</tr>
<tr>
<td>123</td>
<td>B1N1</td>
<td>AGCAGAAGAAACTAAGAGGAAAGAGCCCGCAAGGCTTCTCCACAAAGATGATG</td>
<td>0.0781509</td>
</tr>
<tr>
<td>124</td>
<td>WHSC1</td>
<td>TCTACTGCTATGAGACATGACTTGGAGGCGGATCCGTCAGAGGACCAAGACTGAGAAG</td>
<td>0.0772003</td>
</tr>
<tr>
<td>125</td>
<td>IGKV1D.13</td>
<td>GAAAGCTCTTAAACGTCGCTCTATGATGACTGCTCAGTCTTGGAGAGTGGGTCCTCCAGAAG</td>
<td>0.0764297</td>
</tr>
<tr>
<td>126</td>
<td>SLAMF1</td>
<td>TCTGTGACTGACATCTCTCTCTACAGTATAAGAATGAGAATAGTGGCGAGGACGGTGATTTTCT</td>
<td>0.0753424</td>
</tr>
<tr>
<td>127</td>
<td>ADRB2</td>
<td>CTCTTATTTGCTCAGAGGTAATT1AGCGAGGGGATAGAGGAGATGCTCTAGTTCTT</td>
<td>0.0749854</td>
</tr>
<tr>
<td>128</td>
<td>LO642891</td>
<td>A6CGAAGGCGAATCTGATATCTGCTCTACAGAGAAGATCAAGAAATATTT1GAAATGAA</td>
<td>0.0746785</td>
</tr>
<tr>
<td>129</td>
<td>FBLN5</td>
<td>GGGGAAACCTGCGAGATCGATATGTTAG*TTTT1GGGTACACAGAGAAGCTATGTAACACAA</td>
<td>0.0745848</td>
</tr>
<tr>
<td>130</td>
<td>PNOC</td>
<td>GCCACTGCGCATATCTTTTGTGAAAGACTG*TTTT1TTT1GAC1GAT1GTTIT1TTITAAAAACAC</td>
<td>0.0744541</td>
</tr>
<tr>
<td>131</td>
<td>IGKV2.D.28</td>
<td>GAATT1ACACTGAAATACTGAGGGACGAGGTTGAGGAGTGTGGGTAGTTCTGAGCTATG</td>
<td>0.0739373</td>
</tr>
<tr>
<td>132</td>
<td>IGKV2.D.24</td>
<td>TGCTATAGCTGCTCTGCTCCGATCCGAGCTGATCTGTGATGCTGCGCAGCATGCCACTCTCTCT</td>
<td>0.0737708</td>
</tr>
<tr>
<td>133</td>
<td>IGKV6.D.41</td>
<td>TTACCATAGTGCTGAGGCTGAAGCTGAGATGCTGCAACAAATTACTGCTGACAGGGCAATA</td>
<td>0.0734206</td>
</tr>
<tr>
<td>134</td>
<td>LDLRAD2</td>
<td>GTGGTAGCTACTGAGATCAGGCCACAGCTTTGTATTTGATGTGTTAATAAAAGACACATAT</td>
<td>0.0724355</td>
</tr>
</tbody>
</table>
191 CD27  AGCCACAACTGCAGTCCCATCCTCTTGTCAGGGCCCTTTCCTGTGTACACGTGACAGAGT 0.0510489
192 KCNN3  AGTGACCAAGCCAACACTCTGTTGAGCCTTCCCAAGATGCACGAGTCATGATGATGACTTA 0.050407
193 CSGALNACT1  AGCAGTTCTACTCGATCACCAAGATGCTTCTGAAAATTGCA 0.0503427
194 LG4  AAAGGACCGAGGGTTGATGGCATTCCTTTCCCATTCTAAACTCCACGCAGCCAATCGCT 0.0499844
195 DM D  GCCTTACTATTGTATTATAGCTCTTACTGTTACTCTAAATAAAGCAGCAGCTTATG 0.0493207
196 PRKACB  GCACCTTAGAAGAATGACACCAGAAAACCTTATTATATCACAATATTCTG 0.0490642
197 ATP1B2  CAAGCTTAAAAATCCATAGCTCTCACTCCTGCTCTCAGCTCTCTCTGAGCCCTT 0.0489863
198 TAP2  GCTCAGTA 0.0489432
199 WNT10B  TTTGGAGCATAAGATCTGTATCTAGAAAGAGATCACCCACTCCTATGTACTATCCCCAA 0.0487274
200 ALDH1A1  TGTCAAACCAGCAGAGCAAACTCCTCTCACTGCTCTCCACGTGGCATCTTTAATAAAAGA 0.0485829
201 GFI1  CTGTTGCAAAGATGCTCTTCACTGAGCAAAATGCTATGCTTGACTGACGACGGACTCAAGT 0.0485136
202 LOC439949  GCCTGGAAGGACGTGATTTGCAACATTTGAGAATTTGGCCACCACCAACCCTGATAC 0.0480733
203 KIF19  GGAGGGAATCTGG 0.0480214
204 PRKCB  CCTCACTTTGATGTTG 0.0473384
205 RAPH1  TCATATATCAGGCTATGCAACGTTGCGGAGAGGACCCCCTCCTGCTCCCCCCAAAAGAGA 0.0472965
206 FNBP1  TCATTGCAGAAGAAAGTTTACACGACGTCAAAAAGTGACGTTCATGCTAAGTG 0.0467515
207 LOC647478  GAAGGAGTGTTATTTGCAACATTTGGAATTTGCTATGCTGAGCTCTGATC 0.0464095
208 PID1  TGCTTCTTCTAAAGCTCTTTGTTACATGCTTATAGGGGACCAAGCAGGCACCTAA 0.0461036
209 MAP3K14  CAGGACTACGTGACATCATTAAGATCGTGAATGCTAGCAGGAGGGTGCTGCTAGCAGCAAC 0.0457777
210 CDC40LG  CTTGTGAAACTGTTAGGCTGACGATCTCATTCTCTCTTCTCAGGAAATCTGTATGCTTATGTTAT 0.045739
211 MMRN2  AAATCTTCCACACACCTGATTTCTCTCACTTCTACATCTGATTTGCTGTTTTGCTCCTCCTGGTGTT 0.0455599
212 VWF  TGTCACCATAGTGTTCACTGCTGCTATGCTGACTTCTCTTCTCGATGTTAAGGGGCAGGTGACCTAA 0.0455599
213 SLM01  CAAGGACCTTCCAGAAGT'T'NC1TGCTAGTGAAGCGTAGTGTGAGCTTCTTCTCCCTCCTCTCTCCTCTCCCT 0.0452878
214 P16  AGGGACGAGGGAGGAAAGTAACCTTGACTCTCCTCAATAAACCCTCTCCAACCTGTGGCC 0.0450215
215 TOX  TTTGGAAAACCAAGAAATGCTGATGTGCTCTTTCTTTGGAATATGGAATATATATATAGCAG 0.0449719
216 JSP1  GAGAGGCCTCGGAGAGGGGAAGCCGCGGAAGGAGAAGCCGCGGAAGGAGGAGAGACCT 0.0447145
217 CXorf36  TCTGTGAGCAGCCCTGAGAACGCTGACTCCAGATTTGCTATGCTCACCAGATTGCAAT 0.0444788
218 M1SP  TCTCTCTTGAATGTTTCTGCTCTGGAGCTAAGATGTTTGGGAGGACC 0.0444676
275 ABLI M 1
AAGCAAAACTCTTCTAAGTCCCACTCGTGAATGGCAATTAGAGAAAGGACTGACAGTGGC
0.0322961
276 CARD16
ACCAAGAAGAGATGAGAAAGATAAAGCTCTGAAATGTGAAATGGGATAGGACCAG
0.0322958
277 EBF3
AAATTCCTCAATTTATTTTTAGTGCTAGTaaGGGTTGGATCTAAGAAGAGCTCAATGG
0.0321831
278 PPEF1
GGAAGTCCTAGCAAGCTCTTATGGTAAAGTAGGGTAAAGATTGAAGAGGGTTTG
0.0320775
279 GPR133
GAACATATTTAGTTTTAGATGGAAATTTTTAGAACGTATTTTAAAGGATAGG
0.0320379
280 ELOVL6
TGACAAACAGGAGGCCCCCCATACACTTACCCCTCAATTTTAAGTGTATGAAATACCTG
0.0319714
281 MICAL2
CAGTGATTTTGGCAAGTTGCGGCTTCCCTCGAGAGAATTTAAGGATCGGAATACTTG
0.0316668
282 MT1JP
TGCGCCTTGCGGTCTCTCCATTTATCGCTTGAGATCTCCAGCCTTACCGCGGCTCGAAAT
0.0316355
283 HOXA5
ACCCGAGAAGAGGATGAAATGACATGCTCTTTGGGTTGGGTTGCTCTCTCAGAGAAAT
0.0314651
284 SM R3A
CACCCCTATGGTCCAGGGAGAATTCAATCACACTCTCTTCCTCCTCCTCTTATGCCCAG
0.0314615
285 PPAP2B
TATGGATCAACCACATGCACATCCTTACTACAGAATCCGTCCTTTCATTTCAACTTAG
0.0311292
286 GRM8
AAGACAAAGGACTCTTGAATTTGGAATGGATGTGATGCTGACCCAGATGAGGCCTGGGTGCTAA
0.0308422
287 PALM2
GCAATGCTCAAAATACAAAGTGTTTGGGTGTAATTGGGATACATTAGGAAAGCTC
0.0306482
288 ICAM4
AATTTTTTATTTTTTTTTTTTTTTTTTTTTTAAAGGCTTCTTTTAACCTACTAG
0.0303506
289 DLG5
AGGTTTTCGTTGCTACGACAAACATAGGAAAAATCTGTAGACTACATTTCGAGATTTAACCTG
0.0303506
290 EBF1
ACACAGATGAGATGAGTTAATCACTTCAATTTATTTTATACTTTGTCATGTTAATC
0.0301188
291 FOXN2
AGTCAACAGTTGAGTTGCAACCTCAGTCTTCTTTGGAATTGATGATGAACTATGATGATG
0.0301556
292 TM EM 100
GACTGAAAAACTGCTTTTCTATTACATGAAAAACCTTTGAGGGGCTATGGGAGTCAGAC
0.0300854
293 ANGPT1
GTTTCCAACTTCTTGAGTTGCTAGCTACTTTGCTTTCTCTTTTGGAGAATA
0.029906
294 IGKVI6.21
CGTAGCTCAAGAGTGAATGGGACTGCTTCTTTTGGAGTTGATGAGATGAGATGAGATG
0.029898
295 LOC643431
GTAGACGTGGAGCTCGCTGAAGGTGGCTAGGAGTCAACATTTTCTCTCAGGCTGCTGTAT
0.0298316
296 PRDM 16
CTAGGAGAAGGCCAGTTTGGAGTTAATGCTGATAAGATTAGACTGATGATAACCAATGTAATC
0.0297239
Claims

1. A method of typing a sample from a breast cancer patient, the method comprising:
   providing a sample comprising RNA expression products from breast cancer cells from the patient;
   determining a level of RNA expression for at least two genes that are selected from Table 1 in said sample;
   comparing said determined level of RNA expression of the at least two genes to the level of expression of the genes in a reference sample; and
   typing said sample based on the comparison of the determined levels of RNA expression.

2. The method according to claim 1, whereby said breast cancer is an invasive lobular breast cancer.

3. The method according to claim 1 or claim 2, further comprising determining a level of expression of PDL1, PD1, and/or CTLA4.

4. The method according to any one of claims 1-3, whereby said at least two genes selected from Table 1 comprise GIMAP6 and P2RY14.

5. The method according to any one of the previous claims, whereby a level of RNA expression of at least five genes from Table 1 is determined.

6. The method according to any one of the previous claims, whereby a level of RNA expression of at least GIMAP6, P2RY14, CD79B, GIMAP5, GIMAP8, LOC100290115, IL33, SEL1L3, and SLC9A9 is determined.
7. The method according to any one of the previous claims, whereby a level of RNA expression of all genes from Table 1 is determined.

8. The method according to any of claims 1-7, further comprising determining a metastasizing potential of the sample from the patient.

9. A method of assigning treatment to a patient suffering from breast cancer, comprising
   (a) typing a sample comprising RNA expression products from breast cancer cells from the patient according to the method of any of claims 1-8;
   (b) classifying said sample as Immune Related (IR) or Hormone Response (HR);
   (c) assigning anti-Estrogen Receptor directed therapy to a patient of which the sample is classified as HR.

10. A method of assigning treatment to a patient suffering from breast cancer, comprising
    (a) typing a sample comprising RNA expression products from breast cancer cells from the patient according to the method of any of claims 1-8;
    (b) classifying said sample as Immune Related (IR) or Hormone Response (HR);
    (c) assigning therapy targeting PD1, PDL1 and/or CTLA4 and/or assigning therapy comprising at least one DNA damaging agent to a patient of which the sample is classified as IR.

11. The method according to claim 10, whereby the therapy comprises antibodies to PD1, PDL1 and/or CTLA4.
12. The method according to claim 10 or claim 11, whereby the DNA damaging agent comprises bleomycin, cisplatin and/or a topoisomerase 1 inhibitor.

13. The method according to any one of claims 10-12, wherein the DNA damaging agent comprises topoisomerase 1 inhibitor SN-38.

14. A method of treating a patient suffering from breast cancer, comprising
   (a) typing a sample comprising RNA expression products from breast cancer cells from the patient according to the method of any of claims 1-8;
   (b) classifying said sample as Immune Related (IR) or Hormone Response (HR);
   (c) treating a patient of which the sample is classified as HR with anti-Estrogen Receptor directed therapy.

15. A method of treating a patient suffering from breast cancer, comprising
   (a) typing a sample comprising RNA expression products from breast cancer cells from the patient according to the method of any of claims 1-8;
   (b) classifying said sample as Immune Related (IR) or Hormone Response (HR);
   (c) treating a patient of which the sample is classified as IR with therapy targeting PDL1, PDL1 and/or CTLA4 and/or with therapy comprising at least one DNA damaging agent.
Figure 2

**Bleomycin DNA damage**

- IR: Max concentration = 50 uM
- HR: 

**SN-38 TOP1**

- IR: Max concentration = 0.05 uM
- HR: 

**Cisplatin DNA crosslinker**

- IR: Max concentration = 3 uM
- HR: 

**AT-7519 CDK9**

- IR: Max concentration = 10.24 uM
- HR: 

**FK866 NAMPT**

- IR: Max concentration = 1 uM
- HR: 

**SB-715992 KIF11**

- IR: Max concentration = 0.128 uM
- HR: 

p-values and FDR are given for each comparison.
Figure 3

BCSS ~ Cluster

Logrank test, p = 0.95
Figure 4

The graph illustrates the relationship between the number of genes and cross-validation accuracy. The accuracy increases as the number of genes increases, reaching a peak at around 10 genes. Further increases or decreases in the number of genes result in a decrease in accuracy.
**INTERNATIONAL SEARCH REPORT**

**International application No**

PCT/NL2016/05Q191

**A. CLASSIFICATION OF SUBJECT MATTER**

INV. C12Q1/68

**ADD.**

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

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

C12Q

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

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

EPO-Internal, WPI Data

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>

* 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, cited to establish the publication date of another citation or for other special reasons (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
- **"A"** document member of the same patent family

**Date of the actual completion of the international search**

14 July 2016

**Date of mailing of the international search report**

22/07/2016

**Name and mailing address of the ISA/IB European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016**

**Authorized officer**

Asl und, Fredrick

Form PCT/ISA2/10 (second sheet) (April 2005)
This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. **X** Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:

   The Applicant was asked and confirmed that the restricted search report established for the priority document was deselected.

2. **X** claims Nos.: 1-15 (partially)

   because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

   see FURTHER INFORMATION sheet PCT/ISA/21Q

3. **☐** Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

1. **☐** As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. **☐** As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. **☐** As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. **☐** No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

**Remark on Protest**

**☐** The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

**☐** The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

**☐** No protest accompanied the payment of additional search fees.
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>
Continuation of Box II.1

The Applicant was asked and confirmed that the restricted search report established for the priority document was desired.

-----

Continuation of Box II.2

Claims Nos.: I-15 (partially)

Claim 1 refers to a determination of the RNA expression for at least 2 genes selected from Table 1. As this Table lists 296 genes, there are >87,000 combinations encompassed by the scope leading to a lack of clarity. To provide for clearly defined subject-matter to be searched, the Applicant has requested that the combination of the two genes GIMP6 and P2RY14 (as in claim 4) be searched.

The Applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The Applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the Applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guidelines C-IV, 7.2), should the problems which led to the Article 17(2) declaration be overcome.