(54) **Title:** METHODS

(57) **Abstract:** The invention provides improved methods for the sub-classification cancers into therapeutically relevant subpopulations through the use of branded DNA technology (bDNA).

METHODS

Field of the invention

The invention relates to the treatment, detection and classification of cancer, including the
identification of heterogeneous tumours and in particular relates to breast cancer. It also
relates to identifying patients who are likely to respond to cancer therapy with a PP2A
activator.

The invention defines the use of biomarkers (ERBB2, ESR1, PGR, AURKA, KIF2C and
FOXC1 expression) and the need of the novel biomarkers AURKA and KIF2C to classify
breast cancer patients as Basal or Luminal. In addition, a method is described to further
classify Luminal cases into good (Luminal A-like) or bad (Luminal B-like) prognoses.

The invention also relates to methods useful in predicting if a sample comprises a gene
amplification or gene reduction, or high or low gene expression.

Background

The provision of new treatments for cancer is of high importance, including for cancers
resistant to known treatments. There is also a need for the identification of markers that
allow for the detection, prognosis and classification of cancer, and which can predict
responsiveness of a patient to a given therapy. For example, the deregulation of the
protein phosphatase 2A (PP2A) complex is known to be a common event in cancer (Grech
et al; Tumour Biol; DO1 10.1007/s13277-016-5145-4 (2016)) and PP2A activators have
recently been shown to be useful in treating breast cancers that overexpression AURKA
and/or KIF2C (GB. 1704536.0, the disclosure of which is specifically incorporated herein
by reference).

Due to the genetic instability of tumour cells, genomic rearrangements frequently result in
gene amplification or gene loss. Thus the overexpression of some genes, for instance
HER2 in some breast cancers, is not mediated at the transcriptional level *per se* but is
instead the result of an increase in gene copy number. The identification of gene
amplification or the loss of genes that are directly involved in tumorigenesis or tumour
suppression can be clear indicators of the presence of a tumour, or of a particular tumour
subtype. Additionally the detection of an amplification or loss event of a gene that is not
itself involved in tumorigenesis but which occurs as the result of the type of genomic
instability that is frequently found in tumours can also be a marker of the presence of

tumour cells. For example, a particular indicator gene may be duplicated as part of the translocation process that results in the amplification of an oncogene, and increased expression from the indicator gene may be used as a surrogate to indicate duplication of the oncogene.

Molecular classification of breast cancer interrogates molecular markers to categorise patients into four molecular classes, namely luminal A-like or luminal A, luminal B-like or luminal B, the human epidermal growth factor receptor 2 (HER2)-enriched and the basal types. Luminal A subtype is positive for oestrogen receptor (ER) and/or progesterone receptor (PgR) expression with low expression of Ki-67, while luminal B, apart from having an ER/PgR positive expression, includes also HER2 positive and negative subgroups associated with high Ki-67 expression. The HER2-enriched subtype is well defined, with high expression of HER2 receptor, due to the ERBB2 gene amplification, combined with low or absent ER and PgR. Interestingly, HER2 expression was found to be present in microvesicles originating from tumour cells, that induced activation of mitogenic signals in recipient fibroblasts [10]. Hence, circulating microvesicles such as exosomes are potentially vehicles of early detection and early indicators of metastasis and relapse [11]. Basal type tumours are in general negative for the 3 receptors significantly overlapping with the triple negative breast cancer (TNBC) diagnostic subtype [5,6]. Other markers are used to determine epithelial and mesenchymal characteristics. Fibronectin (FN1) is a main component of breast tissue mesenchymal compartment. Increased FN1 expression is accompanied by high Ki67 staining showing a signature for a more invasive tumour [7,8] and an increased expression is associated with metastasis [9].

The luminal A-like or luminal A, and luminal B-like or luminal B subclasses are classified based immunohistochemical and clinicopathological criteria set out in the St Gallen 2013 conference (Harbeck, N., Thomssen, C., & Gnant, M. (2013). St. Gallen 2013: Brief Preliminary Summary of the Consensus Discussion. *Breast Care*, *8*(2), 102–109. http://doi.org/10.1159/000351193). Alternative classifications have been proposed (Maisonneuve *et al* 2014 Breast Cancer Res 16(3): R65; and Inic et al 2014 Clinical Medicine Insights. Oncology, *8*, 107–111. http://doi.org/10.4137/CMO.S18006).

Each subtype of cancer may be associated with a different response to a particular treatment and so identification of the subtype of cancer is important to ensure that the subject receives the most appropriate treatment.

In the clinical setting, gene amplification events are routinely assayed by either immunohistochemistry (IHC) and in-situ hybridization (ISH). IHC testing can show how much of the particular protein is present on the cancer cell surface which may in some situations be considered to be the most relevant parameter to assay, while ISH testing measures the number of copies of the particular gene inside each cell. There are two main types of ISH tests: fluorescence and bright-field ISH (www.cancer.net).

These are the only two methods currently approved by the FDA in the United States to test for the HER2 status of a tumour (Perez et al 2014 Cancer Treatment Reviews Volume 40, Issue 2, March 2014, Pages 276-284).

However, both IHC and ISH suffer from many problems that make them less than ideal tests to be carried out on large numbers of samples. For example, both tests require a visual interpretation of the degree of hybridisation of a probe, either to the DNA or protein, and are therefore susceptible to variation between pathologists and pathology labs. Both methods require a high quality sample and are affected by degradation, fixation time and care in handling.

ISH is considered to be more accurate than IHC, though it is IHC that is the more widely used since it is more practical.

However, ISH is also susceptible to problems such as loss of the control CEP17 chromosomal region, or the concomitant amplification of CEP17 in ERBB2 amplified cells, both of which would present misleading results.

In addition, some breast cancer cell lines such as MDAMB453 and JIMT1 harbour the ERBB2 gene amplification but show a relatively low expression of HER2 that is undetectable by HER2 immunohistochemistry (Köninki et al., 2010, Kao et al., 2009). Despite the amplification of ERBB2, the JIMT1 ERBB2 gene amplified cell line is resistant to trastuzumab and has reduced levels of the ERBB2 transcript compared to SKBR3 and BT474 HER2 positive cell lines (Tanner et al., 2004, Köninki et al., 2010).

The Cancer Genome Atlas (TCGA) data accessed through the cBioPortal shows that the discrepancy between gene copy number and transcript is also seen in breast cancer patient samples. For example (15.31%) of cases show ERBB2 transcript overexpression while only 119 (12.4%) cases show ERBB2 gene amplification. Hence 4.17% of cases show increased HER2 expression but no gene amplification (false negative cases by

genetic studies) and 1.25% of cases are ERBB2 gene amplified but do not show overexpression (false positive cases by genetic studies) (Figure 1).

The use of whole sections from resected material indicate the presence of heterogenous samples that are missed when diagnosis is done on biopsies (Table 1). A local study assessed 258 positive or equivocal HER2 resection specimens by Immunohistochemistry to assess heterogeneity. Technical staining variation attributed to inadequate specimen fixation, represented by peripheral tissue staining accompanied by a negative central area, was identified in 16 out of 258 (6.20%) sections. 45 (17.44%) resection specimens exhibited HER2 staining heterogeneity. Hence the need of high throughput methodologies to test biopsies and resected material will provide additional information on patient status.

		Diagnostic HER2 result for resection specimen		
		Positive (3+)	Equivocal (2+)	Negative (0/1+)
Diagnostic	Positive (3+)	10	1	1
HER2 result	Equivocal (2+)	1	6	14
for Biopsy specimen	Negative (0/1+)	2	4	72

Table 1: A comparative audit of diagnostic HER2 results on breast biopsy specimens as opposed to the patient-matched breast resection specimen where data was available. Out of 2303 breast cancer cases diagnosed at Mater Dei Hospital, Malta between 2009-2016, only 111 breast cancer cases (4.8%) had a diagnostic result for HER2 for both the biopsy and the main specimen. A discrepancy in the HER2 status was observed in 29/111 (26.13%) of cases (unpublished results).

Thus both IHC and ISH have several drawbacks and are not ideal for the routine and widespread determination of tumour status. Gene amplification and gene expression diagnostic testing using archival material or material that requires transportation to servicing laboratories, needs a more robust and accurate test adapted to current clinical workflow.

Some attempts have been made to produce panels of biomarkers to aid in breast cancer diagnosis and classification. For instance, the Prosigna Breast Cancer Prognostic Gene Signature Assay (NanoString) based on PAM50 biomarker panel was US FDA Approved in September 2013 and CE mark in 2012. However, the Prosigna assay is designed as an

indicator of prognosis of tumours and not classification, therapeutic selection or to detect response[12].

In addition a 12-gene signature (WO2009055823) was developed for Basal versus Luminal classification only as compared to PAM50 classification[18] and to predict therapeutic response to polyamine type chemotherapy.

Summary of the invention

The present invention provides an accurate method or test to allow the classification of a cancer as either cancerous, or a sub-class of that cancer, and in particular relates to cancers or sub-cancers that are known to be associated with a gene amplification event, for example the ERBB2 gene in some breast cancers.

By using an algorithm trained on well-annotated breast cancer samples, the transcription level is used to diagnose the status of the cell with respect to that gene or genes, for example whether a breast cancer sample is HER2 positive or negative. The present method is improved over the currently used ISH method since discrepancies between amplification at the level of chromosomes and levels of transcribed RNA are recorded both in cell lines and patient material; and is improved over the currently used IHC method for at least the reasons given above.

The invention also provides a further method which classifies breast cancer tumours into various sub-classes of breast cancer.

The tests described herein in some circumstances rely on the selection of appropriate normalisation genes. Accordingly, the invention also provides a set of normalisation genes useful for use with the tests described above when applied to breast cancer.

The inventors have previously shown (GB. 1704536.0, data reproduced herein) that cancer cells sensitive to treatment with a PP2A activator exhibit overexpression of the markers AURKA and KIF2C. These include cancer cells exemplified by Triple negative breast cancer (TNBC) cells that do not benefit from targeted therapy and have bad prognosis when current front-line anti-cancer therapies are administered. The markers AURKA and KIF2C further correlate with low PP2A enzymatic activity. The markers AURKA and KIF2C may thus be used to predict responsiveness of a patient to treatment with a PP2A activator. AURKA and KIF2C may also be targeted with antagonists to thereby

treat cancer, based on their correlation with cancer. In addition, the inventors have also shown that the markers AURKA and KIF2C have utility in detecting, prognosing and classifying cancer. The markers may be used to classify cancers with a worse prognosis, in particular basal versus luminal breast cancer subtype and subsets of luminal cases with

5   bad prognosis (Luminal B). The markers may thus be assayed using suitable reagents in kits for detection and classification of cancer and for predicting therapeutic benefit from activation of the PP2A complex.

The present invention aims to address the problems of the prior art discussed above and

10   provides methods (exemplified by the role of HER2 in breast cancer and the classification of breast cancers as HER2 positive (HER2+), triple negative breast cancer (TNBC), Basal, Luminal A or Luminal B) but which can now be generalised and extrapolated to other genes involved in various diseases, for the determination of the status of a given tumour, tumour sample or tumour subsample with regards to a particular gene or set of genes that are

15   quick, can be standardised, are much less susceptible to user interpretation and which advantageously can be used on degraded samples, which makes these methods suitable for use in situations where it isn't possible to handle or store samples under optimum conditions.

20   Optimisation of RNA based assays using archival formalin-fixed paraffin-embedded (FFPE) material is challenging due to variability in surgical tissue processing and degradation of RNA caused by tissue integrity preservation using Formalin [1,2]. To overcome the limitation of performing accurate gene expression studies from archival material, the invention makes use of branched-chain DNA (bDNA) technology, for example

25   the QuantiGene® technology. bDNA technology replaces enzymatic amplification of target template with hybridisation of specific probes and amplification of a reporter signal [3]. The short recognition sequences of the capture and detection probes are designed to hybridise to short fragments of target RNA [4]. In addition, the use of tissue homogenates directly as starting material of this assay, overcomes the inevitable loss of RNA occurring in assays

30   requiring prior RNA extraction and purification and also allows multiple assays to be performed on the same sample. Multiplex technology such as the Luminex technology provides the possibility to multiplex the assay, measuring expression of a panel of targets from low material input.

35   The fact that unlike previous methods the present invention can be used in conjunction with archived, degraded, formalin-fixed and paraffin-embedded samples means that the

present invention can be used to readily identify novel biomarkers by using historic samples.

The invention is further advantageous since the multiplex nature of one embodiment of the invention allows a single assay to be performed which can fully classify a cancer, for example a breast cancer.

In contrast to the biomarker panels and methods of the prior art the methods of the invention (which include in one embodiment the following markers ESR1, HER2, PGR, AURKA, KIF2C, FOXC1) along with particular normalisation genes is able to classify breast cancer in the current diagnostic groups (as determined by IHC and FISH) with high accuracy (see for example Examples 2-7). The potential of this molecular-based assay with quantitative measurements has potential to digitalise the current methodologies of breast cancer diagnosis. Therefore, this assay can assist in the diagnosis of breast cancer as well as determine therapeutic decisions. In addition, based on the earlier finding (GB. 1704536.0) that AURKA and KIF2C are biomarkers of PP2A activity, classification of breast cancer subtypes using these markers set another potential therapeutic subtype, that is considered to benefit from PP2A activation therapy.

The prediction by an algorithm as used herein based on ESR1, ERBB2, PGR, AURKA, KIF2C and FOXC1 expression (6 biomarker panel) allows the definition of Luminal and Basal subtypes and the same measurements can be used to discriminate between Luminal A and Luminal B when prediction is run using the Luminal cases only (see for example Examples 2-4). The use of the methods and panels of the invention reduces the cost of runs, digitalise the workflow, and allows minimal sample requirement through multiplexing of the RNA-based assay.

The compatibility of this assay and algorithm with H&E staining allows for the use of precisely laser microdissected material that allows isolation of specific cellular populations within heterogeneous samples which is considered to result in higher sensitivity due to increased percentage of tumour cell content per sample. In contrast, the prior art methods, for example those of Prosigna requires an unstained macro-dissection FFPE specimen for analysis.

Furthermore, PAM50 algorithms have been trained on microarray and RT-PCR data which require RNA extraction[19]. The methods described herein utilise algorithms that have been

assessed on RNA Sequencing data as well as Quantigene 2.0 data. The latter methodology eliminates extraction bias, and is minimally affected by degraded RNA.

Methods of the invention can also be used to detect heterogeneity within a single tumour
5 sample, for instance, can detect the presence of both luminal and basal subtypes within the same tumour, in some instances without the need of, for instance, microdissection and separate analysis of morphologically distinct tumours identified following staining and microscopy. This is important since whilst the majority of a tumour may be of one sub-type and may respond to a particular treatment, the presence of an additional tumour
10 subtype which may not respond to that treatment indicates that an alternative or additional treatment may be needed.

Detailed description of the invention

15 In a first aspect, the present invention provides a method for classifying a cancer into one or more sub-classes wherein the method involves the use of bDNA technology.

As discussed above, it is well known that although on the face of it a cancer is described as, for example, a breast cancer, within that broad classification exists a heterogenous
20 range of cancers, distinguished by unique phenotypic and genotypic differences and which can each require a different therapy. For example, breast cancer can be classified as HER2+, Basal, Luminal (which encompasses both the Luminal A and Luminal B sub-classes), and each of which responds better to different therapies. For example, HER2+ breast cancers are known to be suitable for treatment with Herceptin (trastuzumab),
25 Kadcyla (Herceptin and emtansine), Nerlynx (neratinib), Perieta (pertuzumab), and Tykerb (lapatinib), all of which take advantage of the overexpression of the HER2 protein to specifically target the cancer cell. Such treatments are considered to be less effective and/or not targeted to the cancer cell if the cell does not either harbour the HER2 gene amplification or the has some equivalent mutation/alteration that increases the expression
30 of HER2. It is important to note that, as stated above, not all gene amplification events of the ERRB2 gene which encodes HER2 leads to an increase in HER2 mRNA and HER2 protein. Since the standard clinical tests generally involves assessment of the DNA gene copy number of ERRB2 such a sample would erroneously be classified as HER2+. In this case a patient may be subjected to treatment with an anti-HER2 therapy which is likely to
35 be ineffective. As also stated above, although some clinical tests look at the actual level of the protein made, for example the HER2 protein, these tests are highly subjective and are subject to inconsistencies between laboratories, degree of fixation and user

interpretation. Similarly, the ER and PR status of breast cancer samples is also assessed by IHC and susceptible to the same problems. The present invention addresses these issues.

Although the present invention is exemplified by the development and optimisation of a method to classify breast cancer samples, for example as HER2+ cancers, or basal cancers or Luminal A or Luminal B cancers, the methods and algorithms described herein can be used to similarly develop and optimise methods to classify other cancers, for example other cancers that have a sub-class associated with a gene amplification. The skilled person will understand for which cancers and sub-classes of cancers the present invention is appropriate, for example by understanding which cancers or sub-classes of cancers are associated with which gene amplifications.

Examples of other cancers and sub-classes of cancers for which the present invention is considered useful are:

RUNX1 gene amplification
Intrachromosomal amplifications of chromosome 21 involves amplification of the gene RUNX1, defining a subgroup of B-cell precursor Acute Lymphoblastic Leukemia (ALL), predicting high relapse rare. Intensifying the therapy in these patients significantly reduce the replication rate [Harrison CJ: Blood spotlight on iAMP21 acute lymphoblastic leukemia (ALL), a high-risk pediatric disease. Blood 125: 1383–1386 (2015)].

MYCN gene amplification
MYCN gene amplification is found in various cancer types including colorectal cancer, neuroblastoma and others. MYCN gene amplification is an independent adverse prognostic factor in neuroblastoma predicting patients associated with rapid disease progression, involving all ages and stages [Thompson D, Vo KT, London WB, Fischer M, Ambros PF, et al: Identification of patient subgroups with markedly disparate rates of MYCN amplification in neuroblastoma: a report from the International Neuroblastoma Risk Group Project. Cancer 122: 935–945 (2016).].

hTERC gene amplification
hTERC gene amplification in liquid-based cervical samples was found in 37% of HPV genotype positive individuals, and 70% of hTERC amplified samples were diagnosed as CIN2+. hTERC amplification significantly improves the specificity and positive predictive value of HPV screening [Zappacosta R, Ianieri M M, Buca D, Repetti E, Ricciardulli A, and

Liberati M. Clinical Role of the Detection of Human Telomerase RNA Component Gene Amplification by Fluorescence in situ Hybridization on Liquid-Based Cervical Samples: Comparison with Human Papillomavirus-DNA Testing and Histopathology. Acta Cytologica 2015;59:345–354].

PD-L1 and PD-L2 expression level

Programmed death-ligand 1 (PD-L1) is a protein that in humans is encoded by the CD274 gene while PD-L2 is encoded by the PDCD1LG2 gene. The binding of PD-L1 or PD-L2 with the PD-1 receptor on T cells induces a signal that inhibits TCR-mediated activation of IL-2 production and T cell proliferation [Sheppard KA, Fitz LJ, Lee JM, Benander C, George JA, Wooters J, Qiu Y, Jussif JM, Carter LL, Wood CR, Chaudhary D (September 2004). "PD-1 inhibits T-cell receptor induced phosphorylation of the ZAP70/CD3zeta signalosome and downstream signaling to PKCtheta". FEBS Letters. 574 (1-3): 37–41; Karwacz K, Bricogne C, MacDonald D, Arce F, Bennett CL, Collins M, Escors D (October 2011). "PD-L1 co-stimulation contributes to ligand-induced T cell receptor down-modulation on CD8+ T cells". EMBO Molecular Medicine. 3 (10): 581–92.]. . This interaction is thought to be one of the causes of how tumour cells might evade detection and destruction by the body's immune system. Gene amplification, gene translocation and gene overexpression have been implicated in the overexpression of these surface bound proteins [Georgiou K Chen L Berglund M Ren W de Miranda N et. al. Genetic basis of PD-L1 overexpression in diffuse large B-cell lymphomas. Blood 2016 vol: 127 (24) pp: 3026-34]. Many PD-L1 inhibitors are already in clinical use with many others in development and are showing good results in clinical trials ["Immune checkpoint inhibitors to treat cancer". www.cancer.org. Retrieved 2018-02-11].

MET amplification

MET is a proto-oncogene that encodes a receptor tyrosine kinase. The aberrant activation of MET signalling in a subgroup of cancers, is an example of how certain cancer become dependent on a single overactive oncogene for their proliferation and survival, a phenomenon that has become known as "oncogene addiction". This activation is usually the result of gene amplification, polysomy, and gene mutations. MET deregulation can be identified in various human malignancies, including cancers of kidney, liver, stomach, breast, and brain [Kawakami H Okamoto I Okamoto W Tanizaki J Nakagawa K et. al. Targeting MET Amplification as a New Oncogenic Driver. Cancers 2014 vol: 6 (3) pp: 1540-52]

AR-V7

Androgen deprivation therapy provides effective, though temporary tumour control in patients with metastatic prostate cancer [G. Attard, J.S. de Bono Translating scientific
5   advancement into clinical benefit for castration-resistant prostate cancer patients Clin Cancer Res, 17 (2011), pp. 3867-3875,]. Androgen deprivation therapy can be achieved through a number of chemotherapeutic protocols. But there is clinical evidence that not all patients benefit from such therapy [H.I. Scher, K. Fizazi, F. Saad, M.-E. Taplin, C.N. Sternberg, K. Miller, et al. Increased survival with enzalutamide in prostate cancer after
10   chemotherapy N Engl J Med, 367 (2012), pp. 1187-1197,] Overexpression of nuclear AR-V7 protein in primary prostate cancer has been identified as an independent negative prognostic marker in men with high-risk disease receiving adjuvant therapy [Chen X Bernemann C Tolkach Y Heller M Nientiedt C et. al. Overexpression of nuclear AR-V7 protein in primary prostate cancer is an independent negative prognostic marker in men
15   with high-risk disease receiving adjuvant therapy Urologic Oncology: Seminars and Original Investigations 2017].

FGFR1 amplifications

The fibroblast growth factor receptor (FGFR) family is a novel potential therapeutic target
20   in cancer. FGFR plays an impoartant role in stimulating cell proliferation and migration as well as in promoting survival of various types of cells [Beenken A, Mohammadi M. The FGF family: biology, pathophysiology and therapy. Nat Rev Drug Discov 2009; 8: 235– 253.] Though different FGFR aberrations, including receptor overexpression through gene amplification or post-transcriptional regulation, FGFR mutations, FGFR translocations,
25   alternative splicing of FGFR, have been identified, all active aberrations constitutively activate downstream pathways and contribute to tumour development. Most of these abnormalities achieve this through the overexpression of FGFR [Dienstmann R, Rodon J, Prat A, et al Genomic aberrations in the FGFR pathway: opportunities for targeted therapies in solid tumors. Ann Oncol 2014; 25: 552–563]. Aberrant FGFR signalling has
30   been observed in many cancers, including NSCLC, bladder, breast, prostate, ovarian, endometrial, gastric, colorectal cancer, head and neck and glioblastoma [Parker BC, Engels M, Annala M, et al Emergence of FGFR family gene fusions as therapeutic targets in a wide spectrum of solid tumours. J Pathol 2014; 232: 4–15]. Inhibition with several small molecules known to target the FGF/FGFR-pathway has shown in vitro anti-tumour
35   activity within tumours with FGFR1-3 aberrations [Greenman C, Stephens P, Smith R, et al Patterns of somatic mutation in human cancer genomes. Nature 2007; 446: 153–158;

Wu YM, Su F, Kalyana-Sundaram S, et al Identification of targetable FGFR gene fusions in diverse cancers. Cancer Discov 2013; 3: 636–647].

5   Other amplifications include PIK3CA in Lung, ovarian and Head and Neck cancers [Meric-Bernstam F et al 2015. A Decision Support Framework for Genomically Informed Investigational Cancer Therapy JNCI J Natl Cancer Inst (2015) 107(7): djv098].

The method of the invention may be used to identify new sub-classes of cancers. For 10   example, the method could be used to assess the RNA expression level of a multitude of samples and analysis of the results can reveal novel therapeutic and/or prognostic tumour sub-classes of cancers of distinct tissue origin, for example new sub-classes of lung cancer, or new sub-classes of colon or breast cancer, each with a unique and distinctive expression pattern.   Methods according to this embodiment are discussed below in 15   subsequent aspects of the invention.

Cancers are generally classified according to one or more various markers that the cancer displays, for example the cancer may have one or two or more genes are expressed to a higher level than the expression level of those genes in another sub-class of the same 20   cancer. For example, although the HER2+ subclass of cancers is typically defined by the presence of an amplification of the ERBB2 gene, it is the important consequence of this amplification, i.e. an increased expression of the ERBB2 gene, i.e. an increased level of the HER2 mRNA and protein that truly makes the HER2+ sub-class of breast cancer a suitable target for certain therapies such as Herceptin.   Similarly, an ER+ cancer is a 25   cancer in which a certain number of the cells in the sample display the ER, as determined by IHC.

There is an extensive list of other classifiers, for example in breast cancer diagnosis E-Cadherin positivity defines cancers as of ductal origin as opposed to being of lobular origin 30   (E-Cadherin negative) ; whilst a mutation in the c-kit gene is associated with c-kit gastrointestinal stromal tumours (GIST).

It will be understood then that the level of RNA associated with each of the key genes that are used to classify the different sub-classes of cancers is a more accurate and informative 35   method of producing clinically relevant classifications than assessment of gene copy number.   The method provided herein, which is quick, sensitive and not subject to

pathologist error or interpretation is also improved over those methods such as IHC which directly assess the resultant protein.

The present invention makes use of the RNA expression levels of at least 1, for example at least 2, optionally at least 3, or at least 4, or at least 5, or at least 6, or at least 7, or at least 8, or at least 9, or at least 10 genes that are known to be associated with various cancer sub-classes. Although the RNA expression level can be assessed by any means, for example by reverse transcription PCR (rtPCR), a preferred embodiment utilises branched-DNA (bDNA) technology. This is advantageous for a number of reasons.

In a first aspect, the present invention provides a method for classifying a cancer into one or more sub-classes wherein the method involves the use of bDNA technology.

bDNA technology replaces enzymatic amplification of target template with hybridisation of specific probes and amplification of a reporter signal [3]. The short recognition sequences of the capture and detection probes are designed to hybridise to short fragments of target RNA [4].

Since the use of bDNA removes the requirement for RNA extraction and allows the direct use of a tissue homogenate from, for example, fresh tissue, fresh-frozen tissue, FFPE tissue sections or a laser dissected, stained or unstained sample, or exosomes, the technical variation obtained by using this assay is much reduced.

The skilled person has the necessary skills and knowledge to be able to design suitable target and amplification probes for use in a bDNA assay, see for example US8426578 and Bushnell et al 1999 Bioinformatics 15: 348-55.

Examples of commercially available kits for performing bDNA assays include the QuantiGene Plex assay from Thermo Fisher; Chiron branched DNA signal amplification (bDNA) assay used for viral molecular diagnositics [VERSANT® HIV-1 RNA 3.0 Assay (bDNA)];   Diacarta   (http://www.diacarta.com/technology/bdna-signal-amplification-technology/); and RNAscope (http://acdbio.com/science/technology-overview).

Briefly, branched DNA uses sequential hybridization of oligonucleotides to a captured target RNA in order to amplify a signal for quantitative measurement. In the context of a multiplex assay, for example using Luminex beads, the sample is first added to a bead mix that consists of both the magnetic Luminex beads as well as a set of probes that are used

to capture the target RNA. During the first incubation, the capture extenders hybridize to the capture probes conjugated to the beads while also hybridizing to the target RNA sequence. This captures the target RNA onto the desired beads through a process called 'cooperative hybridization'. Each bead colour has its own target-specific set of probes, allowing multiple genes to be captured onto different beads.

Also hybridizing to the target RNA are label extenders, which provide the basis for the branched DNA signal amplification structures. These label extenders are always designed in pairs to enhance the specificity of the assay. The third type of probe, the blocking probe, hybridizes to any piece of the target sequence that is not already targeted by the capture extenders or label probes. The purpose of the blocking probes is to form a complete double-stranded piece of RNA, protecting it from RNases and helping to prevent secondary structure within the target region.

The capture extenders, label extenders, and blocking probes all comprise the target-specific probe set which can be designed and provided commercially, for example from Thermo Fisher.

Next, the branched DNA oligonucleotides form the signal amplification structure. First, a pre-amplifier hybridizes to the label extender pairs. In the next incubation, many amplifiers hybridize to each pre-amplifier, and in the following incubation, many label probe oligonucleotides hybridize to each amplifier. The label probe molecule is conjugated with biotin, so when streptavidin-phycoerythrin is added in the last step, a fluorescent signal is created and measured.

Although the invention is herein exemplified by use of the QuantiGene Plex Assay, it is considered that any bDNA assay can be used in the methods of the invention. In one advantageous embodiment, the method involves the detection of RNA levels of relevant genes via the QuantiGene Plex Assay.

Advantageously, in one embodiment the bDNA assay, for example a QuantiGene Plex Assay, is a multiplex assay, which allows the simultaneous detection the expression level of many RNA species. Accordingly, in one embodiment the bDNA assay is a multiplex assay, for example an assay that allows the simultaneous detection of the expression level of more than 1 RNA species, optionally more than 2 RNA species, optionally more than 3 RNA species, optionally more than 4 RNA species, optionally more than 5 RNA species, optionally more than 6 RNA species, optionally more than 7 RNA species, optionally more

than 8 RNA species, optionally more than 9 RNA species, optionally more than 10 RNA species, optionally more than 11 RNA species, optionally more than 12 RNA species, optionally more than 14 RNA species, optionally more than 16 RNA species, optionally more than 18 RNA species, optionally more than 20 RNA species, optionally more than 5   30 RNA species, optionally more than 40 RNA species, optionally more than 50 RNA species.

Multiplex assays of any format are considered useful in the present invention. For example microbead technology, for example the Luminex platform, is considered useful and 10 advantageous. Accordingly in one embodiment the detection of the level of RNA expression is by the use of a multiplex platform, optionally a microbead platform, optionally the Luminex platform.

As described above, in one advantageous and preferred embodiment the detection of the 15 level of RNA expression is by the use of the QuantiGene Plex assay (Thermo Fisher).

In one embodiment the present invention allows the classification of the cancer into at least two sub-classes, optionally at least three sub-classes, optionally at least four-subclasses, optionally at least five or more sub-classes.

20

As discussed above, the sub-classes of cancer may be associated with a gene amplification or gene reduction, or may be associated with a change in RNA expression level that occurs independently of a gene amplification or reduction event. In one embodiment one or more of the sub-classes of cancer is associated with a gene 25 amplification or a gene reduction event. By associated with a gene amplification or gene reduction event we include the meaning that that particular class of cancer typically encompasses cancers with a gene amplification or gene reduction event, but does not mean that all cancers that make up that sub-class necessarily have the gene amplification or gene reduction. For example, the well known HER2+ sub-class of breast cancers 30 typically arises due to a gene amplification of the ERBB2 gene, and currently clinical diagnosis of the HER2+ phenotype often involves detection of a gene amplification event. However, the skilled person will appreciate that there are many factors that can affect the expression level of a particular gene, for example whether a particular gene is overexpressed or not, for example there may be some breast cancers that have a mutation 35 in the ERBB2 gene which results in an increased level of mRNA, for example due to a decreased rate of ERBB2 mRNA turnover. In this case, such a sample may still be considered to be a HER2+ cancer, provided the expression level of ERBB2 reaches a

particular threshold. Typically the threshold is set according to the level of RNA expression of samples that are known to harbour the gene amplification or gene reduction event, for example that are known to be HER2. The skilled person is well equipped to determine the relevant threshold levels of expression using standard lab techniques.

In another embodiment, one or more of the sub-classes of cancer is not associated with a gene amplification or gene reduction event, for example is associated with a change in the RNA expression level of a particular gene due to other mechanisms. For example, such mechanisms can involve a mutation in the gene that is used to classify the cancer, or a mutation in a gene that encodes for a regulatory protein that affects the expression level of the gene that is used to classify the cancer.

Other mechanisms that can result in altered gene expression that is used to classify the cancer include:

altered signal transduction pathway due to increased or decreased proliferation signals or lack of a negative feedback mechanism (exemplified by low PP2A activity);

genetic translocations;

activating mutations;

reduced expression/activity of tumour suppressors due to gene deletion, inactivating mutations; and

epigenetic mechanisms such as hypermethylation.

It is considered that hypermethylation in particular has significant effects on gene expression and the resulting changes in RNA expression can be detected using the methods of the invention.

In one embodiment at least one sub-class of cancer is associated with a gene amplification event. In another embodiment, at least one sub-class of cancer is associated with an overexpression of a particular gene.

Accordingly in one embodiment one or more of the sub-classes of cancer is associated with an increase or decrease in copy number of a gene.

In a further embodiment the cancer can be classified into at least one sub-class of cancer associated with a gene amplification event and at least one sub-class of cancer that is not associated with a gene amplification event, optionally at least one sub-class of cancer that is associated with an overexpression or an underexpression of a particular gene.

The expression level of any number of RNA species may be determined in order to classify the cancer into the various sub-classes. For example, in one embodiment the method comprises the determination of the level of RNA expression of optionally at least one RNA

5   species, optionally more than 2 RNA species, optionally more than 3 RNA species, optionally more than 4 RNA species, optionally more than 5 RNA species, optionally more than 6 RNA species, optionally more than 7 RNA species, optionally more than 8 RNA species, optionally more than 9 RNA species, optionally more than 10 RNA species, optionally more than 11 RNA species, optionally more than 12 RNA species, optionally

10   more than 14 RNA species, optionally more than 16 RNA species, optionally more than 18 RNA species, optionally more than 20 RNA species, optionally more than 30 RNA species, optionally more than 40 RNA species, optionally more than 50 RNA species. Some of the RNA species may be associated with the cancer or sub-class of the cancer, whilst others may not be associated with the cancer or sub-class of the cancer, and are used for

15   normalising the expression levels of the genes that are associated with the cancer or sub-cancer.

The skilled person will be very aware of the significance in choosing the correct normalisation genes. Methods of doing so are described herein, along with a specific and

20   novel set of markers that can used with the method of the invention for the classification of breast cancer. Accordingly, the invention provides particular markers, ACTB; PPIB; HPRT1; and/or TBP, which either alone, in particular combinations, or all together, can be used to normalise the data obtained from the method of the invention when applied to breast cancer. The skilled person will appreciate that the selection of appropriate

25   normalisation genes is affected by cell or tissue type as well as the expression level of the "test" genes, i.e. the genes that are associated with the cancer or sub-cancer.

Accordingly, in one embodiment, the method involves the determination of the expression level:

30   of at least one or more genes associated with the cancer or at least one sub-class of cancer, optionally at least one gene associated with the cancer or at least one sub-class of the cancer, optionally at least two genes associated with the cancer or at least one sub-class of the cancer, optionally at least three genes associated with the cancer or at least one sub-class of the cancer, optionally at least four genes associated with the cancer or

35   at least one sub-class of the cancer, optionally at least five genes associated with the cancer or at least one sub-class of the cancer, optionally at least six genes associated with the cancer or at least one sub-class of the cancer, optionally more than six genes

associated with the cancer or at least one sub-class of the cancer, optionally wherein the at least one or more genes associated with the cancer or at least one sub-class of cancer are selected from the group consisting of ERBB2, ESR1, AURKA, KIF2C, PGR and FOXC1 and/or

5　　　　　　of at least one gene not associated with the cancer or a sub-class of the cancer, of at least two genes not associated with the cancer or a sub-class of the cancer, of at least three genes not associated with the cancer or a sub-class of the cancer, of at least four genes not associated with the cancer or a sub-class of the cancer, optionally wherein the one or more genes has been determined to be suitable for use as a normalising gene,

10　　optionally determined by an algorithm trained on a dataset of known cancer classification, optionally wherein the at least one gene not associated with the cancer or sub-cancer is selected from the group consisting of ACTB; PPIB; HPRT1; and TBP.

It will be appreciated that a test which requires the determination of fewer RNA expression

15　　levels is preferable over a test that requires the determination of more RNA expression levels, provided that the test meets the required level of accuracy. In one embodiment it is possible to classify the cancer into a sub-class of cancer by determining the expression level of just one gene, for example of the HER2 (ERBB2) gene or the ER (ESR1) gene which can be used to classify breast cancer as HER2+ or ER+ using the method of the

20　　invention. In other cases, to accurately distinguish between two sub-classes of cancer, for example Luminal A and Luminal B breast cancer, multiple markers are required, for example when assessed using the present invention, the expression levels of the ERBB2, ESR1, PGR and FOXC1 genes can be used. However, a greater accuracy of classification can be achieved by determining the expression levels of all of the ERBB2, ESR1, PGR,

25　　AURKA, KIF2C and FOXC1 genes.

The skilled person will know what is meant by the genes ERBB2, ESR1, FOXC1, PGR, AURKA, KIF2C, ACTB, PPIB, HPRT1 and TBP. For the avoidance of doubt, the table below provides sequence information for the relevant genes and proteins. However, it will

30　　be understood that proteins or genes with minor variations in the sequences are also encompassed by the terms ERBB2, ESR1, FOXC1, PGR, AURKA, KIF2C, ACTB, PPIB, HPRT1 and TBP. For example, the genes or proteins may contain polymorphisms that are not represented in the sequences provided below. Accordingly, by the terms ERBB2, ESR1, FOXC1, PGR, AURKA, KIF2C, ACTB, PPIB, HPRT1 and TBP we include the

35　　meaning of the relevant gene or protein sequences as provided below, as well as genes or proteins with at least 75% homology to those sequences, for example at least 80% homology to those sequences, for example at least 85% homology to those sequences,

for example at least 90% homology to those sequences, for example at least 95% homology to those sequences, for example at least 98% homology to those sequences, for example at least 99% homology to those sequences, for example 100% homology to those sequences.

5

Name	Protein sequence	Gene sequence
ERBB2	[SEQ ID NO 1] – MELAALCRWGLLLALLPPGAASTQVCTG TDMKLRLPASPETHLDMLRHLYQGCQVV QGNLELTYLPTNASLSFLQDIQEVQGYV LIAHNQVRQVPLQRLRIVRGTQLFEDNY ALAVLDNGDPLNNTTPVTGASPGGLREL QLRSLTEILKGGVLIQRNPQLCYQDTIL WKDIFHKNNQLALTLIDTNRSRACHPCS PMCKGSRCWGESSEDCQSLTRTVCAGGC ARCKGPLPTDCCHEQCAAGCTGPKHSDC LACLHFNHSGICELHCPALVTYNTDTFE SMPNPEGRYTFGASCVTACPYNYLSTDV GSCTLVCPLHNQEVTAEDGTQRCEKCSK PCARVCYGLGMEHLREVRAVTSANIQEF AGCKKIFGSLAFLPESFDGDPASNTAPL QPEQLQVFETLEEITGYLYISAWPDSLP DLSVFQNLQVIRGRILHNGAYSLTLQGL GISWLGLRSLRELGSGLALIHHNTHLCF VHTVPWDQLFRNPHQALLHTANRPEDEC VGEGLACHQLCARGHCWGPGPTQCVNCS QFLRGQECVEECRVLQGLPREYVNARHC LPCHPECQPQNGSVTCFGPEADQCVACA HYKDPPFCVARCPSGVKPDLSYMPIWKF PDEEGACQPCPINCTHSCVDLDDKGCPA EQRASPLTSIISAVVGILLVVVLGVVFG ILIKRRQQKIRKYTMRRLLQETELVEPL TPSGAMPNQAQMRILKETELRKVKVLGS GAFGTVYKGIWIPDGENVKIPVAIKVLR ENTSPKANKEILDEAYVMAGVGSPYVSR LLGICLTSTVQLVTQLMPYGCLLDHVRE NRGRLGSQDLLNWCMQIAKGMSYLEDVR LVHRDLAARNVLVKSPNHVKITDFGLAR LLDIDETEYHADGGKVPIKWMALESILR RRFTHQSDVWSYGVTVWELMTFGAKPYD GIPAREIPDLLEKGERLPQPPICTIDVY MIMVKCWMIDSECRPRFRELVSEFSRMA RDPQRFVVIQNEDLGPASPLDSTFYRSL LEDDDMGDLVDAEEYLVPQQGFFCPDPA PGAGGMVHHRHRSSSTRSGGGDLTLGLE PSEEEAPRSPLAPSEGAGSDVFDGDLGM GAAKGLQSLPTHDPSPLQRYSEDPTVPL PSETDGYVAPLTCSPQPEYVNQPDVRPQ PPSPREGPLPAARPAGATLERPKTLSPG KNGVVKDVFAFGGAVENPEYLTPQGGAA PQPHPPPAFSPAFDNLYYWDQDPPERGA PPSTFKGTPTAENPEYLGLDVPV	[SEQ ID NO 2] – ATGGAGCTGGCGGCCTTGTGCCGCTGGGGGCT CCTCCTCGCCCTCTTGCCCCCCGGAGCCGCGA GCACCCAAGTGTGCACCGGCACAGACATGAAG CTGCGGCTCCCTGCCAGTCCCGAGACCCACCT GGACATGCTCCGCCACCTCTACCAGGGCTGCC AGGTGGTGCAGGGAAACCTGGAACTCACCTAC CTGCCCACCAATGCCAGCCTGTCCTTCCTGCA GGATATCCAGGAGGTGCAGGGCTACGTGCTCA TCGCTCACAACCAAGTGAGGCAGGTCCCACTG CAGAGGCTGCGGATTGTGCGAGGCACCCAGCT CTTTGAGGACAACTATGCCCTGGCCGTGCTAG ACAATGGAGACCCGCTGAACAATACCACCCCT GTCACAGGGGCCTCCCCAGGAGGCCTGCGGGA GCTGCAGCTTCGAAGCCTCACAGAGATCTTGA AAGGAGGGGTCTTGATCCAGCGGAACCCCCAG CTCTGCTACCAGGACACGATTTTGTGGAAGGA CATCTTCCACAAGAACAACCAGCTGGCTCTCA CACTGATAGACACCAACCGCTCTCGGGCCTGC CACCCCTGTTCTCCGATGTGTAAGGGCTCCCG CTGCTGGGGAGAGAGTTCTGAGGATTGTCAGA GCCTGACGCGCACTGTCTGTGCCGGTGGCTGT GCCCGCTGCAAGGGGCCACTGCCCACTGACTG CTGCCATGAGCAGTGTGCTGCCGGCTGCACGG GCCCCAAGCACTCTGACTGCCTGGCCTGCCTC CACTTCAACCACAGTGGCATCTGTGAGCTGCA CTGCCCAGCCCTGGTCACCTACAACACAGACA CGTTTGAGTCCATGCCCAATCCCGAGGGCCGG TATACATTCGGCGCCAGCTGTGTGACTGCCTG TCCCTACAACTACCTTTCTACGGACGTGGGAT CCTGCACCCTCGTCTGCCCCCTGCACAACCAA GAGGTGACAGCAGAGGATGGAACACAGCGGTG TGAGAAGTGCAGCAAGCCCTGTGCCCGAGTGT GCTATGGTCTGGGCATGGAGCACTTGCGAGAG GTGAGGGCAGTTACCAGTGCCAATATCCAGGA GTTTGCTGGCTGCAAGAAGATCTTTGGGAGCC TGGCATTTCTGCCGGAGAGCTTTGATGGGGAC CCAGCCTCCAACACTGCCCCGCTCCAGCCAGA GCAGCTCCAAGTGTTTGAGACTCTGGAAGAGA TCACAGGTTACCTATACATCTCAGCATGGCCG GACAGCCTGCCTGACCTCAGCGTCTTCCAGAA CCTGCAAGTAATCCGGGGACGAATTCTGCACA ATGGCGCCACTCGCTGACCCTGCAAGGGCTG GCATCAGCTGGCTGGGGCTGCGCTCACTGAG GGAACTGGGCAGTGGACTGGCCCTCATCCACC ATAACACCCACCTCTGCTTCGTGCACACGGTG CCCTGGGACCAGCTCTTTCGGAACCCGCACCA AGCTCTGCTCCACACTGCCAACCGGCCAGAGG ACGAGTGTGTGGGCGAGGGCCTGGCCTGCCAC CAGCTGTGCGCCCGAGGGCACTGCTGGGGTCC AGGGCCCACCCAGTGTGTCAACTGCAGCCAGT

19

TCCTTCGGGGCCAGGAGTGCGTGGAGGAATGC
CGAGTACTGCAGGGGCTCCCCAGGGAGTATGT
GAATGCCAGGCACTGTTTGCCGTGCCACCCTG
AGTGTCAGCCCCAGAATGGCTCAGTGACCTGT
TTTGGACCGGAGGCTGACCAGTGTGTGGCCTG
TGCCCACTATAAGGACCCTCCCTTCTGCGTGG
CCCGCTGCCCCAGCGGTGTGAAACCTGACCTC
TCCTACATGCCCATCTGGAAGTTTCCAGATGA
GGAGGGCGCATGCCAGCCTTGCCCCATCAACT
GCACCCACTCCTGTGTGGACCTGGATGACAAG
GGCTGCCCCGCCGAGCAGAGAGCCAGCCCTCT
GACGTCCATCATCTCTGCGGTGGTTGGCATTC
TGCTGGTCGTGGTCTTGGGGGTGGTCTTTGGG
ATCCTCATCAAGCGACGGCAGCAGAAGATCCG
GAAGTACACGATGCGGAGACTGCTGCAGGAAA
CGGAGCTGGTGGAGCCGCTGACACCTAGCGGA
GCGATGCCCAACCAGGCGCAGATGCGGATCCT
GAAAGAGACGGAGCTGAGGAAGGTGAAGGTGC
TTGGATCTGGCGCTTTTGGCACAGTCTACAAG
GGCATCTGGATCCCTGATGGGGAGAATGTGAA
AATTCCAGTGGCCATCAAAGTGTTGAGGGAAA
ACACATCCCCCAAAGCCAACAAAGAAATCTTA
GACGAAGCATACGTGATGGCTGGTGTGGGCTC
CCCATATGTCTCCCGCCTTCTGGGCATCTGCC
TGACATCCACGGTGCAGCTGGTGACACAGCTT
ATGCCCTATGGCTGCCTCTTAGACCATGTCCG
GGAAAACCGCGGACGCCTGGGCTCCCAGGACC
TGCTGAACTGGTGTATGCAGATTGCCAAGGGG
ATGAGCTACCTGGAGGATGTGCGGCTCGTACA
CAGGGACTTGGCCGCTCGGAACGTGCTGGTCA
AGAGTCCCAACCATGTCAAAATTACAGACTTC
GGGCTGGCTCGGCTGCTGGACATTGACGAGAC
AGAGTACCATGCAGATGGGGGCAAGGTGCCCA
TCAAGTGGATGGCGCTGGAGTCCATTCTCCGC
CGGCGGTTCACCCACCAGAGTGATGTGTGGAG
TTATGGTGTGACTGTGTGGGAGCTGATGACTT
TTGGGGCCAAACCTTACGATGGGATCCCAGCC
CGGGAGATCCCTGACCTGCTGGAAAAGGGGGA
GCGGCTGCCCCAGCCCCCCATCTGCACCATTG
ATGTCTACATGATCATGGTCAAATGTTGGATG
ATTGACTCTGAATGTCGGCCAAGATTCCGGGA
GTTGGTGTCTGAATTCTCCCGCATGGCCAGGG
ACCCCCAGCGCTTTGTGGTCATCCAGAATGAG
GACTTGGGCCCAGCCAGTCCCTTGGACAGCAC
CTTCTACCGCTCACTGCTGGAGGACGATGACA
TGGGGGACCTGGTGGATGCTGAGGAGTATCTG
GTACCCCAGCAGGGCTTCTTCTGTCCAGACCC
TGCCCCGGGCGCTGGGGGCATGGTCCACCACA
GGCACCGCAGCTCATCTACCAGGAGTGGCGGT
GGGGACCTGACACTAGGGCTGGAGCCCTCTGA
AGAGGAGGCCCCCAGGTCTCCACTGGCACCCT
CCGAAGGGGCTGGCTCCGATGTATTTGATGGT
GACCTGGGAATGGGGGCAGCCAAGGGGCTGCA
AAGCCTCCCCACACATGACCCCAGCCCTCTAC
AGCGGTACAGTGAGGACCCCACAGTACCCCTG
CCCTCTGAGACTGATGGCTACGTTGCCCCCCT
GACCTGCAGCCCCCAGCCTGAATATGTGAACC
AGCCAGATGTTCGGCCCCAGCCCCCTTCGCCC
CGAGAGGGCCCTCTGCCTGCTGCCCGACCTGC
TGGTGCCACTCTGGAAAGGCCCAAGACTCTCT
CCCCAGGGAAGAATGGGGTCGTCAAAGACGTT

		TTTGCCTTTGGGGGTGCCGTGGAGAACCCCGA GTACTTGACACCCCAGGGAGGAGCTGCCCCTC AGCCCCACCCTCCTCCTGCCTTCAGCCCAGCC TTCGACAACCTCTATTACTGGGACCAGGACCC ACCAGAGCGGGGGGCTCCACCCAGCACCTTCA AAGGGACACCTACGGCAGAGAACCCAGAGTAC CTGGGTCTGGACGTGCCAGTGTGA
ESR1	[SEQ ID NO 3] – MTMTLHTKASGMALLHQIQGNELEPLNR PQLKIPLERPLGEVYLDSSKPAVYNYPE GAAYEFNAAAAANAQVYGQTGLPYGPGS EAAAFGSNGLGGFPPLNSVSPSPLMLLH PPPQLSPFLQPHGQQVPYYLENEPSGYT VREAGPPAFYRPNSDNRRQGGRERLAST NDKGSMAMESAKETRYCAVCNDYASGYH YGVWSCEGCKAFFKRSIQGHNDYMCPAT NQCTIDKNRRKSCQACRLRKCYEVGMMK GGIRKDRRGGRMLKHKRQRDDGEGRGEV GSAGDMRAANLWPSPLMIKRSKKNSLAL SLTADQMVSALLDAEPPILYSEYDPTRP FSEASMMGLLTNLADRELVHMINWAKRV PGFVDLTLHDQVHLLECAWLEILMIGLV WRSMEHPGKLLFAPNLLLDRNQGKCVEG MVEIFDMLLATSSRFRMMNLQGEEFVCL KSIILLNSGVYTFLSSTLKSLEEKDHIH RVLDKITDTLIHLMAKAGLTLQQQHQRL AQLLLILSHIRHMSNKGMEHLYSMKCKN VVPLYDLLLEMLDAHRLHAPTSRGGASV EETDQSHLATAGSTSSHSLQKYYITGEA EGFPATV	[SEQ ID NO 4] – ATGACCATGACCCTCCACACCAAAGCATCTGG GATGGCCCTACTGCATCAGATCCAAGGGAACG AGCTGGAGCCCCTGAACCGTCCGCAGCTCAAG ATCCCCCTGGAGCGGCCCCTGGGCGAGGTGTA CCTGGACAGCAGCAAGCCCGCCGTGTACAACT ACCCCGAGGGCGCCGCCTACGAGTTCAACGCC GCGGCCGCCGCCAACGCGCAGGTCTACGGTCA GACCGGCCTCCCCTACGGCCCCGGGTCTGAGG CTGCGGCGTTCGGCTCCAACGGCCTGGGGGGT TTCCCCCCACTCAACAGCGTGTCTCCGAGCCC GCTGATGCTACTGCACCCGCCGCCGCAGCTGT CGCCTTTCCTGCAGCCCCACGGCCAGCAGGTG CCCTACTACCTGGAGAACGAGCCCAGCGGCTA CACGGTGCGCGAGGCCGGCCCGCCGGCATTCT ACAGGCCAAATTCAGATAATCGACGCCAGGGT GGCAGAGAAAGATTGGCCAGTACCAATGACAA GGGAAGTATGGCTATGGAATCTGCCAAGGAGA CTCGCTACTGTGCAGTGTGCAATGACTATGCT TCAGGCTACCATTATGGAGTCTGGTCCTGTGA GGGCTGCAAGGCCTTCTTCAAGAGAAGTATTC AAGGACATAACGACTATATGTGTCCAGCCACC AACCAGTGCACCATTGATAAAAACAGGAGGAA GAGCTGCCAGGCCTGCCGGCTCCGTAAATGCT ACGAAGTGGGAATGATGAAAGGTGGGATACGA AAAGACCGAAGAGGAGGGAGAATGTTGAAACA CAAGCGCCAGAGAGATGATGGGGAGGGCAGGG GTGAAGTGGGGTCTGCTGGAGACATGAGAGCT GCCAACCTTTGGCCAAGCCCGCTCATGATCAA ACGCTCTAAGAAGAACAGCCTGGCCTTGTCCC TGACGGCCGACCAGATGGTCAGTGCCTTGTTG GATGCTGAGCCCCCGATACTCTATTCCGAGTA TGATCCTACCAGACCCTTCAGTGAAGCTTCGA TGATGGGCTTACTGACCAACCTGGCAGACAGG GAGCTGGTTCACATGATCAACTGGGCGAAGAG GGTGCCAGGCTTTGTGGATTTGACCCTCCATG ATCAGGTCCACCTTCTAGAATGTGCCTGGCTA GAGATCCTGATGATTGGTCTCGTCTGGCGCTC CATGGAGCACCCAGGGAAGCTACTGTTTGCTC CTAACTTGCTCTTGGACAGGAACCAGGGAAAA TGTGTAGAGGGCATGGTGGAGATCTTCGACAT GCTGCTGGCTACATCATCTCGGTTCCGCATGA TGAATCTGCAGGGAGAGGAGTTTGTGTGCCTC AAATCTATTATTTTGCTTAATTCTGGAGTGTA CACATTTCTGTCCAGCACCCTGAAGTCTCTGG AAGAGAAGGACCATATCCACCGAGTCCTGGAC AAGATCACAGACACTTTGATCCACCTGATGGC CAAGGCAGGCCTGACCCTGCAGCAGCAGCACC AGCGGCTGGCCCAGCTCCTCCTCATCCTCTCC CACATCAGGCACATGAGTAACAAAGGCATGGA GCATCTGTACAGCATGAAGTGCAAGAACGTGG TGCCCCTCTATGACCTGCTGCTGGAGATGCTG GACGCCCACCGCCTACATGCGCCCACTAGCCG TGGAGGGGCATCCGTGGAGGAGACGGACCAAA

		GCCACTTGGCCACTGCGGGCTCTACTTCATCG CATTCCTTGCAAAGTATTACATCACGGGGGA GGCAGAGGGTTTCCCTGCCACGGTCTGA
FOXC1	**[SEQ ID NO 5]** – MQARYSVSSPNSLGVVPYLGGEQSYYRA AAAAAGGGYTAMPAPMSVYSHPAHAEQY PGGMARAYGPYTPQPQPKDMVKPPYSYI ALITMAIQNAPDKKITLNGIYQFIMDRF PFYRDNKQGWQNSIRHNLSLNECFVKVP RDDKKPGKGSYWTLDPDSYNMFENGSFL RRRRRFKKKDAVKDKEEKDRLHLKEPPP PGRQPPPAPPEQADGNAPGPQPPPVRIQ DIKTENGTCPSPPQPLSPAAALGSGSAA AVPKIESPDSSSSSLSSGSSPPGSLPSA RPLSLDGADSAPPPPAPSAPPPHHSQGF SVDNIMTSLRGSPQSAAAELSSGLLASA AASSRAGIAPPLALGAYSPGQSSLYSSP CSQTSSAGSSGGGGGGAGAAGGAGGAGT YHCNLQAMSLYAAGERGGHLQGAPGGAG GSAVDDPLPDYSLPPVTSSSSSSLSHGG GGGGGGGGQEAGHHPAAHQGRLTSWYLN QAGGDLGHLASAAAAAAAAGYPGQQQNF HSVREMFESQRIGLNNSPVNGNSSCQMA FPSSQSLYRTSGAFVYDCSKF	**[SEQ ID NO 6]** – ATGCAGGCGCGCTACTCCGTGTCCAGCCCCAA CTCCCTGGGGAGTGGTGCCCTACCTCGGCGGCG AGCAGAGCTACTACCGCGCGGCGGCCGCGGCG GCCGGGGGCGGCTACACCGCCATGCCGGCCCC CATGAGCGTGTACTCGCACCCTGCGCACGCCG AGCAGTACCCGGGCGGCATGGCCCGCGCCTAC GGGCCCTACACGCCGCAGCCGCAGCCCAAGGA CATGGTGAAGCCGCCCTATAGCTACATCGCGC TCATCACCATGGCCATCCAGAACGCCCCGGAC AAGAAGATCACCCTGAACGGCATCTACCAGTT CATCATGGACCGCTTCCCCTTCTACCGGGACA ACAAGCAGGGCTGGCAGAACAGCATCCGCCAC AACCTCTCGCTCAACGAGTGCTTCGTCAAGGT GCCGCGCGACGACAAGAAGCCGGGCAAGGGCA GCTACTGGACGCTGGACCCGGACTCCTACAAC ATGTTCGAGAACGGCAGCTTCCTGCGGCGGCG GCGGCGCTTCAAGAAGAAGGACGCGGTGAAGG ACAAGGAGGAGAAGGACAGGCTGCACCTCAAG GAGCCGCCCCCGCCCGGCCGCCAGCCCCCGCC CGCGCCGCCGGAGCAGGCCGACGGCAACGCGC CCGGTCCGCAGCCGCCGCCCGTGCGCATCCAG GACATCAAGACCGAGAACGGTACGTGCCCCTC GCCGCCCCAGCCCCTGTCCCCGGCCGCCGCCC TGGGCAGCGGCAGCGCCGCCGCGGTGCCCAAG ATCGAGAGCCCCGACAGCAGCAGCAGCAGCCT GTCCAGCGGGAGCAGCCCCCCGGGCAGCCTGC CGTCGGCGCGGCCGCTCAGCCTGGACGGTGCG GATTCCGCGCCGCCGCCGCCCGCGCCCTCCGC CCCGCCGCCGCACCATAGCCAGGGCTTCAGCG TGGACAACATCATGACGTCGCTGCGGGGGTCG CCGCAGAGCGCGGCCGCGGAGCTCAGCTCCGG CCTTCTGGCCTCGGCGGCCGCGTCCTCGCGCG CGGGGATCGCACCCCCGCTGGCGCTCGGCGCC TACTCGCCCGGCCAGAGCTCCCTCTACAGCTC CCCCTGCAGCCAGACCTCCAGCGCGGGCAGCT CGGGCGGCGGCGGCGGCGGCGCGGGGGCCGCG GGGGGCGCGGGCGGCGCCGGGACCTACCACTG CAACCTGCAAGCCATGAGCCTGTACGCGGCCG GCGAGCGCGGGGGCCACTTGCAGGGCGCGCCC GGGGGCGCGGGCGGCTCGGCCGTGGACGACCC CCTGCCCGACTACTCTCTGCCTCCGGTCACCA GCAGCAGCTCGTCGTCCCTGAGTCACGGCGGC GGCGGCGGCGGCGGCGGGGGAGGCCAGGAGGC CGGCCACCACCCTGCGGCCCACCAAGGCCGCC TCACCTCGTGGTACCTGAACCAGGCGGGCGGA GACCTGGGCCACTTGGCGAGCGCGGCGGCGGC GGCGGCGGCCGCAGGCTACCCGGGCCAGCAGC AGAACTTCCACTCGGTGCGGGAGATGTTCGAG TCACAGAGGATCGGCTTGAACAACTCTCCAGT GAACGGGAATAGTAGCTGTCAAATGGCCTTCC CTTCCAGCCAGTCTCTGTACCGCACGTCCGGA GCTTTCGTCTACGACTGTAGCAAGTTTTGA
PGR	**[SEQ ID NO 7]** – MTELKAKGPRAPHVAGGPPSPEVGSPLL CRPAAGPFPGSQTSDTLPEVSAIPISLD GLLFPRPCQGQDPSDEKTQDQQSLSDVE GAYSRAEATRGAGGSSSSPPEKDSGLLD	**[SEQ ID NO 8]** – ATGACTGAGCTGAAGGCAAAGGGTCCCCGGGC TCCCCACGTGGCGGGCGGCCCGCCCTCCCCCG AGGTCGGATCCCCACTGCTGTGTCGCCCAGCC GCAGGTCCGTTCCCGGGGAGCCAGACCTCGGA

SVLDTLLAPSGPGQSQPSPPACEVTSSW CLFGPELPEDPPAAPATQRVLSPLMSRS GCKVGDSSGTAAAHKVLPRGLSPARQLL LPASESPHWSGAPVKPSPQAAAVEVEEE DGSESEESAGPLLKGKPRALGGAAAGGG AAAVPPGAAAGGVALVPKEDSRFSAPRV ALVEQDAPMAPGRSPLATTVMDFIHVPI LPLNHALLAARTRQLLEDESYDGGAGAA SAFAPPRSSPCASSTPVAVGDFPDCAYP PDAEPKDDAYPLYSDFQPPALKIKEEEE GAEASARSPRSYLVAGANPAAFPDFPLG PPPPLPPRATPSRPGEAAVTAAPASASV SSASSSGSTLECILYKAEGAPPQQGPFA PPPCKAPGASGCLLPRDGLPSTSASAAA AGAAPALYPALGLNGLPQLGYQAAVLKE GLPQVYPPYLNYLRPDSEASQSPQYSFE SLPQKICLICGDEASGCHYGVLTCGSCK VFFKRAMEGQHNYLCAGRNDCIVDKIRR KNCPACRLRKCCQAGMVLGGRKFKKFNK VRVVRALDAVALPQPVGVPNESQALSQR FTFSPGQDIQLIPPLINLLMSIEPDVIY AGHDNTKPDTSSSLLTSLNQLGERQLLS VVKWSKSLPGFRNLHIDDQITLIQYSWM SLMVFGLGWRSYKHVSGQMLYFAPDLIL NEQRMKESSFYSLCLTMWQIPQEFVKLQ VSQEEFLCMKVLLLLNTIPLEGLRSQTQ FEEMRSSYIRELIKAIGLRQKGVVSSSQ RFYQLTKLLDNLHDLVKQLHLYCLNTFI QSRALSVEFPEMMSEVIAAQLPKILAGM VKPLLFHKK	CACCTTGCCTGAAGTTTCGGCCATACCTATCT CCCTGGACGGGCTACTCTTCCCTCGGCCCTGC CAGGGACAGGACCCCTCCGACGAAAAGACGCA GGACCAGCAGTCGCTGTCGGACGTGGAGGGCG CATATTCCAGAGCTGAAGCTACAAGGGGTGCT GGAGGCAGCAGTTCTAGTCCCCCAGAAAAGGA CAGCGGACTGCTGGACAGTGTCTTGGACACTC TGTTGGCGCCCTCAGGTCCCGGGCAGAGCCAA CCCAGCCCTCCCGCCTGCGAGGTCACCAGCTC TTGGTGCCTGTTTGGCCCCGAACTTCCCGAAG ATCCACCGGCTGCCCCCGCCACCCAGCGGGTG TTGTCCCCGCTCATGAGCCGGTCCGGGTGCAA GGTTGGAGACAGCTCCGGGACGGCAGCTGCCC ATAAAGTGCTGCCCCGGGGCCTGTCACCAGCC CGGCAGCTGCTGCTCCCGGCCTCTGAGAGCCC TCACTGGTCCGGGGCCCCAGTGAAGCCGTCTC CGCAGGCCGCTGCGGTGGAGGTTGAGGAGGAG GATGGCTCTGAGTCCGAGGAGTCTGCGGGTCC GCTTCTGAAGGGCAAACCTCGGGCTCTGGGTG GCGCGGCGGCTGGAGGAGGAGCCGCGGCTGTC CCGCCGGGGGCGGCAGCAGGAGGCGTCGCCCT GGTCCCCAAGGAAGATTCCCGCTTCTCAGCGC CCAGGGTCGCCCTGGTGGAGCAGGACGCGCCG ATGGCGCCCGGGCGCTCCCCGCTGGCCACCAC GGTGATGGATTTCATCCACGTGCCTATCCTGC CTCTCAATCACGCCTTATTGGCAGCCCGCACT CGGCAGCTGCTGGAAGACGAAAGTTACGACGG CGGGGCCGGGGCTGCCAGCGCCTTTGCCCCGC CGCGGAGTTCACCCTGTGCCTCGTCCACCCCG GTCGCTGTAGGCGACTTCCCCGACTGCGCGTA CCCGCCCGACGCCGAGCCCAAGGACGACGCGT ACCCTCTCTATAGCGACTTCCAGCCGCCCGCT CTAAAGATAAAGGAGGAGGAGGAAGGCGCGGA GGCCTCCGCGCGCTCCCCGCGTTCCTACCTTG TGGCCGGTGCCAACCCCGCAGCCTTCCCGGAT TTCCCGTTGGGGCCACCGCCCCCGCTGCCGCC GCGAGCGACCCCATCCAGACCCGGGGAAGCGG CGGTGACGGCCGCACCCGCCAGTGCCTCAGTC TCGTCTGCGTCCTCCTCGGGGTCGACCCTGGA GTGCATCCTGTACAAAGCGGAGGGCGCGCCGC CCCAGCAGGGCCCGTTCGCGCCGCCGCCCTGC AAGGCGCCGGGCGCGAGCGGCTGCCTGCTCCC GCGGGACGGCCTGCCCTCCACCTCCGCCTCTG CCGCCGCCGCCGGGGCGGCCCCCGCGCTCTAC CCTGCACTCGGCCTCAACGGGCTCCCGCAGCT CGGCTACCAGGCCGCCGTGCTCAAGGAGGGCC TGCCGCAGGTCTACCCGCCCTATCTCAACTAC CTGAGGCCGGATTCAGAAGCCAGCCAGAGCCC ACAATACAGCTTCGAGTCATTACCTCAGAAGA TTTGTTTAATCTGTGGGGATGAAGCATCAGGC TGTCATTATGGTGTCCTTACCTGTGGGAGCTG TAAGGTCTTCTTTAAGAGGGCAATGGAAGGGC AGCACAACTACTTATGTGCTGGAAGAAATGAC TGCATCGTTGATAAAATCCGCAGAAAAACTG CCCAGCATGTCGCCTTAGAAAGTGCTGTCAGG CTGGCATGGTCCTTGGAGGTCGAAAATTTAAA AAGTTCAATAAAGTCAGAGTTGTGAGAGCACT GGATGCTGTTGCTCTCCCACAGCCAGTGGGCG TTCCAAATGAAAGCCAAGCCCTAAGCCAGAGA TTCACTTTTTCACCAGGTCAAGACATACAGTT GATTCCACCACTGATCAACCTGTTAATGAGCA

		TTGAACCAGATGTGATCTATGCAGGACATGAC AACACAAAACCTGACACCTCCAGTTCTTTGCT GACAAGTCTTAATCAACTAGGCGAGAGGCAAC TTCTTTCAGTAGTCAAGTGGTCTAAATCATTG CCAGGTTTTCGAAACTTACATATTGATGACCA GATAACTCTCATTCAGTATTCTTGGATGAGCT TAATGGTGTTTGGTCTAGGATGGAGATCCTAC AAACACGTCAGTGGGCAGATGCTGTATTTTGC ACCTGATCTAATACTAAATGAACAGCGGATGA AAGAATCATCATTCTATTCATTATGCCTTACC ATGTGGCAGATCCCACAGGAGTTTGTCAAGCT TCAAGTTAGCCAAGAAGAGTTCCTCTGTATGA AAGTATTGTTACTTCTTAATACAATTCCTTTG GAAGGGCTACGAAGTCAAACCCAGTTTGAGGA GATGAGGTCAAGCTACATTAGAGAGCTCATCA AGGCAATTGGTTTGAGGCAAAAAGGAGTTGTG TCGAGCTCACAGCGTTTCTATCAACTTACAAA ACTTCTTGATAACTTGCATGATCTTGTCAAAC AACTTCATCTGTACTGCTTGAATACATTTATC CAGTCCCGGGCACTGAGTGTTGAATTTCCAGA AATGATGTCTGAAGTTATTGCTGCACAATTAC CCAAGATATTGGCAGGGATGGTGAAACCCCTT CTCTTTCATAAAAAGTGA
AURKA	**[SEQ ID NO 9] –** MDRSKENCISGPVKATAPVGGPKRVLVT QQFPCQNPLPVNSGQAQRVLCPSNSSQR IPLQAQKLVSSHKPVQNQKQKQLQATSV PHPVSRPLNNTQKSKQPLPSAPENNPEE ELASKQKNEESKKRQWALEDFEIGRPLG KGKFGNVYLAREKQSKFILALKVLFKAQ LEKAGVEHQLRREVEIQSHLRHPNILRL YGYFHDATRVYLILEYAPLGTVYRELQK LSKFDEQRTATYITELANALSYCHSKRV IHRDIKPENLLLGSAGELKIADFGWSVH APSSRRTTLCGTLDYLPPEMIEGRMHDE KVDLWSLGVLCYEFLVGKPPFEANTYQE TYKRISRVEFTFPDFVTEGARDLISRLL KHNPSQRPMLREVLEHPWITANSSKPSN CQNKESASKQS	**[SEQ ID NO 10] –** ATGGACCGATCTAAAGAAAACTGCATTTCAGG ACCTGTTAAGGCTACAGCTCCAGTTGGAGGTC CAAACGTGTTCTCGTGACTCAGCAATTTCCT TGTCAGAATCCATTACCTGTAAATAGTGGCCA GGCTCAGCGGGTCTTGTGTCCTTCAAATTCTT CCCAGCGCATTCCTTTGCAAGCACAAAAGCTT GTCTCCAGTCACAAGCCGGTTCAGAATCAGAA GCAGAAGCAATTGCAGGCAACCAGTGTACCTC ATCCTGTCTCCAGGCCACTGAATAACACCCAA AAGAGCAAGCAGCCCCTGCCATCGGCACCTGA AAATAATCCTGAGGAGGAACTGGCATCAAAAC AGAAAAATGAAGAATCAAAAAAGAGGCAGTGG GCTTTGGAAGACTTTGAAATTGGTCGCCCTCT GGGTAAAGGAAAGTTTGGTAATGTTTATTTGG CAAGAGAAAAGCAAAGCAAGTTTATTCTGGCT CTTAAAGTGTTATTTAAAGCTCAGCTGGAGAA AGCCGGAGTGGAGCATCAGCTCAGAAGAGAAG TAGAAATACAGTCCCACCTTCGGCATCCTAAT ATTCTTAGACTGTATGGTTATTTCCATGATGC TACCAGAGTCTACCTAATTCTGGAATATGCAC CACTTGGAACAGTTTATAGAGAACTTCAGAAA CTTTCAAAGTTTGATGAGCAGAGAACTGCTAC TTATATAACAGAATTGGCAAATGCCCTGTCTT ACTGTCATTCGAAGAGAGTTATTCATAGAGAC ATTAAGCCAGAGAACTTACTTCTTGGATCAGC TGGAGAGCTTAAAATTGCAGATTTTGGGTGGT CAGTACATGCTCCATCTTCCAGGAGGACCACT CTCTGTGGCACCCTGGACTACCTGCCCCCTGA AATGATTGAAGGTCGGATGCATGATGAGAAGG TGGATCTCTGGAGCCTTGGAGTTCTTTGCTAT GAATTTTTAGTTGGGAAGCCTCCTTTTGAGGC AAACACATACCAAGAGACCTACAAAGAATAT CACGGGTTGAATTCACATTCCCTGACTTTGTA ACAGAGGGAGCCAGGGACCTCATTTCAAGACT GTTGAAGCATAATCCCAGCCAGAGGCCAATGC TCAGAGAAGTACTTGAACACCCCTGGATCACA

		GCAAATTCATCAAAACCATCAAATTGCCAAAA CAAAGAATCAGCTAGCAAACAGTCTTAG
KIF2C	[SEQ ID NO 11] – MAMDSSLQARLFPGLAIKIQRSNGLIHS ANVRTVNLEKSCVSVEWAEGGATKGKEI DFDDVAAINPELLQLLPLHPKDNLPLQE NVTIQKQKRRSVNSKIPAPKESLRSRST RMSTVSELRITAQENDMEVELPAAANSR KQFSVPPAPTRPSCPAVAEIPLRMVSEE MEEQVHSIRGSSSANPVNSVRRKSCLVK EVEKMKNKREEKKAQNSEMRMKRAQEYD SSFPNWEFARMIKEFRATLECHPLTMTD PIEEHRICVCVRKRPLNKQELAKKEIDV ISIPSKCLLLVHEPKLKVDLTKYLENQA FCFDFAFDETASNEVVYRFTARPLVQTI FEGGKATCFAYGQTGSGKTHTMGGDLSG KAQNASKGIYAMASRDVFLLKNQPCYRK LGLEVYVTFFEIYNGKLFDLLNKKAKLR VLEDGKQQVQVVGLQEHLVNSADDVIKM IDMGSACRTSGQTFANSNSSRSHACFQI ILRAKGRMHGKFSLVDLAGNERGADTSS ADRQTRMEGAEINKSLLALKECIRALGQ NKAHTPFRESKLTQVLRDSFIGENSRTC MIATISPGISSCEYTLNTLRYADRVKEL SPHSGPSGEQLIQMETEEMEACSNGALI PGNLSKEEEELSSQMSSFNEAMTQIREL EEKAMEELKEIIQQGPDWLELSEMTEQP DYDLETFVNKAESALAQQAKHFSALRDV IKALRLAMQLEEQASRQISSKKRPQ	[SEQ ID NO 12] – ATGGCCATGGACTCGTCGCTTCAGGCCCGCCT GTTTCCCGGTCTCGCTATCAAGATCCAACGCA GTAATGGTTTAATTCACAGTGCCAATGTAAGG ACTGTGAACTTGGAGAAATCCTGTGTTTCAGT GGAATGGGCAGAAGGAGGTGCCACAAAGGGCA AAGAGATTGATTTTGATGATGTGGCTGCAATA AACCCAGAACTCTTACAGCTTCTTCCCTTACA TCCGAAGGACAATCTGCCCTTGCAGGAAAATG TAACAATCCAGAAACAAAAACGGAGATCCGTC AACTCCAAAATTCCTGCTCCAAAGAAAGTCT TCGAAGCCGCTCCACTCGCATGTCCACTGTCT CAGAGCTTCGCATCACGGCTCAGGAGAATGAC ATGGAGGTGGAGCTGCCTGCAGCTGCAAACTC CCGCAAGCAGTTTTCAGTTCCTCCTGCCCCCA CTAGGCCTTCCTGCCCTGCAGTGGCTGAAATA CCATTGAGGATGGTCAGCGAGGAGATGGAAGA GCAAGTCCATTCCATCCGAGGCAGCTCTTCTG CAAACCCTGTGAACTCAGTTCGGAGGAAATCA TGTCTTGTGAAGGAAGTGGAAAAAATGAAGAA CAAGCGAGAAGAGAAGAAGGCCCAGAACTCTG AAATGAGAATGAAGAGAGCTCAGGAGTATGAC AGTAGTTTTCCAAACTGGGAATTTGCCCGAAT GATTAAAGAATTTCGGGCTACTTTGGAATGTC ATCCACTTACTATGACTGATCCTATCGAAGAG CACAGAATATGTGTCTGTGTTAGGAAACGCCC ACTGAATAAGCAAGAATTGGCCAAGAAAGAAA TTGATGTGATTTCCATTCCTAGCAAGTGTCTC CTCTTGGTACATGAACCCAAGTTGAAAGTGGA CTTAACAAAGTATCTGGAGAACCAAGCATTCT GCTTTGACTTTGCATTTGATGAAACAGCTTCG AATGAAGTTGTCTACAGGTTCACAGCAAGGCC ACTGGTACAGACAATCTTTGAAGGTGGAAAAG CAACTTGTTTTGCATATGGCCAGACAGGAAGT GGCAAGACACATACTATGGGCGGAGACCTCTC TGGGAAAGCCCAGAATGCATCCAAAGGGATCT ATGCCATGGCCTCCCGGGACGTCTTCCTCCTG AAGAATCAACCCTGCTACCGGAAGTTGGGCCT GGAAGTCTATGTGACATTCTTCGAGATCTACA ATGGGAAGCTGTTTGACCTGCTCAACAAGAAG GCCAAGCTGCGCGTGCTGGAGGACGGCAAGCA ACAGGTGCAAGTGGTGGGGCTGCAGGAGCATC TGGTTAACTCTGCTGATGATGTCATCAAGATG ATCGACATGGGCAGCGCCTGCAGAACCTCTGG GCAGACATTTGCCAACTCCAATTCCTCCCGCT CCCACGCGTGCTTCCAAATTATTCTTCGAGCT AAAGGGAGAATGCATGGCAAGTTCTCTTTGGT AGATCTGGCAGGGAATGAGCGAGGCGCGGACA CTTCCAGTGCTGACCGGCAGACCCGCATGGAG GGCGCAGAAATCAACAAGAGTCTCTTAGCCCT GAAGGAGTGCATCAGGGCCCTGGGACAGAACA AGGCTCACACCCCGTTCCGTGAGAGCAAGCTG ACACAGGTGCTGAGGGACTCCTTCATTGGGGA GAACTCTAGGACTTGCATGATTGCCACGATCT CACCAGGCATAAGCTCCTGTGAATATACTTTA AACACCCTGAGATATGCAGACAGGGTCAAGGA GCTGAGCCCCCACAGTGGGCCCAGTGGAGAGC AGTTGATTCAAATGGAAACAGAAGAGATGGAA GCCTGCTCTAACGGGGCGCTGATTCCAGGCAA

		TTTATCCAAGGAAGAGGAGGAACTGTCTTCCC AGATGTCCAGCTTTAACGAAGCCATGACTCAG ATCAGGGAGCTGGAGGAGAAGGCTATGGAAGA GCTCAAGGAGATCATACAGCAAGGACCAGACT GGCTTGAGCTCTCTGAGATGACCGAGCAGCCA GACTATGACCTGGAGACCTTTGTGAACAAAGC GGAATCTGCTCTGGCCCAGCAAGCCAAGCATT TCTCAGCCCTGCGAGATGTCATCAAGGCCTTG CGCCTGGCCATGCAGCTGGAAGAGCAGGCTAG CAGACAAATAAGCAGCAAGAAACGGCCCCAGT GA
ACTB	**[SEQ ID NO 13] –** MDDDIAALVVDNGSGMCKAGFAGDDAPR AVFPSIVGRPRHQGVMVGMGQKDSYVGD EAQSKRGILTLKYPIEHGIVTNWDDMEK IWHHTFYNELRVAPEEHPVLLTEAPLNP KANREKMTQIMFETFNTPAMYVAIQAVL SLYASGRTTGIVMDSGDGVTHTVPIYEG YALPHAILRLDLAGRDLTDYLMKILTER GYSFTTTAEREIVRDIKEKLCYVALDFE QEMATAASSSSLEKSYELPDGQVITIGN ERFRCPEALFQPSFLGMESCGIHETTFN SIMKCDVDIRKDLYANTVLSGGTTMYPG IADRMQKEITALAPSTMKIKIIAPPERK YSVWIGGSILASLSTFQQMWISKQEYDE SGPSIVHRKCF	**[SEQ ID NO 14] –** ATGGATGATGATATCGCCGCGCTCGTCGTCGA CAACGGCTCCGGCATGTGCAAGGCCGGCTTCG CGGGCGACGATGCCCCCGGGCCGTCTTCCCC TCCATCGTGGGGCGCCCCAGGCACCAGGGCGT GATGGTGGGCATGGGTCAGAAGGATTCCTATG TGGGCGACGAGGCCCAGAGCAAGAGAGGCATC CTCACCCTGAAGTACCCCATCGAGCACGGCAT CGTCACCAACTGGGACGACATGGAGAAAATCT GGCACCACACCTTCTACAATGAGCTGCGTGTG GCTCCCGAGGAGCACCCCGTGCTGCTGACCGA GGCCCCCCTGAACCCCAAGGCCAACCGCGAGA AGATGACCCAGATCATGTTTGAGACCTTCAAC ACCCCAGCCATGTACGTTGCTATCCAGGCTGT GCTATCCCTGTACGCCTCTGGCCGTACCACTG GCATCGTGATGGACTCCGGTGACGGGGTCACC CACACTGTGCCCATCTACGAGGGGTATGCCCT CCCCCATGCCATCCTGCGTCTGGACCTGGCTG GCCGGGACCTGACTGACTACCTCATGAAGATC CTCACCGAGCGCGGCTACAGCTTCACCACCAC GGCCGAGCGGGAAATCGTGCGTGACATTAAGG AGAAGCTGTGCTACGTCGCCCTGGACTTCGAG CAAGAGATGGCCACGGCTGCTTCCAGCTCCTC CCTGGAGAAGAGCTACGAGCTGCCTGACGGCC AGGTCATCACCATTGGCAATGAGCGGTTCCGC TGCCCTGAGGCACTCTTCCAGCCTTCCTTCCT GGGCATGGAGTCCTGTGGCATCCACGAAACTA CCTTCAACTCCATCATGAAGTGTGACGTGGAC ATCCGCAAAGACCTGTACGCCAACACAGTGCT GTCTGGCGGCACCACCATGTACCCTGGCATTG CCGACAGGATGCAGAAGGAGATCACTGCCCTG GCACCCAGCACAATGAAGATCAAGATCATTGC TCCTCCTGAGCGCAAGTACTCCGTGTGGATCG GCGGCTCCATCCTGGCCTCGCTGTCCACCTTC CAGCAGATGTGGATCAGCAAGCAGGAGTATGA CGAGTCCGGCCCCTCCATCGTCCACCGCAAAT GCTTCTAG
PPIB	**[SEQ ID NO 15] –** MLRLSERNMKVLLAAALIAGSVFFLLLP GPSAADEKKKGPKVTVKVYFDLRIGDED VGRVIFGLFGKTVPKTVDNFVALATGEK GFGYKNSKFHRVIKDFMIQGGDFTRGDG TGGKSIYGERFPDENFKLKHYGPGWVSM ANAGKDTNGSQFFITTVKTAWLDGKHVV FGKVLEGMEVVRKVESTKTDSRDKPLKD VIIADCGKIEVEKPFAIAKE	**[SEQ ID NO 16] –** ATGCTGCGCCTCTCCGAACGCAACATGAAGGT GCTCCTTGCCGCCGCCCTCATCGCGGGGTCCG TCTTCTTCCTGCTGCTGCCGGGACCTTCTGCG GCCGATGAGAAGAAGAAGGGGCCCAAAGTCAC CGTCAAGGTGTATTTTGACCTACGAATTGGAG ATGAAGATGTAGGCCGGGTGATCTTTGGTCTC TTCGGAAAGACTGTTCCAAAAACAGTGGATAA TTTTGTGGCCTTAGCTACAGGAGAGAAAGGAT TTGGCTACAAAAACAGCAAATTCCATCGTGTA ATCAAGGACTTCATGATCCAGGGCGGAGACTT CACCAGGGGAGATGGCACAGGAGGAAAGAGCA TCTACGGTGAGCGCTTCCCCGATGAGAACTTC

		AAACTGAAGCACTACGGGCCTGGCTGGGTGAG CATGGCCAACGCAGGCAAAGACACCAACGGCT CCCAGTTCTTCATCACGACAGTCAAGACAGCC TGGCTAGATGGCAAGCATGTGGTGTTTGGCAA AGTTCTAGAGGGCATGGAGGTGGTGCGGAAGG TGGAGAGCACCAAGACAGACAGCCGGGATAAA CCCCTGAAGGATGTGATCATCGCAGACTGCGG CAAGATCGAGGTGGAGAAGCCCTTTGCCATCG CCAAGGAGTAG
HPRT1	[SEQ ID NO 17] – MATRSPGVVISDDEPGYDLDLFCIPNHY AEDLERVFIPHGLIMDRTERLARDVMKE MGGHHIVALCVLKGGYKFFADLLDYIKA LNRNSDRSIPMTVDFIRLKSYCNDQSTG DIKVIGGDDLSTLTGKNVLIVEDIIDTG KTMQTLLSLVRQYNPKMVKVASLLVKRT PRSVGYKPDFVGFEIPDKFVVGYALDYN EYFRDLNHVCVISETGKAKYKA	[SEQ ID NO 18] – ATGGCGACCCGCAGCCCTGGCGTCGTGATTAG TGATGATGAACCAGGTTATGACCTTGATTTAT TTTGCATACCTAATCATTATGCTGAGGATTTG GAAAGGGTGTTTATTCCTCATGGACTAATTAT GGACAGGACTGAACGTCTTGCTCGAGATGTGA TGAAGGAGATGGGAGGCCATCACATTGTAGCC CTCTGTGTGCTCAAGGGGGGCTATAAATTCTT TGCTGACCTGCTGGATTACATCAAAGCACTGA ATAGAAATAGTGATAGATCCATTCCTATGACT GTAGATTTTATCAGACTGAAGAGCTATTGTAA TGACCAGTCAACAGGGGACATAAAAGTAATTG GTGGAGATGATCTCTCAACTTTAACTGGAAAG AATGTCTTGATTGTGGAAGATATAATTGACAC TGGCAAAACAATGCAGACTTTGCTTTCCTTGG TCAGGCAGTATAATCCAAAGATGGTCAAGGTC GCAAGCTTGCTGGTGAAAAGGACCCCACGAAG TGTTGGATATAAGCCAGACTTTGTTGGATTTG AAATTCCAGACAAGTTTGTTGTAGGATATGCC CTTGACTATAATGAATACTTCAGGGATTTGAA TCATGTTTGTGTCATTAGTGAAACTGGAAAAG CAAAATACAAAGCCTAA
TBP	[SEQ ID NO 19] – MDQNNSLPPYAQGLASPQGAMTPGIPIF SPMMPYGTGLTPQPIQNTNSLSILEEQQ RQQQQQQQQQQQQQQQQQQQQQQQQQQQQ QQQQQQQQQQQQAVAAAAVQQSTSQQATQ GTSGQAPQLFHSQTLTTAPLPGTTPLYP SPMTPMTPITPATPASESSGIVPQLQNI VSTVNLGCKLDLKTIALRARNAEYNPKR FAAVIMRIREPRTTALIFSSGKMVCTGA KSEEQSRLAARKYARVVQKLGFPAKFLD FKIQNMVGSCDVKFPIRLEGLVLTHQQF SSYEPELFPGLIYRMIKPRIVLLIFVSG KVVLTGAKVRAEIYEAFENIYPILKGFR KTT	[SEQ ID NO 20] – ATGGATCAGAACAACAGCCTGCCACCTTACGC TCAGGGCTTGGCCTCCCCTCAGGGTGCCATGA CTCCCGGAATCCCTATCTTTAGTCCAATGATG CCTTATGGCACTGGACTGACCCCACAGCCTAT TCAGAACACCAATAGTCTGTCTATTTTGGAAG AGCAACAAAGGCAGCAGCAGCAACAACAACAG CAGCAGCAGCAGCAGCAGCAGCAACAGCAACA GCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC AGCAGCAGCAGCAGCAGCAGCAGCAACAGGCA GTGGCAGCTGCAGCCGTTCAGCAGTCAACGTC CCAGCAGGCAACACAGGGAACCTCAGGCCAGG CACCACAGCTCTTCCACTCACAGACTCTCACA ACTGCACCCTTGCCGGGCACCACTCCACTGTA TCCCTCCCCCATGACTCCCATGACCCCCATCA CTCCTGCCACGCCAGCTTCGGAGAGTTCTGGG ATTGTACCGCAGCTGCAAAATATTGTATCCAC AGTGAATCTTGGTTGTAAACTTGACCTAAAGA CCATTGCACTTCGTGCCCGAAACGCCGAATAT AATCCCAAGCGGTTTGCTGCGGTAATCATGAG GATAAGAGAGCCACGAACCACGGCACTGATTT TCAGTTCTGGGAAAATGGTGTGCACAGGAGCC AAGAGTGAAGAACAGTCCAGACTGGCAGCAAG AAAATATGCTAGAGTTGTACAGAAGTTGGGTT TTCCAGCTAAGTTCTTGGACTTCAAGATTCAG AATATGGTGGGGAGCTGTGATGTGAAGTTTCC TATAAGGTTAGAAGGCCTTGTGCTCACCCACC AACAATTTAGTAGTTATGAGCCAGAGTTATTT CCTGGTTTAATCTACAGAATGATCAAACCCAG AATTGTTCTCCTTATTTTTGTTTCTGGAAAAG

		TTGTATTAACAGGTGCTAAAGTCAGAGCAGAA ATTTATGAAGCATTTGAAAACATCTACCCTAT TCTAAAGGGATTCAGGAAGACGACGTAA

Table 2 – protein and nucleic acid sequences of the exemplified genes associated with the cancer or at least one sub-class of cancer and normalisation genes

An accuracy of 99.5% is required in order for a test to have clinical significance and to be
5   adopted in practice. By "accuracy" we include the meaning of concordance with samples
of known sub-class as determined by a different method. For example, as described
herein samples are assessed with the method of the invention and compared to samples
of known sub-class, the status of which has in some embodiments been determined
previously by IHC or FISH. An accuracy or concordance of 100% means that all samples
10   that were previously identified as of a particular sub-class are also identified by the
methods described herein as being of that same sub-class. In one embodiment accuracy
is determined using well defined samples such as cell lines and patient samples that are
classified positive or negative to the measured analyte, example HER2 IHC scores of 3+
and 0 or 1+ on FFPE whole sections from tumour resections showing no heterogeneity
15   (only one tumour site). Discrepancies may occur wherein, for example, the status of a
sample has been previously characterised using FISH, i.e. determining gene copy number,
and has been determined as comprising a gene amplification, for example of the ERBB2
gene. If that amplification does not result in the expected increase in RNA level, then the
two results will not correlate. However it is considered that the methods of the present
20   invention provide a more accurate means of classification than at least FISH, and also IHC
due to the inherent variability in a subjective test.

In one embodiment the cancer is breast cancer. In a further embodiment the sub-classes
of cancer comprise at least one or more of HER2+, ER+, Basal, Triple Negative Breast
25   Cancer (TNBC), and Luminal A, Luminal B and HER2 enriched. In another embodiment
the sub-classes are TNBC, ER+ and HER2+. In yet another embodiment the sub-classes
are Basal, Luminal and HER2-enriched. The Luminal sub-class can be further classified
as Luminal A or Luminal B.

30   Accordingly, in one embodiment the method can distinguish, or classify, a breast cancer
as HER2+ or not HER2+. In another embodiment the method can distinguish, or classify,
a breast cancer as ER+ or not ER+. In another embodiment the method can distinguish
or classify a breast cancer as TNBC or not TNBC. In yet another embodiment the method
can distinguish or classify a breast cancer as Basal or not Basal. In another embodiment
35   the method can distinguish or classify a breast cancer as Luminal or not Luminal. In

another embodiment the method can distinguish or classify a breast cancer as Luminal A or not Luminal A. In another embodiment the method can distinguish or classify a breast cancer as Luminal B or not Luminal B. In another embodiment the method can distinguish or classify a breast cancer as HER2-enriched or not Her2-enriched.

In particular embodiments the method can distinguish or classify a breast cancer as TNBC, ER+ and/or HER2+ or not TNBC, ER+ and/or HER2+. In another embodiment the method can distinguish or classify a breast cancer as Basal, Luminal and HER2-enriched or not Basal, Luminal and HER2-enriched. The Luminal sub-class can be further classified as Luminal A or Luminal B.

In yet another embodiment the method can distinguish or classify a breast cancer as HER2+; ER+, HER2+ and ER+; or TNBC. In yet another embodiment the method can classify a breast cancer as Luminal or Basal breast cancer. Further, in another embodiment, the method can classify a breast cancer as Luminal A or Luminal B breast cancer. In a preferred embodiment the method can classify a breast cancer as HER2+; ER+; HER2+ and ER+; TNBC; Basal; Luminal A; and/ or Luminal B.

In one embodiment, the method comprises the determination of the level of RNA expression of ERBB2.

In one embodiment, the method comprises the determination of the level of RNA expression of ESR1.

In one embodiment, the method comprises the determination of the level of RNA expression of AURKA.

In one embodiment, the method comprises the determination of the level of RNA expression of KIF2C.

In one embodiment, the method comprises the determination of the level of RNA expression of PGR.

In one embodiment, the method comprises the determination of the level of RNA expression of FOXC1.

As discussed above, any determination of the expression level of a gene associated with a cancer or sub-class of a cancer requires normalisation against appropriately selected normalising genes. The inventors have shown that normalisation of expression data to at least any one of ACTB, HPRT1, TBP and PPIB; preferably all of ACTB, HPRT1, TBP and PPIB allows a highly accurate classification of breast cancers. Accordingly, in any embodiment of the invention, the expression levels as determined by the bDNA are

normalised to at least any one of ACTB, HPRT1, TBP and PPIB; preferably all of ACTB, HPRT1, TBP and PPIB.

Accordingly, in one embodiment the method comprises the determination of the level of RNA expression of any one or more of ACTB, PPIB, HPRT1 and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of ERBB2, ACTB, PPIB, HPRT1 and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of ESR1, ACTB, PPIB, HPRT1 and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of ESR1, ERBB2, ACTB, PPIB, HPRT1 and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of ERBB2; AURKA and/or KIF2C; ACTB; PPIB; HPRT1; and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of ESR1; AURKA and/or KIF2C; ACTB; PPIB; HPRT1; and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of ERBB2; ESR1; AURKA and/or KIF2C; ACTB; PPIB; HPRT1; and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of ERBB2; ESR1; AURKA; KIF2C; PGR; FOXC1; ACTB; PPIB; HPRT1; and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of ERBB2; ESR1; PGR; FOXC1; ACTB; PPIB; HPRT1; and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of AURKA; KIF2C; PGR; FOXC1; ACTB; PPIB; HPRT1; and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of PGR; FOXC1; ACTB; PPIB; HPRT1; and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of PGR; ACTB; PPIB; HPRT1; and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of ESR1; PGR; ACTB; PPIB; HPRT1; and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of AURKA; KIF2C; ACTB; PPIB; HPRT1; and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of ESR1; ERBB2; PGR; ACTB; PPIB; HPRT1; and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of ESR1; ERBB2; AURKA; ACTB; PPIB; HPRT1; and TBP.

In one embodiment the method comprises the determination of the level of RNA expression of ESR1; ERBB2; KIF2C; ACTB; PPIB; HPRT1; and TBP.

It will be appreciated that whilst it is possible for the skilled person to compare the expression levels of various genes in cancers from different sub-classes to identify which genes are expressed to which levels in each sub-class, the use of machine learning greatly enhances both the speed and accuracy of such determinations.

Accordingly in one embodiment the method further comprises the use of machine learning to classify the cancer into the sub-classes. The skilled person will be aware of suitable programmes and algorithms that can analyse the data, for example in one embodiment the machine learning is an algorithm, optionally is an algorithm selected from the group consisting of the Neural Network algorithm, Decision Tree, Random Forest and Support Vector Machine. In a preferred embodiment the machine learning is the Neural Network algorithm. Accordingly in one embodiment the method comprises the use of the Neural Network algorithm, Decision Tree, Random Forest and/or Support Vector Machine. In a preferred embodiment the method comprises the use of the Neural Network algorithm. In another preferred embodiment the method comprises the use of RapidMiner Studio Community software.

Such algorithms require training on a defined data set so that the algorithm can learn which features are associated with each other, for example which level of HER2 expression is

31

associated with a declaration of HER2+ status; which level of ER expression is associated with a declaration of ER+ status; which expression levels of certain genes are associated with the Luminal A and Luminal B sub-classes etc. Accordingly in one embodiment the algorithm, for example the Neural Network, has been trained on a suitable data set.

Suitable datasets are considered to include at least large databases of cancers samples, for example those found within the The Cancer Genome Atlas and Oncomine portals. An example of just one of the datasets that are available through such portals is the Breast Invasive Carcinoma (TCGA Cell, 2015) N=818 dataset. For example, such databases have details or the classification of samples, for examples samples may have been classed as various sub-classes of cancer based on FISH or IHC data. In a preferred embodiment the dataset contains details of samples that have been classified using a clinically relevant method, for example IHC and FISH, preferably without equivocal or heterogenous cases. The algorithm may be trained on all samples that samples that have been classified using a clinically relevant method, for example IHC and FISH, preferably without equivocal or heterogenous cases, or it may be necessary to remove certain samples prior to analysis, for example it may be necessary to first identify samples of a certain sub-class, for example HER2+ samples, and remove them from the dataset prior to training the algorithm for, for example Luminal vs Basal classification.

In one embodiment the algorithm used in the present invention is any one or more of:

Algorithm 1 – trained on normalised expression of ERBB2 in:
        a sample set comprising HER2+ samples as defined by IHC and FISH;
        and a sample set not comprising HER2+ samples as defined by IHC and FISH;

Algorithm 2 - trained on normalised expression of ESR1 in:
        a sample set comprising ER+ samples as defined by IHC;
        and a sample set not comprising ER+ samples as defined by IHC;

Algorithm 3 - trained on normalised expression of ERBB2 and ESR1 in:
        a sample set comprising HER2+ and/or ER+ samples as defined by IHC and FISH;
        and a sample set not comprising HER2+ and ER+ samples as defined by IHC and FISH;

Algorithm 4 – trained on normalised expression of ESR1, ERBB2, AURKA and/or KIF2C in:

a sample set comprising HER2+ and/or ER+ samples as defined by IHC and FISH; and a sample set not comprising HER2+ and/or ER+ samples as defined by IHC and FISH;

5   Algorithm 5 - trained on normalised expression of ERBB2 samples in a dataset from The Cancer Genome Atlas (TGCA) comprising
       HER2+ samples as defined by IHC and FISH; and
       samples defined as not HER2+ by IHC and FISH;

10   Algorithm 6 - trained on normalised expression of ERBB2 and ESR1 in samples in a dataset from the TCGA, comprising
       HER2+ and/or ER+ samples as defined by IHC and FISH; and
       Samples defined as not HER2+ and not ER+ by IHC and FISH;

15   Algorithm 7 - trained on normalised expression of ERBB2 and ESR1 in samples in a dataset from the TCGA which had HER2-enriched samples as defined by PAM50 and HER2+ as defined by IHC/FISH removed.

     Algorithm 8 - trained on normalised expression of ERBB2, ESR1, PGR, AURKA, KIF2C
20   and FOXC1 in samples in a dataset from the TCGA which had HER2-enriched samples as defined by PAM50 and HER2+ samples as defined by IHC/FISH removed.

     Algorithm 9 – trained on normalised expression of ERBB2, ESR1, FOXC1, PGR normalised expression in:
25   samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and
          samples in a dataset that comprises breast cancer cases that were classified as not Luminal by both PAM50 and the Algorithm 8;

30   Algorithm 10 – trained on normalised expression of ERBB2, ESR1, FOXC1, PGR, AURKA and KIF2C in:
          samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and
          samples in a dataset that comprises breast cancer cases that were classified as
35   not Luminal by both PAM50 and the Algorithm 8;

Algorithm 11 – trained on normalised expression of AURKA, KIF2C, FOXC1 and PGR in:

     samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and

     samples in a dataset that comprises breast cancer cases that were classified as not Luminal by both PAM50 and the Algorithm 8;

Algorithm 12 - Trained on normalised expression of FOXC1 and PGR in:

     samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and

     samples in a dataset that comprises breast cancer cases that were classified as not Luminal by both PAM50 and the Algorithm 8;

optionally wherein all expression levels were normalised to the expression levels of ACTB; PPIB; HPRT1; and/or TBP, preferably wherein all expression levels were normalised to the expression levels of ACTB; PPIB; HPRT1; and TBP.

Algorithm 1 identifies samples as HER2 positive or HER2 Negative.

Algorithm 2 identifies samples as ER positive or ER Negative.

Algorithm 3 identifies samples as HER2 positive and/or ER positive or TNBC.

Algorithm 4 identifies samples as HER2 positive and/or ER positive or TNBC.

Algorithm 5 identifies samples as HER2 positive or HER2 Negative.

Algorithm 6 identifies samples as HER2 positive and/or ER positive or TNBC.

Algorithm 7 identifies samples as Luminal or Basal.

Algorithm 8 identifies samples as Luminal or Basal.

Algorithm 9 identifies samples as Luminal A or Luminal B.

Algorithm 10 identifies samples as Luminal A or Luminal B.

Algorithm 11 identifies samples as Luminal A or Luminal B.

Algorithm 12 identifies samples as Luminal A or Luminal B.

In particular embodiments the method classifies the cancer as a HER2+ cancer or a HER2 negative cancer, optionally wherein the method comprises the determination of the level of expression of the ERBB2 gene, optionally further comprises normalisation of the level of expression of the ERBB2 gene to the expression levels of ACTB; PPIB; HPRT1; and

5   TBP, optionally further comprise the use of any one or more of Algorithm 1, Algorithm 3, Algorithm 4 and Algorithm 5 as defined above.

In another embodiment, the method classifies the cancer as an ER+ cancer or an ER negative cancer, optionally wherein the method comprises the determination of the level

10   of expression of the ESR1 gene, optionally further comprises normalisation of the level of expression of the ESR1 gene to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of any one or more of Algorithm 2, Algorithm 3, Algorithm 4 and Algorithm 6 as defined above.

15   In a further embodiment the method classifies the cancer as a HER2+, ER+ or TNBC, optionally wherein the method comprises the determination of the level of expression of the ERBB2, ESR1, AURKA and/or KIF2C genes, optionally further comprises normalisation of the level of expression of the ERBB2, ESR1, AURKA and/or KIF2C genes to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the

20   use of any one or more of Algorithm 3, Algorithm 4 and Algorithm 6 as defined above.

In yet another embodiment the method classifies the cancer as a Basal cancer or a Luminal cancer, optionally wherein the method comprises the determination of the level of expression of the ERBB2 and ESR1 genes, optionally further comprises normalisation of

25   the level of expression of the ERBB2 and ESR1 genes to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of Algorithm 7 as defined above.

In another embodiment the method classifies the cancer as a Basal cancer or a Luminal

30   cancer, optionally wherein the method comprises the determination of the level of expression of the ERBB2, ESR1, PGR, AURKA, KIF2C and FOXC1 genes, optionally further comprises normalisation of the level of expression of the ERBB2, ESR1, PGR, AURKA, KIF2C and FOXC1 genes to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of Algorithm 8 as defined above.

35

In a further embodiment the method classifies the cancer as a Luminal A or a Luminal B cancer, optionally wherein the method comprises the determination of the level of

expression of the ERBB2, ESR1, PGR and FOXC1 genes, optionally further comprises normalisation of the level of expression of the ERBB2, ESR1, PGR and FOXC1 genes, to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of Algorithm 9 as defined above.

In yet another embodiment the method classifies the cancer as a Luminal A or a Luminal B cancer, optionally wherein the method comprises the determination of the level of expression of the ERBB2, ESR1, PGR, AURKA, KIF2C and FOXC1 genes, optionally further comprises normalisation of the level of expression of the ERBB2, ESR1, PGR, AURKA, KIF2C and FOXC1 genes, to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of Algorithm 10 as defined above.

In another embodiment the method classifies the cancer as a Luminal A or a Luminal B cancer, optionally wherein the method comprises the determination of the level of expression of the PGR, AURKA, KIF2C and FOXC1 genes, optionally further comprises normalisation of the level of expression of the PGR, AURKA, KIF2C and FOXC1 genes, to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of Algorithm 11 as defined above.

In another embodiment the method classifies the cancer as a Luminal A or a Luminal B cancer, optionally wherein the method comprises the determination of the level of expression of the PGR and FOXC1 genes, optionally further comprises normalisation of the level of expression of the PGR and FOXC1 genes, to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of Algorithm 12 as defined above.

In a further embodiment the method classifies the cancer as HER2+, ER+, TNBC, Luminal, Basal, Luminal A and Luminal B by a) Identifying the HER2+ class as defined diagnostically by IHC (accuracy compared to IHC/FISH positive cases); b) Define between Luminal and Basal (concordance with PAM50); c) Luminal subclassification into Luminal A or Luminal B (prognostic significance); wherein part a) provides information to further classify the clinically actionable classes: HER+, ER+ and TNBC.

The skilled person will be aware of the algorithm parameters that must be defined using appropriate training sets. An example of suitable settings for the Neural Network Algorithm computed using RapidMiner Studio Community software is as follows:

500 training cycles

0.3 learning rate

0.2 momentum

Shuffle

5   Normalise

Error Epsilon: 1.0E-5.

See for example Figure 15.

Accordingly, through determination of the RNA expression level of the genes described herein, along with normalisation to the specific set of normalising genes as described

10   herein, and the use of a trained algorithm as described herein, the method of the invention is able to classify a breast cancer as either HER2+, ER+, Basal, Luminal A, Luminal B or TNBC, optionally wherein the method comprises the determination of the level of expression of the ERBB2, ESR1, PGR, AURKA, KIF2C and FOXC1 genes, optionally further comprises normalisation of the level of expression of the ERBB2, ESR1, PGR,

15   AURKA, KIF2C and FOXC1 genes, to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of the Algorithms as defined above.

Suitable combinations of algorithms are as follows:

20   a) Algorithm 1 and Algorithm 2 together; or Algorithm3; or Algorithm 4; or Algorithm 6; to classify the ER+, HER2+ and TNBC diagnostic classification. A preferred algorithm is Algorithm 4 (see Table 4).

b) the HER2+ diagnostic class (by IHC/FISH) can be defined by Algorithm 1 or 5 A

25   preferred algorithm is Algorithm 1.

c) the ER+ diagnostic class (by IHC) can be defined by Algorithm 2.

d) The ER+ and TNBC case set (for example as defined by any of algorithms 1 and 2; 3

30   or 4) or the HER2 negative case set (for example as defined by any of algorithms 1 or 5) can be re-classified into the molecular classes Luminal and Basal using Algorithms 7 or 8. A preferred algorithm is Algorithm 8 (see Table 10).

e) The Luminal class (as defined in (c) above) can be re-classified into Luminal A and

35   Luminal B using any of Algorithms 9-12. A preferred algorithm is Algorithm 10 (see Table 12).

A preferred sequential application of algorithms to expression data obtained from a single sample is as follows (summarised in Figure 4):

1.   Algorithm 4 – Identify HER2+ (Table 4)

2.   Algorithm 8 – identify Basal (Table 10)

3.   Algorithm 10 – identify Luminal A and Luminal B (Table 12)

The method of the invention is considered to be an *in vitro* method, wherein the method is performed on a sample obtained from a subject. However, there may be some situations whereby the method is suitable for *in vivo* use and *in vivo* detection of the cancer and/or sub-class of cancer, for example where the means required for performing the method are incorporated into an implantable sensor. Such *in vivo* use is also encompassed by the present invention and the method of the invention is in one embodiment an *in vivo* method. However in a preferred embodiment the method of the invention in an *in vitro* method.

It is considered that any sample type in which RNA (intact or degraded) is present, will work with the present invention. Data provided herein demonstrate that the method of the invention can classify breast cancer with a high degree of accuracy when the sample is a Hematoxylin and Eosin stained sample and/or is a sample comprising totally degraded RNA, i.e. a sample with RNA with a RIN value of around 2.0. Data also provided herein demonstrates that the use of the method with a sample comprising exosomes is appropriate.

The sample may be from any organism, for example any mammal, for example a human, a dog, a primate, cattle or any virus and other microorganisms. Preferably the sample is from a human.

The sample may be a fresh sample, for example may be a fresh biopsy of a tumour, or may be a fresh blood or plasma sample. The sample may be a sample that comprises exosomes. The sample may also be not a fresh sample, for example may be an archived sample, for example a sample that has been stained and/or frozen and/or fixed for example fixed with formalin and/or embedded for example embedded in paraffin.

The sample may be a tissue sample obtained from a subject; a cell line; a liquid biopsy, for example a blood sample or a plasma sample. In a preferred embodiment the liquid biopsy comprises circulating tumour cells. In another embodiment the liquid biopsy comprises exosomes.

In one embodiment the sample is a homogenate or lysate, for example a homogenate or lysate of a homogenous tumour. In another embodiment the sample is a homogenate or lysate of a heterogeneous tumour.

5

In one embodiment, the sample is an archived/historical sample, for example the sample is between 1 month-100 years old, optionally wherein the sample is at least 1 month old, optionally wherein the sample is at least 2 months old, or at least 3 months old, or at least 4 months old, or at least 5 months old, or at least 6 months old, or at least 7 months old,

10   or at least 8 months old, or at least 9 months old, or at least 10 months old, or at least 11 months old, or at least 12 months old for example 1 year old, or at least 1.5 years old, or at least 2 years old, or at least 3 years old, or at least 4 years old, or at least 5 years old, or at least 6 years old, or at least 7 years old, or at least 8 years old, or at least 9 years old, or at least 10 years old, or at least 11 years old, or at least 12 years old, or at least 13

15   years old, or at least 14 years old, or at least 15 years old, or at least 16 years old, or at least 17 years old, or at least 18 years old, or at least 19 years old, or at least 20 years old, or at least 25 years old, or at least 30 years old, or at least 40 years old, or at least 50 years old, or at least 60 years old, or at least 70 years old, or at least 80 years old, or at least 90 years old or at least 100 years old.

20

In another embodiment the sample is a fresh sample, for example the sample is less than 1 month old, optionally less than 4 weeks old, optionally less than 21 days old, optionally less than 14 days old, optionally less than 7 days old, optionally less than 6 days old, optionally less than 5 days old, optionally less than 4 days old, optionally less than 3 days

25   old, optionally less than 2 days old optionally less than 1 day old, optionally less than 18 hours old, optionally less than 12 hours old, optionally less than 6 hours old, optionally less than 5 hours old, optionally less than 4 hours old, optionally less than 3 hours old, optionally less than 2 hours old, optionally less than 1 hour old, optionally less than 30 minutes old, optionally less than 15 minutes old, optionally less than 10 minutes old, optionally less than

30   5 minutes old.

As discussed above, the present inventors have found that the present invention, involving the use of bDNA technology, is able to accurately classify cancer samples even when the sample comprises degraded RNA. Accordingly in one embodiment the sample comprises

35   RNA, including degraded RNA, for example wherein the sample has a RIN of 8 or less, for example less than 7.5, for example less than 7.0, for example less than 6.5, for example less than 6.0, for example less than 5.5, for example less than 5.0, for example less than

4.5, for example less than 4.0, for example less than 3.5, for example less than 3.2 , for example less than 3.0, for example less than 2.8, for example less than 2.6, for example less than 2.4, for example less than 2.2, for example less than 2.0, for example less than 1.8, for example less than 1.6, for example less than 1.4, for example less than 1.2, for example 1.0. The skilled person will be aware of the RIN parameter, for example it is discussed in https://www.agilent.com/cs/library/applications/5989-1165EN.pdf.

As demonstrated herein, the method of the present invention is not affected by formalin fixation of paraffin embedding. Accordingly, in one embodiment the sample has been formalin-fixed (FF). In the same or alternative embodiment the sample has been paraffin-embedded (PE). In a further embodiment the sample is a formalin-fixed paraffin-embedded sample (FFPE).

Also as demonstrated herein, the method of the present invention is not affected by staining, for example is not affected by Haematoxylin and Eosin staining. Accordingly in one embodiment the sample is a stained sample, for example is a Haematoxylin and Eosin stained sample.

It is considered useful if the sample is a selected part of the sample that is taken from the subject. As discussed above, tumours are often heterogeneous, often comprising numerous individual morphologically distinct tumours. For example, a breast cancer sample may contain several different tumours within it, for example may comprise both HER2+ and HER2 negative tumours. Whilst analysis of the entire tumour sample is considered to be useful, it is considered more useful if the individual tumours can be isolated and assessed separately. For example, analysing a whole tumour sample that actually contains both HER2+ and HER2- microtumours may give an overall HER2+ phenotype. This may direct the clinician to administer Herceptin therapy. Whilst Herceptin may be effective against the HER2+ microtumour, the HER2 negative microtumour is unlikely to be affected and may continue to grow unchecked.

Accordingly, an advantage of the fact that the present invention is not affected by staining is that it allows the staining of a sample, for example with Haematoxylin and Eosin, and the dissection of the individual morphologically distinct tumours followed by direct determination of RNA expression level, according to the present invention, classification of each individual microtumour that was isolated and analysed.

The skilled person will be well aware of suitable techniques for staining the samples and subsequent identification and isolation of morphologically distinct tumours. For example, in one embodiment the sample is a laser dissected sample or a macro dissected sample.

5   In another embodiment the sample is a sample of a morphologically distinct tumour within a larger tumour.

It is clear then that the present invention has the advantage of being able to detect morphologically distinct tumours of different sub-classes.

10

Accordingly, the invention further provides a method for the detection of heterogeneity in a tumour. Preferences for features discussed in relation to the first aspect of the invention also apply here. For example in one embodiment the method for the detection of heterogeneity in a tumour involves the detection of the RNA expression level of particular

15   genes, for instance the ERBB2 and ESR1 gene, via bDNA technology, for example via the QuantiGene Plex Assay, wherein the expression levels have been normalised to ACTB, PPIB, HPRT1, and TBP.

In one embodiment the method comprises:

20   the identification of morphologically distinct tumours identified in a sample obtained from a patient;
dissection, optionally laser dissection or macro dissection of the morphologically distinct tumours into two or more samples; and
performing a method according to the first aspect of the invention separately on each

25   sample.

In other embodiments, the method comprises performing a method according to the first aspect of the invention separately on more than one sample obtained from the tumour. For example, it may be preferable for samples to be taken from several different sites in

30   the tumour. The advantages of the present invention which include a rapid and accurate assessment of cancer type or cancer sub-class permits the routine classification of multiple samples from the same tumour. The skilled person can then readily determine if a subject has a tumour that has heterogeneity, and what that heterogeneity is, for example if it is a tumour that has HER2+ and HER2 negative microtumours, or both Luminal and Basal

35   microtumours. It is considered that such an approach would greatly inform the clinical practitioner's therapeutic strategy.

In addition, the present invention also provides the use of bDNA technology to predict the presence of a gene amplification.

It will be appreciated that the various methods described herein are amenable to being packaged as a kit or part of a kit. The invention therefore also provides kits that are suitable for use in the methods of the present invention.

In one embodiment the invention provides a kit for use in one or more methods of the invention, said kit comprising bDNA probes directed towards at least two of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1, ACTB, PPIB, HPRT1, and TBP.

For example the kit may contain bDNA probes directed towards ERBB2 and ACTB, PPIB, HPRT1, and TBP; or ESR1 and ACTB, PPIB, HPRT1, and TBP; or ERBB2, ESR1 and ACTB, PPIB, HPRT1, and TBP; or all of ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1, ACTB, PPIB, HPRT1, and TBP.

The kit may additionally or alternatively contain bDNA probes directed towards any two or more of ACTB, PPIB, HPRT1, and TBP; any three or more of ACTB, PPIB, HPRT1, and TBP; or all four of ACTB, PPIB, HPRT1, and TBP.

The kit may additionally or alternatively contain bDNA probes directed towards:
Any one or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1;
Any two or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1;
Any three or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1;
Any four or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1;
Any five or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1; or
All six of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1.

The kit may additionally or alternatively contain bDNA probes to AURKA and/or KIF2C.

The kit may also additionally or alternatively contain means for staining the sample to identify morphologically distinct tumours within the sample, for example may contain Haematoxylin & Eosin.

Any part of the kits described above are also for use in the diagnosis or prognosis of cancer.

The present invention also provides the use of any or part of the kits as described above in a method of manufacture of a composition for use in diagnosing or prognosing cancer.

5   As discussed above, the present methods are suitable for use on archived samples which means that a wealth of data can be mined using the present invention. The invention also therefore provides a method of validating a potential gene amplification or gene reduction as a biomarker, the method comprising the use of any one or more of the methods of the invention.

10   It will be apparent to the skilled person that the present invention can be used in a method of diagnosis or prognosis. For examples the methods described herein can be used to diagnose cancer, for example to diagnose breast cancer or can be used to diagnose a particular sub-class of cancer, for example to diagnose a subject has having a HER2+ breast cancer and/or an ER+ breast cancer and/or a TNBC and/or a Basal breast cancer

15   and/or a Luminal breast cancer, for example a Luminal A or a Luminal B breast cancer.

Similarly, the methods described herein can be used in a method for aiding in determining the prognosis of a subject, for example where a subject has been diagnosed with breast cancer, the methods of the invention can be used to aid in determining the prognosis of

20   the subject by classifying the sub-type or sub-types of cancer that the subject has. The skilled person will be aware that a diagnosis of Basal breast cancer is associated with a worse prognosis than HER2-enriched, for example, and that HER2-enriched is associated with a worse prognosis than Luminal B, for example, and that Luminal B is associated with a worse prognosis than Luminal A.

25
The skilled person will also understand that TNBC is associated with a worse prognosis than HER2+, and that HER2+ is associated with a worse prognosis than ER+.

30   It will be appreciated, particularly in view of the ability to use the present invention on archived samples that the present invention is suitable for research use. Accordingly in one embodiment the methods of the present invention are for research use.

It will be evident from the discussions above that knowledge of the cancer sub-class or

35   sub-classes that a subject has can inform treatment strategy. Accordingly, the present invention provides a method of selecting a suitable treatment strategy wherein the method comprises any of the methods of the invention as described above. For instance, where

the method is used to classify breast cancers, a diagnosis of a HER2+ breast cancer can be used to select the subject for treatment with Herceptin (trastuzumab), Kadcyla (Herceptin and emtansine), Nerlynx (neratinib), Perieta (pertuzumab), and/or Tykerb (lapatinib); a diagnosis of ER+ breast cancer can be used to select the subject for treatment with Tamoxifen, Aromatase Inhibitors, and/or SERMs; and a diagnosis of Basal breast cancer can be used to select the subject for treatment with other chemotherapeutic agents.

Accordingly, the invention also provides Herceptin (trastuzumab), Kadcyla (Herceptin and emtansine), Nerlynx (neratinib), Perieta (pertuzumab), and/or Tykerb (lapatinib) for use in the treatment of a subject with breast cancer wherein a sample from the subject has been identified as HER2+ by use of any of the preceding methods. Similarly the invention provides a method for treating HER2+ cancer wherein the subject has been diagnosed as having a HER2+ cancer or microtumour, wherein said method comprises administration of any one or more of Herceptin (trastuzumab), Kadcyla (Herceptin and emtansine), Nerlynx (neratinib), Perieta (pertuzumab), and/or Tykerb (lapatinib) to said subject.

Accordingly, the invention also provides Tamoxifen, Aromatase Inhibitors, and/or SERMs for use in the treatment of a subject with breast cancer wherein a sample from the subject has been identified as ER+ by use of any of the preceding methods.

The invention also provides various algorithms that have been trained on different datasets. Preferences for the algorithms as given above, for instance a preferred algorithm is the Neural Network algorithm. Accordingly the invention provides any one or more of Algorithms 1-12 as defined above. In one embodiment the algorithm is the Neural Network algorithm.

Since it has been previously shown that an overexpression of AURKA and/or KIF2C makes cancers susceptible to treatment with a PP2A activator, the invention also provides a method of treating a patient having a cancer comprising an overexpression of AURKA and/or KIF2C, comprising administering to the patient a PP2A activator and thereby treating the cancer, wherein the overexpression of AURKA and/or KIF2C has been determined by use of any of the methods disclosed herein. The invention further provides a method of treating cancer in a patient, the method comprising administering an AURKA antagonist and/or a KIF2C antagonist to the patient and thereby treating the cancer wherein the patient has a cancer with an overexpression of AURKA and/or KIF2C, wherein the overexpression of AURKA and/or KIF2C has been determined by use of any of the

methods disclosed herein. The invention additionally provides a method of treating cancer in a patient, or predicting sensitivity of a cancer to PP2A activators, the method comprising (a) measuring the amount of AURKA and/or KIF2C RNA in the cancer according to any of the methods described herein and optionally (b) if the cancer comprises an overexpression

5   of AURKA and/or KIF2C, administering to the patient a PP2A activator and thereby treating the cancer.

The invention also provides a PP2A activator for use in treating cancer in a patient, wherein the cancer comprises an overexpression of AURKA and/or KIF2C, wherein the

10   overexpression of AURKA and/or KIF2C has been determined by use of any of the methods disclosed herein, and use of a PP2A activator in the manufacture of a medicament for treating cancer in a patient, wherein the cancer comprises an overexpression of AURKA and/or KIF2C wherein the overexpression of AURKA and/or KIF2C has been determined by use of any of the methods disclosed herein. The invention

15   further provides an AURKA antagonist and/or a KIF2C antagonist for use in treating cancer in a patient, and use of an AURKA antagonist and/or a KIF2C antagonist in the manufacture of a medicament for treating cancer in a patient wherein the cancer has an overexpression of AURKA and/or KIF2C, wherein the overexpression of AURKA and/or KIF2C has been determined by use of any of the methods disclosed herein.

20

The invention additionally provides a method for detecting cancer in a patient, the method comprising measuring expression of AURKA and KIF2C in the patient in accordance with any of the methods described herein, wherein overexpression of AURKA and KIF2C indicates that the patient comprises a cancer. The invention further provides a method for

25   prognosing a cancer in a patient, the method comprising determining whether or not the cancer comprises an overexpression of AURKA and/or KIF2C in accordance with any of the methods described herein, wherein an overexpression of AURKA and/or KIF2C in the cancer indicates that the patient has a worse prognosis than in the situation of normal expression of AURKA and/or KIF2C. The invention also provides a method for determining

30   whether or not a patient having or suspected of having or being at risk of developing cancer will respond to treatment with a PP2A activator, which method comprises measuring expression of AURKA and/or KIF2C in the individual in accordance with any of the methods described herein, and thereby predicting whether or not the patient will respond to treatment with a PP2A activator.

35

The invention further provides a method for classifying a cancer in a patient, the method comprising measuring expression of AURKA and/or KIF2C in the patient in accordance

with any of the methods described herein, and classifying the cancer as of a particular subtype based on the expression. The invention also provides a kit for detecting a cancer comprising a deregulation of PP2A, comprising reagents suitable for detecting expression of AURKA and/or KIF2C. The invention also provides a system for detecting, classifying

5   or prognosing cancer in a patient, or for predicting responsiveness of a cancer patient to treatment with a PP2A activator, the system comprising: (a) a measuring module for determining expression of AURKA and/or KIF2C in the patient, (b) a storage module configured to store control data and output data from the measuring module, (c) a computation module configured to provide a comparison between the value of the output

10   data from the measuring module and the control data; and (d) an output module configured to display whether or not the patient has cancer based on the comparison, wherein an overexpression of AURKA and/or KIF2C in the patient indicates the presence of cancer, classifies the cancer, indicates a worse prognosis of cancer, or predicts that the patient will respond to treatment with a PP2A activator.

15

Figure Legends

Figure 1

A boxplot of ERBB2 mRNA transcript expression across cases grouped by the ERBB2

20   putative gene copy number. N = 960 where 147 (15.31%) cases show ERBB2 transcript overexpression while only 119 (12.4%) cases show ERBB2 gene amplification. Hence 4.17% of cases show increased HER2 expression but no gene amplification (false negative cases by genetic studies) and 1.25% of cases are ERBB2 gene amplified but do not show overexpression (false positive cases by genetic studies). Data retrieved from the

25   cBIOportal from the TCGA Breast Cancer dataset (TCGA, *Provisional*) on 13th June, 2017.

Figure 2

ERBB2 raw expression data against HER2 positive total RNA input derived from BT474 cell line. [A] LX200 – shows saturation at approximately 100ng of input RNA [B] Magpix –

30   shows signal saturation at approximately 180ng of RNA.

Figure 3

Quantigene results of Breast Cancer receptor status and normalising genes across a range of RNA concentration. The x-axis represents the concentration of RNA (μg/ml) and

35   the y-axis represent the Mean Fluorescence Intensity (MFI) measured using the Magpix Luminex instrument.   Gene expression was measured using the breast cancer cell line BT474, positive for both oestrogen receptor and HER2 (ERBB2). [A] Shows the full range

of MFIs measures; [B] zooms into the low MFIs to show the oestrogen receptor expression and normalising genes.

Figure 4

Algorithm flow chart from sample processing down to classification. [C] Shows the preferred algorithm 4 for defining HER2+ cases. [D] Includes the preferred algorithm 8 to define Basal and Luminal cases. [E] subclassifies Luminal cases into Luminal A and Luminal B using the preferred algorithm 10.

Figure 5

Distribution of expression levels of the TBP normalising gene of 98 breast cancer samples. [A] Raw MFI data of TBP [B] TBP MFI data expressed as Log10 and [C] TBP MFI Log10 data expressed as z-scores (MFI = Median Fluorescence Intensity)

Figure 6

Kaplan-Meier overall survival curve derived from the cBIOportal for the TCGA Breast cancer cohort for Luminal A and Luminal B as annotated with PAM50. Log Rank (Mantel-Cox) χ^2 = 6.310, p-value = 0.012

Figure 7

Kaplan-Meier overall survival curve derived from the cBIOportal for the TCGA Breast cancer cohort for Luminal A and Luminal B as predicted by the Neural Network based on the expression data of ESR1, ERBB2, PGR and FOXC1 (Algorithm 9). Log Rank (Mantel-Cox) χ^2 = 1.394, p-value = 0.238

Figure 8

Kaplan-Meier overall survival curve derived from the cBIOportal for the TCGA Breast cancer cohort for Luminal A and Luminal B as predicted by the Neural Network based on the expression data of AURKA, KIF2C, ESR1, ERBB2, PGR and FOXC1 (Algorithm 10). Log Rank (Mantel-Cox) χ^2 = 3.953, p-value = 0.047

Figure 9

Kaplan-Meier overall survival curve derived from the cBIOportal for the TCGA Breast cancer cohort for Luminal A and Luminal B as predicted by the Neural Network based on the expression data of PGR, FOXC1, AURKA and KIF2C (Algorithm 11). Log Rank (Mantel-Cox) χ^2 = 3.906, p-value = 0.048

Figure 10

Kaplan-Meier overall survival curve derived from the cBIOportal for the TCGA Breast cancer cohort for Luminal A and Luminal B as predicted by the Neural Network based on the expression data of PGR and FOXC1 (Algorithm 12). Log Rank (Mantel-Cox) χ^2 = 0.95,

5   p-value = 0.330.

Figure 11

Prediction of breast cancer molecular classification using known and novel luminal / basal classifiers and HER2 expression. The Principal component analysis (PCA) plot shows

10   prediction of 12-gene annotated Basal (red), Luminal (blue) and HER2-enriched (green) samples using Quantigene 6-plex expression data from FFPE patient material. The table shows a 100% concordance of the HER2 enriched with IHC and FISH results, basal subtype representing the Triple negative breast cancer (TNBC) patients and a high representation of ER positivity in the Luminal group as expected.

15

Figure 12

RNA Integrity numbers and concentration for a ladder, MDAMB453 RNA and degraded RNA controls, a BT474 RNA degradation array and 3 RNA samples extracted from patient derived breast cancer from FFPE blocks.

20

Figure 13

Raw expression profiles across degraded BT474 RNA samples for [A] ERBB2 expression as MFI and [B] normalising gene expression as MFI.

25   Figure 14

Normalised ERBB2 expression profile across degraded BT474 RNA samples for Relative expression between different genes in the RNA profile is maintained across the RNA degradation gradient.

30   Figure 15

Structure of data processing module. Expression Data from FFPE patient data [Localdata] is filtered for outliers using the Filter Examples operator. ERBB2 expression data and HER2 diagnostic (IHC/FISH) status is retrieved by the Select Attributes operator. The Set Role operator defines the HER2 diagnostic status as the variable to predict. The

35   streamlined data is used to train the Neural Network operator which is applied using the Apply model operator to the BT474 RNA degradation series data that is processed in an

identical fashion. The output from the Apply model operator provides a label for the unlabelled data along with a confidence value for the prediction.

Figure 16

Correlation between an expression profile derived from an unstained tumour section as opposed to a stained tumour section. [Pearson Correlation p-value = 5.34E-26]

Figure 17

Laser Microdissection of FFPE tissues. **[A]**: H&E stained slide; **[B]**: Immunohistochemical staining for ER expression; **[C]**: HER2 immunohistochemical staining; **[D]**: Unstained 20μm section on laser microdissection membrane slides. The yellow arrow indicates an area of invasive tumour that is not clearly demarcated due to lack of staining. **[E]**: A 20μm section stained with H&E for better delineation of areas of interest. White arrows indicate laser dissection trail while the Red arrow shows the laser focus during dissection. All illustrations were captured at 10x magnification.

Figure 18

Expression of **[A]**: receptor status and **[B]**: mesenchymal marker FN1, in breast tumour compared to matched normal. [K-W: Kruskal Wallis Test Statistic]

Figure 19

Case Study: Tumour heterogeneity. Morphologically distinct tumours were microdissected and treated as distinct samples. **[A]**: shows the master scan of the H&E section. **[B] & [C]**: show a 10x and 40x magnification respectively for each tumour morphology identified; **[D]**: Immunohistochemical staining for ER at 10x magnification; **[E]**: Immunohistochemical staining for Ki67 at 10x magnification showing a higher mitotic activity in tumour 1; **[F]**: Normalised expression levels for the ESR1 gene in each tumour showing relatively high and equal expression between tumours as expected from the immunohistochemical result; **[G]**: Normalised expression levels of FN1, a mesenchymal marker, where increased FN1 expression is accompanied by high Ki67 staining showing a signature for a more invasive tumour. The inverse is observed in tumour 2 which appears to be a slower proliferating tumour with a lower malignant potential represented by reduced FN1 expression.

Figure 20

Table indicating sources of material and equipment used in the Examples.

Figure 21

MTT assays for cell viability of various breast cancer cell lines at increasing concentrations of FTY720 ranging from 0 (untreated) to 5μM. Cells were left to adhere for 24 hours following seeding, then treated for 48 hours in duplicate experiments. The cell viability of the cell lines at each dose is expressed as a percentage of the untreated cells. Error bars are expressed as percentages of the untreated cell viability and represent the standard deviation from triplicate assay values for biological replicates. **[A]** Sensitive basal and triple negative breast cancer (TNBC) cell lines; **[B]** Basal and triple negative cell lines that are not sensitive to FTY720; **[C]** Normal-like cell lines: HB-2 and MCF10A; **[D]** MCF-7, MDAMB453 Luminal, ER positive cell lines; and **[E]** Luminal, HER2 positive SKBR-3 and BT-474.

Figure 22

Distribution of gene expression levels derived from Quantigene assay intensity data, normalised to housekeeping genes, comparing sensitive versus non-sensitive cell lines. RNA expression was measured by Quantigene and normalised to housekeeping gene expression of each cell line respectively. **[A]** Expression levels of PP2A complex subunits: PPP2CA (white), PPP2R2A (hatched) and the inhibitory subunit, CIP2A (black). The adjacent table shows the statistical significance of the difference in the median expression between the two categories of cell lines that are sensitive or not sensitive to FTY720 [* shows significant P-values to 95% confidence interval]. **[B]** expression levels of AURKA (black) and KIF2C (white). The adjacent table shows a statistically significant difference in expression between sensitive and non-sensitive cell lines for AURKA and KIF2C gene expression. Statistical analysis was done using the Mann-Whitney U test and a p-value smaller than 0.05 was considered significant. The expression of CIP2A, AURKA and KIF2C are significantly higher in the FTY720-senstive cell lines (p values of <0.05, <0.02, <0.001 respectively). Sensitive cell lines thus include the TNBC cell lines MDAMB231, BT-20 and Hs578T, while non-sensitive cell lines include HCC1937, MDAMB436, MDAMB468, MDAMB453, BT-474, MCT-7 and SKBR-3.

Figure 23

AURKA (white) and KIF2C (black) expression levels derived from RNASeq dataset downloaded from TCGA data portal (n=72), **[A]** Distribution of AURKA and KIF2C expression across breast cancer subtypes as defined by the PAM50 annotation, compared to normal tissues. [Normal N = 82; Basal N = 97; HER2-Enriched N = 82; Luminal B N = 129; Luminal A N = 228]. **[B]** Comparison of expression in patient tumour tissue and matched normal tissue. The statistical significance of the differential expression between

tumours and matched normal tissue is shown in the adjacent table using the Related samples Wilcoxon signed rank test. The tumour versus normal column shows the direction of significant difference where applicable. [n = 72; RNASeqV2 normalised expression as downloaded from TCGA data portal]. **[C]** Comparison of expression in patient tumour tissue and matched normal tissue across different PAM50 defined patient subtypes (N = 55).

Figure 24

Percent patients with amplifications or an expression level (z-score) greater than 2 (AMP; EXP>2) for AURKA (white bars) and KIF2C (black bars) in tumours of different origin. The analysis of data was done using the data portal (cBioPortal for Cancer Genomics (Cerami et al., 2012), available at http://www.cbioportal.org. The normalised RNASeqV2 data was used from the TCGA dataportal (https://tcga-data.nci.nih.gov/tcga/).

Figure 25

Expression of the selected probes (genes) in the colorectal cancer cell lines COLO320. The x-axis represents the expression in exosomes derived from COLO320 and the y-axis the expression measured using RNA from originator cells.

Figure 26

Expression of AURKA (Log_{10}) in cell line derived exosomes can be used to normalise and characterise HER2 expression into positive or negative.

Further embodiments of the invention provide:

1.   A method of treating a patient having a cancer comprising an overexpression of AURKA and/or KIF2C, comprising administering to the patient a PP2A activator and thereby treating the cancer.

2.   The method according to embodiment 1, wherein said patient has an overexpression of
AURKA and KIF2C.

3.   A method of treating cancer in a patient, the method comprising administering an AURKA antagonist and/or a KIF2C antagonist to the patient and thereby treating the cancer.

4.   The method according to embodiment 3, further comprising administering a PP2A
activator.

5.   The method according to any one of the preceding embodiments, which is for
treating
breast cancer.

6.   The method according to embodiment 5, which is for treating basal breast cancer.

7.   The method according to any one of the preceding embodiments, wherein the
cancer has
an underexpression of one or more of HER2, ER and PR.

8.   A method according to embodiment 7, wherein the cancer has an underexpression
of
HER2, ER and PR.

9.   A method according to any one of the preceding embodiments, wherein the patient
is
selected for said treatment on the basis of having a cancer comprising an overexpression
of AURKA and/or KIF2C.

10.   A method according to any one of the preceding embodiments, wherein the patient
is
selected for said treatment on the basis of the cancer having an underexpression of one
or more of HER2, ER and PR.

11.   A method according to embodiment 9 or 10 wherein said selection for treatment is
carried
out on diagnosis of the cancer, during treatment of the cancer, or following resistance to a
cancer therapy.

12.   A method according to any one of the preceding embodiments, wherein the PP2A
activator
or the AURKA antagonist and/or KIF2C antagonist is a small molecule, a protein, an
antibody, a polynucleotide, an oligonucleotide, an antisense RNA, a small interfering RNA
(siRNA) or a small hairpin RNA (shRNA).

13.   A method according to any one of the preceding embodiments, wherein the PP2A activator

and/or the AURKA antagonist and/or KIF2C antagonist is administered in combination

with another cancer therapy.

14.   A method according to any one of the preceding embodiments, wherein the patient

is

human.

15.   A method of treating cancer in a patient, the method comprising (a) measuring the amount of AURKA and/or KIF2C in the cancer and (b) if the cancer comprises an overexpression of AURKA and/or KIF2C, administering to the patient a PP2A activator and thereby treating the cancer.

16.   A PP2A activator for use in treating cancer in a patient, wherein the cancer comprises an overexpression of AURKA and/or KIF2C.

17.   An AURKA antagonist and/or a KIF2C antagonist for use in treating cancer in a patient.

18.   Use of a PP2A activator in the manufacture of a medicament for treating cancer in a patient, wherein the cancer comprises an overexpression of AURKA and/or KIF2C.

19.   Use of an AURKA antagonist and/or a KIF2C antagonist in the manufacture of a medicament for treating cancer in a patient.

20.   A method for detecting cancer in a patient, the method comprising measuring expression of AURKA and KIF2C in the patient, wherein overexpression of AURKA and KIF2C indicates that the patient comprises a cancer.

21.   A method for prognosing a cancer in a patient, the method comprising determining whether or not the cancer comprises an overexpression of AURKA and/or KIF2C, wherein an overexpression of AURKA and/or KIF2C in the cancer indicates that the patient has a worse prognosis than in the situation of normal expression of AURKA and/or KIF2C.

22.   *A method for determining whether or not a patient having or suspected of having or being at risk of developing cancer will respond to treatment with a PP2A activator, which method comprises measuring expression of AURKA and/or KIF2C in the individual, and thereby predicting whether or not the patient will respond to treatment with a PP2A*

5   *activator.*

23.   *A method for classifying a cancer in a patient, the method comprising measuring expression of AURKA and/or KIF2C in the patient, and classifying the cancer as of a particular subtype based on the expression.*

10

24.   *The method according to any one of embodiments 20 to 23, which further comprises*

measuring expression of one or more of HER2, ER and PR in the patient.

15   25.   *The method according to any one of embodiments 20 to 24, which comprises determining whether there is a deregulation of PP2A in the patient.*

26.   *The method according to any one of embodiments 20 to 25, wherein the cancer is breast cancer.*

20

27.   *The method according to embodiment 23 or any of embodiments 24 to 26 as dependent on embodiment*

23, comprising classifying the cancer as a luminal or basal breast cancer, wherein overexpression of AURKA and/or KIF2C classifies the cancer as a basal breast cancer,

25   *and wherein normal expression of AURKA and/or KIF2C classifies the cancer as a luminal breast cancer.*

28.   *The method according to embodiment 27, further comprising measuring expression of one*

30   *or more of HER2, ER and PR to further classify the cancer.*

29.   *A kit for treating cancer comprising (a) one or more reagents suitable for measuring expression of AURKA and/or KIF2C and (b) a PP2A activator.*

35   30.   *A kit for detecting a cancer comprising a deregulation of PP2A, comprising reagents suitable for detecting expression of AURKA and/or KIF2C.*

31.   The kit according to embodiment 30, which comprises reagents suitable for detecting

expression of AURKA and KIF2C.

5   32.   The kit according to embodiment 30 or 31 further comprising reagents suitable for detecting expression of an endogenous inhibitor of PP2A or a PP2A subunit.

33.   The kit according to any one of embodiments 30-32 wherein the reagents suitable for

10   detecting expression are selected from nucleic acid probes or primers and antibodies.

34.   A system for detecting, classifying or prognosing cancer in a patient, or for predicting responsiveness of a cancer patient to treatment with a PP2A activator, the system comprising:

15

(a) a measuring module for determining expression of AURKA and/or KIF2C in the patient,

(b) a storage module configured to store control data and output data from the measuring module,

20   (c) a computation module configured to provide a comparison between the value of

the output data from the measuring module and the control data; and

(d) an output module configured to display whether or not the patient has cancer based on the comparison,

25   wherein an overexpression of AURKA and/or KIF2C in the patient indicates the presence of cancer, classifies the cancer, indicates a worse prognosis of cancer, or predicts that the patient will respond to treatment with a PP2A activator.

The listing or discussion of an apparently prior-published document in this specification

30   should not necessarily be taken as an acknowledgement that the document is part of the state of the art or is common general knowledge.

Preferences and options for a given aspect, feature or parameter of the invention should, unless the context indicates otherwise, be regarded as having been disclosed in

35   combination with any and all preferences and options for all other aspects, features and parameters of the invention. For example, the invention provides a method for classifying a cancer as a Luminal A or a Luminal B cancer, and involves the determination of the level

of expression of the ERBB2, ESR1, PGR, AURKA, KIF2C and FOXC1 genes in a formalin-fixed paraffin-embedded sample (FFPE) of a morphologically distinct tumours identified in a sample obtained from a patient wherein the sample is 5 months old.

5 ***Examples***

Here we describe a technology that provides advantages to measure gene amplifications using RNA-based multiplex measurements that can be adapted to various actionable gene amplifications summarised below.

10 **Example 1 - Selection of appropriate normalisation genes**

Assay Linearity – HER2

This exercise aims to identify the dynamic range for HER2 on both the LX200 and Magpix. This establishes the ideal RNA concentration for the detection of these expressions (linear
15 phase).

Selection of normalising genes using Cell line assays

Well annotated Breast cancer cell lines were cultured and RNA isolated. RNA was quantified using Agilent Bioanalyser and diluted to prepare a range of concentrations. A
20 multi-plex Luminex based RNA assay was used to study the expression of HER2 (ERBB2) and ESR1 (oestrogen receptor) across the various concentrations to assess sensitivity of the assay. Normalising genes were included to select the best combination to ensure proper threshold settings for the simultaneous measurement of the lowly expressing ESR1 and the amplified HER2 gene.

25

As indicated in Figure 2, the normalising genes ACTB and GAPDH are expressed at high levels using the cell line BT474 and at a concentration of 8 µg/ml (400ng total RNA input) the signal reaches saturation and hence can result in false negative when normalising the HER2 (ERBB2) expression. To utilise these normalising genes, the concentration should
30 not exceed 4µg/ml (200ng input RNA). GAPDH shows a significantly lower expression in patient material with some extreme outliers, becoming among the lowest expressing normalising gene within the analysed normalising geneset. GAPDH does not follow the same trends as the other normalising genes analysed and thus was excluded from further analysis. The normalising genes ACTB, PPIB, HPRT1 and TBP were selected to
35 normalise the MFI. This allowed us to normalise the lowly expressed oestrogen receptor gene and the highly expressed amplified HER2 in one run.

Example 2 - Algorithm description

See Figure 2.

Sample population and assay methodology

A local breast cancer dataset (N=98) was assembled from breast cancer diagnosed
between 2007-2012. Equal numbers from each breast cancer class (ER, HER2, TNBC
and ER and HER2 positive) were recruited. Haematoxylin & Eosin (H&E) slides were
examined to identify the tumour area. The tumour area was isolated by microdissection or
macrodissection depending on the tumour morphology, with the aim of having the highest
tumour content within the sample. The dissected section was stained with H&E to confirm
tumour cell content. Following sample lysis using a ratio of 2.5mm^3 per 300µL of
Homogenizing mix, RNA expression is measured using a custom designed Invitrogen
Quantigene 2.0 10-plex assay.

[A] Data quality control

Quality control starts by assessing the bead count per region. If the bead count is less than
30, sample is regarded as inadequate and a repeat would be necessary. Low bead counts
are generally caused by bead clumping which can be avoided by obtaining a cleaner
sample (less paraffin and tissue fragments) or by diluting the sample.

The Limit of Detection (LOD) is determined as the average of the blank + 3 Standard
Deviations of the same blanks for each gene. A minimum of 3 blank wells are required per
assay. Sample Median Fluorescence Intensity (MFI) expression values lower than the
LOD are regarded as undetected for each respective gene.

[B] Detecting outliers

Normalising gene data is expressed as a log base 10, resulting in a normally distributed
dataset.

The z-score of the log is computed based on the average and standard deviation derived
from the population set (N=98; N=94 complete with HER2 and ER IHC data). A threshold
to select outliers was set based on concordance for the classification of HER2 status
defined by Immunohistochemistry (IHC) and Fluorescence in situ Hybridisation (FISH)
(N=28/28 HER2 positive), using ERBB2 expression normalised to normalising gene sets.
The same exercise was performed using the ESR1 expression to predict the Oestrogen
receptor (ER) status (N=42/43 ER positive). The best classification with the minimal loss
of samples as outliers was selected:

Outlier threshold for saturated or too dilute samples: N of outliers = 12; N; N of adequate samples = 86, 82 are annotated by the ER and HER2 IHC.

-1.6 < TBP log z-score < 1.6

5   _Classification of Breast Cancer Classes_

The Breast cancer classes (TNBC, ER+ and HER2+) can be predicted using ESR1 and ERBB2 expression data normalised to 4 normalising genes (ACTB, PPIB, HPRT1, and TBP) with up to 98.78% accuracy using the Neural network algorithm (1 out of 82 cases is misclassified).

10

Breast Cancer Classification by IHC (ER, HER2)		_true TNBC_	_true ER_	_true HER2_	_class precision_
Neural Net prediction based on ESR1 and ERBB2 expression	pred. TNBC	23	0	0	100.00%
	pred. ER	1	29	0	96.67%
	pred. HER2	0	0	29	100.00%
	class recall	95.83%	100.00%	100.00%	

Table 3: Crosstab of Breast cancer classification defined by IHC and FISH compared to Neural Network prediction based on ESR1 and ERBB2 normalised expression. Classification accuracy of 98.78%, N=82

15

Adding AURKA and/or KIF2C gives a 100% accuracy for this classification when using Neural Net algorithm.

Breast Cancer Classification by IHC (ER, HER2)		_true TNBC_	_true ER_	_true HER2_	_class precision_
Neural Net prediction based on ESR1, ERBB2, (AURKA and / or KIF2C) expression	pred. TNBC	24	0	0	100.00%
	pred. ER	0	29	0	100.00%
	pred. HER2	0	0	29	100.00%
	class recall	100.00%	100.00%	100.00%	

20

Table 4: Crosstab of Breast cancer classification defined by IHC and FISH compared to Neural Network prediction based on ESR1, ERBB2 and AURKA and/or KIF2C normalised expression. Classification accuracy of 100%, N=82

25   _TCGA – PAM50 and Diagnostic IHC breast cancer classification_

RNASeq V2 data of N = 911

- Breast cancer cases annotated with PAM50 (N=512)
- annotated with complete IHC and FISH data (N=470)
- annotated with complete IHC and FISH data and PAM50 (N=316)

Data was retrieved on 16th January, 2016 from The Cancer Genome Atlas Portal (https://tcga-data.nci.nih.gov/tcga/) from the Breast Invasive Carcinoma (TCGA, Provisional) cohort. Samples derived from normal tissue was omitted for this analysis.

Concordance of PAM50 with HER2 classification and Breast cancer classes as defined by IHC (N = 317).

HER2 Status by IHC and FISH		HER2 Negative	HER2 Positive	Class Precision
PAM50	Basal or Luminal	248	24	91.2%
	HER2-Enriched	12	33	73.3%
	Class recall	95.4%	57.9%	

Table 5: Crosstab of HER2 Status defined by IHC and FISH compared to PAM50 HER2-Enriched or otherwise. Classification concordance of 88.64%, N=317

HER2 cases were selected using the Neural Network algorithm based on the ERBB2 gene expression. Other algorithms such as Decision Tree, Random Forest and Support Vector Machine can be applied here with as similar prediction (Respective concordance of 95.0%, 95.3% and 94.0%).

HER2 status by IHC		HER2 Negative	HER2 Positive	Class precision
Neural Net prediction based on ERBB2 expression	pred. HER2 negative	328	18	94.8%
	pred. HER2 positive	3	52	94.6%
	Class recall	99.1%	74.3%	

Table 6: Crosstab of Breast cancer classification of HER2 status defined by IHC and FISH compared to Neural Net prediction based on the normalised expression of ERBB2 and shown as pred. HER2 negative and pred. HER2 positive. Classification concordance of 94.76%, N=401

Breast Cancer Class by IHC/FISH		ER	HER2	TNBC	Class precision
PAM50	Luminal	195	22	2	89.0%
	Basal	5	2	46	86.8%
	HER2-Enriched	9	32	3	72.7%
	Class recall	93.3%	57.1%	90.2%	

Table 7: Crosstab of Breast cancer classification defined by IHC and FISH compared to PAM50 Luminal, Basal and HER2-Enriched. Classification concordance of 86.39%, N=316

Breast Cancer Class by IHC/FISH		ER	HER2	TNBC	Class precision
Prediction by Neural Network based on ESR1 and ERBB2 expression	pred. ER	263	16	2	93.6%
	pred. HER2	2	52	0	96.3%
	pred. Triple Negative	8	1	58	86.6%

Table 8: Crosstab of Breast cancer classification defined by IHC and FISH compared to Neural Net prediction based on the normalised expression of ESR1 and ERBB2 and shown as pred. ER, pred. HER2 and pred. TNBC. Classification concordance of 92.79%, N=402

[C] **Algorithm to select HER2 positive patients:**

HER2 status as determined by IHC and FISH can be predicted perfectly by Neural Network based on the expression of ERBB2 normalised expression. A Neural Network model trained on available dataset can be used to predict status of unknown cases.

Neural Network prediction of TCGA IHC/FISH data based on ERBB2 normalised expression (ACTB, PPIB, HPRT1, TBP).

Prediction of N = 401 breast cancer for HER2 status (Table 65)

[D] **Luminal Basal Classification**

The HER2 cases N=55/401 were removed from the TCGA dataset as HER2 positive cases. The PAM50 HER2-Enriched were also omitted to obtain a reference dataset specific for the Luminal/Basal classification based on the PAM50 classification (N = 372).

Cases that were regarded as HER2 positive by IHC/FISH but not with the Neural Network algorithm or the PAM50 (N=9) were also removed as potential confounders.
Final dataset = 363

5

PAM50		Luminal	Basal	class precision
Prediction by Neural Network based on ESR1 and ERBB2 expression	pred. Luminal	281	3	98.9%
	pred. Basal	7	72	91.1%
	class recall	97.6%	96.0%	

Table 9: Crosstab of the molecular classification of breast cancer with PAM50 as reference. A Neural Net prediction based on the normalised expression of ESR1 and ERBB2 is shown as pred. Luminal and pred. Basal. Classification concordance of 97.25%

10

When including also PGR, AURKA, KIF2C and FOXC1 normalised expression as predicting variables, the algorithm provides a more accurate classification of Luminal and Basal patients (Table 10).

15

PAM50		Luminal	Basal	class precision
Prediction by Neural Network based on ESR1, ERBB2, PGR, AURKA, KIF2C and FOXC1 expression	pred. Luminal	286	3	98.96%
	pred. Basal	2	72	97.30%
	class recall	99.31%	96.00%	

Table 10: Crosstab of the molecular classification of breast cancer with PAM50 as reference. A Neural Net prediction based on the normalised expression of ESR1, ERBB2, PGR, AURKA, KIF2C and FOXC1 is shown as pred. Luminal and pred. Basal. Classification concordance of 98.62%

20

[E] Dissecting Luminal Breast Cancer into Luminal A and Luminal B

Breast cancer cases that were classified as Luminal by both PAM50 and the Neural Network Algorithm were selected to train the algorithm on but also assess the performance as the patient set subclasses are known. (N=286 of which 3 cases had no survival data).

PAM50 Luminal A vs Luminal B

Survival Data Processing Summary

PAM50	Total N	N of Events	Censored	
			N	Percent
Luminal A	191	25	166	86.9%
Luminal B	92	18	74	80.4%
Overall	283	43	240	84.8%

AURKA and KIF2C are required for clinically significant Luminal A and Luminal B categorisation

Algorithm 9: Gene signature: ERBB2, ESR1, FOXC1, PGR (Normalised to ACTB, HPRT1 and TBP)

PAM50		*true*	*true*	*class*
		Luminal A	*Luminal B*	*precision*
Prediction by Neural Network based on ESR1, ERBB2, PGR and FOXC1 expression	pred. Luminal A	167	42	**79.90%**
	pred. Luminal B	26	50	**65.79%**
	class recall	**86.53%**	**54.35%**	

Table 11: Crosstab of the Luminal classification of breast cancer with PAM50 as reference. A Neural Net prediction based on the normalised expression of ESR1, ERBB2, PGR and FOXC1 is shown as pred. Luminal A and pred. Luminal B. Classification concordance of 76.14%, N=285

Survival Data Processing Summary

Neural Net Prediction	Total N	N of Events	Censored	
			N	Percent
Luminal A	207	30	177	85.5%
Luminal B	76	13	63	82.9%
Overall	283	43	240	84.8%

Algorithm 10: Neural Net Prediction using gene signature: ERBB2, ESR1, FOXC1, PGR, AURKA, KIF2C (Normalised to ACTB, HPRT1 and TBP)

	PAM50	true Luminal A	true Luminal B	class precision
Prediction by Neural Network based on ESR1, ERBB2, PGR, AURKA, KIF2C and FOXC1 expression	pred. Luminal A	171	16	91.44%
	pred. Luminal B	22	76	77.55%
	class recall	88.60%	82.61%	

Table 12: Crosstab of the Luminal classification of breast cancer with PAM50 as reference. A Neural Net prediction based on the normalised expression of AURKA, KIF2C, ESR1, ERBB2, PGR and FOXC1 is shown as pred. Luminal A and pred. Luminal B. Classification concordance of 86.67%, N=285

Case Processing Summary

Neural Net Prediction	Total N	N of Events	Censored	
			N	Percent
Luminal A	175	24	151	86.3%
Luminal B	107	19	88	82.2%
Overall	282	43	239	84.8%

Algorithm 11: Gene signature: PGR, FOXC1, AURKA and KIF2C (Normalised to ACTB, HPRT1 and TBP)

PAM50		true Luminal A	true Luminal B	class precision
	pred. Luminal A	180	26	87.38%

| *Prediction by Neural Network based on PGR, AURKA, KIF2C and FOXC1 expression* | *pred. Luminal B* | 13 | 66 | **83.54%** |
| | *class recall* | **93.26%** | **71.74%** | |

Table 13: Crosstab of the Luminal classification of breast cancer with PAM50 as reference. A Neural Net prediction based on the normalised expression of AURKA, KIF2C, PGR and FOXC1 is shown as pred. Luminal A and pred. Luminal B. Classification concordance of 86.32%, N=285

Case Processing Summary

Neural Net Prediction	Total N	N of Events	Censored N	Percent
Luminal A	177	23	154	87.0%
Luminal B	105	20	85	81.0%
Overall	282	43	239	84.8%

Algorithm 12: Gene signature: PGR and FOXC1 (Normalised to ACTB, HPRT1 and TBP)

PAM50		*true Luminal A*	*true Luminal B*	*class precision*
Prediction by Neural Network based on PGR and FOXC1 expression	*pred. Luminal A*	171	49	**77.73%**
	pred. Luminal B	22	43	**66.15%**
	class recall	**88.60%**	**46.74%**	

Table 14: Crosstab of the Luminal classification of breast cancer with PAM50 as reference. A Neural Net prediction based on the normalised expression of PGR and FOXC1 is shown as pred. Luminal A and pred. Luminal B. Classification concordance of 75.09%, N=285

Case Processing Summary

Neural Net Prediction	Total N	N of Events	Censored N	Percent
Luminal A	218	30	188	86.2%
Luminal B	65	13	52	80.0%
Overall	283	43	240	84.8%

Example 3

APPENDIX A – Neural Network computation details

5

Neural Net details (computed using RapidMiner Studio Community software):
All dataset was used to train and as an unlabelled dataset.

500 training layers
10   0.3 learning rate
0.2 momentum
Shuffle
Normalise
Error Epsilon: 1.0E-5

15

Example 4 - Algorithms

20   Algorithm 1 – trained on normalised expression of ERBB2 in:
        a sample set comprising HER2+ samples as defined by IHC and FISH;
        and a sample set not comprising HER2+ samples as defined by IHC and FISH;

Algorithm 2 - trained on normalised expression of ESR1 in:
25   a sample set comprising ER+ samples as defined by IHC;
        and a sample set not comprising ER+ samples as defined by IHC;

Algorithm 3 - trained on normalised expression of ERBB2 and ESR1 in:
        a sample set comprising HER2+ and/or ER+ samples as defined by IHC and FISH;
30   and a sample set not comprising HER2+ and ER+ samples as defined by IHC and
        FISH;

Algorithm 4 – trained on normalised expression of ESR1, ERBB2, AURKA and/or KIF2C
in:

a sample set comprising HER2+ and/or ER+ samples as defined by IHC and FISH; and a sample set not comprising HER2+ and/or ER+ samples as defined by IHC and FISH;

5   Algorithm 5 - trained on normalised expression of ERBB2 samples in a dataset from The Cancer Genome Atlas (TGCA) comprising
         HER2+ samples as defined by IHC and FISH; and
         samples defined as not HER2+ by IHC and FISH;

10   Algorithm 6 - trained on normalised expression of ERBB2 and ESR1 in samples in a dataset from the TCGA, comprising
         HER2+ and/or ER+ samples as defined by IHC and FISH; and
         Samples defined as not HER2+ and not ER+ by IHC and FISH;

15   Algorithm 7 - trained on normalised expression of ERBB2 and ESR1 in samples in a dataset from the TCGA which had HER2-enriched samples as defined by PAM50 and HER2+ as defined by IHC/FISH removed.

     Algorithm 8 - trained on normalised expression of ERBB2, ESR1, PGR, AURKA, KIF2C
20   and FOXC1 in samples in a dataset from the TCGA which had HER2-enriched samples as defined by PAM50 and HER2+ samples as defined by IHC/FISH removed.

     Algorithm 9 – trained on normalised expression of ERBB2, ESR1, FOXC1, PGR normalised expression in:
25   samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and
         samples in a dataset that comprises breast cancer cases that were classified as not Luminal by both PAM50 and the Algorithm 8;

30   Algorithm 10 – trained on normalised expression of ERBB2, ESR1, FOXC1, PGR, AURKA and KIF2C in:
         samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and
         samples in a dataset that comprises breast cancer cases that were classified as
35   not Luminal by both PAM50 and the Algorithm 8;

Algorithm 11 – trained on normalised expression of AURKA, KIF2C, FOXC1 and PGR in:

samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and

samples in a dataset that comprises breast cancer cases that were classified as
5 not Luminal by both PAM50 and the Algorithm 8;

Algorithm 12 - Trained on normalised expression of FOXC1 and PGR in:

samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and

10   samples in a dataset that comprises breast cancer cases that were classified as not Luminal by both PAM50 and the Algorithm 8;

optionally wherein all expression levels were normalised to the expression levels of ACTB; PPIB; HPRT1; and/or TBP, preferably wherein all expression levels were normalised to
15 the expression levels of ACTB; PPIB; HPRT1; and TBP.

Algorithm 1 identifies samples as HER2 positive or HER2 Negative.

Algorithm 2 identifies samples as ER positive or ER Negative.

20 Algorithm 3 identifies samples as HER2 positive and/or ER positive or TNBC.

Algorithm 4 identifies samples as HER2 positive and/or ER positive or TNBC.

Algorithm 5 identifies samples as HER2 positive or HER2 Negative.

Algorithm 6 identifies samples as HER2 positive and/or ER positive or TNBC.

Algorithm 7 identifies samples as Luminal or Basal.

25 Algorithm 8 identifies samples as Luminal or Basal.

Algorithm 9 identifies samples as Luminal A or Luminal B.

Algorithm 10 identifies samples as Luminal A or Luminal B.

Algorithm 11 identifies samples as Luminal A or Luminal B.

Algorithm 12 identifies samples as Luminal A or Luminal B

30

Example 5 - Concordance with receptor status using FFPE material

A Neural Net algorithm (as described in Algorithm section Example 2) has been trained based on a subset of breast cancer patients that have been previously analysed and
35 classified into Luminal, Basal and HER2-Enriched by PAM50 (94.81% overall concordance). The reference classification was defined using a patented 12-gene signature that was validated on Quantigene 2.0 (WO2009055823).

Likewise, a Neural Net algorithm trained using 6-gene Quantigene data (ERBB2, ESR1, AURKA, KIF2C, FOXC1) against the 12-gene signature (WO2009055823) molecular classification as reference, can use predict Luminal, Basal and HER2-Enriched molecular

5   classification with 100% concordance. The 6-gene signature presented here, can be applied to the TCGA breast cancer data to predict PAM50 with 95.03% concordance when using the Neural Net algorithm.

Normalisation using various normalising genes, shows that a combination of 4 normalising genes (ACTB, PPIB, HPRT1 and TBP) will allow a 100% concordance of HER2

10 overexpression when compared to IHC and FISH results.

Example 6 - Quantigene expression analysis using degraded RNA

15 Sample used:

•   BT474 RNA – a HER2 positive, ER positive breast cancer cell line derived from a primary infiltrative ductal carcinoma of luminal type

•   MDAMB453 RNA – derived from a Luminal breast cancer cell line with amplified ERBB2 gene that is not overexpressed.

20 •   Breast cancer tumours in paraffin blocks (Formalin Fixed Paraffin Embedded – FFPE)

Methodology:

RNA degradation was performed by heating RNA at 95°C (as suggested by TATAA

25 Biocenter – AIT 2016) over the following time points:

1.   Untreated,
2.   1 minute,
3.   5 minutes,
4.   10 minutes
30 5.   30minutes
6.   60 minutes
7.   120 minutes

RNA degradation was significant at 5 minutes until complete degradation at 30 minutes

35 where it becomes comparable to degradation seen in tissue FFPE blocks, also in terms of RIN (Figure 12).

Sample	453 UnTx	453 120m	BT474 UnTX	BT474 1m	BT474 5m	BT474 10m	BT474 30m	BT474 60m	BT474 120m	6708/0 09	10535/ 031	11274/ 0323
Conc	86	134	112	82	115	168	216	117	112	444	578	239
RIN	8.9	2.5	7.2	7.1	3.2	2.1	2.1	2.4	2.5	2.2	2.2	2.3

Table 15

Stable RNA Expression profiles in heat degraded RNA – see Figure 14.

RAW data is defined by MFI data with subtracted the blank average and any signals below (blank average + 3SDs) are indicated as 0 (undetected)

On analysis of the 40-gene plex a significant reduction in signal is observed when RNA was heat denatured for over 60 minutes.

When assessing normalised data this effect seems to be overcome. The housekeeping gene RNA and the target RNA seem to bear the same effect from degradation, hence data derived from degraded RNA can yield the same information when normalised.

The Neural Network algorithm was trained based on patient FFPE data (N=82, 29/82 are HER2 positive) without outliers using the HER2 IHC/FISH annotation. This algorithm is applied on the BT474 degraded RNA series data to predict HER2 status based on ERBB2 gene expression (Figure 15). The Neural network can accurately determine the HER2 status across the RNA degradation series using ERBB2 or ERBB2/ESR1 (Table 16).

	BT474 RNA DEGRADATION SERIES	NEURAL NET ALGORITHM HER2 PREDICTION	CONFIDENCE
REPLICATE 1	Stable RNA	Positive	0.999
	5min @ 95°C	Positive	0.999
	10min @ 95°C	Positive	0.999
	30min @ 95°C	Positive	0.999
	60min @ 95°C	Positive	0.999
	120min @ 95°C	Positive	0.999
REPLICATE 2	Stable RNA	Positive	0.999
	5min @ 95°C	Positive	0.999
	10min @ 95°C	Positive	0.999
	30min @ 95°C	Positive	0.999
	60min @ 95°C	Positive	0.999

| | 120min @ 95°C | Positive | 0.999 |

Table 16: Neural Network algorithm HER2 status prediction results for the BT474 heat degraded RNA.

Example 7 - Subtyping Breast Cancer Cell lines in FFPE

To provide a proof of principle for the performance of Quantigene on degraded RNA samples, cell lines were subjected to the same tissue processing that a tumour tissue would undergo during the diagnostic process. The paraffinized cell lines were cored, lysed and analysed using Quantigene.

Cell blocks were produced to represent well annotated cell lines that are formalin fixed, and embedded in paraffin (FFPE-cell line blocks). Using a 12-plex Quantigene assay, normalising genes and classifier genes are measured in the FFPE-cell blocks to provide quality control material for the patient classification runs.

Samples used:

Patient-derived breast cancer samples from FFPE n = 82 (12 more cases did not pass QC)

Breast Cancer Cell Lines (BCCL) FFPE = 11 (2 HER2 positive + 9 HER2 Negative)

Total N = 93

<u>Neural Network Prediction of BCCL in FFPE</u>

The Neural Network algorithm trained on the patient set (N=82) can predict the HER2 status of the BCCL FFPEs with 100% accuracy. A larger dataset will be required to validate and provide a more robust algorithm for class prediction.

BREAST CANCER CELL LINES (BCCL)	HER2 STATUS	HER2 ALGORITHM PREDICTION
SKBR3 FFPE	Positive	Positive
BT474 FFPE	Positive	Positive
MCF7 FFPE	Negative	Negative
MCF10A FFPE	Negative	Negative

HB2 FFPE	Negative	Negative
436 FFPE	Negative	Negative
231 FFPE	Negative	Negative
468 FFPE	Negative	Negative
HCC1937 FFPE	Negative	Negative
HS578T FFPE	Negative	Negative
BT20 FFPE	Negative	Negative

Table 17: Neural Network Predicted HER2 status based on Breast Cancer cell line (BCCL) FFPE compared with the actual HER2 status confirmed by IHC.

	true Negative	*true HER2pos*	*class precision*
pred. Negative	62	0	100.00%
pred. HER2pos	0	31	100.00%
class recall	100.00%	100.00%	

Table 18: Neural Network Performance in determining the HER2 status on breast cancer FFPE (N=82) and cell line FFPE dataset (N = 11). The algorithm can detect the HER2 status of the available dataset with complete accuracy.

Example 8

Optimisation of RNA based assays using archival formalin-fixed paraffin-embedded (FFPE) material is challenging due to variability in surgical tissue processing and degradation of RNA caused by tissue integrity preservation using formalin [1,2]. To overcome the limitation of performing accurate gene expression studies from archival material, our group used the branched-chain DNA (bDNA) multiplex magnetic bead assay. bDNA technology replaces enzymatic amplification of target template with hybridisation of specific probes and amplification of a reporter signal [3]. The short recognition sequences of the capture and detection probes are designed to hybridise to short fragments of target RNA [4]. In addition, the use of tissue homogenates directly as starting material of this assay, overcomes the inevitable loss of RNA occurring in assays requiring prior RNA extraction and purification. Signal amplification, the use of short recognition sequences and the lack of a purification step, contribute to reduce technical variation of the assay. The technology provides the possibility to multiplex the assay (up to 80 RNA targets), measuring

expression of a panel of targets from low material input. This protocol describes the preparation and staining of tissue samples for laser microdissection. Staining on membrane slides facilitate imaging of tumour and histological architecture providing accurate selection and profiling of (1) tumour and normal ducts in breast tissue and (2)

5   malignant cell clones within heterogeneous tumours.

Molecular classification of Breast cancer interrogates molecular markers to categorise patients into molecular classes, namely luminal classes, the human epidermal growth factor receptor 2 (HER2)-enriched and the basal types. The HER2-enriched subtype is

10   well defined, with high expression of HER2 receptor, due to the ERBB2 gene amplification, combined with low or absent ER and PgR. Luminal subtype is generally positive for oestrogen receptor (ER) and the Basal type tumours are in general negative for the 3 receptors significantly overlapping with the triple negative breast cancer (TNBC) diagnostic subtype [5,6]. Other markers are used to determine epithelial and mesenchymal

15   characteristics. Fibronectin (FN1) is a main component of breast tissue mesenchymal compartment. Increased FN1 expression is accompanied by high Ki67 staining showing a signature for a more invasive tumour [7,8] and an increased expression is associated with metastasis [9]. Interestingly, FN1 was found to be present in microvesicles originating from tumour cells, that induced activation of mitogenic signals in recipient fibroblasts [10]. Hence,

20   circulating microvesicles such as exosomes are potentially vehicles of early detection and early indicators of metastasis and relapse [11].

Tumour area selection for breast cancer transcriptional subtyping has recurrently been done by macrodissection [12,13]. To overcome tissue heterogeneity and increase sensitivity,

25   we have reliably combined classical tissue staining with multiplex molecular profiling methods. As a proof of principle, two distinct breast cancer clones have been defined by their epithelial mesenchymal signature and metastatic potential. The workflow of the described protocol can be easily translated to the current clinical setup and used to selectively isolate and characterise tissue subtypes using targeted mRNA profiling.

30

The described method has been applied for the simultaneous measurement of 40 transcripts in H&E stained (**Figure 16**), microdissected (**Figure 17**) highly degraded FFPE material. Using this method, we show accurate characterisation of receptor status (**Figure 18A**), classification of tumours into luminal and basal molecular subtypes [17] and differential

35   expression of the mesenchymal marker, FN1 when comparing tumour and matched control tissue (**Figure 18B**), in the various receptor positive and negative subtypes.

A bead-based multiplex branched DNA (bDNA) assay has been optimised to quantify gene expression on degraded RNA derived from formalin fixed paraffin embedded (FFPE) breast cancer tissue and normal breast ducts. Optimising the assay, involved developing an algorithm to classify breast cancer patients in luminal and basal subtypes utilising a 8 well-known biomarkers and 5 potential normalising genes. Data normalisation was done using permutations of the normalising genes. The selection of the normalising genes was based on the best prediction of receptor status using the Luminal/Basal classifier genes. To classify Luminal/Basal subtypes from FFPE tissues, the normalising genes selected were Beta-actin (ACTB), Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) and Hypoxanthine Phosphoribosyltransferase 1 (HPRT1).

The method can be adapted for use in other diagnostic and research areas following adequate selection of the normalising gene set. One important application of this method in the research sector is the measurement of biomarkers in archival material that is well annotated with clinical outcomes. This provides the possibility to validate potential predictive markers in retrospective studies, quickly and accurately, avoiding long-term prospective studies awaiting disease-free survival and overall survival data. Figures 25 and 26 show the correlation of RNA expression level of genes from colorectal cell lines with the expression level in exosomes. The data shows that AURKA can be used as a normalising gene to characterise transcript expression of diagnostic genes such as HER2 within cancer-derived exosomes. HER2 expression is in concordance with HER2 amplification. Similarly, this can be applied to other diagnostic gene amplifications and/or overexpression. The use of liquid biopsies and robust gene expression assays will allow high throughput multiplex assays adapted for patient management during treatment, providing the means to follow treatment efficacy, potential relapse due to resistance to therapy and the metastatic capacity of the tumour.

This method has also a wide range of possible applications in the diagnosis of tumours and is adapted to the current diagnostic workflow. The main advantages of the method in the diagnostic field include (1) implementation of high throughput assays, (2) excluding subjectivity and equivocal results originating from image-based measurements; (3) accurate detection of multiple targets simultaneously, enhancing accuracy and minimising use of precious patient samples and (4) there is no need for highly specialised facilities and human resources. The optimised sampling process, together with the low input of material required for the bead-based multiplex assay, allows further investigation of tumour heterogeneity by the use of laser microdissection to accurately separate multiple foci of malignant tissue from the same patient section and compare multiple gene expression

between them and also with matched normal tissue (**Figure 19**). Low material input is vital for diagnostic application on tumour biopsies that provide limited tumour tissue. The capacity of the assay to measure gene expression from degraded RNA samples allows easy transportation of samples for analysis within an institution or to servicing laboratories.

5   In addition, whole section analysis was also possible using H&E stained material (**Figure 16**).

For the success of this protocol, it is imperative to (1) ensure proper sampling of the tumour site/s that are lysed for the assay and (2) develop well optimised and validated data

10   normalisation algorithms, for each gene expression panel and/or individual prognostic or predictive biomarkers. The former depends on the technical experience of the technician/scientist performing the sampling. It is recommended to take an additional core and prepare a tissue microarray (TMA) in the same format of the multiplex magnetic bead assay runs (96 well format). This will provide an archive of tumour sites as a replica of

15   samples used for the RNA-based assay. TMAs can also be assessed with other techniques for follow-up research or validation of results. The development of normalisation algorithms is dependent on the material being investigated and the normalising genes selected for normalisation. Different panels of normalising genes are selected based on the level and variability of expression in the sample analysed and this

20   varies between cancer tissues from different origins, exosomes from plasma or circulating tumour cells. Validation of the assay includes sample processing since various preparations will also result in different normalisation algorithms.

To summarise, the use of bDNA technology in combination with magnetic bead technology

25   and the selection of the proper panel of target genes, will provide the added advantage of measuring gene expression directly in tissue lysates derived from small amounts of patient material, including microdissected material, exosomes and circulating tumour cells. In addition to detection of tumour heterogeneity, the proper use of panels has the potential to detect tumour derived exosomes for early diagnostics and early detection of relapses.

30   Since there is no need of any nucleic acid amplification step, signal amplification using the bDNA technology, combined with the bead-based multiplex measures multiple gene expression in clinically-annotated archival material providing a resource for biomarker validation.

35   **REFERENCES:**

1   Abrahamsen, H. N., Steiniche, T., Nexo, E., Hamilton-Dutoit, S. J. & Sorensen, B. S. Towards quantitative mRNA analysis in paraffin-embedded tissues using real-time

reverse transcriptase-polymerase chain reaction: a methodological study on lymph nodes from melanoma patients. *J Mol Diagn.* **5** (1), 34-41, doi:S1525-1578(10)60449-7 [pii] 10.1016/S1525-1578(10)60449-7, (2003).

2   Macabeo-Ong, M. *et al.* Effect of duration of fixation on quantitative reverse transcription polymerase chain reaction analyses. *Mod Pathol.* **15** (9), 979-987 (2002).

3   Yang, W. *et al.* Direct quantification of gene expression in homogenates of formalin-fixed, paraffin-embedded tissues. *Biotechniques.* **40** (4), 481-486 (2006).

4   Flagella, M. *et al.* A multiplex branched DNA assay for parallel quantitative gene expression profiling. *Anal Biochem.* **352** (1), 50-60, doi:S0003-2697(06)00117-5 [pii] 10.1016/j.ab.2006.02.013, (2006).

5   Goldhirsch, A. *et al.* Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol.* **22** (8), 1736-1747, doi:10.1093/annonc/mdr304 mdr304 [pii], (2011).

6   Schnitt, S. J. Classification and prognosis of invasive breast cancer: from morphology   to   molecular   taxonomy.   *Mod   Pathol.*   **23   Suppl   2**   S60-64, doi:10.1038/modpathol.2010.33
modpathol201033 [pii], (2010).

7   Bae, Y. K. *et al.* Fibronectin expression in carcinoma cells correlates with tumor aggressiveness and poor clinical outcome in patients with invasive breast cancer. *Hum Pathol.* **44** (10), 2028-2037, doi:10.1016/j.humpath.2013.03.006
S0046-8177(13)00118-4 [pii], (2013).

8   Park, J. & Schwarzbauer, J. E. Mammary epithelial cell interactions with fibronectin stimulate   epithelial-mesenchymal   transition.   *Oncogene.*   **33**   (13),   1649-1657, doi:10.1038/onc.2013.118
onc2013118 [pii], (2014).

9   Zhou, Z. *et al.* MRI detection of breast cancer micrometastases with a fibronectin-targeting contrast agent. *Nat Commun.* **6** 7984, doi:10.1038/ncomms8984
ncomms8984 [pii], (2015).

10   Antonyak, M. A. *et al.* Cancer cell-derived microvesicles induce transformation by transferring tissue transglutaminase and fibronectin to recipient cells. *Proc Natl Acad Sci U S A.* **108** (12), 4852-4857, doi:10.1073/pnas.1017667108
1017667108 [pii], (2011).

11   Moon, P. G. *et al.* Fibronectin on circulating extracellular vesicles as a liquid biopsy to detect breast cancer. *Oncotarget.* **7** (26), 40189-40199, doi:10.18632/oncotarget.9561 9561 [pii], (2016).

12   Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol.* **27**, doi:10.1200/jco.2008.18.1370, (2009).

13   Nielsen, T. *et al.* Analytical validation of the PAM50-based Prosigna Breast Cancer Prognostic Gene Signature Assay and nCounter Analysis System using formalin-fixed paraffin-embedded breast tumor specimens. *BMC Cancer.* **14** 177-177, doi:10.1186/1471-2407-14-177, (2014).

14   Fischer, A. H., Jacobson, K. A., Rose, J. & Zeller, R. Hematoxylin and Eosin Staining of Tissue and Cell Sections. *Cold Spring Harbor Protocols.* **2008** (5), pdb.prot4986, doi:10.1101/pdb.prot4986, (2008).

15   Molecular Machines and Industries. H&E Staining Kit Plus for Laser Microdissection and more, Prod. No 70302.

16   Lin, F. & Prichard, J. *Handbook of Practical Immunohistochemistry: Frequently asked questions.* (Springer, 2011).

17   Grech, G. *et al.* Molecular Classification of Breast Cancer Patients Using Formalin-fixed Paraffin-embedded Derived RNA Samples. *Journal of Molecular Biomarkers & Diagnosis.* **7** (S8), doi:doi:10.4172/2155-9929.S8-016, (2016).

18   MACKAY A *et al* 2011. Microarray-Based Class Discovery for Molecular Classification of Breast Cancer: Analysis of Interobserver Agreement. JNCI: Journal of the National Cancer Institute, 103, 662-673.

19   PEROU C *et al* 2010. Clinical implementation of the intrinsic subtypes of breast cancer. The Lancet Oncology, 11, 718-719.

Example 9

Identification of markers of sensitivity to PP2A activators

Data analysis of well annotated, publicly available databases (TCGA) showed that PP2A deregulation is a common event in breast cancer, with 59.6% of basal breast cancer patients showing a low expression of one of the PP2A complex components or a high expression of the inhibitory subunits.

The main aim of the study was to identify genes whose expression depends directly on PP2A activity. In the ideal scenario, these novel biomarkers would reflect PP2A activity regardless of the mechanism driving the activity regulation.

5   We first established cell line models to investigate possible markers of sensitivity to PP2A activators. Basal breast cancer cell line MDAMB231 was sensitive to the PP2A activator, 20 FTY720, and showed a suppressed PP2A activity while the MDAMB453 Luminal cell line was not sensitive to FTY720 and had relatively high PP2A activity. Drug sensitivity is the inverse of percent viability.

10

In more detail, the dose-dependent effect of FTY720 on breast cancer cell lines provided information on the sensitivity of the cell line to PP2A activation. Cytotoxicity 25 results for the 12 breast cancer cell lines tested are illustrated in Figure 21 where the % cell viability is expressed as a percentage of the parallel control culture (without drug) against FTY720

15   concentration. The average vehicle control (VC) % cell viability for each cell line was never under 85% viability compared to the untreated ruling out any interference by the vehicle. The 10µM and the 25µM doses were considered to be cytotoxic as cellular 30 morphology was generally altered, and hence are not shown.

20   The three sensitive cell lines, defined by % cell viability under 50% (IC50) at a dose lower than 5µM, were represented by a Triple Negative phenotype (Figure 21A). BT-20 was the most sensitive at 0.5µM, followed by MDAMB231 and Hs578T showing and IC50 at 2.5µM FTY720 and 5µM FTY720, respectively. From this experiment, the effective dose of FTY720, that is not lethal, was set at 5µM for assessing the effect of 5 FTY720 on RNA

25   expression. MCF10A is a cell-line derived from normal epithelium. This cell line was used as a baseline control in the study of the PP2A mechanism in breast cancer cell lines.

PP2A biomarker discovery and evaluation

30   The knowledge gained from the cellular models implied the need to characterise novel biomarkers for PP2A activity. The PP2A deregulation criteria was designed based on this information to probe datasets for gene candidates.

Commonly deregulated regulatory subunits of PP2A were identified. Preliminary 15
35   analysis of the PP2A regulators was done on TCGA data to assess possible pathways for deregulation within breast cancer.

Initial analysis of transcript expression and protein expression of PP2A regulators in breast cancer cellular models did not identify clear markers that consistently reflect FTY720 sensitivity. Moreover, the assessment of PP2A regulators and phosphotargets of 20 PP2A, in local patients and a set of tissue arrays (TMAs), highlighted the complexity and
5   heterogeneity of the feedback mechanism. Biomarkers that directly measure the activity of the complex or the specific function of the variety of complexes were required to predict sensitivity to drugs at a tissue level.

A number of candidate PP2A biomarker genes were identified. As outlined in 25 various
10   studies, a biomarker needs to be well defined in its objectives, use and target population (Altar et al., 2007; Pepe, Feng, Janes, Bossuyt, & Potter, 2008). The approach for biomarker discovery involved the use of breast cancer data in the TCGA data portals in the context of PP2A deregulation. Genes were shortlisted using network analysis and known biological evidence connecting biomarkers to the PP2A pathway.
15

Following discovery, the viability of using these biomarkers in the same cohort based on RNASeqV2 expression was assessed. Validation analysis was performed using the TCGA RNASeqV2 data analysed manually providing a better insight into the data. The in silico analysis was then compared to observations within the cellular models and more
20   importantly a cohort of local patients (N = 95 tumours, 3 Relapse tumours, 7 in situ, non-invasive tumours, 36 Normal breast tissue, of which 33 were patient-matched with 5 tumours).

The set of biomarkers was validated and evidence-based selection of biomarker
25   candidates, identified AURKA and KIF2C as strong predictors of sensitivity to FTY720. These genes were used to classify patients and cell lines into those that are eligible to PP2A activation and those that are not predicted to have PP2A deregulation. The novel 10 therapeutic class was mainly represented by the Basal or Triple Negative groups and was associated with higher tumour aggressivity.
30

In more detail, the expression of the PP2A complex and regulator genes was measured in the 12 breast cancer cell lines when untreated, but also when treated with FTY720 (5μM).
15
Figure 22 shows that Cip2a (KIAA1524) RNA expression correlates with sensitivity of
35   breast cancer cell lines to FTY720 (Figure 22). MDA-MB-231 exemplifies FTY720-sensitive breast cancer cell lines, showing a low PP2A enzymatic activity and a high Cip2a RNA expression relative to MCF10A (a cell line derived from normal epithelium).

AURKA and KIF2C RNA expression were found to be positively correlated with 20 FTY720-sensitive breast cancer cell lines (Figure 22B). Of interest expression also correlated positively with PP2A endogenous inhibitors, cip2a and SET (Table 19).

		AURKA	KIF2C
PPP2CA	Spearman Correlation	-0.056	-0.181
	Sig. (2-tailed)	0.107	< 0.001*
PPP2R2A	Spearman Correlation	-0.262	-0.209
	Sig. (2-tailed)	< 0.001*	0.005*
CIP2A	Spearman Correlation	0.750	0.773
	Sig. (2-tailed)	< 0.001*	< 0.001*

Table 19: Spearman correlation analysis between the expression of the PP2A complex subunits and the 2 biomarkers.

The normalised RNASeqV2 data as downloaded from the TCGA database was used. [N=873 tumour cases; Sig. (2-tailed) stands for two-tailed significance, significance at the 95% confidence interval is highlighted]

In addition, the expression of the genes within breast cancer tissue was also compared to the respective patient matched control tissue (Figure 23B). AURKA and KIF2C RNA expression was found to be significantly higher in tumour material relative to 5 matched normal tissues. This is also seen when patients are subdivided by the hormone status (Figure 23C).

AURKA and KIF2C are thus preferred as they are exclusive for aggressive tumours, providing evidence that they are directly involved in the malignant phenotype (Table 18).

PP2A deregula-tion biomarkers	Basal (N=99) %	HER2-en-riched (N=58) %	Luminal B (N=133) %	Luminal A (N=235) %
AURKA: EXP>2	47	36	30	3
KIF2C: EXP>2	44	7	5	0
Total cases with deregulation	56	36	30	3

Table 18: Occurrence of overexpression of the candidate biomarkers over the PAM50 breast cancer subtypes expressed as a percentage of the subtype population (N = 525). PAM50 annotation retrieved from data generated by TCGA Research Network: http://cancergenome.nih.gov/ and data generated from the cBioPortal accessible through http://www.cBioPortal.org/ (Cerami, et al., 2012; Gao, et al., 2013).

KIF2C and AURKA expression also correlated with percentage cell viability (Table 19) during FTY720 treatment in FTY720-sensitive cell lines. There was no correlation, when non sensitive cell lines were used. AURKA and KIF2C also correlates with CIP2A 5 across treatment.. This supports the use of KIF2C and AURKA as biomarkers to identify novel therapeutic groups within breast cancer patients eligible for PP2A activation therapy.

		CIP2A	KIF2C	AURKA	Viability (24hours)
CIP2A	Spearman Cor-relation	n/a	0.677**	0.808**	0.671**
	Sig. (2-tailed)		0.04	0.0002	0.004
KIF2C	Spearman Cor-relation	0.677**	n/a	0.887**	0.879**
	Sig. (2-tailed)	0.004		0.000005	0.00001
AURKA	Spearman Cor-relation	0.808**	0.887**	n/a	0.799**
	Sig. (2-tailed)	0.0002	0.000005		0.0002

Table 19: Spearman correlation analysis between the expression of AURKA, KIF2C and CIP2A following treatment at 0.1, 1 and 5µM FTY720 in the FTY720-sensitive cell lines. 10 Correlation of expression with cellular viability is included.

The normalised Quantigene expression levels were used.

[N=16 conditions (4 sensitive cell lines at 4 treatment conditions); Sig. (2-tailed) stands for two-tailed significance, significance at the 95% confidence interval is highlighted]

PP2A activity biomarkers are correlated with protein expression of inhibitors and pS6K

AURKA expression was also found to correlate strongly with cytoplasmic cip2a staining (Table 20). AURKA significantly correlates with nuclear pS6K, establishing an association between AURKA expression and growth factor (PI3K/mTOR) signaling attenuation as part of the negative feedback driven by PP2A.

Protein Biomarker		pAKT		pS6K	
		Cytoplasmic	Nuclear	Cytoplasmic	Nuclear
AURKA	Spearman's Rho	0.124	-0.058	0.032	0.272
	P	0.344	0.661	0.806	0.035*
KIF2C	Spearman's Rho	-0.138	-0.238	0.079	0.198
	P	0.293	0.067+	0.548	0.129
Protein Biomarker		Cytoplasmic CIP2A	Nuclear SET	Cytoplasmic SET	Cytoplasmic α4
AURKA	Spearman's Rho	0.378	0.197	0.249	0.295
	P	0.003*	0.132	0.055+	0.025
KIF2C	Spearman's Rho	0.227	0.005	-0.033	0.129
	P	0.083+	0.967	0.800	0.336

Table 20: Spearman correlations between PP2A activity biomarker expression and the protein expression of inhibitors of PP2A and downstream phospho-proteins with defined localisation in the breast cancer cell lines.

Expression of PP2A activity biomarkers is normalised to the housekeeping genes, while protein expression was scored using the H-score (0-300) for each cellular compartment and then expressed as positive or negative based on thresholds set on significance with survival and/or correlation with histopathological factors [* significant at 95% confidence interval; + significant at 90% confidence interval]

Hence, in this study we identified a novel subgroup of basal-type patients, using KIF2C and AURKA as biomarkers of sensitivity to the PP2A activator FTY720. The sensitivity of this class to FTY720 is supported following proper and expected classification of the sensitive cell lines within this subgroup.

References for Example 9

KÖNINKI, K., BAROK, M., TANNER, M., STAFF, S., PITKÄNEN, J., HEMMILÄ, P.,
ILVESARO, J. & ISOLA, J. 2010. Multiple molecular mechanisms underlying
trastuzumab and lapatinib resistance in JIMT-1 breast cancer cells. *Cancer Letters,* 294,
211-219.

KAO, J., SALARI, K., BOCANEGRA, M., CHOI, Y.-L., GIRARD, L., GANDHI, J., KWEI,
K. A., HERNANDEZ-BOUSSARD, T., WANG, P., GAZDAR, A. F., MINNA, J. D. &
POLLACK, J. R. 2009. Molecular Profiling of Breast Cancer Cell Lines Defines Relevant
Tumor Models and Provides a Resource for Cancer Gene Discovery. *PLoS ONE,* 4,
e6146.

TANNER, M., KAPANEN, A. I., JUNTTILA, T., RAHEEM, O., GRENMAN, S., ELO, J.,
ELENIUS, K. & ISOLA, J. 2004. Characterization of a novel cell line established from a
patient with Herceptin-resistant breast cancer. *Molecular Cancer Therapeutics,* 3, 1585-
1592.

Example 10

The expression of genes in RNA isolated from colorectal cancer cell lines was compared
to the expression in exosomes derived from matched originator cells. **Figure 1** shows that
using the colorectal cancer cell line, COLO320 the expression of these genes correlated
($r^2=0.9198$), hence exosomes represent the expression in originator cells.

Cell lines were incubated for 48 hours at 0% FBS to produce exosomes. Measuring the
proliferation marker AURKA indicates that originator cells are proliferating and provides an
indication of the exosome quantity within the sample. Therefore, AURKA can be used as
a normalising gene to characterise transcript expression of diagnostic genes such as
HER2 within cancer-derived exosomes. HER2 expression is in concordance with HER2
amplification. Similarly, this can be applied to other diagnostic gene amplifications and/or
overexpression.

Claims

5   1.   A method for classifying a cancer into one or more sub-classes wherein the method involves the use of bDNA technology.

2.   The method of claim 1 wherein the method classifies the cancer in respect of at least two sub-classes, optionally at least three sub-classes, optionally at least four-
10   subclasses, optionally at least five or more sub-classes.

3.   The method of claim 1 wherein one or more of the sub-classes is associated with an increase or decrease in copy number of a gene.

15   4.   The method of any of claims 1-3 wherein the method comprises the determination of the level of RNA expression of at least one RNA species, optionally more than 2 RNA species, optionally more than 3 RNA species, optionally more than 4 RNA species, optionally more than 5 RNA species, optionally more than 6 RNA species, optionally more than 7 RNA species, optionally more than 8 RNA species, optionally more than 9 RNA
20   species, optionally more than 10 RNA species, optionally more than 11 RNA species, optionally more than 12 RNA species, optionally more than 14 RNA species, optionally more than 16 RNA species, optionally more than 18 RNA species, optionally more than 20 RNA species, optionally more than 30 RNA species, optionally more than 40 RNA species, optionally more than 50 RNA species.

25
5.   The method of any of claims 1-4 wherein the method comprises the determination of the level of RNA expression:

of at least one or more genes associated with the cancer or at least one sub-class of
30   cancer, optionally at one gene associated with the cancer or at least one sub-class of the cancer, optionally at least two genes associated with the cancer or at least one sub-class of the cancer, optionally at least three genes associated with the cancer or at least one sub-class of the cancer, optionally at least four genes associated with the cancer or at least one sub-class of the cancer, optionally at least five genes associated with the cancer
35   or at least one sub-class of the cancer, optionally at least six genes associated with the cancer or at least one sub-class of the cancer, optionally more than six genes associated with the cancer or at least one sub-class of the cancer,

optionally wherein the at least one or more genes associated with the cancer or at least one sub-class of cancer are selected from the group consisting of ERBB2, ESR1, AURKA, KIF2C, PGR and FOXC1; and/or

5 of at least one gene not associated with the cancer or a sub-class of the cancer, of at least two genes not associated with the cancer or a sub-class of the cancer, of at least three genes not associated with the cancer or a sub-class of the cancer, of at least four genes not associated with the cancer or a sub-class of the cancer, optionally wherein the one or more genes has been determined to be suitable for use as a normalising gene, optionally 10 determined by an algorithm trained on a dataset of known cancer classification optionally wherein the at least one gene not associated with the cancer or sub-cancer is selected from the group consisting of ACTB; PPIB; HPRT1; and TBP.

6. The method of any of claims 1-5 wherein the cancer is breast cancer, optionally 15 wherein the sub-classes of cancer comprise at least one or more of HER2+, ER+, Basal, Triple Negative Breast Cancer (TNBC) and Luminal, optionally wherein sub-class Luminal comprises the sub-classes Luminal A and Luminal B.

7. The method of any of claims 1-6 wherein the method comprises the determination 20 of the level of RNA expression of any one or more of ERBB2, ESR1, AURKA, KIF2C and PGR, FOXC1.

8. The method of any of claims 1-7 wherein the method comprises the determination of the level of RNA expression of any one or more of ACTB, HPRT1, TBP and PPIB. 25

9. The method of any of claims 1-8 wherein the method comprises the determination of the level of RNA expression of ERBB2, ACTB, PPIB, HPRT1 and TBP.

10. The method of any of claims 1-8 wherein the method comprises the determination 30 of the level of RNA expression of ESR1, ACTB, PPIB, HPRT1 and TBP.

11. The method of any of claims 1-8 wherein the method comprises the determination of the level of RNA expression of ESR1, ERBB2, ACTB, PPIB, HPRT1 and TBP.

35 12. The method of any of claims 1-8 wherein the method comprises the determination of the level of RNA expression of ERBB2; AURKA and/or KIF2C; ACTB; PPIB; HPRT1; and TBP.

84

13.   The method of any of claims 1-8 wherein the method comprises the determination of the level of RNA expression of ESR1; AURKA and/or KIF2C; ACTB; PPIB; HPRT1; and TBP.

14.   The method of any of claims 1-8 wherein the method comprises the determination of the level of RNA expression of ERBB2; ESR1; AURKA and/or KIF2C; ACTB; PPIB; HPRT1; and TBP.

15.   The method of any of claims 1-8 wherein the method comprises the determination of the level of RNA expression of ERBB2; ESR1; AURKA; KIF2C; PGR; FOXC1; ACTB; PPIB; HPRT1; and TBP.

16.   The method of any of claims 1-8 wherein the method comprises the determination of the level of RNA expression of ERBB2; ESR1; PGR; FOXC1; ACTB; PPIB; HPRT1; and TBP.

17. The method of any of claims 1-8 wherein the method comprises the determination of the level of RNA expression of AURKA; KIF2C; PGR; FOXC1; ACTB; PPIB; HPRT1; and TBP.

18. The method of any of claims 1-8 wherein the method comprises the determination of the level of RNA expression of PGR; FOXC1; ACTB; PPIB; HPRT1; and TBP.

19.   The method of any of claims 1-18 wherein the method further comprises the use of machine learning to classify the cancer into the sub-classes,
optionally wherein the machine learning is an algorithm, optionally is an algorithm selected from the group consisting of the Neural Network algorithm, Decision Tree, Random Forest and Support Vector Machine.

20.   The method of claim 19 wherein the algorithm is any one of the following:

Algorithm 1 – trained on normalised expression of ERBB2 in:
        a sample set comprising HER2+ samples as defined by IHC and FISH;
        and a sample set not comprising HER2+ samples as defined by IHC and FISH;

Algorithm 2 - trained on normalised expression of ESR1 in:

a sample set comprising ER+ samples as defined by IHC;

and a sample set not comprising ER+ samples as defined by IHC;

Algorithm 3 - trained on normalised expression of ERBB2 and ESR1 in:

a sample set comprising HER2+ and/or ER+ samples as defined by IHC and FISH;

and a sample set not comprising HER2+ and ER+ samples as defined by IHC and FISH;

Algorithm 4 – trained on normalised expression of ESR1, ERBB2, AURKA and/or KIF2C in:

a sample set comprising HER2+ and/or ER+ samples as defined by IHC and FISH;

and a sample set not comprising HER2+ and/or ER+ samples as defined by IHC and FISH;

Algorithm 5 - trained on normalised expression of ERBB2 samples in a dataset from The Cancer Genome Atlas (TGCA) comprising

HER2+ samples as defined by IHC and FISH; and

samples defined as not HER2+ by IHC and FISH;

Algorithm 6 - trained on normalised expression of ERBB2 and ESR1 in samples in a dataset from the TCGA, comprising

HER2+ and/or ER+ samples as defined by IHC and FISH; and

Samples defined as not HER2+ and not ER+ by IHC and FISH;

Algorithm 7 - trained on normalised expression of ERBB2 and ESR1 in samples in a dataset from the TCGA which had HER2-enriched samples as defined by PAM50 and HER2+ as defined by IHC/FISH removed.

Algorithm 8 - trained on normalised expression of ERBB2, ESR1, PGR, AURKA, KIF2C and FOXC1 in samples in a dataset from the TCGA which had HER2-enriched samples as defined by PAM50 and HER2+ samples as defined by IHC/FISH removed.

Algorithm 9 – trained on normalised expression of ERBB2, ESR1, FOXC1, PGR normalised expression in:

samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and

86

samples in a dataset that comprises breast cancer cases that were classified as not Luminal by both PAM50 and the Algorithm 8;

Algorithm 10 – trained on normalised expression of ERBB2, ESR1, FOXC1, PGR, AURKA
5   and KIF2C in:
      samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and
      samples in a dataset that comprises breast cancer cases that were classified as not Luminal by both PAM50 and the Algorithm 8;

10

Algorithm 11 – trained on normalised expression of AURKA, KIF2C, FOXC1 and PGR in:
      samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and
      samples in a dataset that comprises breast cancer cases that were classified as
15   not Luminal by both PAM50 and the Algorithm 8;

Algorithm 12 - Trained on normalised expression of FOXC1 and PGR in:
      samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and
20   samples in a dataset that comprises breast cancer cases that were classified as not Luminal by both PAM50 and the Algorithm 8;

optionally wherein all expression levels were normalised to the expression levels of ACTB; PPIB; HPRT1; and/or TBP, preferably wherein all expression levels were normalised to
25   the expression levels of ACTB; PPIB; HPRT1; and TBP.
optionally wherein the algorithm is the Neural Network algorithm.

21.   The method of any of the preceding claims wherein the method classifies the cancer as a HER2+ cancer or a HER2 negative cancer, optionally wherein the method
30   comprises the determination of the level of expression of the ERBB2 gene, optionally further comprises normalisation of the level of expression of the ERBB2 gene to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of any one or more of Algorithm 1, Algorithm 3, Algorithm 4, Algorithm 5, and Algorithm 6 as defined in claim 20.

35

22.   The method of any of the preceding claims wherein the method classifies the cancer as an ER+ cancer or an ER negative cancer, optionally wherein the method

comprises the determination of the level of expression of the ESR1 gene, optionally further comprises normalisation of the level of expression of the ESR1 gene to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of any one or more of Algorithm 2, Algorithm 3, Algorithm 4, Algorithm 5 and Algorithm 6 as defined in
5   claim 20.

23.   The method of any of the preceding claims wherein the method classifies the cancer as a TNBC cancer or not a TNBC negative cancer, optionally wherein the method comprises the determination of the level of expression of the ERBB2 and ESR1 genes,
10   optionally further comprises normalisation of the level of expression of the ERBB2 and ESR1 genes to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of any one or more of Algorithm 1 plus Algorithm 2; Algorithm 3; Algorithm 4; Algorithm 5; and Algorithm 6 as defined in claim 20.

15   24.   The method of any of the preceding claims wherein the method classifies the cancer as a Basal cancer or a Luminal cancer, optionally wherein the method comprises the determination of the level of expression of the ERBB2 and ESR1 genes, optionally further comprises normalisation of the level of expression of the ERBB2 and ESR1 genes to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the
20   use of Algorithm 7 as defined in claim 20.

25.   The method of any of the preceding claims wherein the method classifies the cancer as a Basal cancer or a Luminal cancer, optionally wherein the method comprises the determination of the level of expression of the ERBB2, ESR1, PGR, AURKA, KIF2C
25   and FOXC1 genes, optionally further comprises normalisation of the level of expression of the ERBB2, ESR1, PGR, AURKA, KIF2C and FOXC1 genes to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of Algorithm 8 as defined in claim 20.

30   26.   The method of any of the preceding claims wherein the method classifies the cancer as a Luminal A or a Luminal B cancer, optionally wherein the method comprises the determination of the level of expression of the ERBB2, ESR1, PGR and FOXC1 genes, optionally further comprises normalisation of the level of expression of the ERBB2, ESR1, PGR and FOXC1 genes, to the expression levels of ACTB; PPIB; HPRT1; and TBP,
35   optionally further comprise the use of Algorithm 9 as defined in claim 20.

27. The method of any of the preceding claims wherein the method classifies the cancer as a Luminal A or a Luminal B cancer, optionally wherein the method comprises the determination of the level of expression of the ERBB2, ESR1, PGR, AURKA, KIF2C and FOXC1 genes, optionally further comprises normalisation of the level of expression of the ERBB2, ESR1, PGR, AURKA, KIF2C and FOXC1 genes, to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of Algorithm 10 as defined in claim 20.

28. The method of any of the preceding claims wherein the method classifies the cancer as a Luminal A or a Luminal B cancer, optionally wherein the method comprises the determination of the level of expression of the PGR, AURKA, KIF2C and FOXC1 genes, optionally further comprises normalisation of the level of expression of the PGR, AURKA, KIF2C and FOXC1 genes, to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of Algorithm 11 as defined in claim 20.

29. The method of any of the preceding claims wherein the method classifies the cancer as a Luminal A or a Luminal B cancer, optionally wherein the method comprises the determination of the level of expression of the PGR and FOXC1 genes, optionally further comprises normalisation of the level of expression of the PGR and FOXC1 genes, to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of Algorithm 12 as defined in claim 20.

30. The method of any of the preceding claims wherein the Algorithm is the Neural Network Algorithm computed using RapidMiner Studio Community software and trained with the following parameters:

500 training layers
0.3 learning rate
0.2 momentum
Shuffle
Normalise
Error Epsilon: 1.0E-5.

31. The method of any of the preceding claims wherein the method classifies the cancer as either HER2+, ER+, Basal, Luminal A, Luminal B or TNBC, optionally wherein the method comprises the determination of the level of expression of the ERBB2, ESR1, PGR, AURKA, KIF2C and FOXC1 genes, optionally further comprises normalisation of the

level of expression of the ERBB2, ESR1, PGR, AURKA, KIF2C and FOXC1 genes, to the expression levels of ACTB; PPIB; HPRT1; and TBP, optionally further comprise the use of any one or more of the Algorithms as defined in claim 20.

32.   The method of any one of the preceding claims wherein the detection of the level of RNA expression is by the use of a multiplex platform, optionally a microbead platform, optionally the Luminex platform.

33.   The method of any of the preceding claims wherein the detection of the level of RNA expression is by the use of the QuantiGene Plex assay (ThermoFisher).

34.   The method of any of the preceding claims wherein the method is performed on a sample obtained from a subject.

35.   The method of claim 34 wherein the sample is an archived/historical sample, optionally wherein the sample is between 1 month-100 years old, optionally wherein the sample is at least 1 month old, optionally wherein the sample is at least 2 months old, or at least 3 months old, or at least 4 months old, or at least 5 months old, or at least 6 months old, or at least 7 months old, or at least 8 months old, or at least 9 months old, or at least 10 months old, or at least 11 months old, or at least 12 months old for example 1 year old, or at least 1.5 years old, or at least 2 years old, or at least 3 years old, or at least 4 years old, or at least 5 years old, or at least 6 years old, or at least 7 years old, or at least 8 years old, or at least 9 years old, or at least 10 years old, or at least 11 years old, or at least 12 years old, or at least 13 years old, or at least 14 years old, or at least 15 years old, or at least 16 years old, or at least 17 years old, or at least 18 years old, or at least 19 years old, or at least 20 years old, or at least 25 years old, or at least 30 years old, or at least 40 years old, or at least 50 years old, or at least 60 years old, or at least 70 years old, or at least 80 years old, or at least 90 years old or at least 100 years old.

36. The method of claim 34 wherein the sample is a fresh sample, optionally wherein the sample is less than 1 month old, optionally less than 4 weeks old, optionally less than 21 days old, optionally less than 14 days old, optionally less than 7 days old, optionally less than 6 days old, optionally less than 5 days old, optionally less than 4 days old, optionally less than 3 days old, optionally less than 2 days old optionally less than 1 day old, optionally less than 18 hours old, optionally less than 12 hours old, optionally less than 6 hours old, optionally less than 5 hours old, optionally less than 4 hours old, optionally less than 3 hours old, optionally less than 2 hours old, optionally less than 1 hour old, optionally

less than 30 minutes old, optionally less than 15 minutes old, optionally less than 10 minutes old, optionally less than 5 minutes old.

37.   The method of any of claims 34-36 wherein the sample comprise degraded RNA, optionally wherein the sample has a RIN of 3.2 or less, optionally less than 3.0, optionally less than 2.8, optionally less than 2.6, optionally less than 2.4, optionally less than 2.2, optionally less than 2.0, optionally less than 1.8, optionally less than 1.6, optionally less than 1.4, optionally less than 1.2, optionally less than 1.0, optionally less than 0.8, optionally less than 0.6, optionally less than 0.4, optionally less than 0.2.

38.   The method of any of claims 34-37 wherein the sample is a tissue sample obtained from a subject; a cell line; a liquid biopsy, optionally a blood sample a plasma sample, optionally wherein the liquid biopsy comprises circulating tumour cells.

39. The method of any of claims 34-38 wherein the sample has been fresh frozen.

40.   The method of any of claims 34-39 wherein the sample has been formalin-fixed (FF).

41.   The method of any of claims 34-40 wherein the sample has been paraffin-embedded (PE).

42.   The method of any of claims 34-41 wherein the sample is a formalin-fixed paraffin-embedded sample (FFPE).

43.   The method of any of claims 34-42 wherein the sample is a homogenate or lysate, optionally a homogenate or lysate of a heterogeneous tumour.

44.   The method of any of claims 34-43 wherein the sample is a laser dissected sample.

45.   The method of any of claims 34-44 wherein the sample is a Haematoxylin and Eosin stained sample.

46.   The method of any of claims 34-45 wherein the sample is a sample of a morphologically distinct tumour within a larger tumour.

47.   A method for the detection of heterogeneity in a tumour, optionally a breast cancer tumour, wherein the method comprises a method according to any of claims 1-46.

48.   The method of claim 47 wherein the method comprises:

a)

        i) the identification of morphologically distinct tumours identified in a sample obtained from a patient;

        ii) dissection, optionally laser dissection or macro dissection of the morphologically distinct tumours into two or more samples; and

        iii) performing a method according to any of the preceding claims separately on each sample; and/or

b)   performing a method according to any of the preceding claims separately on more than one sample of the tumour, optionally on more than one sample of the tumour wherein each sample has been taken from a different site in the tumour.

49.   The use of bDNA technology to predict the presence of a gene amplification or gene reduction.

50.   A kit for use in the method of any one or more of claims 1-49, said kit comprising bDNA probes directed towards at least two of:

ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1, ACTB, PPIB, HPRT1, and TBP.

51.   The kit for use according to claim 50 wherein the kit comprises bDNA probes directed towards:
any two or more of ACTB, PPIB, HPRT1, and TBP;
any three or more of ACTB, PPIB, HPRT1, and TBP; or
all four of ACTB, PPIB, HPRT1, and TBP.

52.   The kit for use according to claim 50 or 51 wherein the kit comprises bDNA probes directed towards:
Any two or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1;
Any three or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1;
Any four or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1;
Any five or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1; or
All six of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1.

53.   The kit for use according to any of claims 50-52 wherein the kit comprises means for staining the sample to identify morphologically distinct tumours within the sample, optionally comprises Haematoxylin & Eosin.

54.   A method of validating a potential gene copy number variation as a biomarker, the method comprising the use of any one or more of the methods according to claims 1-49.

55.   The method of any of claims 1-49 for use in a method of diagnosis or prognosis

56.   The method of any of the preceding claims for research use.

57.   A method of selecting a suitable treatment strategy wherein the method comprises the method of any of the preceding claims.

58.   An Algorithm trained on any one or more of:

i) normalised expression of ERBB2 in:
        a sample set comprising HER2+ samples as defined by IHC and FISH;
        and a sample set not comprising HER2+ samples as defined by IHC and FISH;

ii) normalised expression of ESR1 in:
        a sample set comprising ER+ samples as defined by IHC;
        and a sample set not comprising ER+ samples as defined by IHC;

iii) normalised expression of ERBB2 and ESR1 in:
        a sample set comprising HER2+ and/or ER+ samples as defined by IHC and FISH;
        and a sample set not comprising HER2+ and ER+ samples as defined by IHC and
        FISH;

iv) normalised expression of ESR1, ERBB2, AURKA and/or KIF2C in:
        a sample set comprising HER2+ and/or ER+ samples as defined by IHC and FISH;
        and a sample set not comprising HER2+ and/or ER+ samples as defined by IHC
        and FISH;

v) normalised expression of ERBB2 samples in a dataset from The Cancer Genome Atlas (TGCA) comprising:

HER2+ samples as defined by IHC and FISH; and

samples defined as not HER2+ by IHC and FISH;

vi) normalised expression of ERBB2 and ESR1 in samples in a dataset from the TCGA, comprising:

HER2+ and/or ER+ samples as defined by IHC and FISH; and

Samples defined as not HER2+ and not ER+ by IHC and FISH;

vii) normalised expression of ERBB2 and ESR1 in samples in a dataset from the TCGA which had HER2-enriched samples as defined by PAM50 and HER2+ as defined by IHC/FISH removed;

viii) normalised expression of ERBB2, ESR1, PGR, AURKA, KIF2C and FOXC1 in samples in a dataset from the TCGA which had HER2-enriched samples as defined by PAM50 and HER2+ samples as defined by IHC/FISH removed;

ix) normalised expression of ERBB2, ESR1, FOXC1, PGR in:

samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and

samples in a dataset that comprises breast cancer cases that were classified as not Luminal by both PAM50 and the Algorithm 8;

x) normalised expression of ERBB2, ESR1, FOXC1, PGR, AURKA and KIF2C in:

samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and

samples in a dataset that comprises breast cancer cases that were classified as not Luminal by both PAM50 and the Algorithm 8;

xi) normalised expression of AURKA, KIF2C, FOXC1 and PGR in:

samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and

samples in a dataset that comprises breast cancer cases that were classified as not Luminal by both PAM50 and the Algorithm 8;

xii) normalised expression of FOXC1 and PGR in:

samples in a dataset that comprises breast cancer cases that were classified as Luminal by both PAM50 and the Algorithm 8; and

samples in a dataset that comprises breast cancer cases that were classified as not Luminal by both PAM50 and the Algorithm 8;

optionally wherein all expression levels were normalised to the expression levels of ACTB;
5   PPIB; HPRT1; and/or TBP, preferably wherein all expression levels were normalised to the expression levels of ACTB; PPIB; HPRT1; and TBP,

optionally wherein the algorithm is the Neural Network algorithm.

10   59.   A HER2 specific therapy, optionally Herceptin (trastuzumab), Kadcyla (Herceptin and emtansine), Nerlynx (neratinib), Perieta (pertuzumab), and/or Tykerb (lapatinib), for use in the treatment of a subject with breast cancer wherein a sample from the subject has been identified as HER2+ by use of any of the preceding methods.

15   60.   An ER+ specific therapy, optionally Tamoxifen, Aromatase Inhibitors, and/or SERMs, for use in the treatment of a subject with breast cancer wherein a sample from the subject has been identified as ER+ by use of any of the preceding methods.

61.   A panel of biomarkers comprising
20   any one or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1;
any two or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1;
Any three or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1;
Any four or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1;
Any five or more of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1; or
25   All six of: ERBB2, ESR1, PGR, AURKA, KIF2C, FOXC1;

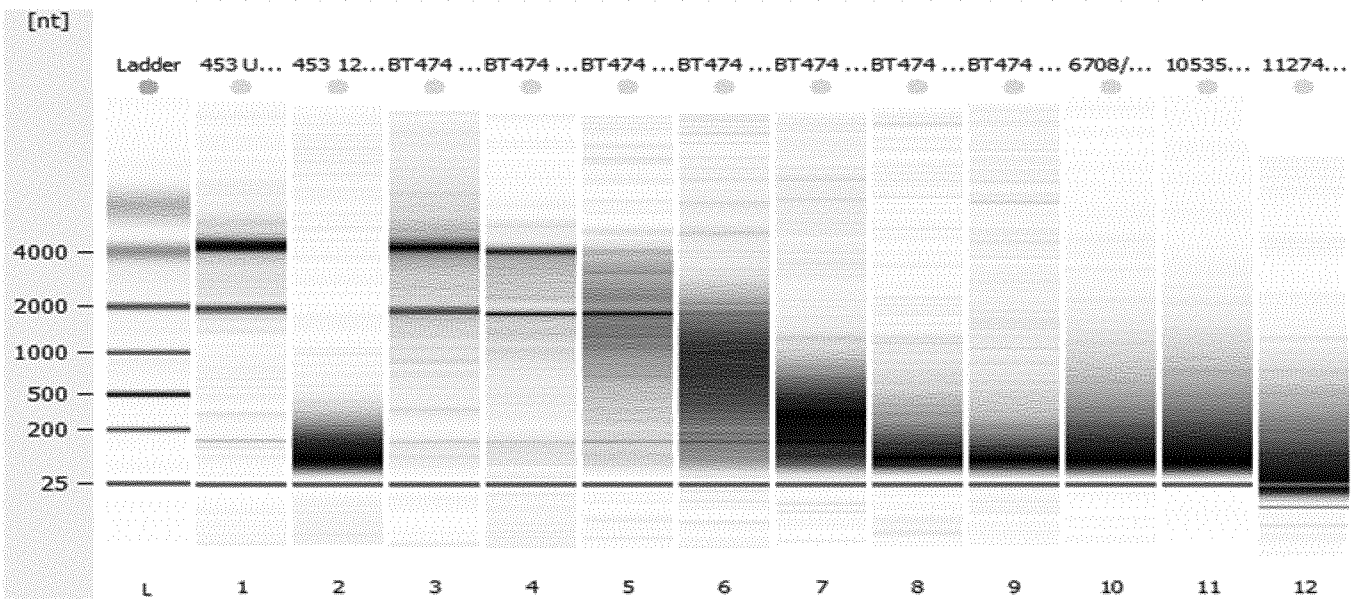optionally wherein the panel comprises AURKA and/or KIF2C.

30

35

Figure 1
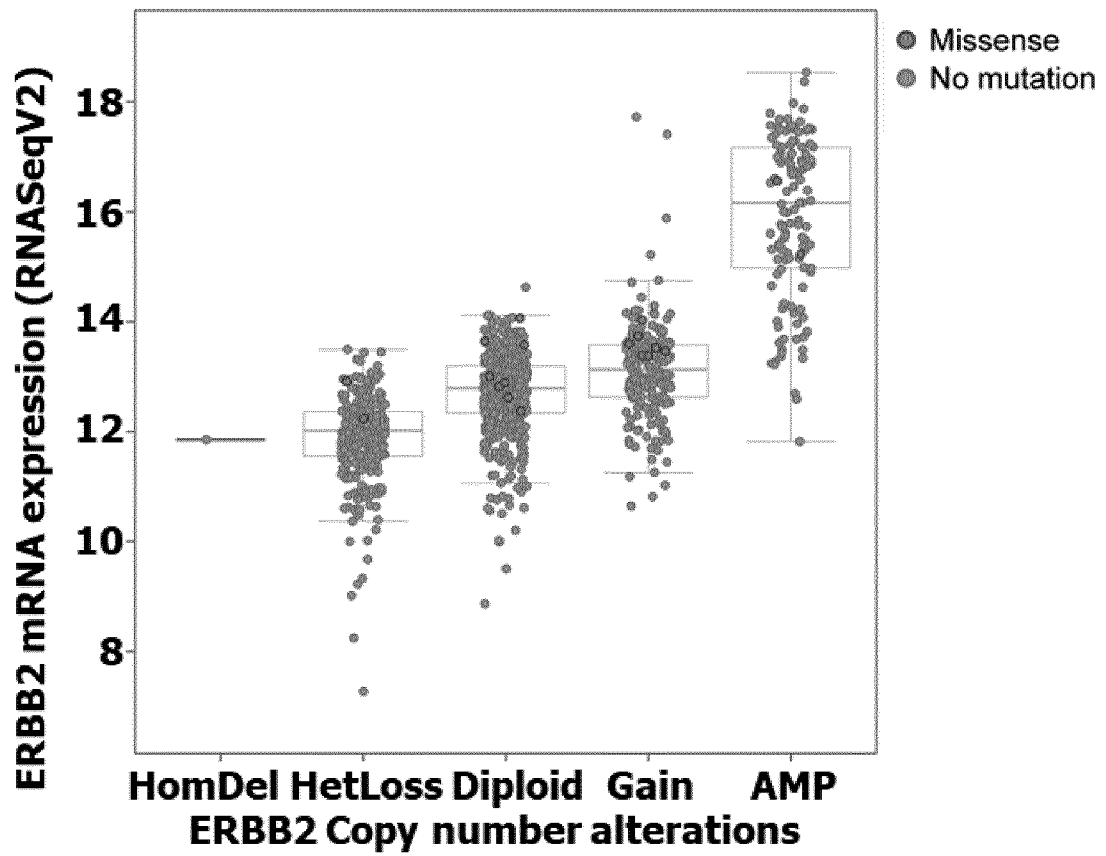
Figure 2

A: LX200 RAW
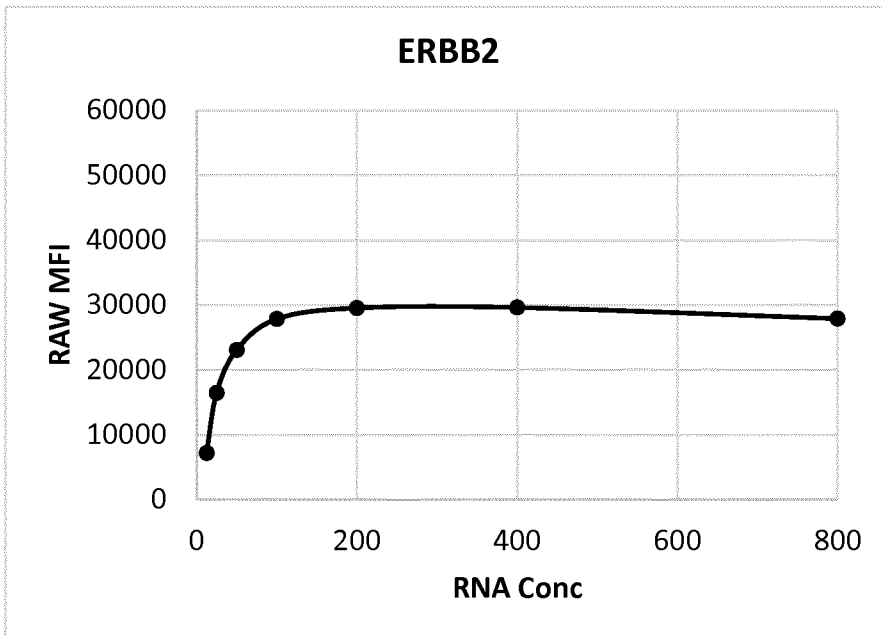
B: MAGPIX RAW
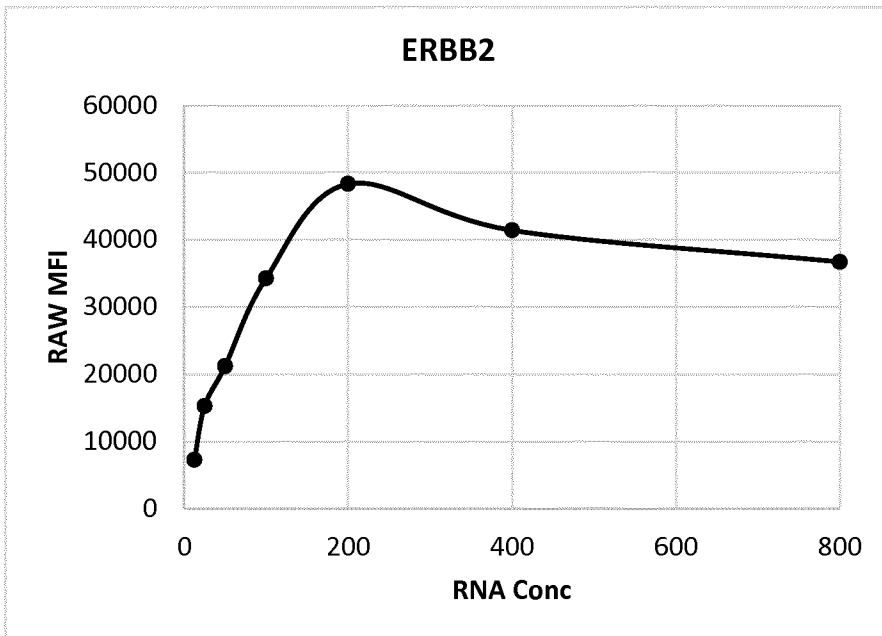
Figure 3

A

B

Figure 4

Figure 5

Mean = 1269.8
Std. Dev. = 1848.503
N = 98

TBP Expression (Raw)

Mean = 2.8203
Std. Dev. = .5032
N = 98

Log(10) TBP Expression

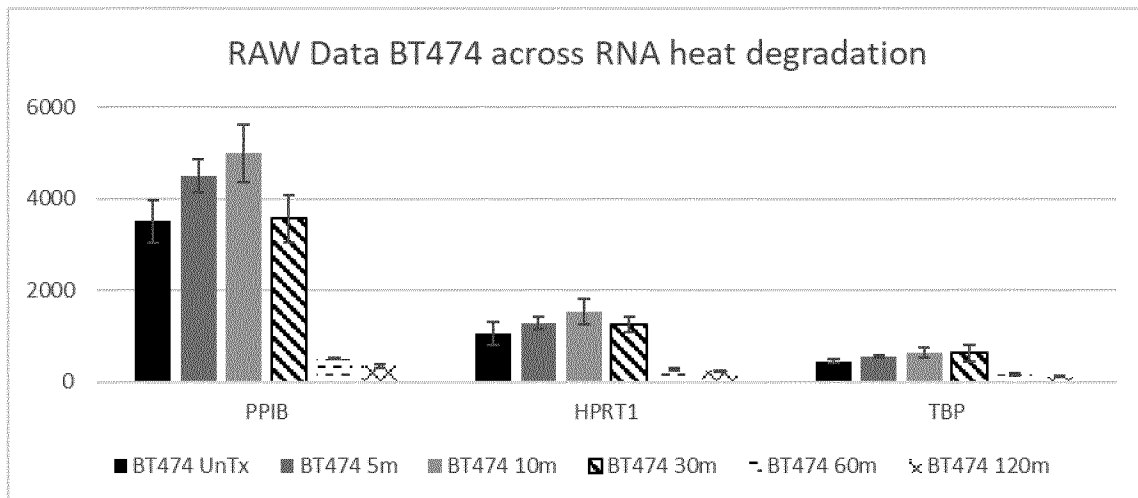Mean = 4.08E-11
Std. Dev. = 1.0000
N = 98

Log(10) TBP Expression z-score

5/28
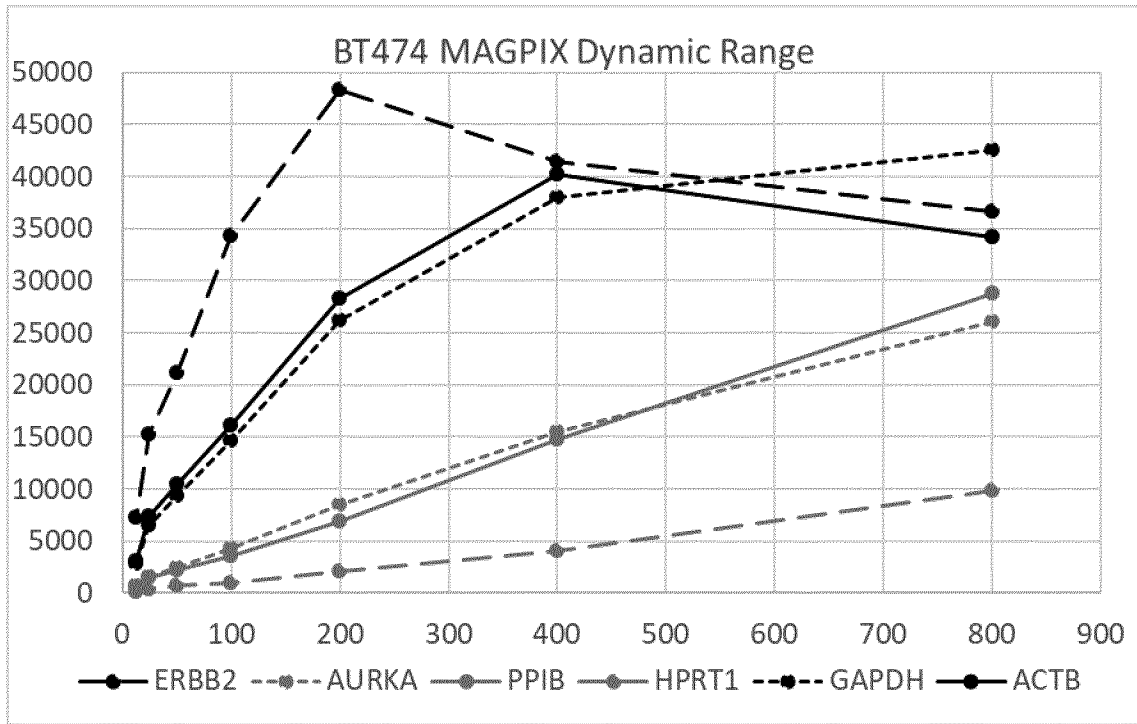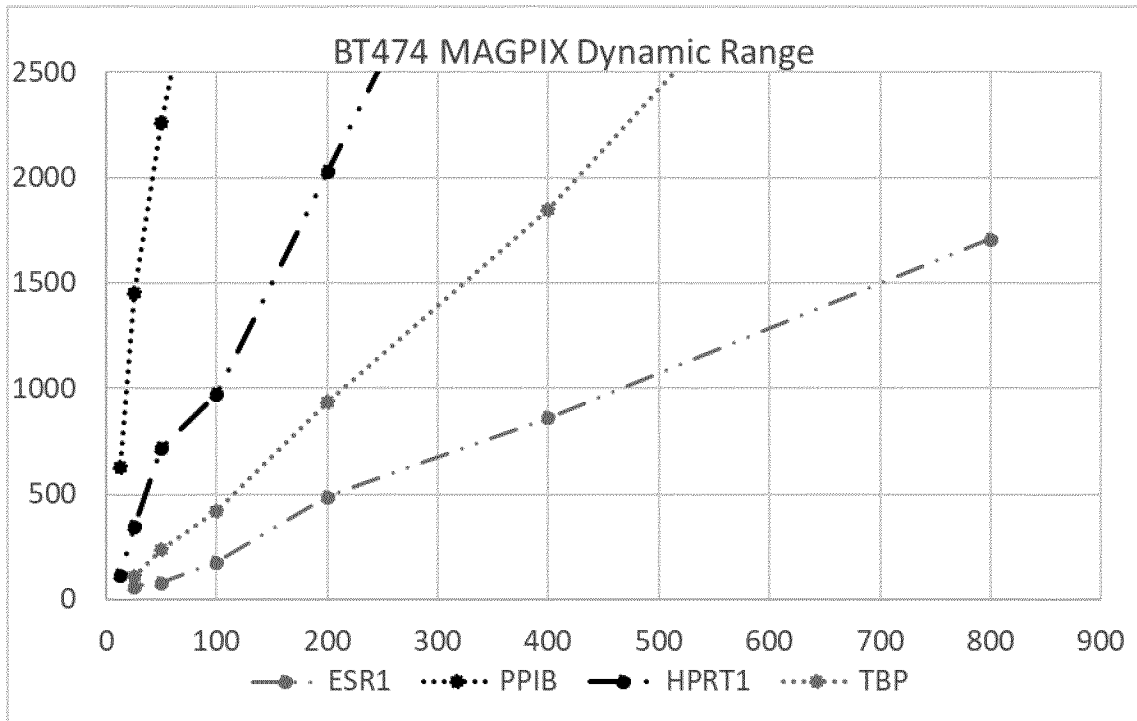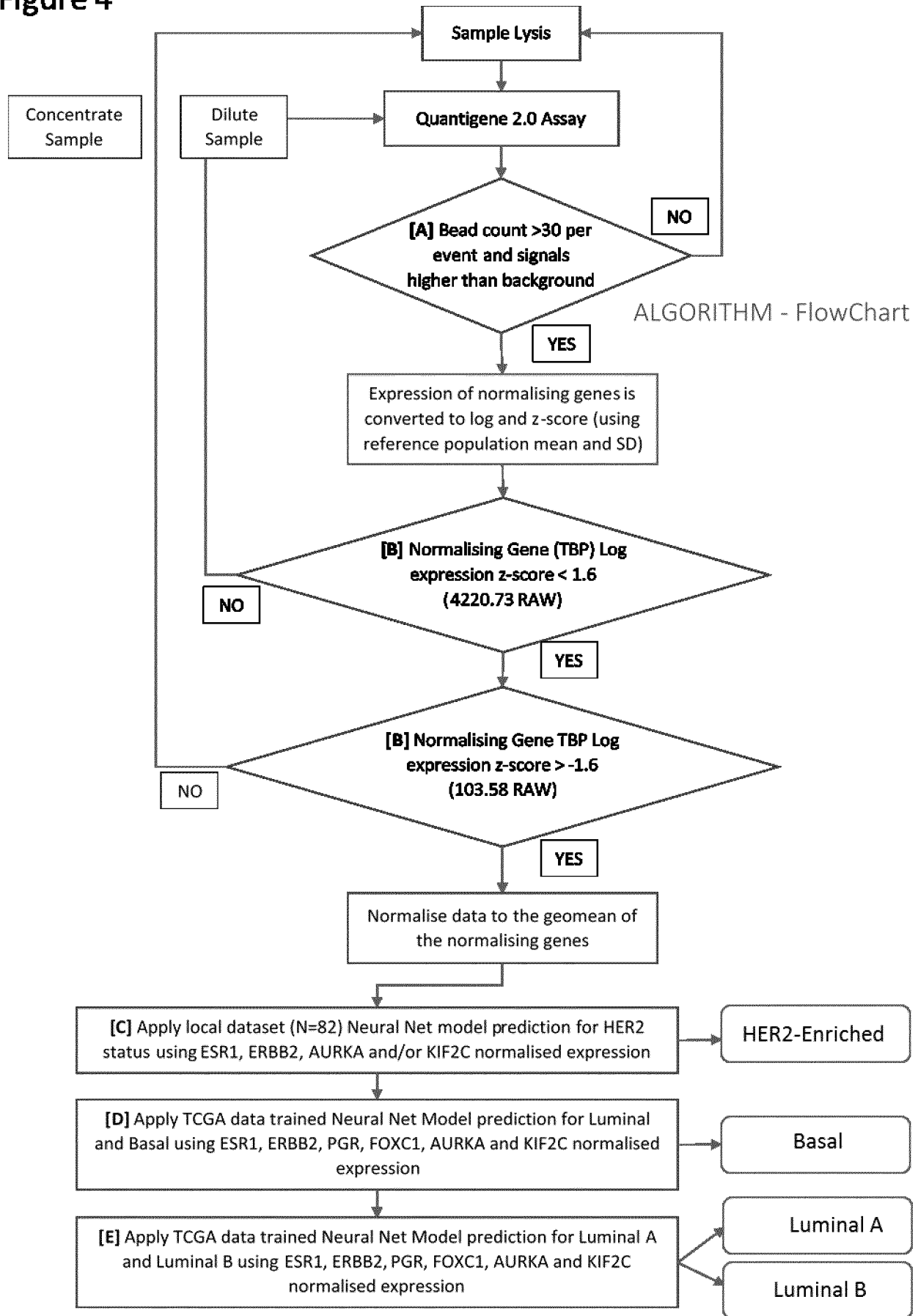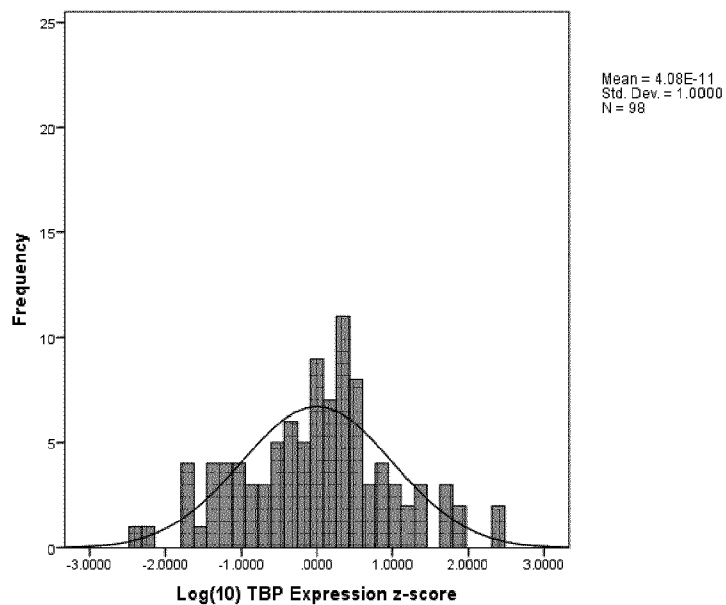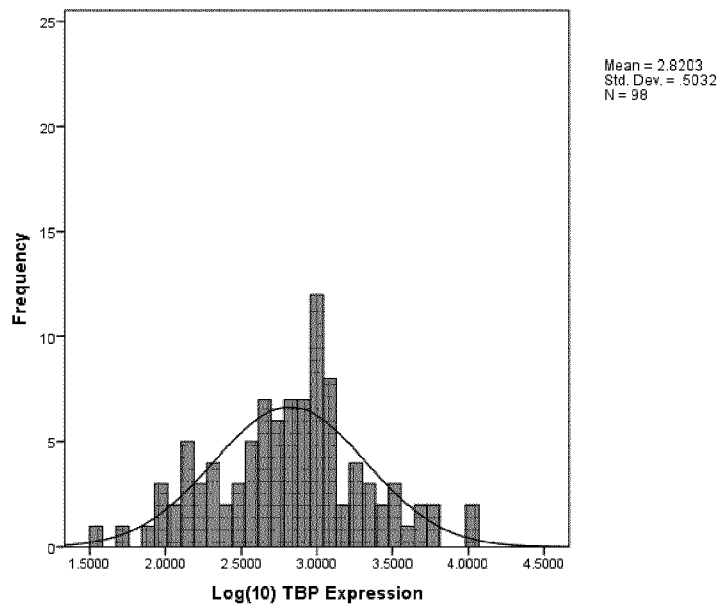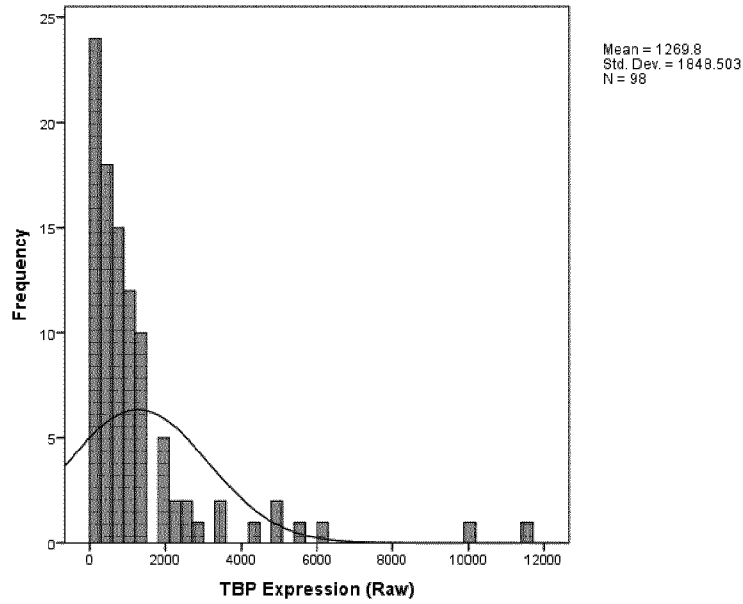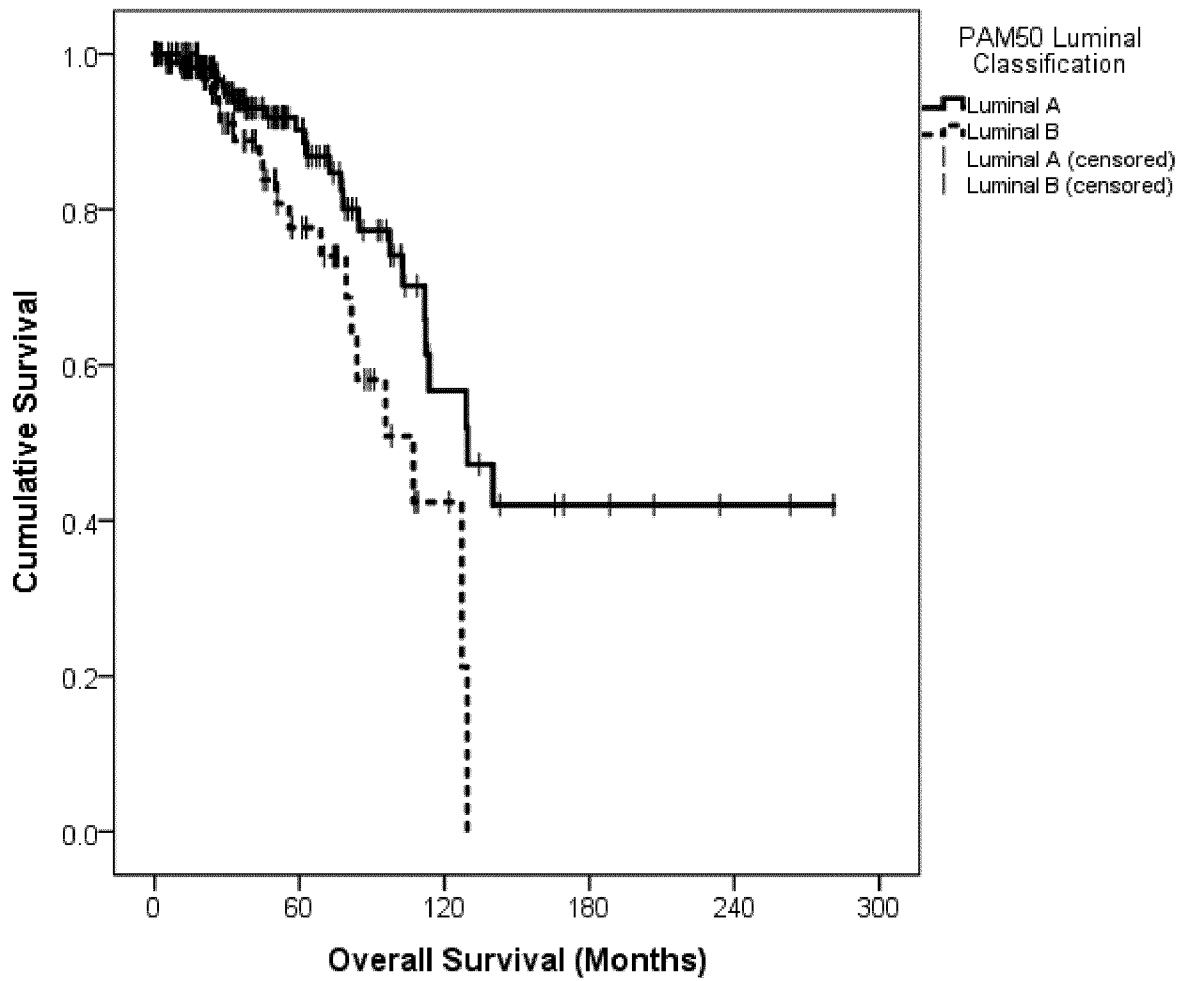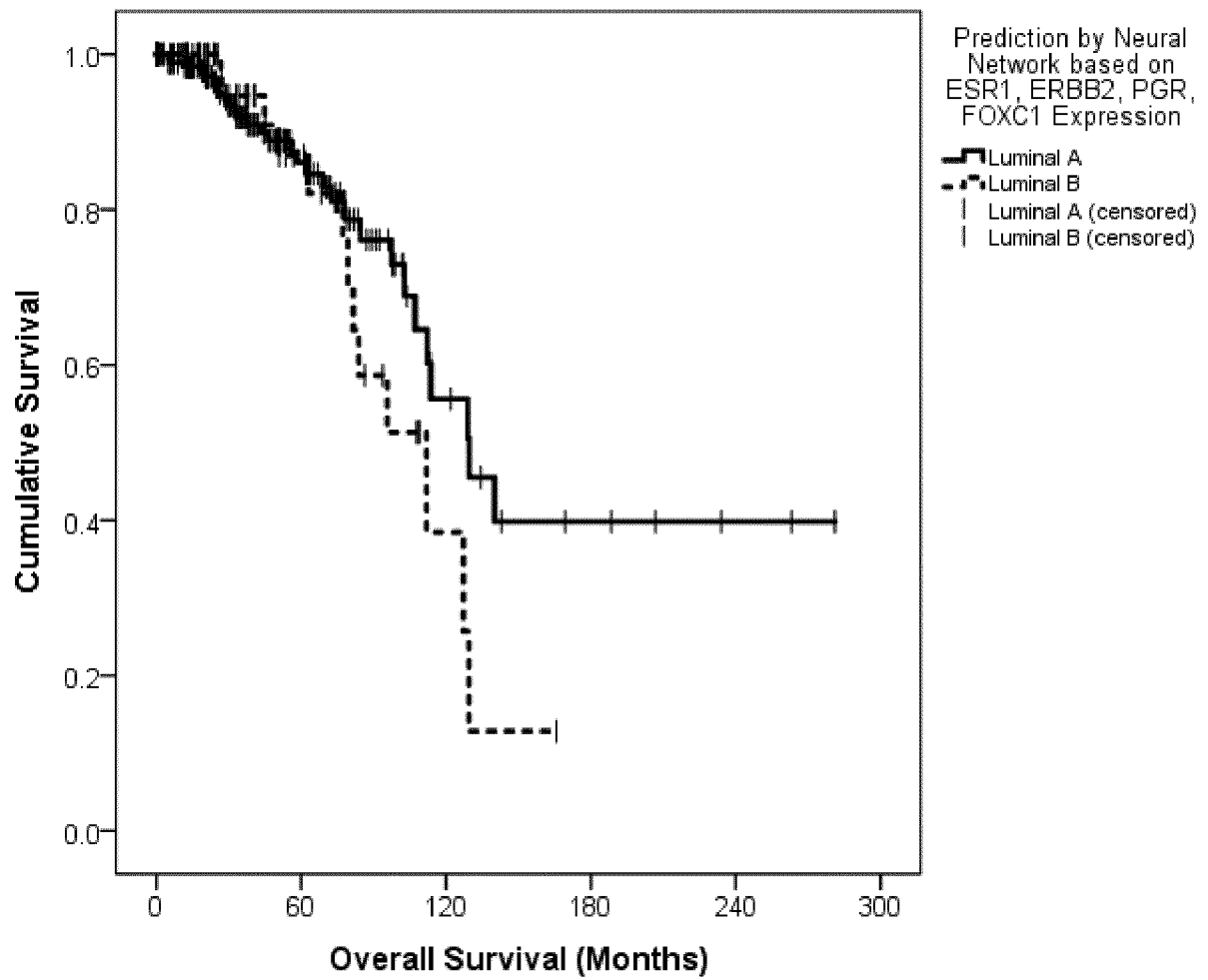
Figure 6

Figure 7

Figure 8
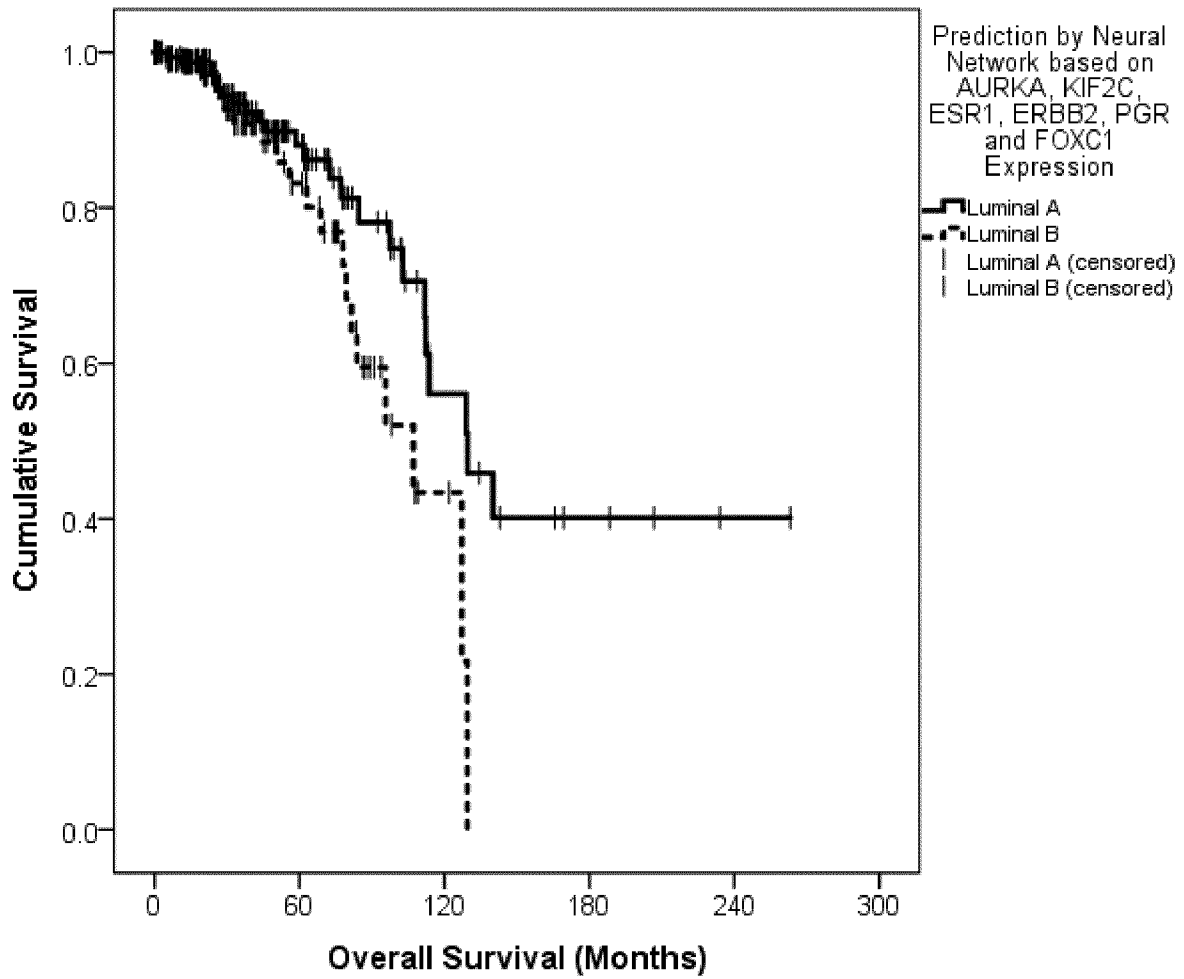
Figure 9

Figure 10

Figure 11

Molecular Class: ○ Luminal   ● Basal   ◉ HER2 Enriched
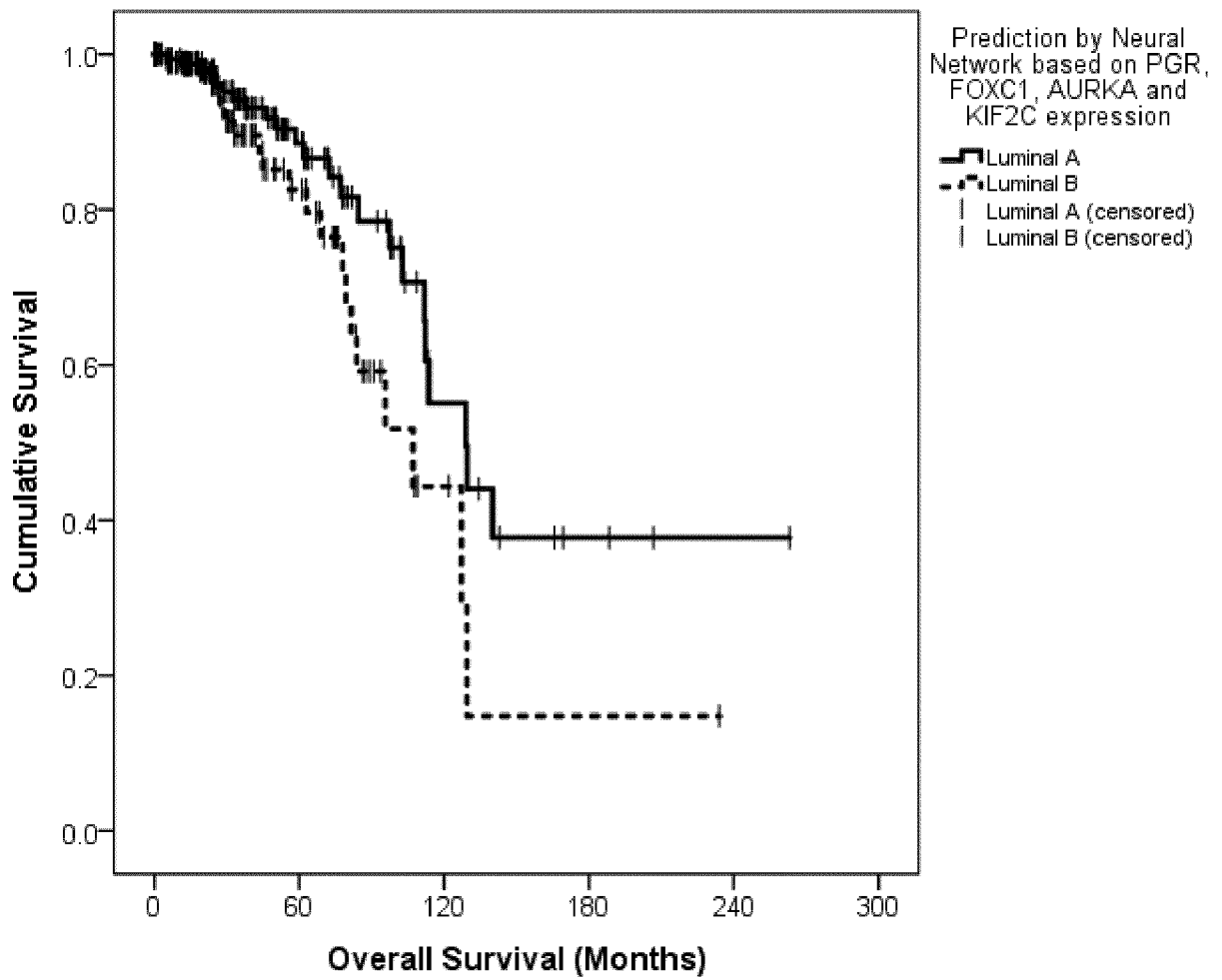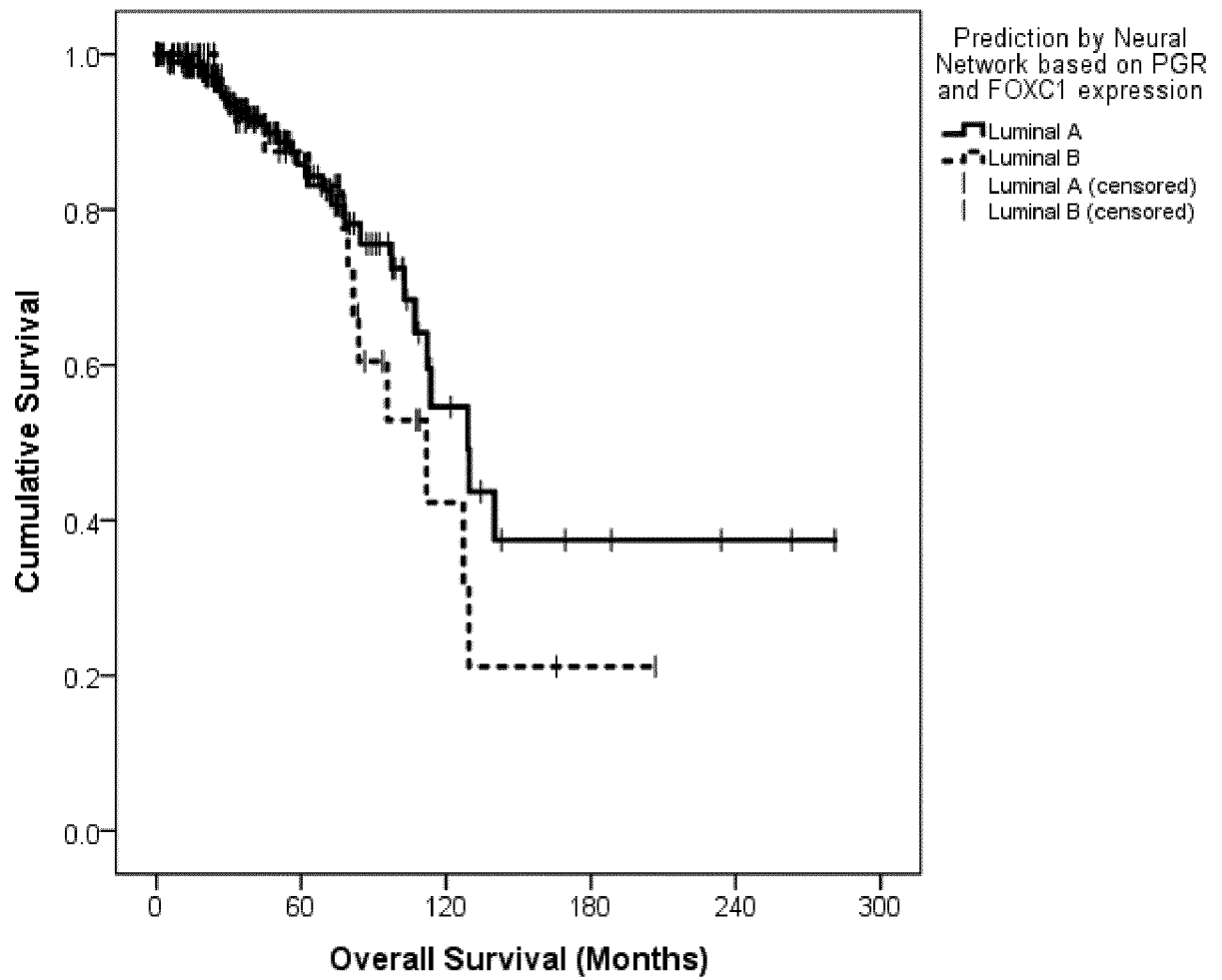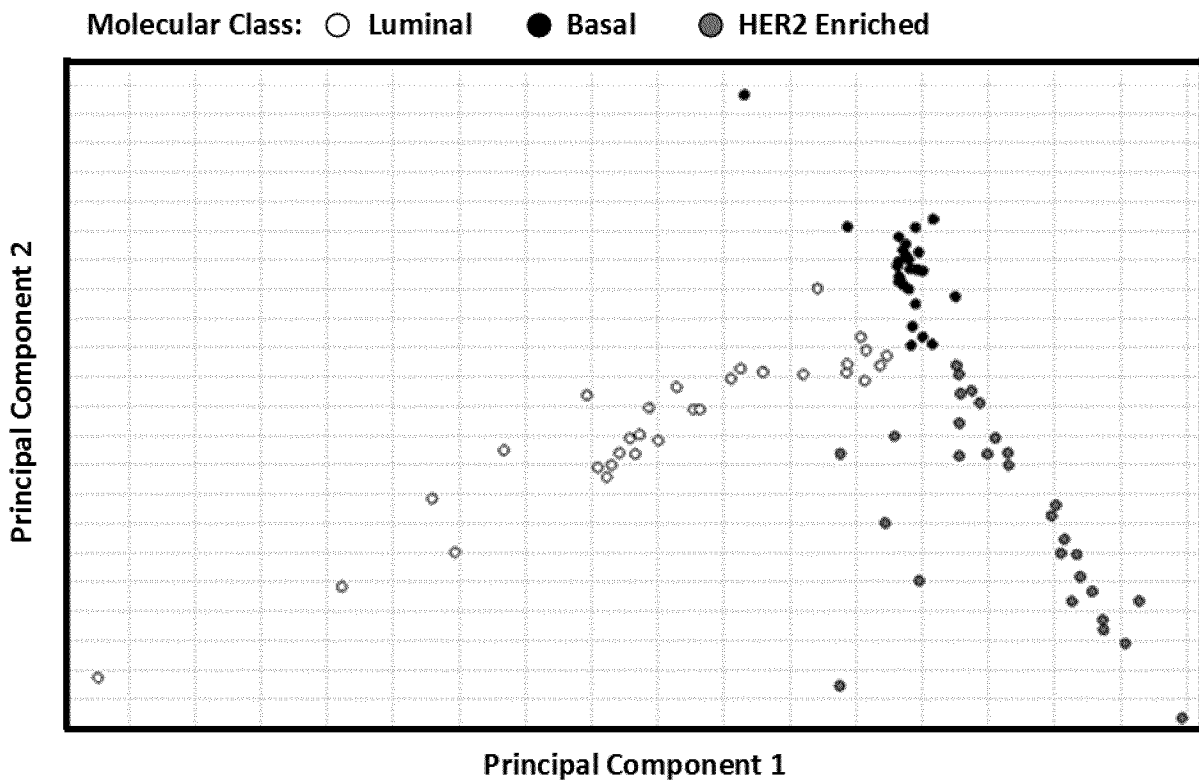
Concordance of IHC with the Expression algorithm	HER2 Positive by IHC/FISH (N=29)	ER positive by IHC (N=31)	TNBC by IHC (N=25)
	100% HER2 Enriched	96.8% Luminal	100% Basal

Figure 12

Sample	MDA. MB453	MDA. MB453 120m	BT474	BT474 1m	BT474 5m	BT474 10m	BT474 30m	BT474 60m	BT474 120m	6708/00	10535/03	11274/03
Conc	86	134	112	82	115	168	216	117	112	444	578	239
RIN	8.9	2.5	7.2	7.1	3.2	2.1	2.1	2.4	2.5	2.2	2.2	2.3

Figure 13

A

B

Figure 14

Normalised Data BT474 across degradation
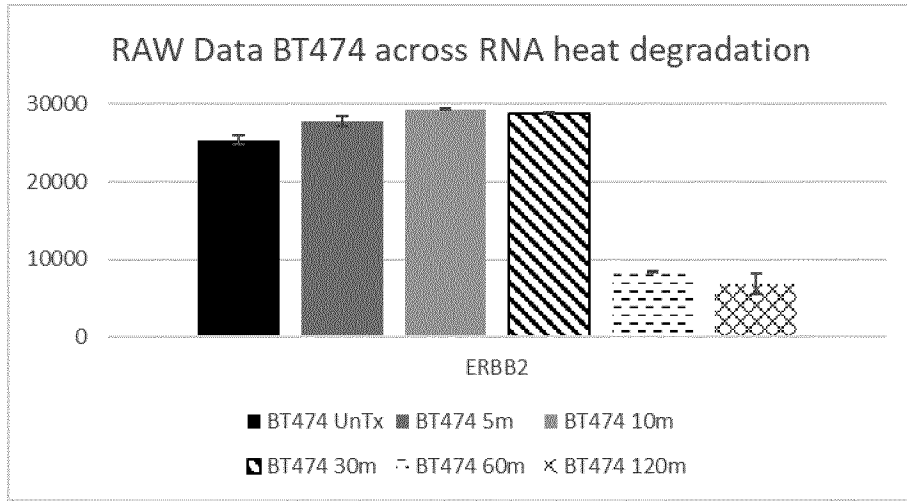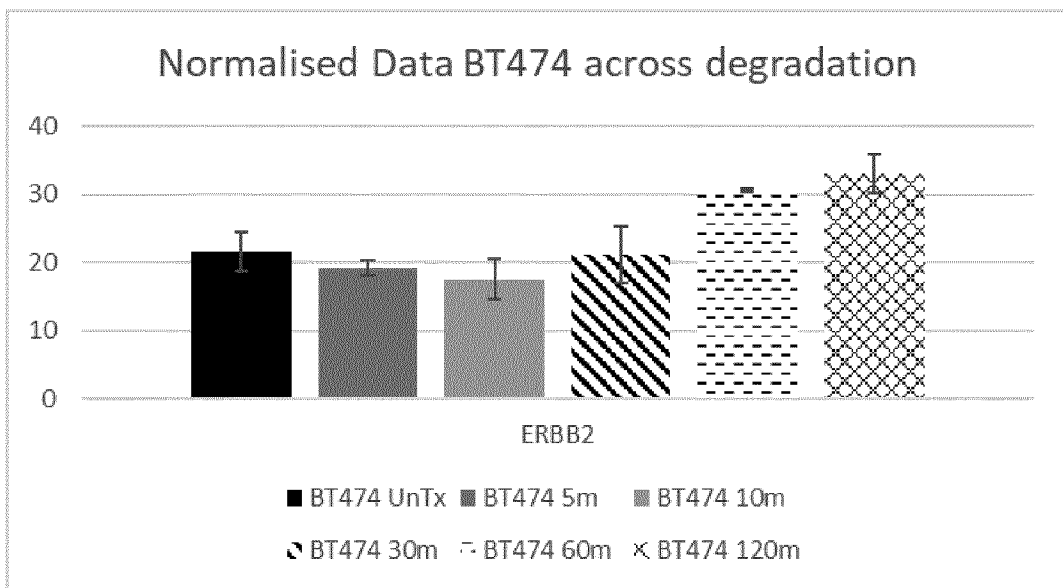
Figure 15

Figure 16

Figure 17

HER2 positive tumours (n=8)	ER positive tumours (n=8)	TNBC (n=9)	
A. Receptor Status			

K-W χ2 = 11.294
p-value = 0.001

K-W χ2 = 0.540
p-value = 0.462

K-W χ2 = 4.306
p-value = 0.038

K-W χ2 = 0.912
p-value = 0.339
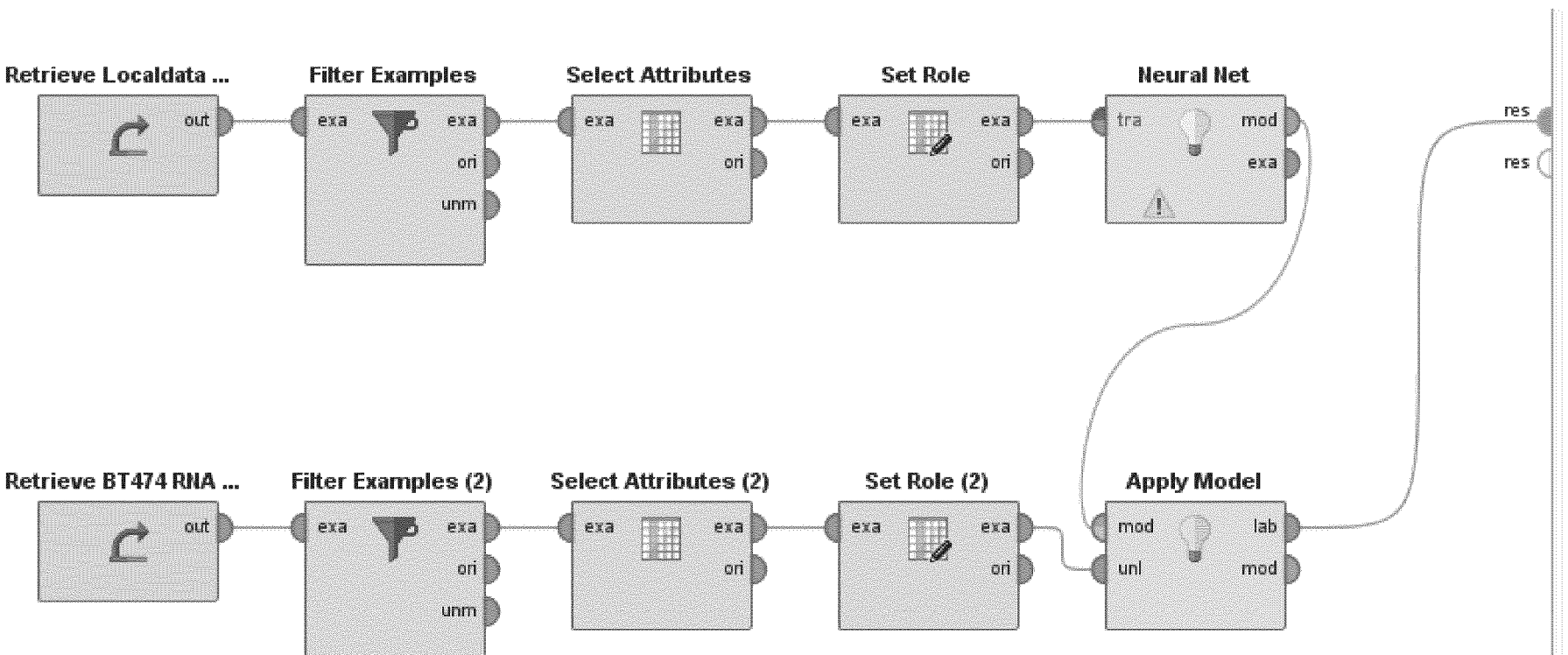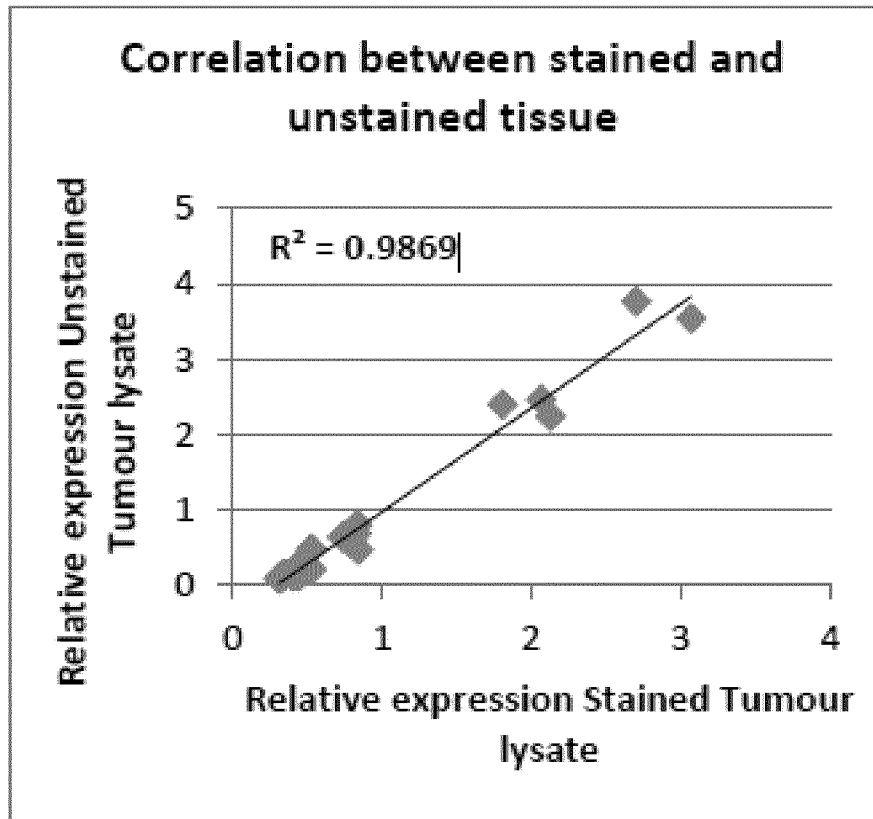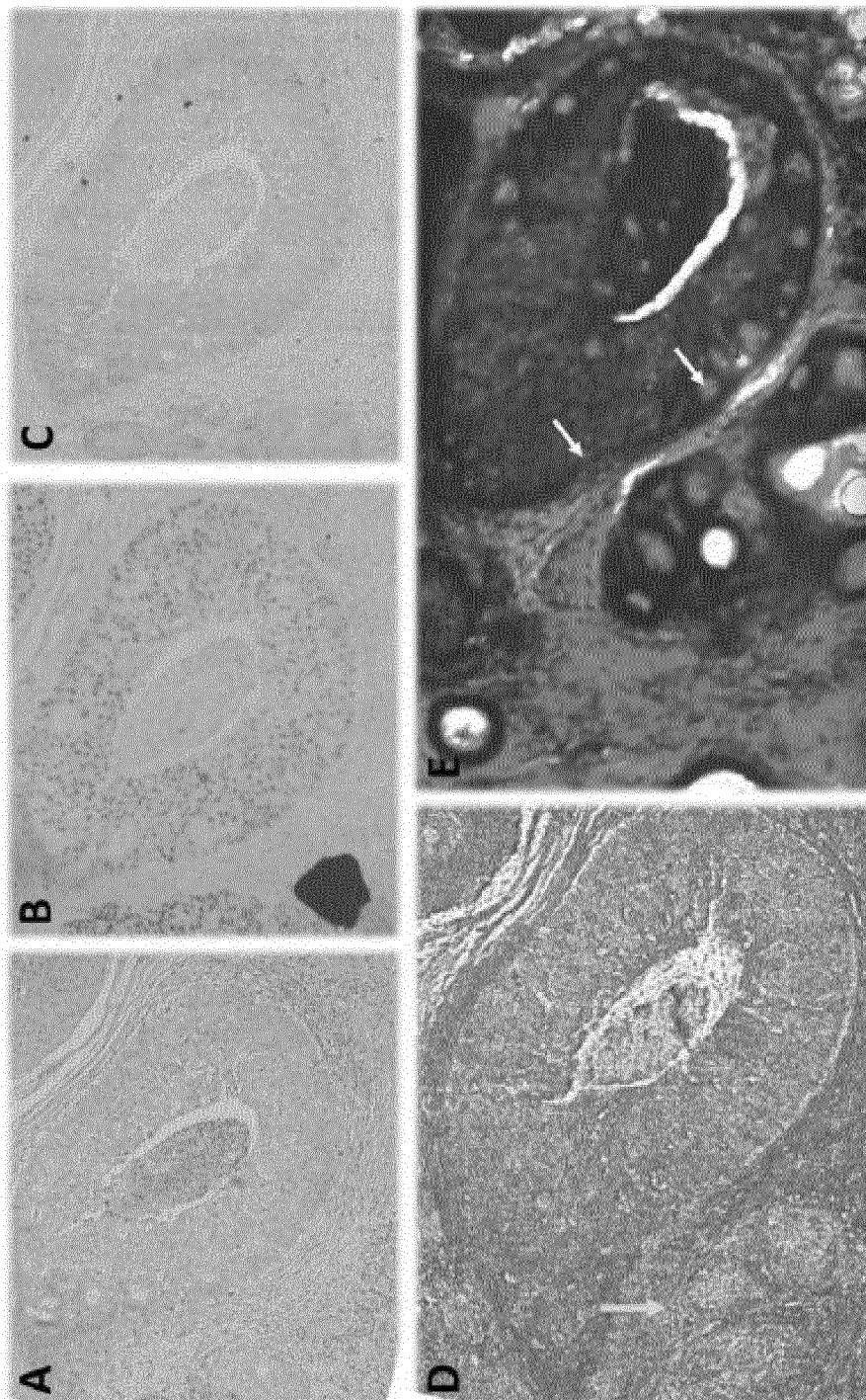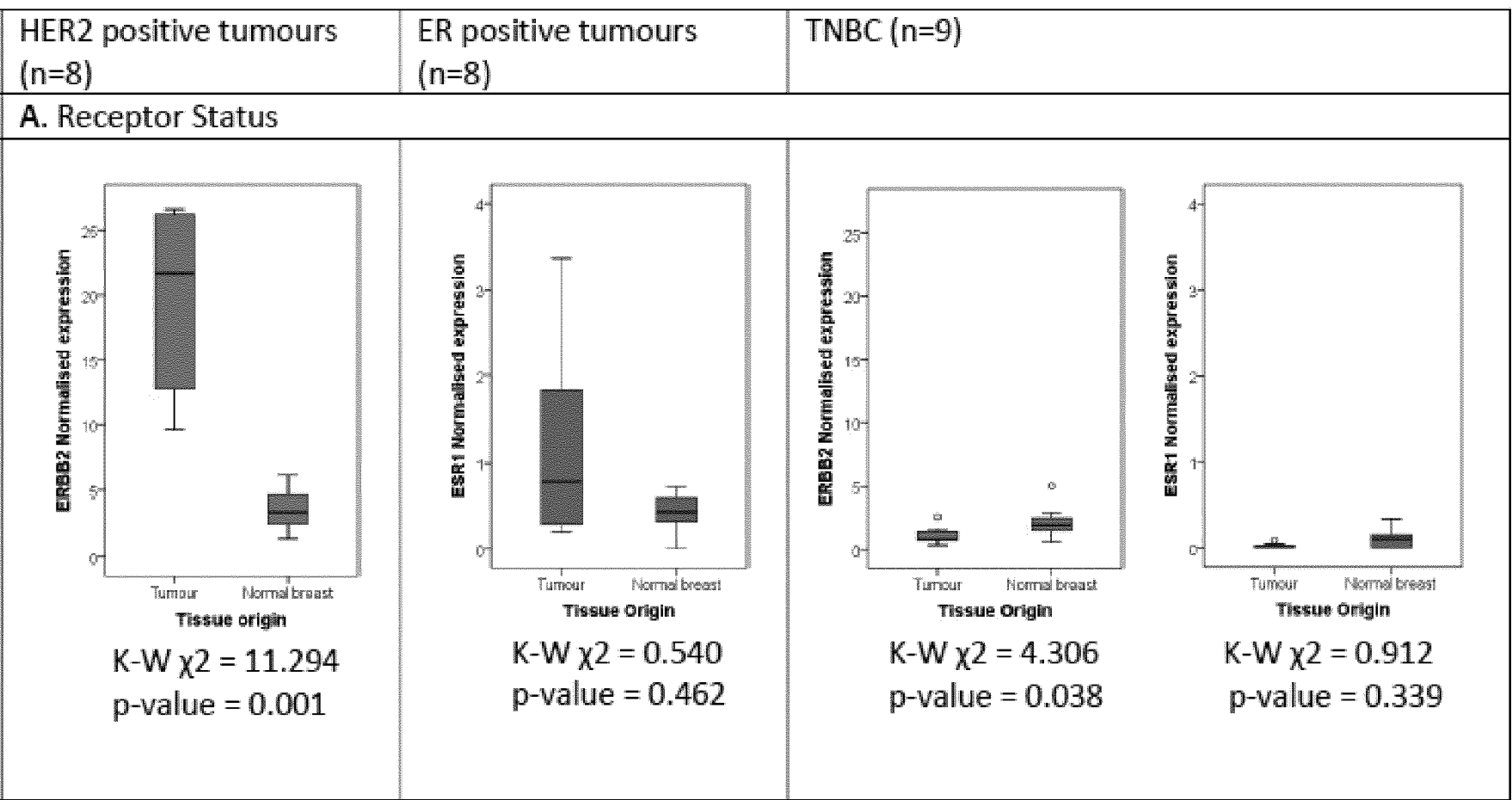
B. Expression of mesenchymal marker, Fibronectin (FN1)

K-W χ2 = 11.294
p-value = 0.001

K-W χ2 = 5.338
p-value = 0.021

K-W χ2 = 3.290
p-value = 0.070
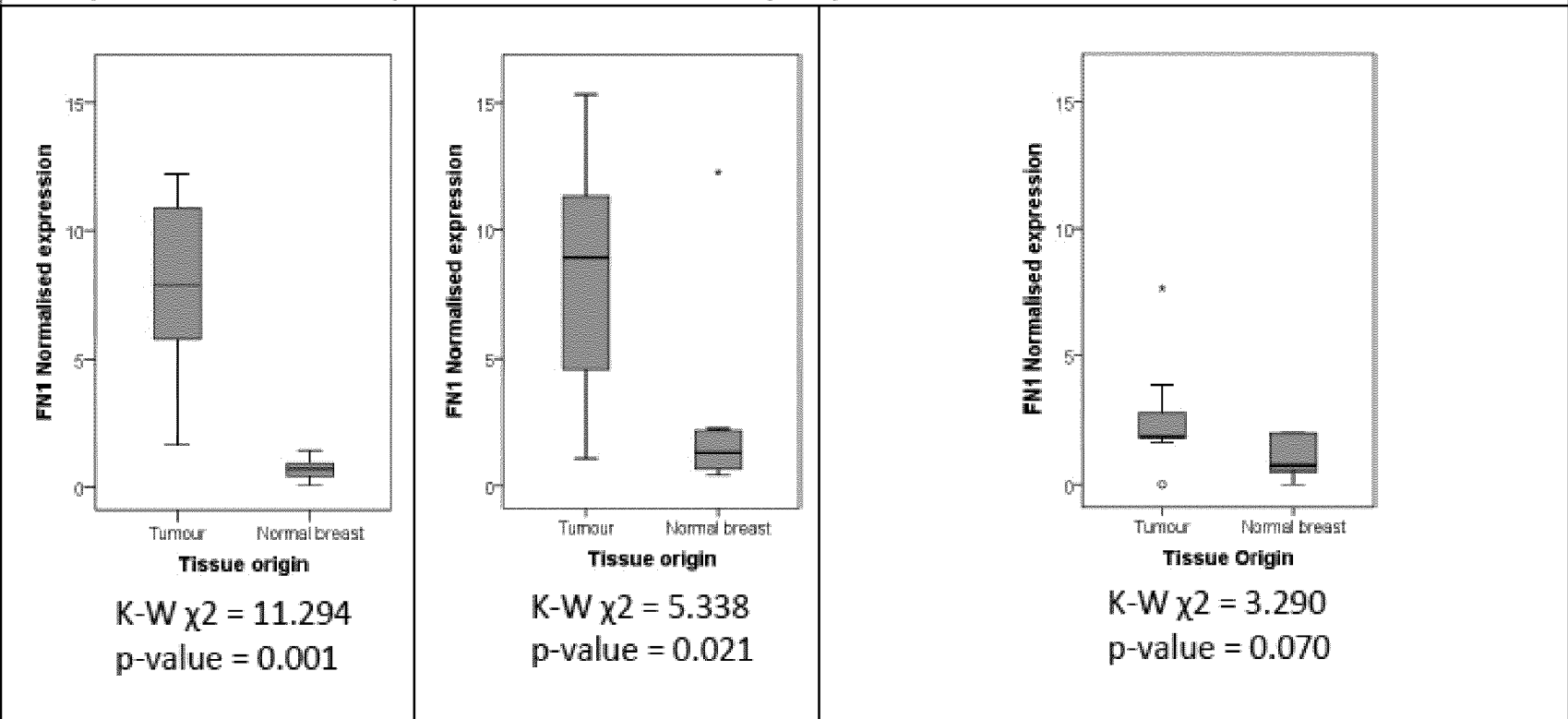
Figure 19

Figure 20

21/28
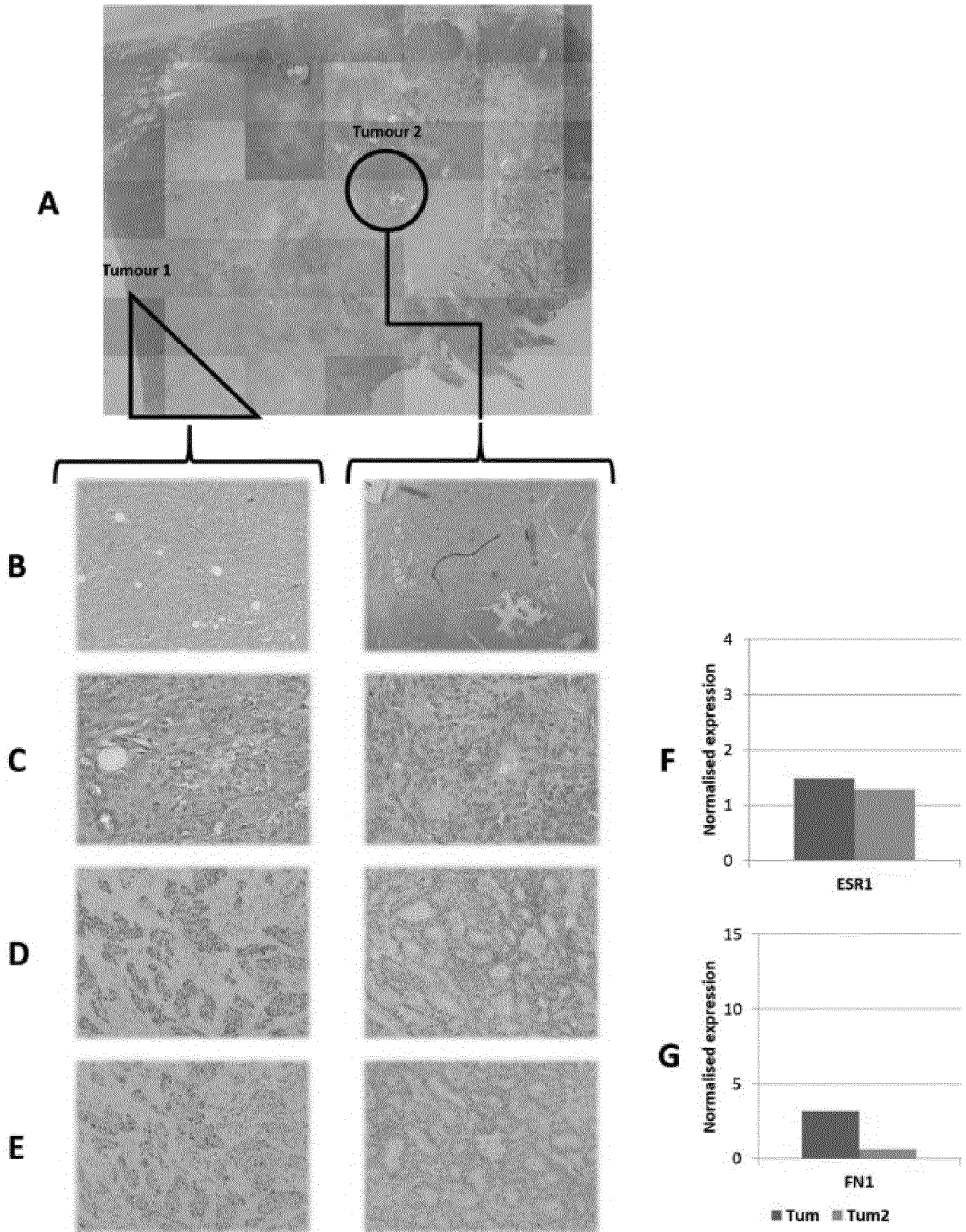
| Name of Material/ Equipment	Company	Catalog Number
Microtome	Leica	RM2235
Heamatoxylin Mayer's	Sigma	MHS16-500mL
Eosin Y Aquaeous solution	Sigma	HT110216-500mL
Nikon Eclipse Ti-E Inverted microscope	Nikon	Ti-E
Laser Microdissection membranes	Molecular Machines &Industries	S0103
mmi CellCamera 1.4	Molecular Machines &Industries	MX4285c-ACK07
mmi Cellcut Plus®	Molecular Machines &Industries	
Diffuser caps	Molecular Machines &Industries	50210
mmi Celltools® Software v.4.01rcl	Molecular Machines &Industries	
Eppendorf Thermomixer® comfort	Eppendorf	5355000038
1.5mL heating block for Eppendorf Thermomixer	Eppendorf	Cat no: 022670522
96-well plate heating block for Eppendorf Thermomixer	Eppendorf	Cat no: 022670565
Labnet Vortemp 56®	Labnet	52056A-220
LX200 100/200®	Luminex	
QuantiGene Sample Processing Kit - FFPE Tissues	Affymetrix	QS0109
QuantiGene Plex 2.0 Assay Kit (Magnetic Separation)	Affymetrix	QP1011
Thermaseal RTS Sealing Film	Thermaseal	765246
Hand-Held Magnetic Plate Washer	Affymetrix	QP1011
QuantiGene® Incubator Temperature Validation Kit	Affymetrix/Panomics	QS0517
Proteinase K (50µg/µL)	Affymetrix	14622
QuantiGene Plex 2.0 Sets	Affymetrix	Various
Multi Speed Vortex	Kisker Biotech	MSV-3500
Sonicator	Silvercrest	
RNASEZAP	Sigma	R2020-250ML

Figure 21
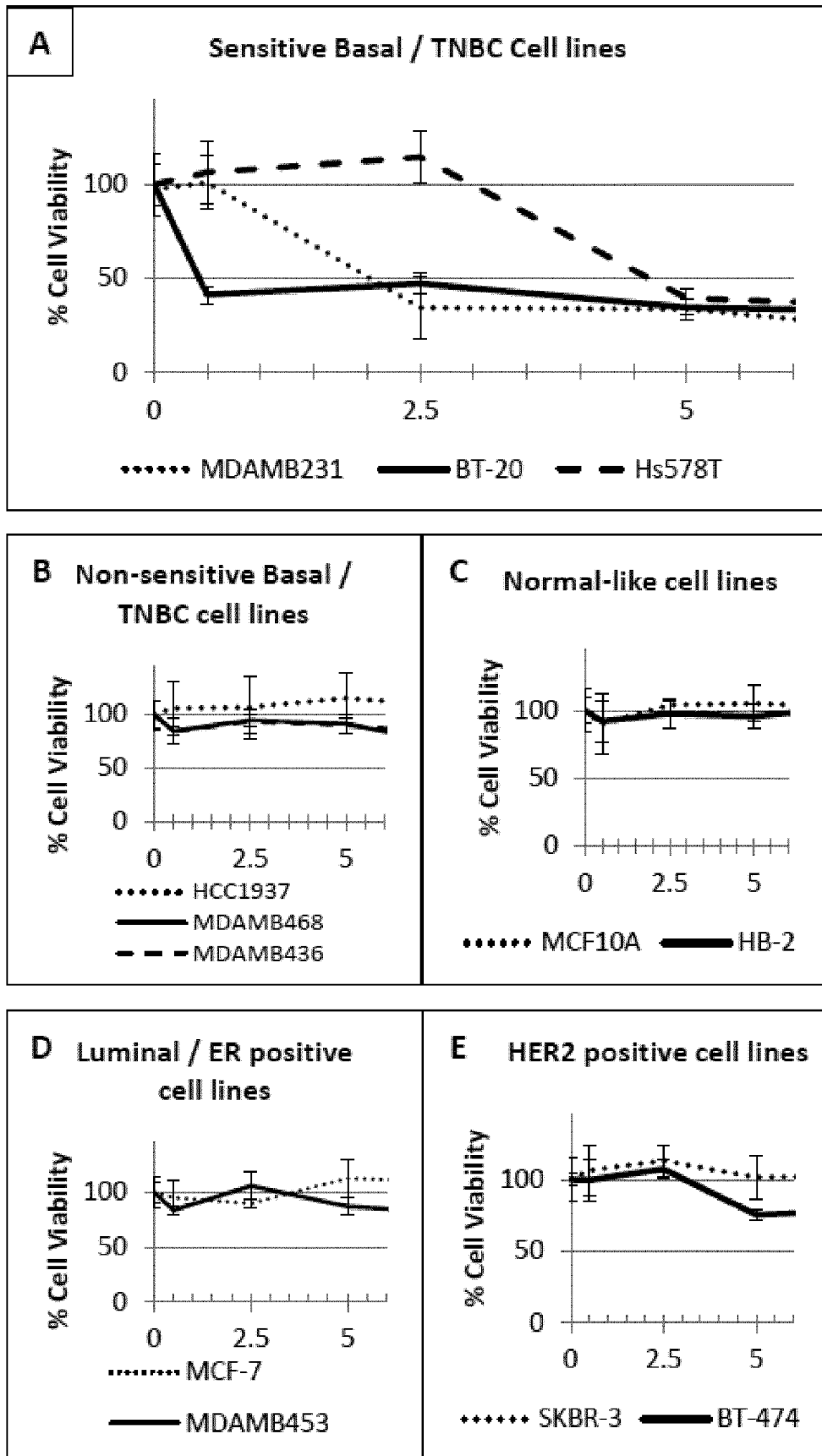
Figure 22

A
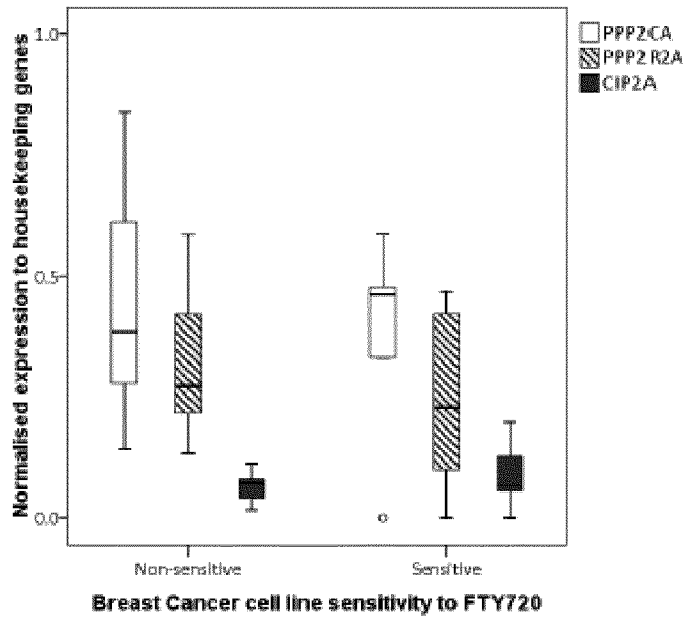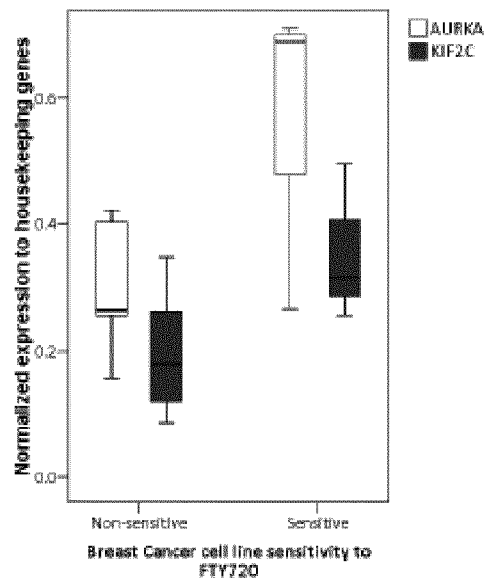

Distribution of PP2A complex and regulator expression between sensitive and non-sensitive cell lines

Gene	Mann-Whitney U test, *P*
PPP2CA	0.256
PPP2R2A	0.310
CIP2A	0.040*

B

Distribution of PP2A complex and regulator expression between sensitive and non-sensitive cell lines

Candidate Biomarkers	Mann-Whitney U test, *P*
AURKA	0.014*
KIF2C	< 0.001*

Figure 23 (part 1 of 2)

A

B

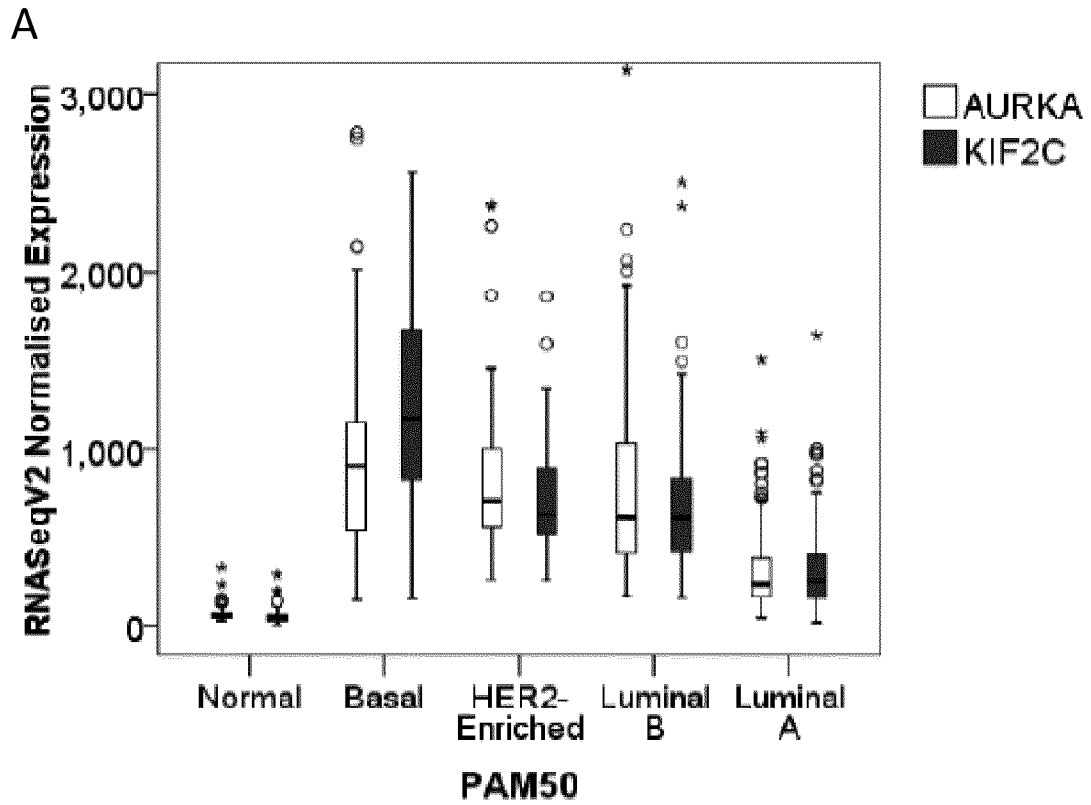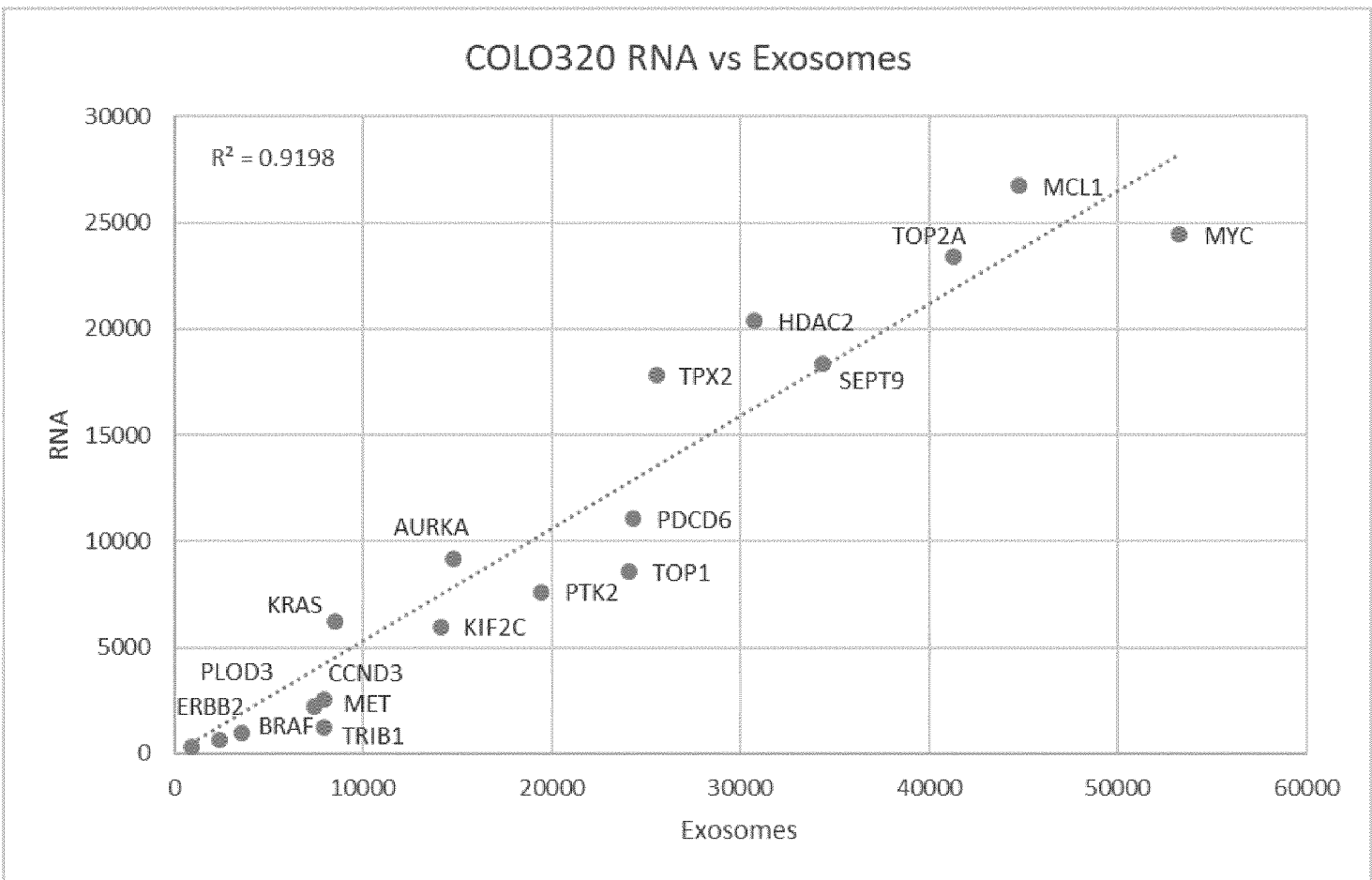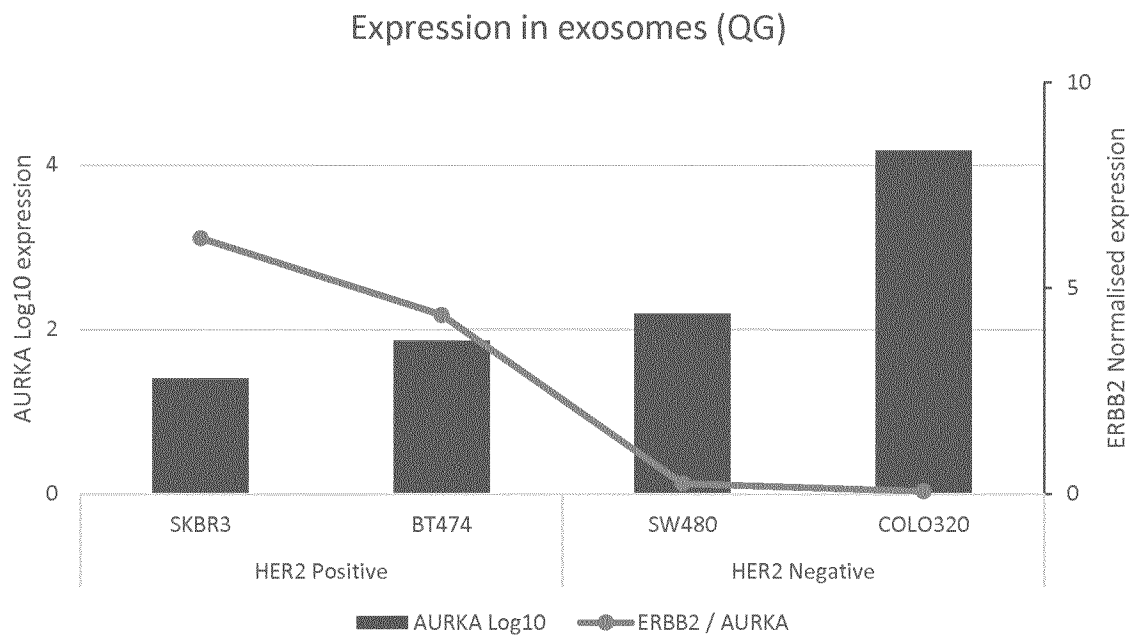| Candidate Biomarkers	*Wilcoxon Signed Rank test, P*
AURKA	< 0.001*
KIF2C	< 0.001*
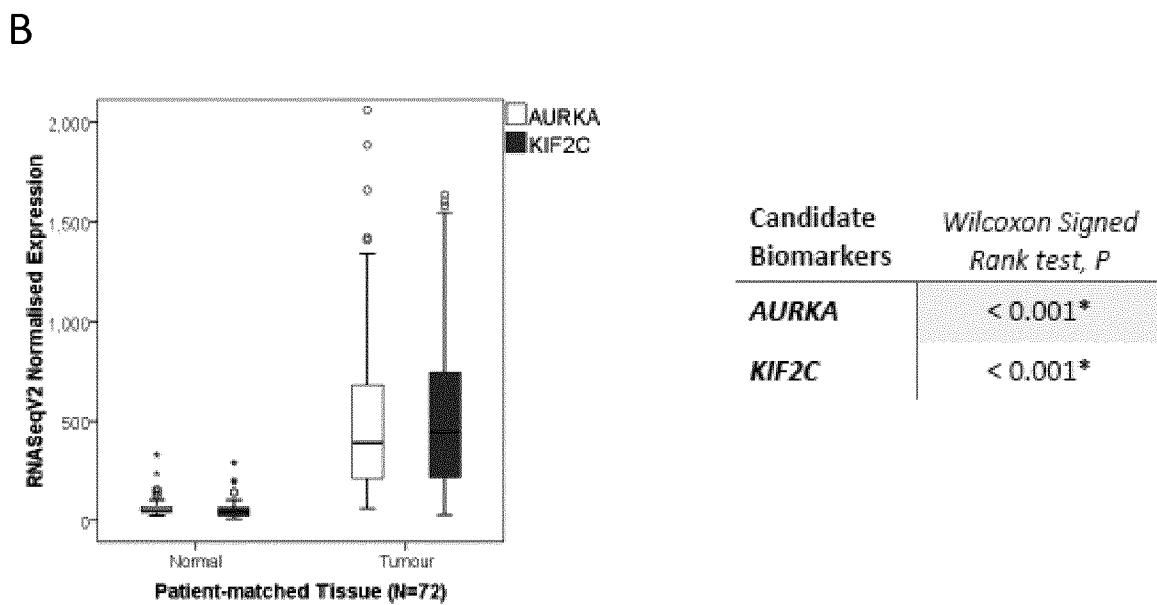
Figure 23 (part 2 of 2)

C
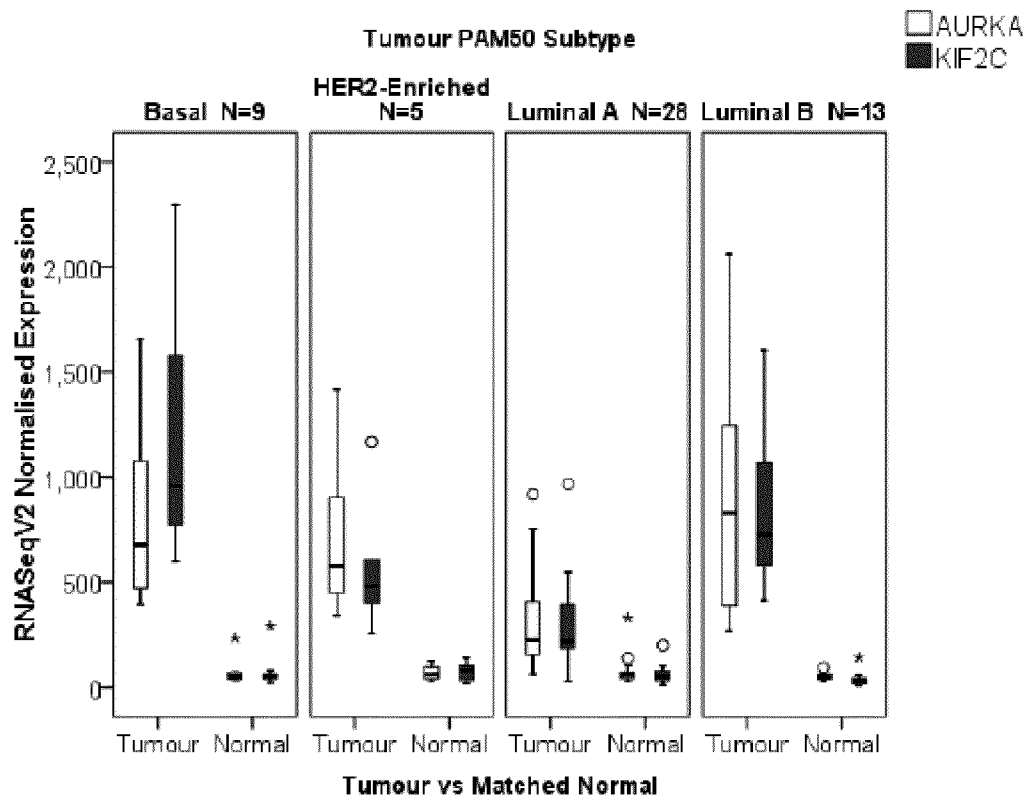
Figure 24
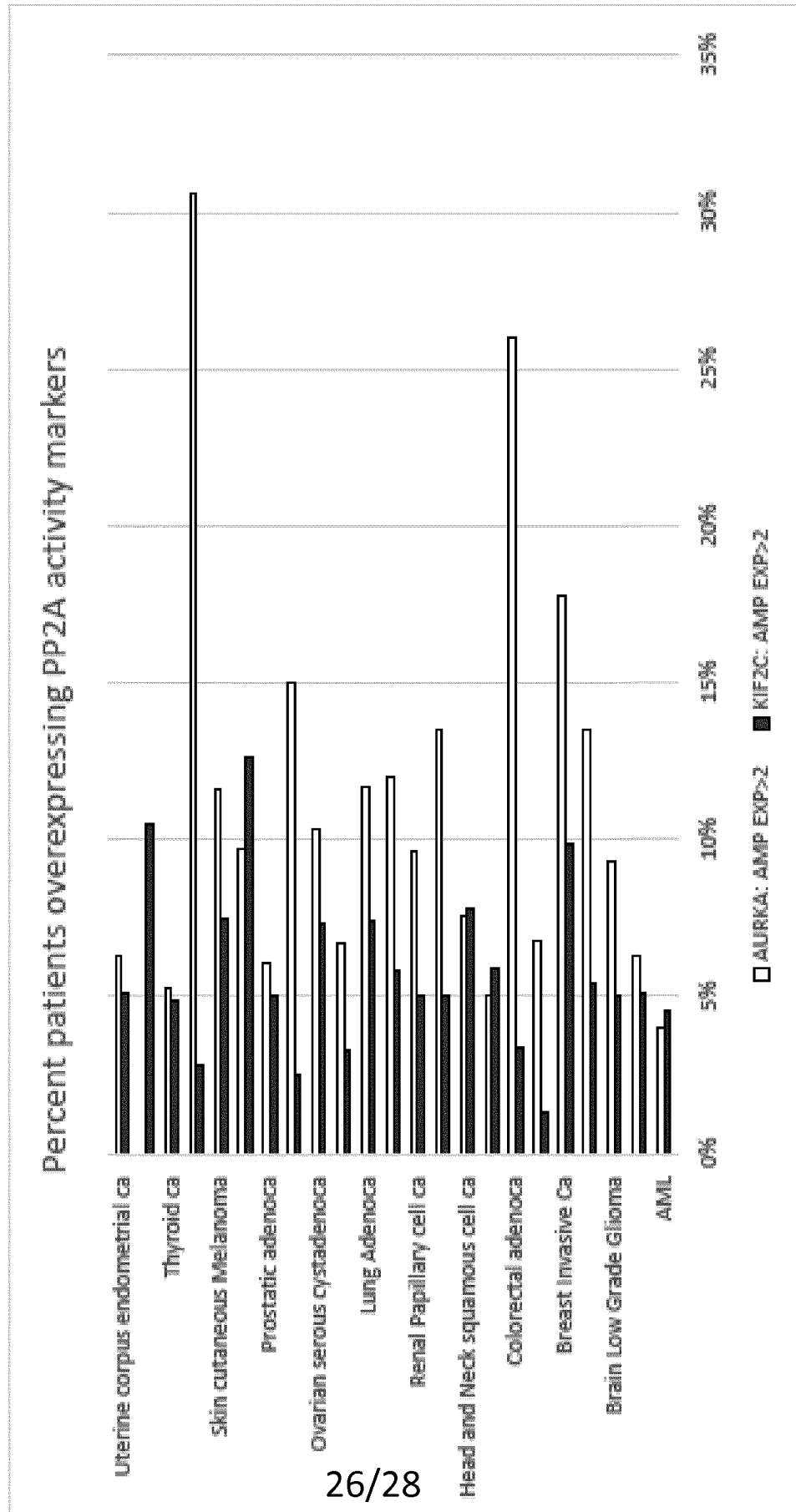
Figure 25

COLO320 RNA vs Exosomes

$R^2 = 0.9198$

Figure 26

Expression in exosomes (QG)

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/6851   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, 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	US 2008/286827 A1 (STONE STEVE [US] ET AL) 20 November 2008 (2008-11-20) paragraphs [0005] - [0007], [0015], [0017] - [0019], [0024], [0049], [0056], [0127], [0173]; table 1 -----	1-61
X	WO 2015/073949 A1 (UNIV NORTH CAROLINA [US]) 21 May 2015 (2015-05-21) paragraphs [0008] - [0012], [0016], [0048], [0071], [0079] - [0082]; table 3 -----	1-61
A	WO 2016/144265 A1 (AGENCY SCIENCE TECH & RES [SG]) 15 September 2016 (2016-09-15) claims 1-23, Fig. 1, 9, para. 28-31 and 60 ----- -/--	1-61

| X | Further documents are listed in the continuation of Box C. | X | 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	Date of mailing of the international search report
28 March 2019	10/04/2019

| 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 Lapopin, Laurence |

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation).	DOCUMENTS CONSIDERED TO BE RELEVANT	
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ZHANG QUAN ET AL: "Combined analysis of rearrangement of ALK, ROS1, somatic mutation of EGFR, KRAS, BRAF, PIK3CA, and mRNA expression of ERCC1, TYMS, RRM1, TUBB3, EGFR in patients with non-small cell lung cancer and their clinical significance", CANCER CHEMOTHERAPY AND PHARMACOLOGY, SPRINGER VERLAG, BERLIN, vol. 77, no. 3, 3 February 2016 (2016-02-03), pages 583-593, XP035872960, ISSN: 0344-5704, DOI: 10.1007/S00280-016-2969-Y [retrieved on 2016-02-03] p 586, col. 1 "mRNA expression analysis" -----	1-61
A	US 2012/029051 A1 (SMITH ANJA [US] ET AL) 2 February 2012 (2012-02-02) paragraph [0117] -----	1-61

1

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 2008286827	A1	20-11-2008	NONE		
WO 2015073949	A1	21-05-2015	NONE		
WO 2016144265	A1	15-09-2016	CN	107429295 A	01-12-2017
			EP	3268494 A1	17-01-2018
			SG	11201707325R A	30-10-2017
			US	2018230544 A1	16-08-2018
			WO	2016144265 A1	15-09-2016
US 2012029051	A1	02-02-2012	AR	082411 A1	05-12-2012
			AU	2011282902 A1	31-01-2013
			BR	112013001918 A2	24-05-2016
			CA	2804847 A1	02-02-2012
			CL	2013000237 A1	21-03-2014
			CN	103068981 A	24-04-2013
			EP	2598639 A2	05-06-2013
			JP	2013539456 A	24-10-2013
			KR	20130095737 A	28-08-2013
			RU	2013102545 A	10-09-2014
			TW	201219041 A	16-05-2012
			US	2012029051 A1	02-02-2012
			US	2014031414 A1	30-01-2014
			UY	33533 A	29-02-2012
			WO	2012015775 A2	02-02-2012
			ZA	201300531 B	26-03-2014