

MOLECULAR DIAGNOSTIC TEST FOR CANCER

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 61/732,761 entitled "Molecular Diagnostic Test for Cancer" and filed on December 3, 2012.

FIELD OF THE INVENTION

[0002] The present invention relates to a molecular diagnostic test useful for diagnosing cancers from different anatomical sites that includes the use of a common subtype related to angiogenesis. The invention includes the derivation of a gene classification model from gene expression levels. One application is the stratification of response to and selection of patients for cancer therapeutic drug classes and thus guide patient treatment selection. Another application is the stratification of cancer patients into those that respond and those that do not respond to anti-angiogenic therapeutics. The present invention provides a test that can guide therapy selection as well as selecting patient groups for enrichment strategies during clinical trial evaluation of novel therapeutics. The invention can be used as a prognostic indicator for certain cancers including ovarian cancer, breast cancer, colon, prostate, lung and glioblastoma. The angiogenesis subtype can be identified from fresh/frozen (FF) or formalin fixed paraffin embedded FFPE patient samples.

BACKGROUND

[0003] The pharmaceutical industry continuously pursues new drug treatment options that are more effective, more specific or have fewer adverse side effects than currently administered drugs. Drug therapy alternatives are constantly being developed because genetic variability within the human population results in substantial differences in the effectiveness of many established drugs. Therefore, although a wide variety of drug therapy options are currently available, more therapies are always needed in the event that a patient fails to respond.

[0004] Traditionally, the treatment paradigm used by physicians has been to prescribe a first-line drug therapy that results in the highest success rate possible for treating a disease. Alternative drug therapies are then prescribed if the first is ineffective. This paradigm is clearly not the best treatment method for certain diseases. For example, in diseases such as cancer, the first treatment is often the most important and offers the best opportunity for successful therapy, so there exists a heightened need to choose an initial drug that will be the most effective against that particular patient's disease.

[0005] Ovarian cancer is the leading cause of death among all gynecological cancers in western countries. This high death rate is due to the diagnosis at an advanced stage in most patients. Epithelial ovarian cancer (EOC) constitutes 90% of ovarian malignancies and is classified into distinct histologic categories including serous, mucinous, endometrioid, clear cell, transitional, mixed, and undifferentiated subtypes. There is increasing evidence that these differed histologies arise from different aetiologies. There have been recent advances in the methodology used to classify epithelial ovarian cancer (McCluggage, W.G. "Morphological subtypes of ovarian carcinoma: a review with emphasis on new developments and pathogenesis," *PATHOLOGY* 2011 Aug;43(5):420-32). One of the consequences of this is that many tumors that would previously been classified as endometrioid are now being classified as serous.

[0006] The current standard treatment for ovarian cancer is debulking surgery and standard platinum taxane based cytotoxic chemotherapy. However, not all patients respond to this, and of those that do, approximately 70% will experience a recurrence. Specific targeted therapies for ovarian cancer based on histological or molecular classification have not yet reached the marketplace. Similarly for other types of cancer, there is still no accurate way of selecting appropriate cytotoxic chemotherapeutic agents.

[0007] The advent of microarrays and molecular genomics has the potential for a significant impact on the diagnostic capability and prognostic classification of disease, which may aid in the prediction of the response of an individual patient to a defined therapeutic regimen. Microarrays provide for the analysis of large amounts of genetic information, thereby providing a genetic fingerprint of an individual. There is much enthusiasm that this technology will ultimately provide the necessary tools for custom-made drug treatment regimens.

[0008] Currently, healthcare professionals have few mechanisms to help them identify cancer patients who will benefit from chemotherapeutic agents. Identification of the optimal first-line drug has been difficult because methods are not available for accurately predicting which drug treatment would be the most effective for a particular cancer's physiology. This deficiency results in relatively poor single agent response rates and increased cancer morbidity and death. Furthermore, patients often needlessly undergo ineffective, toxic drug therapy.

[0009] Angiogenesis is a key component of neo-vascularisation of tumors and essential to tumorigenesis and metastasis. As such, it is a key area for therapeutic intervention and has been correlated to poor prognosis and reduced survival. This has promoted the development of a number of agents that target angiogenesis related processes and pathways, including the market leader and first FDA-approved anti-angiogenic, bevacizumab (Avastin), produced by Genentech/Roche.

[0010] Treatment regimens that include bevacizumab have demonstrated broad clinical activity¹⁻¹⁰. However, no overall survival (OS) benefit has been shown after the addition of bevacizumab to cytotoxic chemotherapy in most cancers^{8, 12-13}. This suggests that a substantial proportion of tumours are either initially resistant or quickly develop resistance to VEGF blockade (the mechanism of action of bevacizumab). In fact, 21% of ovarian, 10% of renal and 33% of rectal cancer patients show partial regression when receiving bevacizumab monotherapy, suggesting that bevacizumab may be active in small subgroups of patients, but that such incremental benefits do not reach significance in unselected patients. As such, the use of a biomarker of response to bevacizumab would improve assessment of treatment outcomes and thus enable the identification of patient subgroups that would receive the most clinical benefit from bevacizumab treatment. This would be particularly relevant in the case of metastatic breast cancer, where the absence of a clinically beneficial biomarker has undermined the use of bevacizumab. Thus far, no such biomarker has been clinically validated to predict bevacizumab efficacy. Hypertension and VEGF polymorphisms are so far the only biomarkers to show potential, but important questions remain about their use in a clinical setting.

[0011] Another approach to anti-angiogenic therapy is simultaneous targeting of multiple angiogenic pathways rather than selective targeting of the VEGF pathway. Theoretically, multitargeted anti-angiogenic agents should more completely inhibit angiogenesis than agents such as bevacizumab and thus may produce greater therapeutic benefit. It has been postulated that in some tumors, angiogenesis may require VEGF only in the early stages of disease but is driven by additional angiogenic pathways as the disease progresses. Therefore, by targeting multiple pathways, it may be possible to counteract compensatory escape mechanisms that could lead to resistance to VEGF inhibition.

[0012] As for other types of cancer there is still no accurate way of selecting which patients will or will not respond to standard of care with an anti-angiogenic therapeutic or single agent anti-angiogenic therapy.

[0013] What is therefore needed is a molecular diagnostic test that would facilitate the stratification of patients based upon their predicted response to anti-angiogenic therapeutics, either in combination with standard of care or as a single-agent therapeutic. This would allow for the rapid identification of those patients who should receive alternative therapies. Such a molecular diagnostic test should be predictive of therapeutic responsiveness across different cancer types with sufficient accuracy.

SUMMARY OF THE INVENTION

[0014] Disclosed are methods of using a collection of biomarkers expressed in cancer such that when some or all of the transcripts are under-expressed, they identify a subtype of cancer that has an up-regulation or down-regulation in molecular signaling relating to angiogenesis. The collection of biomarkers may be defined by an expression signature, and the expression signature is used to assign a cumulative score to the measured expression values of the collection of biomarkers. The invention also provides methods for indicating responsiveness or non-responsiveness to anti-angiogenic agents as well as identifying patients with a good or poor prognosis, using the collection of biomarkers and expression signatures disclosed herein. In different aspects, the biomarkers and expression signatures may form the basis of a single parameter or multiparametric predictive test that could be delivered using methods known in the art such as microarray, Q-PCR, immunohistochemistry, ELISA or other technologies that can quantify mRNA or protein expression.

[0015] In one example embodiments, the present invention provides an expression signature defining a cancer subtype associated with up-regulation of angiogenesis related biomarkers ("angiogenesis"). In another example embodiment, the present invention provides an expression signature defining a cancer subtype association with down-regulation of angiogenesis related biomarkers ("non-angiogenesis"). In addition, the cancer subtypes described herein are common to many types of cancer and are not limited to a single cancer disease type. Accordingly, the expression signatures of the present invention are not limited to a single cancer type. In certain exemplary embodiments, the expression signature comprises two or more biomarkers selected from the biomarkers listed in Tables 1A-1C. In another exemplary embodiment, the angiogenesis expression signature comprises two or more biomarkers selected from the biomarkers listed in Tables 2A or 2B. In another exemplary embodiment, the non-angiogenesis expression signature comprises two or more biomarkers listed in Table 2C. In another exemplary embodiment, the angiogenesis expression signature comprises the biomarkers listed in Table 2A or 2B and their corresponding weights as determined using a PLS classifier. In another exemplary embodiment, the non-angiogenesis signature comprises the biomarkers listed in Table 2C and their corresponding weights as determined using a PLS classifier.

[0016] In one embodiment of the invention, the expression signatures are useful for evaluating a cancer tumor's responsiveness to anti-angiogenic therapeutics. The use of the invention to determine a tumor's responsiveness to anti-angiogenic therapeutics is not limited to a single cancer type. In one example embodiment, the cancer may be ovarian cancer, breast cancer, colon cancer, colorectal cancer, lung cancer, prostate cancer, or glioblastoma. The

present invention relates to prediction of response to drugs using at least or up to different 10 classifications of response, such as overall survival, progression free survival, radiological response, as defined by RECIST, complete response, partial response, stable disease and serological markers such as, but not limited to, PSA, CEA, CA125, CA15-3 and CA19-9. The invention described herein is not limited to any one drug; it can be used to identify responders and non responders to any of a range of drugs currently in use, under development and novel, that directly or indirectly affect or target angiogenic processes. In one embodiment, the present invention may be used to evaluate adjuvant or neoadjuvant bevacizumab or dasatinib, either as single agents, or in combination with standard of care therapy. In another embodiment, the present invention may be used to evaluate Avastin, VEGF-TRAP, treatment in ovarian cancer.

[0017] In another embodiment of the invention, the expression signatures disclosed herein may be used to determine a patient's clinical prognosis. For example, patients identified as having a cancer subtype associated with down-regulation of angiogenesis related biomarkers may exhibit a longer survival rate than a cancer subtype related to up-regulation of angiogenesis related biomarkers. The use of the invention to determine an individual's clinical prognosis is not limited to a single cancer type. In one example embodiment, the cancer may be ovarian cancer, colon cancer, colorectal cancer, breast cancer, lung cancer, prostate cancer, or glioblastoma. The present invention relates to prediction of clinical prognosis using at least progression free survival. Additional prognostic factors that may be considered are ethnicity and race, age, stage of disease, histology, tumor grade, tumor makers (for example, CA125), site-specific surgical treatment, size of residual disease, and tumor response

[0018] The predictive and prognostic uses of the expression signatures disclosed herein may be achieved using a single expression signature or multiple expression signatures. For example, one expression signature may be a non-angiogenesis signature and a second expression signature may be an angiogenesis signature. In one example embodiment, a non-angiogenesis expression signature is used to determine a patient's clinical prognosis and an angiogenesis signature is used to determine a patient's predicted responsiveness to a given class of therapeutic agent or treatment regimen. In one example embodiment, the angiogenesis signature comprises one or more biomarkers from Table 2A or Table 2B, and the non-angiogenesis signature comprises one or more biomarkers from Table 2C. In another example embodiment, the angiogenesis signature comprises the biomarkers in Table 2A or Table 2B, and the non-angiogenesis signature comprises the biomarkers in Table 2A.

[0019] In another aspect, the present invention relates to kits for conventional diagnostic uses listed above such as qPCR, microarray, and immunoassays such as immunohistochemistry,

ELISA, Western blot and the like. Such kits include appropriate reagents and directions to assay the expression of the genes or gene products and quantify mRNA or protein expression.

[0020] Also disclosed are methods for identifying human tumors with or without the non-angiogenesis phenotype. In certain exemplary embodiments, such methods may be used to identify patients that are sensitive to and respond to drugs that inhibit, either directly or indirectly, processes relating to angiogenesis. In certain other exemplary embodiments, such methods may be used to identify patients that are resistant to or do not respond to drugs that inhibit, either directly or indirectly, processes relating to angiogenesis.

[0021] This invention also relates to guiding effective treatment of patients. Further, methods relating to selection of patient treatment regimens and selecting patients for clinical trials of current, or developmental stage drugs that directly or indirectly affect angiogenesis are provided.

[0022] In addition, methods that accommodate the use of archived formalin fixed paraffin-embedded (FFPE) biopsy material, as well as fresh/frozen (FF) tissue, for assay of all transcripts, and are therefore compatible with the most widely available type of biopsy material, are described herein. A biomarker expression level may be determined using RNA obtained from FFPE tissue, fresh frozen tissue or fresh tissue that has been stored in solutions such as RNeasy®.

BRIEF DESCRIPTION OF DRAWINGS

[0023] Figure 1 provides a heatmap representing the hierarchical agglomerative clustering analysis of the most variable probe sets across 199 serous samples of the Almac Diagnostics' epithelial ovarian cancer sample set. The functional analysis of the probe set clusters is summarized on the right hand side of the image. The legend across the top of the image indicates the classifier group each sample was assigned to for classifier generation (i.e. Class labels).

[0024] Figures 2A and 2B represents the functional analysis results of the angiogenesis probe set cluster in the 199 serous only samples in an epithelial ovarian cancer training set using a functional enrichment analysis. Figure 2A shows a histogram representing the significance of the top 10 enriched Gene Ontology Biological processes. Red bars indicate significance of a process at a p-value of 0.05 after false discovery rate correction. Figure 2B represents the subset of the Gene Ontology Biological processes tree where processes include one or more genes encoded by the probe sets in cluster 3. Red coloured processes indicate significance of that process at a p-value of 0.05 after false discovery rate correction. Black coloured processes

include one or more genes encoded by the probe sets in angiogenesis probe set cluster, but are not significant.

[0025] Figure 3 represents the functional enrichment results of the genes within an exemplary 25-gene expression signature that identifies the molecular subtype related to angiogenesis. Red bars indicate significance of a process at a p-value of 0.05 after False Discovery Rate correction.

[0026] Figure 4 provides Kaplan-Meier curves for recurrence-free survival (time-to-event in weeks) following initial surgical resection from patients with high-grade glioblastoma without prior treatment (Phillips HS, Kharbanda S, Chen R, Forrest WF *et al.* "Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis," *CANCER CELL* 2006 Mar;9(3):157-73. PMID: 16530701; Costa BM, Smith JS, Chen Y, Chen J *et al.* "Reversing HOXA9 oncogene activation by PI3K inhibition: epigenetic mechanism and prognostic significance in human glioblastoma," *CANCER RES* 2010 Jan 15;70(2):453-62. PMID: 20068170).

[0027] Figure 5 provides a diagram of a ROC curve of the classification performance of an exemplary 25-gene classifier model within 16 prostate cell-lines following treatment with Dasatinib. The AUC is approximately 0.84 following application of the classifier model. The 95% confidence limits were determined using 1000 bootstrap iterations (Wang XD, Reeves K, Luo FR, Xu LA *et al.* "Identification of candidate predictive and surrogate molecular markers for dasatinib in prostate cancer: rationale for patient selection and efficacy monitoring," *GENOME BIOL* 2007;8(11):R255. PMID: 18047674).

[0028] Figure 6 provides a heatmap representing the hierarchical agglomerative clustering analysis of the most variable genes across 265 serous samples of an epithelial ovarian cancer sample set reclassified according to updated pathological classification criteria. The functional analysis of the probe set clusters is summarized on the right hand side of the image. The legend across the top of the image indicates the classifier group each sample was assigned to for classifier generation (i.e. Class labels) of both "angiogenesis" and "non-angiogenesis" subtypes.

[0029] Figures 7A and 7B represents the functional analysis results of the angiogenesis probe set cluster of the 265 serous only samples in an epithelial ovarian cancer training set using a functional enrichment tool (FET) algorithm. Figure 7A shows a histogram representing the significance of the top 10 enriched Gene Ontology biological processes. Red bars indicate significance of a process at a p-value of 0.05 after False Discovery Rate correction. Figure 7B represents the subset of the Gene Ontology Biological processes tree where processes include one or more genes encoded by the probe sets in cluster 2 (angiogenesis probe set cluster). Red

coloured processes indicate significance of that process at a p-value of 0.05 after false discovery rate correction. Black coloured processes include one or more genes encoded by the probe sets in the angiogenesis cluster, but are not significant.

[0030] Figure. 8 represents the functional enrichment results of the genes within an exemplary 45-gene classifier model that identifies the molecular sub-type related to angiogenesis. Red bars indicate significance of a process at a p-value of 0.05 after False Discovery Rate correction.

[0031] Figure 9 provides a diagram of a ROC curve of the classification performance of the 45-gene classifier model within 16 prostate cell-lines following treatment with Dasatinib. The AUC is ~0.95 following application of the classifier model. The 95% confidence limits were determined using 1000 bootstrap iterations.

[0032] Figure 10 provides Kaplan-Meier curves for progression-free survival (in weeks) for the non-angiogenesis sample group versus the “other” samples not labelled as “non-angiogenesis” sample group in the reclassified ovarian sample set.

[0033] Figures 11A and 11B provide heatmaps representing the hierarchical agglomerative clustering analysis of the most variable genes across 51 ER negative samples (Figure 11A) of a breast cancer sample set and the most variable genes across 56 ER positive samples of a breast cancer sample set (Figure 11B). The functional analysis of the probe set clusters is summarized on the right hand side of the image for those clusters showing vasculature development/angiogenesis or immune response/IFN signalling.

[0034] Figure 12 provides Kaplan-Meier curves for progression-free survival (in weeks) for the samples identified as “non-angiogenesis” using the example 63 gene “non-angiogenesis” signature, versus the “other” samples not classified as “non-angiogenesis” using the example 63 gene signature, in the 265 reclassified ovarian sample set.

[0035] Figure 13 provides Kaplan-Meier curves for progression-free survival (in weeks) for the samples identified as “non-angiogenesis” using the example 63 gene “non-angiogenesis” signature, versus the “other” samples not classified as “non-angiogenesis” using the example 63 gene signature, in the independent 265 epithelial and serous ovarian sample set.

[0036] Figure 14A and 14B are graphs demonstrating certain classification performance benchmarks of an example non-angiogenesis signature as applied to colorectal cancer samples.

[0037] Figure 15 provides a diagram of a ROC curve of the classification performance of the 45-gene classifier model within 16 prostate cell-lines following treatment with Dasatinib. The AUC is ~0.95 following application of the classifier model. The 95% confidence limits were determined using 1000 bootstrap iterations (Wang XD, Reeves K, Luo FR, Xu LA *et al.* “Identification of candidate predictive and surrogate molecular markers for dasatinib in prostate

cancer: rationale for patient selection and efficacy monitoring,” GENOME BIOL 2007;8(11):R255. PMID: 18047674).

[0038] Figure 16 provides a heatmap representing the hierarchical agglomerative clustering analysis of the most variable probe sets across 199 serous samples of the Almac Diagnostics’ epithelial ovarian cancer sample set. The functional analysis of the probe set clusters is summarized on the right hand side of the image. The legend across the top of the image indicates the classifier group each sample was assigned to for classifier generation (i.e. Class labels).

[0039] Figure 17 provides a process flow diagram example Kaplan Meier curves for the use of a non-angiogenesis signature to classify a patient based on prognosis and then an angiogenesis signature to determine responsiveness to a given therapeutic regimen in accordance with example embodiments.

[0040] Figure 18A and B are graphs demonstrating certain classification performance benchmarks of an example 25 angiogenesis signature as applied to colorectal cancer samples.

[0041] Figure 19A and B are graphs demonstrating certain classification performance benchmarks of a 45 gene angiogenesis signature as applied to colorectal cancer samples.

DETAILED DESCRIPTION OF THE INVENTION

[0042] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes IX*, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, *Dictionary of Microbiology and Molecular Biology* 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), and March, *Advanced Organic Chemistry Reactions, Mechanisms and Structure* 4th ed., John Wiley & Sons (New York, N.Y. 1992).

[0043] The singular terms “a,” “an,” and “the” include plural referents unless context clearly indicates otherwise. Similarly, the word “or” is intended to include “and” unless the context clearly indicates otherwise. The term “comprises” means “includes.” In case of conflict, the present specification, including explanations of terms, will control.

[0044] As used herein terms “marker panel,” “expression classifier,” “classifier,” “expression signature,” or “signature” may be used interchangeably.

[0045] All publications, published patent documents, and patent applications cited in this application are indicative of the level of skill in the art(s) to which the application pertains. All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

Overview

[0046] A major goal of current research efforts in cancer is to increase the efficacy of perioperative systemic therapy in patients by incorporating molecular parameters into clinical therapeutic decisions. Pharmacogenetics/genomics is the study of genetic/genomic factors involved in an individuals' response to a foreign compound or drug. Agents or modulators which have a stimulatory or inhibitory effect on expression of a biomarker of the invention can be administered to individuals to treat (prophylactically or therapeutically) cancer in the patient. It is ideal to also consider the pharmacogenomics of the individual in conjunction with such treatment. Differences in metabolism of therapeutics may possibly lead to severe toxicity or therapeutic failure by altering the relationship between dose and blood concentration of the pharmacologically active drug. Thus, understanding the pharmacogenomics of an individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the level of expression of a biomarker of the invention in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

[0047] The present invention relates to a molecular diagnostic tests useful for diagnosing cancers from different anatomical sites that includes the use of one or more common subtypes related to angiogenesis. The invention includes expression signatures that identify a subject as responsive or non-responsive to anti-angiogenic therapeutics and/or having a good or poor clinical prognosis. The expression signature is derived by obtaining the expression profiles of samples from a sample set of known pathology and/or clinical outcome. The samples may originate from the same sample tissue type or different tissue types. As used herein an "expression profile" comprises a set of values representing the expression level for each biomarker analyzed from a given sample.

[0048] The expression profiles from the sample set are then analyzed using a mathematical model. Different mathematical models may be applied and include, but are not limited to, models from the fields of pattern recognition (Duda *et al.* Pattern Classification, 2nd ed., John Wiley, New York 2001), machine learning (Schölkopf *et al.* Learning with Kernels,

MIT Press, Cambridge 2002, Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford 1995), statistics (Hastie *et al.* The Elements of Statistical Learning, Springer, New York 2001), bioinformatics (Dudoit *et al.*, 2002, J. Am. Statist. Assoc. 97:77-87, Tibshirani *et al.*, 2002, Proc. Natl. Acad. Sci. USA 99:6567-6572) or chemometrics (Vandeginste, *et al.*, Handbook of Chemometrics and Qualimetrics, Part B, Elsevier, Amsterdam 1998). The mathematical model identifies one or more biomarkers expressed in the sample set that are most predictive of a given disease phenotype. These one or more biomarkers define an expression signature. Accordingly, an expression signature includes the biomarkers identified as most predictive of a given disease phenotype. In certain exemplary embodiments, the mathematical model defines a variable, such as a weight, for each identified biomarker. In certain exemplary embodiments, the mathematical model defines a decision function. The decision function may further define a threshold score which separates the sample set into two disease phenotypes such as, but not limited to, samples that are responsive and non-responsive to anti-angiogenic therapeutics. In one exemplary embodiment, the decision function and expression signature are defined using a linear classifier.

[0049] To classify new samples using a defined expression signature, the biomarkers defined by the expression signature are isolated and an expression profile of the biomarker(s) determined. The new sample biomarker expression profile is analyzed with the same mathematical model used to define the expression signature. In certain exemplary embodiments, the mathematical model defines an expression score for the new sample. The expression score may be determined by combining the expression values of the biomarkers with corresponding scalar weights using non-linear, algebraic, trigonometric or correlative means to derive a single scalar value. The expression score is compared to the threshold score and the sample classified based on whether the expression score is greater than, or equal to, or less than the threshold score. In one exemplary embodiment, a sample expression value greater than the reference expression value indicates a patient will be responsive to an anti-angiogenic therapeutic. In another exemplary embodiment, a sample expression score below the threshold score indicates the patient will not be responsive to an anti-angiogenic therapeutic. In another exemplary embodiment, a sample expression score below the threshold expression score indicates the patient has a cancer type, or is at risk of developing a cancer type, that is not responsive to an anti-angiogenic therapeutic. In another exemplary embodiment, a sample expression score above the reference expression score indicates the patient has a cancer type, or is at risk of developing a cancer type, that is responsive to an anti-angiogenic therapeutic. In another example embodiment, a sample expression score above the threshold score indicates the patient has a cancer sub-type with a good clinical prognosis. In another example embodiment, a

sample expression score below the threshold score indicates a patient with a cancer subtype with a poor clinical prognosis. Where an expression signature is derived from a tissue sample set comprising one type of cancer tissue, the expression signature is not limited to identifying the same cancer sub-type only in tissues of the same cancer type, but may be used with other cancer types that share the same cancer sub-type. For example, where an expression signature is derived from ovarian cancer samples, the expression signature may be used to identify a similar angiogenesis sub-type in different cancers such as glioblastoma, breast cancer, lung cancer, colon cancer, or prostate cancer.

[0050] One application of the expression signatures disclosed herein is the identification of patients with good and poor prognosis. By examining the expression of a collection of the identified biomarkers in a tumor, it is possible to determine the likely clinical outcomes of a patient. By examining the expression of a collection of biomarkers, it is therefore possible to identify those patients in most need of more aggressive therapeutic regimens and likewise eliminate unnecessary therapeutic treatments or those unlikely to significantly improve a patient's clinical outcome.

[0051] A patient may be considered to have a "good prognosis" where, for example, the survival rate associated with the cancer subtype is greater than a survival rate associated with other related cancer subtypes. In certain embodiments, a "good prognosis" indicates at least an increased expected survival time. This may be based upon a classification as responsive to an anti-angiogenic therapeutic agent as described herein. The increased expected survival time may be as compared to classification as non-responsive to the anti-angiogenic therapeutic agent.

[0052] A patient may be considered to have a "poor prognosis" or "bad prognosis" where, for example, the survival rate associated with the cancer subtype is less than the survival rate associated with other related cancer subtypes.

[0053] Another application of the expression signatures disclosed herein is the stratification of response to, and selection of patients for therapeutic drug classes that encompass anti-angiogenic therapies. By examining the expression of a collection of the identified biomarkers in a tumor, it is possible to determine which therapeutic agent or combination of agents will be most likely to reduce the growth rate of a cancer. It is also possible to determine which therapeutic agent or combination of agents will be the least likely to reduce the growth rate of a cancer. By examining the expression of a collection of biomarkers, it is therefore possible to eliminate ineffective or inappropriate therapeutic agents. Importantly, in certain embodiments, these determinations can be made on a patient-by-patient basis or on an agent-by-agent basis. Thus, one can determine whether or not a particular therapeutic regimen is likely to benefit a particular patient or type of patient, and/or whether a particular regimen should be

continued. The present invention provides a test that can guide therapy selection as well as selecting patient groups for enrichment strategies during clinical trial evaluation of novel therapeutics. For example, when evaluating a putative anti-angiogenic agent or treatment regime, the expression signatures and methods disclosed herein may be used to select individuals for clinical trials that have cancer subtypes that are responsive to anti-angiogenic agents.

[0054] A cancer is "responsive" to a therapeutic agent if its rate of growth is inhibited as a result of contact with the therapeutic agent, compared to its growth in the absence of contact with the therapeutic agent. Growth of a cancer can be measured in a variety of ways. For instance, the size of a tumor or measuring the expression of tumor markers appropriate for that tumor type.

[0055] A cancer is "non-responsive" to a therapeutic agent if its rate of growth is not inhibited, or inhibited to a very low degree, as a result of contact with the therapeutic agent when compared to its growth in the absence of contact with the therapeutic agent. As stated above, growth of a cancer can be measured in a variety of ways, for instance, the size of a tumor or measuring the expression of tumor markers appropriate for that tumor type. The quality of being non-responsive to a therapeutic agent is a highly variable one, with different cancers exhibiting different levels of "non-responsiveness" to a given therapeutic agent, under different conditions. Still further, measures of non-responsiveness can be assessed using additional criteria beyond growth size of a tumor such as, but not limited to, patient quality of life, and degree of metastases.

[0056] Another application of the expression signatures disclosed herein is the combined identification of prognosis of a patient along with identification of a predicted responsiveness of the patient to a given therapeutic. In certain example embodiments, a non-angiogenesis expression signature defined below is used to identify patients with a good prognosis from those with a poor prognosis, and an angiogenesis expression signature is used to predict the patient's response to a given therapeutic class of drugs.

[0057] The angiogenesis subtype can be identified from a fresh/frozen (FF) or formalin fixed paraffin embedded (FFPE) patient sample. In one exemplary embodiment, the cancer type is ovarian cancer, breast cancer, colon cancer, colorectal cancer, lung cancer, prostate cancer, or glioblastoma. In another exemplary embodiment, the cancer type is a ovarian cancer. In a further exemplary embodiment, the cancer type is breast cancer. In another exemplary embodiment, the cancer type is lung cancer. In another exemplary embodiment, the cancer type is colon cancer. In another exemplary embodiment, the cancer type is prostate cancer. In another exemplary embodiment, the cancer type is glioblastoma

Identifying Expression Signatures

[0058] The expression signatures of the present invention are identified by analyzing the expression profiles of certain biomarkers in a patient sample set. Biomarkers suitable for use in the present invention include DNA, RNA, and proteins. The biomarkers are isolated from a patient sample and their expression levels determined to derive a set of expression profiles for each sample analyzed in the patient sample set. In certain example embodiments the expression signature is an angiogenesis expression signature. An angiogenesis expression signature relates to an angiogenesis phenotype observed in cancer tissues, the phenotype characterized by an up-regulation of biomarkers associated with angiogenesis and vascular development. In certain other example embodiments, the expression signature is a non-angiogenesis expression signature. A non-angiogenesis expression signature relates to a phenotype observed in cancer tissues, the phenotype characterized by a down-regulation of biomarkers associated with angiogenesis and vascular development.

a. Expression Profiles

[0059] In certain embodiments, the expression profile obtained is a genomic or nucleic acid expression profile, where the amount or level of one or more nucleic acids in the sample is determined. In these embodiments, the sample that is assayed to generate the expression profile employed in the diagnostic or prognostic methods is one that is a nucleic acid sample. The nucleic acid sample includes a population of nucleic acids that includes the expression information of the phenotype determinative biomarkers of the cell or tissue being analyzed. In some embodiments, the nucleic acid may include RNA or DNA nucleic acids, e.g., mRNA, cRNA, cDNA etc., so long as the sample retains the expression information of the host cell or tissue from which it is obtained. The sample may be prepared in a number of different ways, as is known in the art, e.g., by mRNA isolation from a cell, where the isolated mRNA is used as isolated, amplified, or employed to prepare cDNA, cRNA, etc., as is known in the field of differential gene expression. Accordingly, determining the level of mRNA in a sample includes preparing cDNA or cRNA from the mRNA and subsequently measuring the cDNA or cRNA. The sample is typically prepared from a cell or tissue harvested from a subject in need of treatment, e.g., via biopsy of tissue, using standard protocols, where cell types or tissues from which such nucleic acids may be generated include any tissue in which the expression pattern of the to be determined phenotype exists, including, but not limited to, disease cells or tissue, body fluids, etc.

[0060] The expression profile may be generated from the initial nucleic acid sample using any convenient protocol. While a variety of different manners of generating expression profiles are known, such as those employed in the field of differential gene expression/biomarker

analysis, one representative and convenient type of protocol for generating expression profiles is array-based gene expression profile generation protocols. Such applications are hybridization assays in which a nucleic acid that displays "probe" nucleic acids for each of the genes to be assayed/profiled in the profile to be generated is employed. In these assays, a sample of target nucleic acids is first prepared from the initial nucleic acid sample being assayed, where preparation may include labeling of the target nucleic acids with a label, e.g., a member of a signal producing system. Following target nucleic acid sample preparation, the sample is contacted with the array under hybridization conditions, whereby complexes are formed between target nucleic acids that are complementary to probe sequences attached to the array surface. The presence of hybridized complexes is then detected, either qualitatively or quantitatively. Specific hybridization technology which may be practiced to generate the expression profiles employed in the subject methods includes the technology described in U.S. Pat. Nos. 5,143,854; 5,288,644; 5,324,633; 5,432,049; 5,470,710; 5,492,806; 5,503,980; 5,510,270; 5,525,464; 5,547,839; 5,580,732; 5,661,028; 5,800,992; the disclosures of which are herein incorporated by reference; as well as WO 95/21265; WO 96/31622; WO 97/10365; WO 97/27317; EP 373 203; and EP 785 280. In these methods, an array of "probe" nucleic acids that includes a probe for each of the biomarkers whose expression is being assayed is contacted with target nucleic acids as described above. Contact is carried out under hybridization conditions, e.g., stringent hybridization conditions as described above, and unbound nucleic acid is then removed. The resultant pattern of hybridized nucleic acids provides information regarding expression for each of the biomarkers that have been probed, where the expression information is in terms of whether or not the gene is expressed and, typically, at what level, where the expression data, i.e., expression profile, may be both qualitative and quantitative.

b. Diseases and Sample Tissue Sources

[0061] In certain exemplary embodiments, the patient sample set comprises cancer tissue samples, such as archived samples. The patient sample set is preferably derived from cancer tissue samples having been characterized by prognosis, likelihood of recurrence, long term survival, clinical outcome, treatment response, diagnosis, cancer classification, or personalized genomics profile. As used herein cancer includes, but is not limited to, leukemia, brain cancer, prostate cancer, liver cancer, ovarian cancer, stomach cancer, colorectal cancer, throat cancer, breast cancer, skin cancer, melanoma, lung cancer, sarcoma, cervical cancer, testicular cancer, bladder cancer, endocrine cancer, endometrial cancer, esophageal cancer, glioma, lymphoma, neuroblastoma, osteosarcoma, pancreatic cancer, pituitary cancer, renal cancer, and the like. As used herein, colorectal cancer encompasses cancers that may involve cancer in tissues of both the rectum and other portions of the colon as well as cancers that may be individually classified

as either colon cancer or rectal cancer. In one embodiment, the methods described herein refer to cancers that are treated with anti-angiogenic agents, anti-angiogenic targeted therapies, inhibitors of angiogenesis signaling, but not limited to these classes. These cancers also include subclasses and subtypes of these cancers at various stages of pathogenesis. In certain exemplary embodiments, the patient sample set comprises ovarian cancer samples. In another exemplary embodiment, the patient sample set comprises breast cancer samples. In yet another exemplary embodiment, the patient sample set comprises glioblastoma samples.

[0062] "Biological sample", "sample", and "test sample" are used interchangeably herein to refer to any material, biological fluid, tissue, or cell obtained or otherwise derived from an individual. This includes blood (including whole blood, leukocytes, peripheral blood mononuclear cells, buffy coat, plasma, and serum), sputum, tears, mucus, nasal washes, nasal aspirate, breath, urine, semen, saliva, meningeal fluid, amniotic fluid, glandular fluid, lymph fluid, nipple aspirate, bronchial aspirate, synovial fluid, joint aspirate, ascites, cells, a cellular extract, and cerebrospinal fluid. This also includes experimentally separated fractions of all of the preceding. For example, a blood sample can be fractionated into serum or into fractions containing particular types of blood cells, such as red blood cells or white blood cells (leukocytes). If desired, a sample can be a combination of samples from an individual, such as a combination of a tissue and fluid sample. The term "biological sample" also includes materials containing homogenized solid material, such as from a stool sample, a tissue sample, or a tissue biopsy, for example. The term "biological sample" also includes materials derived from a tissue culture or a cell culture. Any suitable methods for obtaining a biological sample can be employed; exemplary methods include, e.g., phlebotomy, swab (e.g., buccal swab), and a fine needle aspirate biopsy procedure. Samples can also be collected, e.g., by micro dissection (e.g., laser capture micro dissection (LCM) or laser micro dissection (LMD)), bladder wash, smear (e.g., a PAP smear), or ductal lavage. A "biological sample" obtained or derived from an individual includes any such sample that has been processed in any suitable manner after being obtained from the individual, for example, fresh frozen or formalin fixed and/or paraffin embedded. The methods of the invention as defined herein may begin with an obtained sample and thus do not necessarily incorporate the step of obtaining the sample from the patient. The methods may be in vitro methods performed on an isolated sample.

[0063] As used herein, the term "patient" includes human and non-human animals. The preferred patient for treatment is a human. "Patient," "individual" and "subject" are used interchangeably herein.

c. Biomarkers

[0064] As used herein, the term “biomarker” can refer to a gene, an mRNA, cDNA, an antisense transcript, a miRNA, a polypeptide, a protein, a protein fragment, or any other nucleic acid sequence or polypeptide sequence that indicates either gene expression levels or protein production levels. When a biomarker indicates or is a sign of an abnormal process, disease or other condition in an individual, that biomarker is generally described as being either over-expressed or under-expressed as compared to an expression level or value of the biomarker that indicates or is a sign of a normal process, an absence of a disease or other condition in an individual. “Up-regulation”, “up-regulated”, “over-expression”, “over-expressed”, and any variations thereof are used interchangeably to refer to a value or level of a biomarker in a biological sample that is greater than a value or level (or range of values or levels) of the biomarker that is typically detected in similar biological samples from healthy or normal individuals. The terms may also refer to a value or level of a biomarker in a biological sample that is greater than a value or level (or range of values or levels) of the biomarker that may be detected at a different stage of a particular disease.

[0065] “Down-regulation”, “down-regulated”, “under-expression”, “under-expressed”, and any variations thereof are used interchangeably to refer to a value or level of a biomarker in a biological sample that is less than a value or level (or range of values or levels) of the biomarker that is typically detected in similar biological samples from healthy or normal individuals. The terms may also refer to a value or level of a biomarker in a biological sample that is less than a value or level (or range of values or levels) of the biomarker that may be detected at a different stage of a particular disease.

[0066] Further, a biomarker that is either over-expressed or under-expressed can also be referred to as being “differentially expressed” or as having a “differential level” or “differential value” as compared to a “normal” expression level or value of the biomarker that indicates or is a sign of a normal process or an absence of a disease or other condition in an individual. Thus, “differential expression” of a biomarker can also be referred to as a variation from a “normal” expression level of the biomarker.

[0067] The terms “differential biomarker expression” and “differential expression” are used interchangeably to refer to a biomarker whose expression is activated to a higher or lower level in a subject suffering from a specific disease, relative to its expression in a normal subject, or relative to its expression in a patient that responds differently to a particular therapy or has a different prognosis. The terms also include biomarkers whose expression is activated to a higher or lower level at different stages of the same disease. It is also understood that a differentially expressed biomarker may be either activated or inhibited at the nucleic acid level or protein level, or may be subject to alternative splicing to result in a different polypeptide

product. Such differences may be evidenced by a variety of changes including mRNA levels, miRNA levels, antisense transcript levels, or protein surface expression, secretion or other partitioning of a polypeptide. Differential biomarker expression may include a comparison of expression between two or more genes or their gene products; or a comparison of the ratios of the expression between two or more genes or their gene products; or even a comparison of two differently processed products of the same gene, which differ between normal subjects and subjects suffering from a disease; or between various stages of the same disease. Differential expression includes both quantitative, as well as qualitative, differences in the temporal or cellular expression pattern in a biomarker among, for example, normal and diseased cells, or among cells which have undergone different disease events or disease stages.

[0068] In certain exemplary embodiments, the biomarker is an RNA transcript. As used herein “RNA transcript” refers to both coding and non-coding RNA, including messenger RNAs (mRNA), alternatively spliced mRNAs, ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNAs (snRNA), and antisense RNA. Measuring mRNA in a biological sample may be used as a surrogate for detection of the level of the corresponding protein and gene in the biological sample. Thus, any of the biomarkers or biomarker panels described herein can also be detected by detecting the appropriate RNA. Methods of biomarker expression profiling include, but are not limited to quantitative PCR, NGS, northern blots, southern blots, microarrays, SAGE, immunoassays (ELISA, EIA, agglutination, nephelometry, turbidimetry, Western blot, immunoprecipitation, immunocytochemistry, flow cytometry, Luminex assay), and mass spectrometry. The overall expression data for a given sample may be normalized using methods known to those skilled in the art in order to correct for differing amounts of starting material, varying efficiencies of the extraction and amplification reactions.

[0069] In certain exemplary embodiments, biomarkers useful for distinguishing between cancer types that are responsive and non-responsive to anti-angiogenic therapeutics can be determined by identifying biomarkers exhibiting the highest degree of variability between samples in the patient data set as determined using the expression detection methods and patient sample sets discussed above. Standard statistical methods known in the art for identifying highly variable data points in expression data may be used to identify the highly variable biomarkers. For example, a combined background and variance filter to the patient data set. The background filter is based on the selection of probe sets with expression E and expression variance var_E above the thresholds defined by background standard deviation σ_{Bg} (from the Expression Console software) and quantile of the standard normal distribution z_α at a specified significance α probe sets were kept if:

$$E > \log_2((z_\alpha \sigma_{Bg})); \log_2((\text{var}_E) > 2 [\log_2(\sigma_{Bg}) - E - \log_2(\log(2))])$$

[0070] where a defines a significance threshold. In certain exemplary embodiment, the significance threshold is $6.3 \cdot 10^{-5}$. In another exemplary embodiment, the significance threshold may be between $1.0 \cdot 10^{-7}$ to $1.0 \cdot 10^{-3}$.

[0071] In certain exemplary embodiments, the highly variable biomarkers may be further analyzed to group samples in the patient data set into subtypes or clusters based on similar gene expression profiles. For examples, biomarkers may be clustered based on how highly correlated the up-regulation or down-regulation of their expression is to one another. Various clustering analysis techniques known in the art may be used. In one exemplary embodiment, hierarchical agglomerative clustering is used to identify the cancer subtypes. To determine the biological relevance of each subtype, the biomarkers within each cluster may be further mapped to their corresponding genes and annotated by cross-reference to one or more gene ontology databases containing information on biological activity and biological pathways associated with the gene. In one exemplary embodiment, biomarker in clusters enriched for angiogenesis, vasculature development and immune response general functional terms are grouped into a putative angiogenesis sample group and used for expression signature generation. In another exemplary embodiment, biomarkers in clusters that are up regulated and enriched for angiogenesis, vasculature development and immune response general functional terms are grouped into a putative angiogenesis sample group and used for expression signature generation. In another exemplary embodiment, biomarkers in clusters that are down regulated and enriched for angiogenesis, vasculature development and immune response general functional terms are grouped into a putative angiogenesis sample group and used for expression signature generation. Further details for conducting functional analysis of biomarker clusters is provided in the Examples section below.

[0072] In one exemplary embodiment, the biomarkers useful in deriving an expression signature for use in the present invention are those biomarkers listed in Table 1A, Table 1B, Table 1C, or a combination thereof. These biomarkers are identified as having predictive value to determine a patient response to a therapeutic agent and/or a prognostic value in identifying individuals with a good or poor clinical prognosis.

[0073] In certain example embodiments, their expression correlates with the response, or lack thereof, to an agent, and more specifically, an anti-angiogenic therapeutic agent. By examining the expression of a collection of the identified biomarkers in a tumor, it is possible to determine which therapeutic agent or combination of agents will be most likely to reduce the growth rate of a cancer. By examining a collection of identified biomarkers in a tumor, it is also possible to determine which therapeutic agent or combination of agents will be the least likely to reduce the growth rate of a cancer. By examining the expression of a collection of biomarkers,

it is therefore possible to eliminate ineffective or inappropriate therapeutic agents. Importantly, in certain embodiments, these determinations can be made on a patient-by-patient basis or on an agent-by-agent basis. Thus, one can determine whether or not a particular therapeutic regimen is likely to benefit a particular patient or type of patient, and/or whether a particular regimen should be continued.

[0074] In certain other example embodiments, the expression of the biomarkers disclosed herein correlated with a patient's overall clinical prognosis. By examining the expression of a collection of biomarkers identified in a tumor, it is possible to determine whether the individual has a cancer subtype associated with good clinical prognosis or poor clinical prognosis. Importantly, in certain embodiments, these determinations can be made on a patient-by-patient basis. Thus, one of ordinary skill in the art can use predicted prognosis to help select appropriate treatment regimens to treat the underlying disease while eliminating those treatment regimens most likely to produce undesired or medically unwarranted adverse side effects.

[0075] The SEQ ID NOs listed in Table 1A, Table 1B, and Table 1C refer to probe set identifiers used to measure the expression levels of the genes on an exemplary transcriptome array. Expression signatures of the present invention have been cross-validated using expression data from different arrays with different probe sets as detailed further in the Examples section below. Accordingly, the expression signatures and methods disclosed herein are not limited to expression values measured using the probe sets disclosed herein.

Table 1A: Genes in Clusters of Figure 1

SEQ NO:	Orientation	Gene Symbol
1	Sense	STAT1
2	Sense	PDGFC
3	Sense	TGFB3
4	Sense	RAC2
5	Sense	MARCKS
6	Sense	ALOX5
7	Sense	COL8A1
8	Sense	CTSS
9	Sense	HAVCR2
10	Sense	RAB31
11	Sense	KCNAB2
12	Sense	THBS1
13	Sense	SH3BP4

14	Sense	CTGF
15	Sense	CTGF
16	Sense	VCAN
17	Sense	IGKC
18	Sense	IGKC
19	Sense	IGKC
20	Sense	SGK1
21	Sense	NFATC1
22	Sense	HMHA1
23	Sense	FCGR1C /// FCGR1A /// FCGR1B
24	Sense	EDA2R
25	Sense	COL8A1
26	Sense	COL12A1
27	Sense	HLA-B
28	Sense	HLA-F
29	Sense	HLA-C
30	Sense	SPP1
31	Sense	MYO1F
32	Sense	SPARC
33	Sense	SPARC
34	Sense	UBD /// GABBR1
35	Sense	CCND1
36	Sense	COL1A1
37	Sense	EGR1
38	Sense	EGR1
39	Sense	TNFSF10
40	Sense	SULF2
41	Sense	CERCAM
42	Sense	ATF3
43	Sense	MIR21
44	Sense	BASP1
45	Sense	IFIT2
46	Sense	SULF1

47	Sense	IGLC2 /// IGLC3
48	Sense	IGLC2 /// IGLC3
49	Sense (Fully0Exonic)	IGLC2 /// IGLC3
50	Sense	IGLC2 /// IGLC3
51	Sense	IGLC2 /// IGLC3
52	Sense	IGLC1
53	Sense	IGLC1
54	Sense	IGLC2 /// IGLC3
55	Sense	ANGPTL2
56	Sense	COL5A2
57	Sense	IGJ
58	Sense	THY1
59	Sense	NDN
60	Sense	RGS2
61	Sense	MEIS3P1 /// MEIS3P2
62	Sense	GBP2
63	Sense	CSF1R
64	Sense	C1R
65	Sense	FAT1
66	Sense	COL1A1
67	Sense	RHOB
68	Sense	MMP11
69	Sense	GADD45B
70	Sense	MMP14
71	Sense	MMP14
72	Sense	IGHG4
73	Sense	DDX60L
74	Sense	SPP1
75	Sense	ROR2
76	Sense	CTSK
77	Sense	FCGR2B
78	Sense	PTAFR
79	Sense	ICAM1
80	Sense	HCLS1

81	No Transcript Match	---
82	Sense	SLFN11
83	No Transcript Match	---
84	Sense	JAM3
85	Sense	TMEM49
86	Sense	TMEM49
87	Sense	LTBP2
88	Sense	IRS1
89	Sense	COL5A2
90	Sense	C17orf91
91	Sense	GPNMB
92	Sense	FAM198B
93	Sense	MICAL2
94	Sense	TMEM2
95	Sense	CHST15
96	Sense	SECTM1
97	Sense	DCN
98	Sense	VCAM1
99	Sense	TNFAIP3
100	Sense	C1QA
101	Sense	C1QA
102	Sense	FBXO32
103	Sense	COL12A1
104	Sense	CPE
105	Sense	CIITA
106	Sense	GAS7
107	Sense	COL3A1
108	Sense	FN1
109	Sense	IFI30
110	Sense	ITGB2
111	Sense	ELN
112	Sense	CMTM3
113	Sense	ANTXR1
114	Sense	ARHGDIB

115	Sense	LAPTM5
116	Sense	SOX4
117	Sense	IFI44L
118	Sense	IL4I1
119	Sense	ANTXR2
120	Sense	IGLC2 /// IGLC3
121	Sense	EPSTI1
122	Sense	BIRC3
123	Sense	IGLC2 /// IGLC3
124	Sense	BST2
125	Sense	TNFSF10
126	Sense	COL10A1
127	Sense	IGLC2 /// IGLC3
128	Sense	FBP1
129	Sense	RHOBTB3
130	Sense	CDK6
131	Sense	CD74
132	Sense	ISM1
133	Sense	C1QC
134	Sense	BIN2
135	Sense	CSRNP1
136	Sense	TYROBP
137	Sense	C1QTNF3
138	Sense	DCN
139	Sense	IGFBP4
140	Sense	AOAH
141	Sense	SIRPA
142	Sense	FOSB
143	Sense	CCDC80
144	Sense	IGLC1
145	Sense	HCST
146	Sense	IFI35
147	Sense	BIRC3
148	Sense	COL3A1

149	Sense	IFITM2
150	Sense	ZFP36
151	Sense	MMP11
152	Sense	COL1A2
153	Sense	HLA-DPA1
154	Sense	TWIST1
155	Sense	ZNF154
156	Sense	EGR1
157	Sense	IGLC2 /// IGLC3
158	Sense	TNFSF10
159	Sense	IGKC
160	Sense	IGHG1 /// IGHG4
161	Sense	GBP5
162	Sense	COL1A2
163	Sense	APOC1
164	No Transcript Match	---
165	Sense	COL3A1
166	AntiSense	PXDN
167	AntiSense	EGR1
168	Sense	GBP3
169	Sense	ISG15
170	Sense	---
171	Sense	KIAA0146
172	Sense	CMAH
173	Sense	APBB2
174	Sense	TPM1
175	No Transcript Match	---
176	Sense	DMD
177	No Genome Match	---
178	Sense	IL10RA
179	Sense	---
180	No Transcript Match	---
181	Sense	DUSP1
182	Sense	GBP1

183	Sense	PARVG
184	Sense	MAFF
185	Sense	PDGFC
186	Sense	MSN
187	Sense	RSAD2
188	Sense	TPM1
189	Sense	EMB
190	Sense	C6orf155
191	Sense	FOS
192	Sense	DEXI
193	Sense	RNF19A
194	Sense	FBXO32
195	Sense	DPYSL3
196	Sense	PRICKLE1
197	AntiSense	EGR1
198	AntiSense	NRP2
199	Sense	B2M
200	AntiSense	MIR21
201	Sense	MMP2
202	Sense	CDR1
203	Sense	HLA-B
204	Sense	CTGF
205	Sense	DCN
206	Sense	SOD2
207	Sense	FN1
208	Sense	COL8A2
209	Sense	SGK1
210	Sense	TIMP3
211	Sense	ACTA2
212	Sense	SRGN
213	Sense	LOXL1
214	Sense	CCR1
215	Sense	GBP1
216	Sense	CDH11

217	Sense	FCGR3A
218	Sense	LUM
219	Sense	NNMT
220	Sense	COL1A2
221	Sense	RGS1
222	Sense	GJA1
223	Sense	SPARCL1
224	Sense	DAB2
225	AntiSense	CTHRC1
226	Sense	RGS16
227	Sense	FBLN1
228	Sense	SPP1
229	Sense	CTSB
230	Sense	SPP1
231	Sense	SDC1
232	Sense	PLAU
233	Sense	PDGFRA
234	Sense	SERPINF1
235	Sense	BGN
236	Sense	COL6A3
237	AntiSense	C3
238	AntiSense	C3
239	AntiSense	SPP1
240	AntiSense	HLA-DQA1
241	AntiSense	GAS1
242	Sense	VCAN
243	AntiSense	---
244	Sense	IGHG4 /// IGHG2 /// IGHG1 /// IGHGP
245	Sense	IGHG2
246	Sense	C3orf26
247	AntiSense	ATF3
248	AntiSense	ATF3
249	AntiSense	SULF1

250	Sense	FN1
251	AntiSense	CALD1
252	AntiSense	CALD1
253	Sense	TMEM49
254	Sense	TMEM49
255	Sense	CHD5
256	AntiSense	EGR1
257	AntiSense	SNAI2
258	AntiSense	ITPRIPL2
259	AntiSense	GABBR1 /// UBD
260	AntiSense	GABBR1 /// UBD
261	AntiSense	TWIST1
262	AntiSense	TWIST1
263	AntiSense	BATF2
264	AntiSense	NFKBIZ
265	Sense	C3orf26
266	AntiSense	LOXL1
267	Sense	---
268	AntiSense	TIMP2
269	AntiSense	FN1
270	AntiSense	COL1A1
271	AntiSense	DCN
272	Sense	TREH
273	AntiSense	UBE2L6
274	AntiSense	APOL1
275	AntiSense	BIRC3
276	AntiSense	BIRC3
277	Sense	LILRB4
278	Sense	FGD2
279	Sense	TMEM49
280	Sense	NCF4
281	Sense	COL10A1
282	Sense	GAL3ST4
283	Sense	HCK

284	Sense	TAGLN
285	Sense	TWIST1
286	Sense	HCLS1
287	Sense	LPAR6
288	Sense	ITGB2
289	Sense	LST1
290	Sense	HLA-B
291	Sense	C17orf91
292	Sense	ZC3H12A
293	Sense	KLF10
294	Sense	BASP1
295	Sense	BASP1

Table 1B: Genes in Clusters of Figure 6

SEQ NO:	Orientation	Gene Symbol
2	Sense (Fully Exonic)	PDGFC
3	Sense (Fully Exonic)	TGFB3
4	Sense (Fully Exonic)	RAC2
5	Sense (Fully Exonic)	MARCKS
6	Sense (Fully Exonic)	ALOX5
7	Sense (Fully Exonic)	COL8A1
11	Sense (Fully Exonic)	KCNAB2
12	Sense (Fully Exonic)	THBS1
14	Sense (Fully Exonic)	CTGF
15	Sense (Fully Exonic)	CTGF
16	Sense (Fully Exonic)	VCAN
18	Sense (Fully Exonic)	IGKC
19	Sense (Fully Exonic)	IGKC
21	Sense (includes Intronic)	NFATC1
22	Sense (Fully Exonic)	HMHA1
23	Sense (Fully Exonic)	FCGR1B
24	Sense (Fully Exonic)	EDA2R
25	Sense (Fully Exonic)	COL8A1
26	Sense (Fully Exonic)	COL12A1
27	Sense (Fully Exonic)	HLA-B
28	Sense	HLA-F
37	Sense (Fully Exonic)	EGR1

40	Sense (Fully Exonic)	SULF2
41	Sense (Fully Exonic)	CERCAM
42	Sense (Fully Exonic)	ATF3
43	Sense (Fully Exonic)	MIR21
45	Sense (Fully Exonic)	IFIT2
47	Sense (Fully Exonic)	IGLC3
48	Sense (Fully Exonic)	IGLC3
49	Sense (Fully Exonic)	IGLC3
50	Sense (Fully Exonic)	IGLC3
51	Sense (Fully Exonic)	IGLC3
54	Sense (Fully Exonic)	IGLC3
55	Sense (Fully Exonic)	ANGPTL2
56	Sense (Fully Exonic)	COL5A2
58	Sense (Fully Exonic)	THY1
59	Sense (Fully Exonic)	NDN
60	Sense (Fully Exonic)	RGS2
61	Sense (Fully Exonic)	MEIS3P2
62	Sense (Fully Exonic)	GBP2
65	Sense (Fully Exonic)	FAT1
66	Sense (Fully Exonic)	COL1A1
68	Sense (Fully Exonic)	MMP11
69	Sense (Fully Exonic)	GADD45B
71	Sense (Fully Exonic)	MMP14
72	Sense (Fully Exonic)	IGHG4
80	Sense (Fully Exonic)	HCLS1
83	No Transcript match	
84	Sense (Fully Exonic)	JAM3
86	Sense (Fully Exonic)	TMEM49
87	Sense (Fully Exonic)	LTBP2
88	Sense (Fully Exonic)	IRS1
90	Sense (Fully Exonic)	C17orf91
91	Sense (Fully Exonic)	GPNUMB
92	Sense (Fully Exonic)	FAM198B
95	Sense (Fully Exonic)	CHST15
97	Sense (Fully Exonic)	DCN
98	Sense (Fully Exonic)	VCAM1
105	Sense (Fully Exonic)	CIITA
106	Sense (Fully Exonic)	GAS7
107	Sense (Fully Exonic)	COL3A1

110	Sense (Fully Exonic)	ITGB2
111	Sense (Fully Exonic)	ELN
112	Sense (Fully Exonic)	CMTM3
113	Sense (Fully Exonic)	ANTXR1
118	Sense (Fully Exonic)	IL4I1
119	Sense (Fully Exonic)	ANTXR2
120	Sense (Fully Exonic)	IGLC2 /// IGLC3
123	Sense (Fully Exonic)	IGLC3
124	Sense (Fully Exonic)	BST2
126	Sense (Fully Exonic)	COL10A1
127	Sense (Fully Exonic)	IGLC3
128	Sense (Fully Exonic)	FBP1
129	Sense (Fully Exonic)	RHOBTB3
131	Sense (Fully Exonic)	CD74
132	Sense (Fully Exonic)	ISM1
135	Sense (Fully Exonic)	CSRNP1
138	Sense (Fully Exonic)	DCN
139	Sense (Fully Exonic)	IGFBP4
143	Sense (Fully Exonic)	CCDC80
148	Sense (Fully Exonic)	COL3A1
150	Sense (Fully Exonic)	ZFP36
151	Sense (Fully Exonic)	MMP11
152	Sense (Fully Exonic)	COL1A2
153	Sense (Fully Exonic)	HLA-DPA1
154	Sense (Fully Exonic)	TWIST1
155	Sense (Fully Exonic)	ZNF154
157	Sense (Fully Exonic)	IGLC3
159	Sense (Fully Exonic)	IGKC
160	Sense (Fully Exonic)	IGHG1
162	Sense (Fully Exonic)	COL1A2
163	Sense (Fully Exonic)	APOC1
167	AntiSense	EGR1
171	Sense (Fully Exonic)	KIAA0146
174	Sense (Fully Exonic)	TPM1
176	Sense (includes Intronic)	DMD
180	No Transcript match	
181	Sense (Fully Exonic)	DUSP1
182	Sense (Fully Exonic)	GBP1
185	Sense (includes Intronic)	PDGFC
186	Sense (includes Intronic)	MSN
188	Sense (includes Intronic)	TPM1
189	Sense (Fully Exonic)	EMB

191	Sense (Fully Exonic)	FOS
195	Sense (includes Intronic)	DPYSL3
197	AntiSense	EGR1
198	AntiSense	NRP2
201	Sense (Fully Exonic)	MMP2
204	Sense (Fully Exonic)	CTGF
211	Sense (Fully Exonic)	ACTA2
213	Sense (Fully Exonic)	LOXL1
216	Sense (Fully Exonic)	CDH11
218	Sense (Fully Exonic)	LUM
219	Sense (Fully Exonic)	NNMT
222	Sense (Fully Exonic)	GJA1
225	AntiSense	CTHRC1
229	Sense (Fully Exonic)	CTSB
232	Sense (Fully Exonic)	PLAU
233	Sense (Fully Exonic)	PDGFRA
242	Sense (Fully Exonic)	VCAN
243	AntiSense	--
244	Sense (Fully Exonic)	IGHG4 /// IGHG2 /// IGHG1 ///I GHGP
245	Sense (Fully Exonic)	IGHG2
246	Sense (includes Intronic)	C3orf26
247	AntiSense	ATF3
248	AntiSense	ATF3
250	Sense (Fully Exonic)	FN1
251	AntiSense	CALD1
252	AntiSense	CALD1
256	AntiSense	EGR1
261	AntiSense	TWIST1
262	AntiSense	TWIST1
263	AntiSense	BATF2
264	AntiSense	NFKBIZ
265	Sense (includes Intronic)	C3orf26
266	AntiSense	LOXL1
267	Sense (includes Intronic)	--
269	AntiSense	FN1
270	AntiSense	COL1A1
272	Sense (Fully Exonic)	TREH
274	AntiSense	APOL1
281	Sense (Fully Exonic)	COL10A1
282	Sense (Fully Exonic)	GAL3ST4

284	Sense (Fully Exonic)	TAGLN
285	Sense (Fully Exonic)	TWIST1
286	Sense (Fully Exonic)	HCLS1
288	Sense (Fully Exonic)	ITGB2
290	Sense (Fully Exonic)	HLA-B
291	Sense (Fully Exonic)	C17orf91
296	Sense (Fully Exonic)	FBLIM1
297	Sense (Fully Exonic)	COL15A1
298	Sense (Fully Exonic)	AQP7P3
299	AntiSense	IGFBP5
300	Sense (Fully Exonic)	FANK1
301	AntiSense	INS
302	Sense (Fully Exonic)	COL27A1
303	Sense (Fully Exonic)	COL5A1
304	Sense (Fully Exonic)	PRICKLE2
305	Sense (Fully Exonic)	N/A
306	Sense (Fully Exonic)	GXYLT2
307	Sense (includes Intronic)	KLF12
308	No Transcript match	
309	Sense (Fully Exonic)	FBXO32
310	No Transcript match	
311	Sense (Fully Exonic)	ASAH2B
312	AntiSense	PPFIBP1
313	AntiSense	XIST
314	Sense (Fully Exonic)	IGFBP6
315	Sense (Fully Exonic)	ROBO1
316	Sense (Fully Exonic)	TPM1
317	AntiSense	N/A
318	AntiSense	PLEKHG1
319	Sense (Fully Exonic)	NR2F1
320	Sense (Fully Exonic)	NPDC1
321	AntiSense	INS
322	Sense (Fully Exonic)	TRAF5
323	Sense (Fully Exonic)	CALD1
324	Sense (includes Intronic)	CHRM3
325	Sense (Fully Exonic)	AMOTL1
326	Sense (includes Intronic)	COL12A1
327	Sense (Fully Exonic)	PLXNA4
328	Sense (includes Intronic)	TMEM43

329	Sense (includes Intronic)	RORA
330	AntiSense	INS
331	Sense (Fully Exonic)	TSPAN18
332	No Transcript match	
333	Sense (Fully Exonic)	TNC
334	Sense (Fully Exonic)	TYRO3
335	AntiSense	EFNA5
336	Sense (Fully Exonic)	MYL9
337	Sense (Fully Exonic)	MIR198
338	Sense (includes Intronic)	N/A
339	Sense (includes Intronic)	PLA2R1
340	Sense (Fully Exonic)	COL14A1
341	Sense (Fully Exonic)	NRP1
342	Sense (Fully Exonic)	FSCN1
343	Sense (includes Intronic)	PDGFD
344	No Transcript match	
345	Sense (includes Intronic)	DOCK4
346	Sense (Fully Exonic)	TRIM13
347	Sense (Fully Exonic)	IGFBP5
348	Sense (Fully Exonic)	C19orf63
349	AntiSense	KLF6
350	AntiSense	TRIO
351	Sense (Fully Exonic)	COL4A1
352	Sense (Fully Exonic)	EPDR1
353	Sense (Fully Exonic)	FNDC1
354	Sense (Fully Exonic)	IL1R1
355	Sense (Fully Exonic)	CES4
356	Sense (Fully Exonic)	GPR176
357	Sense (includes Intronic)	GXYLT2
358	AntiSense	WHSC1L1
359	Sense (Fully Exonic)	N/A
360	Sense (Fully Exonic)	RGN
361	Sense (includes Intronic)	CA3
362	Sense (Fully Exonic)	TIMP3
363	Sense (Fully Exonic)	EFNA5
364	Sense (Fully Exonic)	RASGRF2
365	Sense (includes Intronic)	RELL1
366	AntiSense	ACSS3
367	Sense (Fully Exonic)	STMN3

368	Sense (Fully Exonic)	N/A
369	AntiSense	C7orf29
370	Sense (Fully Exonic)	HOXC6
371	Sense (Fully Exonic)	KLF8
372	Sense (includes Intronic)	SERINC5
373	Sense (Fully Exonic)	AKT3
374	Sense (Fully Exonic)	TGFB2
375	AntiSense	WNT5A
376	No Transcript match	
377	No Transcript match	
378	AntiSense	IGFBP7
379	No Transcript match	
380	Sense (includes Intronic)	SULT1C4
381	Sense (Fully Exonic)	AASS
382	Sense (Fully Exonic)	HEPH
383	Sense (Fully Exonic)	ADH5
384	Sense (Fully Exonic)	TIMP2
385	Sense (Fully Exonic)	EMP1
386	Sense (Fully Exonic)	CXCL14
387	Sense (Fully Exonic)	ZNF548
388	Sense (Fully Exonic)	SGCB
389	Sense (includes Intronic)	ASH2L
390	Sense (includes Intronic)	SERINC5
391	No Genome match	
392	Sense (Fully Exonic)	TMEM159
393	Sense (includes Intronic)	RBMS3
394	Sense (Fully Exonic)	TMEM49
395	Sense (includes Intronic)	RORA
396	No Transcript match	
397	AntiSense	ZNF608
398	No Genome match	
399	Sense (Fully Exonic)	ADAMTS2
400	Sense (Fully Exonic)	APCDD1
401	AntiSense	GXYLT2
402	Sense (Fully Exonic)	XIST
403	Sense (Fully Exonic)	MBNL2
404	Sense (Fully Exonic)	SHF
405	Sense (includes Intronic)	APBB2
406	No Transcript match	

407	Sense (Fully Exonic)	COL14A1
408	Sense (Fully Exonic)	IGFBP5
409	Sense (Fully Exonic)	CREB5
410	AntiSense	INS
411	Sense (Fully Exonic)	BAHCC1
412	Sense (Fully Exonic)	RFXAP
413	Sense (Fully Exonic)	INS
414	Sense (Fully Exonic)	DDR2
415	Sense (Fully Exonic)	CA12
416	Sense (Fully Exonic)	RHOB
417	Sense (Fully Exonic)	N/A
418	Sense (Fully Exonic)	SNORD116-4
419	Sense (Fully Exonic)	MEG3
420	Sense (Fully Exonic)	WNT4
421	Sense (Fully Exonic)	FBLN2
422	AntiSense	DAAM1
423	No Transcript match	
424	Sense (Fully Exonic)	CHN1
425	Sense (includes Intronic)	APBB2
426	Sense (Fully Exonic)	PTRF
427	AntiSense	IGF1
428	Sense (Fully Exonic)	UST
429	Sense (Fully Exonic)	SMARCA1
430	Sense (includes Intronic)	N/A
431	Sense (Fully Exonic)	IGLC3
432	AntiSense	INS
433	Sense (Fully Exonic)	KANK4
434	AntiSense	IGF1
435	Sense (Fully Exonic)	CYP27A1
436	AntiSense	EIF2B5
437	No Transcript match	
438	Sense (Fully Exonic)	SNRNP25
439	Sense (Fully Exonic)	SETD7
440	Sense (Fully Exonic)	MSX1
441	Sense (Fully Exonic)	HOPX
442	Sense (Fully Exonic)	NID2
443	Sense (Fully Exonic)	IGF1
444	Sense (Fully Exonic)	PSD3
445	Sense (Fully Exonic)	FGFR1

446	Sense (Fully Exonic)	ETV1
447	Sense (Fully Exonic)	ZNF655
448	No Genome match	
449	AntiSense	INS
450	Sense (Fully Exonic)	SFRP2
451	Sense (Fully Exonic)	SPAG16
452	AntiSense	NR2F2
453	Sense (includes Intronic)	SYNPO2
454	Sense (Fully Exonic)	FAM101B
455	AntiSense	IGF2
456	Sense (Fully Exonic)	CA3
457	Sense (Fully Exonic)	XIST
458	No Transcript match	
459	Sense (Fully Exonic)	WNT7A
460	Sense (includes Intronic)	N/A
461	Sense (Fully Exonic)	FGFR1
462	AntiSense	FXVD6
463	Sense (Fully Exonic)	FGFR1
464	Sense (includes Intronic)	IGFBP7
465	Sense (Fully Exonic)	TIMP2
466	Sense (Fully Exonic)	DUSP1
467	Sense (includes Intronic)	SERINC5
468	No Transcript match	
469	Sense (Fully Exonic)	ABLIM1
470	Sense (Fully Exonic)	ARL4A
471	AntiSense	SH3TC2
472	AntiSense	NR2F2
473	Sense (Fully Exonic)	ENG
474	Sense (Fully Exonic)	MGP
475	Sense (Fully Exonic)	MEG3
476	AntiSense	FAM115A
477	Sense (Fully Exonic)	EGR1
478	Sense (Fully Exonic)	SNORD116-3
479	Sense (Fully Exonic)	AEBP1
480	Sense (includes Intronic)	SDK1
481	Sense (Fully Exonic)	ENC1
482	Sense (Fully Exonic)	SNORD116-7
483	Sense (Fully Exonic)	N/A
484	Sense (Fully Exonic)	APOD

485	AntiSense	N/A
486	AntiSense	GAS1
487	Sense (Fully Exonic)	VPS36
488	No Transcript match	
489	Sense (Fully Exonic)	SPHK2
490	Sense (Fully Exonic)	SNORD116-8
491	Sense (Fully Exonic)	MYO10
492	Sense (Fully Exonic)	HOXC6
493	Sense (Fully Exonic)	RNF149
494	Sense (Fully Exonic)	BTG2
495	Sense (includes Intronic)	MAP3K1
496	Sense (Fully Exonic)	SNORD116-23
497	Sense (includes Intronic)	ACSL4
498	Sense (Fully Exonic)	CYP27C1
499	Sense (includes Intronic)	COL12A1
500	Sense (Fully Exonic)	IGFBP5
501	Sense (Fully Exonic)	DUSP4
502	Sense (Fully Exonic)	PFKFB3
503	Sense (Fully Exonic)	SDC2
504	AntiSense	FXYP6
505	Sense (Fully Exonic)	COL5A1
506	Sense (Fully Exonic)	MARCKS
507	Sense (Fully Exonic)	IRS2
508	Sense (Fully Exonic)	N/A
509	AntiSense	FSCN1
510	Sense (Fully Exonic)	FYN
511	Sense (Fully Exonic)	IGFBP5
512	Sense (Fully Exonic)	NUDT4P1
513	Sense (Fully Exonic)	NFKBIZ
514	Sense (Fully Exonic)	N/A
515	Sense (Fully Exonic)	C7orf41
516	Sense (Fully Exonic)	MEG3
517	Sense (Fully Exonic)	N/A
518	Sense (Fully Exonic)	PLEKHG1
519	Sense (Fully Exonic)	ZNF827
520	Sense (Fully Exonic)	ZNF175
521	Sense (Fully Exonic)	XIST
522	Sense (includes Intronic)	GSN
523	Sense (includes Intronic)	RORA

524	Sense (Fully Exonic)	CA13
525	AntiSense	TMX4
526	Sense (Fully Exonic)	KIT
527	Sense (includes Intronic)	WDR78
528	Sense (Fully Exonic)	ECEL1
529	Sense (Fully Exonic)	XIST
530	Sense (Fully Exonic)	PROCR
531	Sense (Fully Exonic)	C9orf167
532	Sense (Fully Exonic)	MUC6
533	Sense (includes Intronic)	P4HA2
534	Sense (Fully Exonic)	FAM69C
535	Sense (Fully Exonic)	NOX4
536	Sense (includes Intronic)	N/A
537	No Transcript match	
538	Sense (Fully Exonic)	SMOX
539	Sense (Fully Exonic)	KIAA0922
540	No Transcript match	
541	Sense (Fully Exonic)	XIST
542	Sense (Fully Exonic)	NPAS2
543	Sense (Fully Exonic)	NAV1
544	Sense (includes Intronic)	N/A
545	Sense (Fully Exonic)	HLA-A
546	Sense (Fully Exonic)	FAM46C
547	Sense (Fully Exonic)	N/A
548	Sense (Fully Exonic)	SLAMF7
549	Sense (Fully Exonic)	FCER1G
550	Sense (Fully Exonic)	C1S
551	Sense (Fully Exonic)	NUPR1
552	AntiSense	C1QC
553	AntiSense	SAT1
554	Sense (Fully Exonic)	SOD2
555	Sense (Fully Exonic)	IRF1
556	Sense (Fully Exonic)	SFN
557	AntiSense	LTB
558	Sense (Fully Exonic)	ARID5A
559	Sense (Fully Exonic)	BST2
560	Sense (Fully Exonic)	HLA-F
561	Sense (Fully Exonic)	XAF1
562	Sense (Fully Exonic)	TCOF1

563	Sense (Fully Exonic)	RPL23AP1
564	Sense (Fully Exonic)	IL1RN
565	Sense (Fully Exonic)	IFIT5
566	Sense (Fully Exonic)	B2M
567	AntiSense	GBP1
568	Sense (Fully Exonic)	HLA-F
569	Sense (Fully Exonic)	DGKA
570	Sense (Fully Exonic)	XBP1
571	Sense (Fully Exonic)	PLCG2
572	Sense (Fully Exonic)	FAM46C
573	No Genome match	
574	Sense (Fully Exonic)	TREM2
575	Sense (Fully Exonic)	LGALS9
576	Sense (Fully Exonic)	HLA-DPB1
577	AntiSense	ODF3B
578	Sense (Fully Exonic)	MX1
579	Sense (Fully Exonic)	STAT1
580	Sense (Fully Exonic)	CTSB
581	Sense (Fully Exonic)	FAM26F
582	Sense (includes Intronic)	PARP14
583	AntiSense	SAT1
584	Sense (Fully Exonic)	CTSS
585	No Transcript match	
586	Sense (Fully Exonic)	CTSB
587	Sense (Fully Exonic)	ADAM8
588	Sense (includes Intronic)	B2M
589	Sense (Fully Exonic)	FLVCR2
590	Sense (Fully Exonic)	TYROBP
591	AntiSense	SAMD9L
592	Sense (Fully Exonic)	SAMD9L
593	Sense (Fully Exonic)	SIGLEC1
594	Sense (Fully Exonic)	MMP7
595	Sense (Fully Exonic)	APOL1
596	Sense (Fully Exonic)	CYLD
597	Sense (Fully Exonic)	HLA-B
598	Sense (Fully Exonic)	SAT1
599	Sense (Fully Exonic)	C1QB
600	Sense (Fully Exonic)	HLA-DMB
601	Sense (Fully Exonic)	NLRC5

602	Sense (Fully Exonic)	FAM20A
603	AntiSense	N/A
604	Sense (Fully Exonic)	STAT1
605	Sense (includes Intronic)	STAT1
606	Sense (Fully Exonic)	STAT1
607	AntiSense	N/A
608	Sense (Fully Exonic)	DERL3
609	Sense (Fully Exonic)	HLA-F
610	Sense (Fully Exonic)	MAFB
611	Sense (Fully Exonic)	CD4
612	Sense (Fully Exonic)	HLA-A
613	Sense (Fully Exonic)	UBE2L6
614	Sense (Fully Exonic)	C1QC
615	Sense (Fully Exonic)	CD163
616	Sense (Fully Exonic)	LRMP
617	Sense (Fully Exonic)	C11orf17
618	Sense (Fully Exonic)	XAF1
619	Sense (Fully Exonic)	GLRX
620	Sense (Fully Exonic)	IFIH1
621	Sense (Fully Exonic)	CD44
622	Sense (Fully Exonic)	LITAF
623	Sense (Fully Exonic)	CCDC69
624	Sense (Fully Exonic)	GBP5
625	Sense (Fully Exonic)	PML
626	Sense (Fully Exonic)	SAMD9
627	Sense (Fully Exonic)	CBR3
628	Sense (Fully Exonic)	RASGRP2
629	Sense (Fully Exonic)	FCGR2A
630	Sense (Fully Exonic)	BST2
631	Sense (Fully Exonic)	HLA-A

Table 1C: Angiogenesis and immune response cluster genes of Figs 11A and 11B

SEQ No.	ER group	Gene Symbol	Orientation
632	ER pos	STAT1	Sense (Fully Exonic)
709	ER pos	STAT1	Sense (Fully Exonic)
803	ER neg	HOXD3	Sense (includes Intronic)
804	ER neg	LCP1	Sense (Fully Exonic)
875	ER neg	SPON1	Sense (Fully Exonic)

633	ER pos	IGF1	Sense (Fully Exonic)
805	ER neg	FHL1	Sense (Fully Exonic)
806	ER neg	MS4A1	Sense (Fully Exonic)
876	ER neg	FN1	Sense (Fully Exonic)
710	ER pos	FN1	Sense (Fully Exonic)
634	ER pos	HLA-DPA1	Sense (Fully Exonic)
877	ER neg	ADAM12	Sense (Fully Exonic)
711	ER pos	ITGBL1	Sense (Fully Exonic)
712	ER pos	POSTN	Sense (Fully Exonic)
635	ER pos	IGKV3-20 /// IGKV3D-20	AntiSense
748	ER neg,ER pos	DDX58	Sense (Fully Exonic)
636	ER pos	SLAMF7	Sense (Fully Exonic)
878	ER neg	AGPAT4	Sense (Fully Exonic)
959	ER neg,ER pos	FAM46C	Sense (Fully Exonic)
713	ER pos	GXYLT2	Sense (includes Intronic)
714	ER pos	SNX5	Sense (Fully Exonic)
637	ER pos		No Genome match
879	ER neg	PRICKLE1	Sense (Fully Exonic)
922	ER neg,ER pos	IL7R	Sense (Fully Exonic)
750	ER neg,ER pos	CMPK2	Sense (Fully Exonic)
638	ER pos	KLHL36	Sense (Fully Exonic)
807	ER neg	CD274	Sense (Fully Exonic)
880	ER neg	NRP2	AntiSense
924	ER neg,ER pos	POSTN	Sense (includes Intronic)
639	ER pos	CAMK2N1	AntiSense
881	ER neg	COL1A2	Sense (Fully Exonic)
715	ER pos	N/A	AntiSense
716	ER pos		No Transcript match
717	ER pos	COL12A1	Sense (Fully Exonic)
640	ER pos	MIR143	Sense (Fully Exonic)
808	ER neg		No Genome match
960	ER neg,ER pos	DNM3	AntiSense
718	ER pos	BNC2	Sense (includes Intronic)

641	ER pos	BCL11B	Sense (includes Intronic)
809	ER neg	CYBB	Sense (Fully Exonic)
882	ER neg	COL12A1	Sense (Fully Exonic)
642	ER pos	OR2I1P	AntiSense
752	ER neg,ER pos	EPSTI1	Sense (Fully Exonic)
926	ER neg,ER pos	GXYLT2	Sense (Fully Exonic)
643	ER pos	SAMD9L	Sense (Fully Exonic)
810	ER neg	SLC2A3	Sense (Fully Exonic)
927	ER neg,ER pos	IGKJ3 /// IGKJ4	Sense (Fully Exonic)
644	ER pos	THBS1	Sense (Fully Exonic)
811	ER neg	GLUL	Sense (Fully Exonic)
812	ER neg		No Transcript match
813	ER neg	THSD7A	Sense (includes Intronic)
928	ER neg,ER pos	COL11A1	Sense (Fully Exonic)
645	ER pos		No Transcript match
814	ER neg	SPP1	AntiSense
815	ER neg	TIMP3	Sense (Fully Exonic)
883	ER neg	FBN1	Sense (Fully Exonic)
646	ER pos	IGFBP5	Sense (Fully Exonic)
816	ER neg	IKZF1	Sense (Fully Exonic)
817	ER neg	CTGF	AntiSense
884	ER neg	COL5A2	AntiSense
647	ER pos	PALMD	Sense (Fully Exonic)
818	ER neg	IGKV4-1	Sense (Fully Exonic)
756	ER neg,ER pos	N/A	Sense (Fully Exonic)
648	ER pos	NLRC3	Sense (Fully Exonic)
819	ER neg	TNFRSF13B	Sense (Fully Exonic)
885	ER neg	APOC1	Sense (Fully Exonic)
649	ER pos	ELN	Sense (Fully Exonic)
757	ER neg,ER pos		No Genome match
650	ER pos	ELN	Sense (Fully Exonic)
758	ER neg,ER pos	IGKC	Sense (Fully Exonic)
651	ER pos	CIITA	Sense (Fully Exonic)

788	ER neg,ER pos	CCR7	Sense (Fully Exonic)
719	ER pos	STAT1	Sense (Fully Exonic)
720	ER pos	SDC1	Sense (Fully Exonic)
721	ER pos	ISG15	Sense (Fully Exonic)
722	ER pos	APOC1	Sense (Fully Exonic)
652	ER pos	INS /// IGF2	Sense (Fully Exonic)
820	ER neg	LPL	Sense (Fully Exonic)
962	ER neg,ER pos	IFI27	Sense (Fully Exonic)
723	ER pos	CTSK	Sense (Fully Exonic)
724	ER pos	PLEK	Sense (Fully Exonic)
725	ER pos	MIR198	Sense (Fully Exonic)
759	ER neg,ER pos	IGJ	AntiSense
653	ER pos	PXDN	Sense (Fully Exonic)
790	ER neg,ER pos	ENO2	Sense (Fully Exonic)
726	ER pos	CILP	Sense (Fully Exonic)
726	ER neg,ER pos	GBP4	Sense (Fully Exonic)
654	ER pos	GBP1	Sense (Fully Exonic)
886	ER neg	SULF1	Sense (Fully Exonic)
655	ER pos	EFEMP1	Sense (Fully Exonic)
761	ER neg,ER pos	MX1	Sense (Fully Exonic)
656	ER pos	CD109	Sense (Fully Exonic)
887	ER neg		No Transcript match
762	ER neg,ER pos	CXCL10	Sense (Fully Exonic)
657	ER pos	FYB	Sense (Fully Exonic)
791	ER neg,ER pos	IFI44	Sense (Fully Exonic)
727	ER pos	COL5A1	Sense (Fully Exonic)
658	ER pos	TIMP3	Sense (Fully Exonic)
888	ER neg	CADM1	AntiSense
659	ER pos	STAT1	AntiSense
889	ER neg	MXRA5	AntiSense
660	ER pos	CCDC152	Sense (Fully Exonic)
821	ER neg	MX1	AntiSense
890	ER neg	C1QC	AntiSense

792	ER neg,ER pos	N/A	Sense (includes Intronic)
728	ER pos	FNDC1	AntiSense
661	ER pos	HOXC5 /// HOXC10	AntiSense
822	ER neg	GABBR1 /// UBD	AntiSense
823	ER neg	CYP1B1	AntiSense
824	ER neg	SPP1	Sense (Fully Exonic)
824	ER neg,ER pos	FYB	Sense (Fully Exonic)
662	ER pos		No Transcript match
825	ER neg	GBP5	Sense (Fully Exonic)
937	ER neg,ER pos	IGHA1	Sense (Fully Exonic)
663	ER pos	CAMK2N1	Sense (Fully Exonic)
826	ER neg	SASH3	Sense (Fully Exonic)
827	ER neg	COL14A1	Sense (Fully Exonic)
793,	ER neg,ER pos	CD79A /// ARHGEF1	Sense (Fully Exonic)
729	ER pos	LAIR1	Sense (Fully Exonic)
664	ER pos	APOBEC3G	Sense (Fully Exonic)
828	ER neg	GPX8	Sense (Fully Exonic)
829	ER neg	RTP4	Sense (Fully Exonic)
830	ER neg	IFIH1	Sense (Fully Exonic)
891	ER neg	IFI6	Sense (Fully Exonic)
730	ER pos	IFI6	Sense (Fully Exonic)
665	ER pos	C1orf113	Sense (Fully Exonic)
831	ER neg	FAM26F	Sense (Fully Exonic)
832	ER neg	PARP9	Sense (Fully Exonic)
833	ER neg	ARHGAP9	Sense (Fully Exonic)
765	ER neg,ER pos	FNDC1	Sense (Fully Exonic)
666	ER pos	C13orf33	Sense (Fully Exonic)
766	ER neg,ER pos	GBP5	Sense (Fully Exonic)
667	ER pos	VNN2	Sense (Fully Exonic)
892	ER neg	RSAD2	Sense (Fully Exonic)
668	ER pos	PTPRC	Sense (Fully Exonic)
893	ER neg	LRRC15	Sense (Fully Exonic)
731	ER pos	CTHRC1	Sense (Fully Exonic)
669	ER pos	SMARCA1	Sense (Fully Exonic)

894	ER neg	STAT1	Sense (Fully Exonic)
767	ER neg,ER pos	IGJ	Sense (Fully Exonic)
670	ER pos	DACT3	Sense (Fully Exonic)
834	ER neg	PHLDB2	Sense (Fully Exonic)
967	ER neg,ER pos	IL2RG	Sense (Fully Exonic)
732	ER pos	C1orf162	Sense (Fully Exonic)
768	ER neg,ER pos	CECR1	Sense (Fully Exonic)
671	ER pos	BTN3A2 /// BTN3A3	Sense (Fully Exonic)
835	ER neg	CES1 /// CES4	Sense (Fully Exonic)
836	ER neg	TBC1D10C	Sense (Fully Exonic)
837	ER neg	CCDC80	Sense (Fully Exonic)
895	ER neg	GAS7	Sense (Fully Exonic)
672	ER pos	LPL	Sense (Fully Exonic)
896	ER neg	CD163	Sense (Fully Exonic)
795	ER neg,ER pos	LYZ	Sense (Fully Exonic)
733	ER pos	NAPSB	Sense (Fully Exonic)
673	ER pos	OAS2	Sense (Fully Exonic)
769	ER neg,ER pos	ZNF469	Sense (Fully Exonic)
674	ER pos	KIF26A	Sense (Fully Exonic)
770	ER neg,ER pos	IGHG1 /// IGHG2 /// IGHG3	Sense (Fully Exonic)
675	ER pos	N/A	Sense (Fully Exonic)
838	ER neg	DDX60L	Sense (Fully Exonic)
839	ER neg	SCD5	Sense (Fully Exonic)
840	ER neg	ANPEP	Sense (Fully Exonic)
841	ER neg	FABP4	Sense (Fully Exonic)
771	ER neg,ER pos	IL10RA	Sense (Fully Exonic)
676	ER pos	CXCL10	Sense (Fully Exonic)
897	ER neg	APOC1	Sense (Fully Exonic)
677	ER pos	APOD	Sense (Fully Exonic)
842	ER neg	APOD	Sense (Fully Exonic)
843	ER neg	DST	Sense (Fully Exonic)
772	ER neg,ER pos	CD2	Sense (Fully Exonic)
678	ER pos	CD22	Sense (Fully Exonic)

844	ER neg	CD52	Sense (Fully Exonic)
898	ER neg	CTSK	Sense (Fully Exonic)
773,	ER neg,ER pos	CCR7	Sense (Fully Exonic)
679	ER pos	CTGF	Sense (Fully Exonic)
899	ER neg	EDN2	Sense (Fully Exonic)
680	ER pos	EGR1	Sense (Fully Exonic)
845	ER neg	ENO2	Sense (Fully Exonic)
900	ER neg	IFI6	Sense (Fully Exonic)
681	ER pos	HLA-DOA	Sense (Fully Exonic)
846	ER neg	TNC	Sense (Fully Exonic)
847	ER neg	IDO1	Sense (Fully Exonic)
848	ER neg	ID1	Sense (Fully Exonic)
901	ER neg	IRF1	Sense (Fully Exonic)
774,	ER neg,ER pos	ITGAL	Sense (Fully Exonic)
682	ER pos	KRT81	Sense (Fully Exonic)
902	ER neg	LUM	Sense (Fully Exonic)
734	ER pos	MFAP2	Sense (Fully Exonic)
683	ER pos	CXCL9	Sense (Fully Exonic)
775,	ER neg,ER pos	MMP13	Sense (Fully Exonic)
684	ER pos	MSX1	Sense (Fully Exonic)
903	ER neg	MX1	Sense (Fully Exonic)
685	ER pos	TWIST1	Sense (Fully Exonic)
904	ER neg	OAS2	Sense (Fully Exonic)
686	ER pos	PIK3CG	Sense (Fully Exonic)
905	ER neg	PLAU	Sense (Fully Exonic)
735	ER pos	COL10A1	Sense (Fully Exonic)
687	ER pos	PSMB9	Sense (Fully Exonic)
849	ER neg	FBP1	Sense (Fully Exonic)
850	ER neg	PTN	Sense (Fully Exonic)
776	ER neg,ER pos	RAC2	Sense (Fully Exonic)
688	ER pos	CCL5	Sense (Fully Exonic)
851	ER neg	CCL5	Sense (Fully Exonic)
852	ER neg	CCL21	Sense (Fully Exonic)
906	ER neg	THBS2	Sense (Fully Exonic)
689	ER pos	TAP1	Sense (Fully Exonic)

853	ER neg	IGFBP5	Sense (Fully Exonic)
854	ER neg	IGF2	Sense (Fully Exonic)
855	ER neg	SELL	Sense (Fully Exonic)
777,	ER neg,ER pos	GJB2	Sense (Fully Exonic)
690	ER pos	F2RL2	Sense (Fully Exonic)
907	ER neg	VCAN	Sense (Fully Exonic)
691	ER pos	ALDH1A3	Sense (Fully Exonic)
856	ER neg	EGR3	Sense (Fully Exonic)
796	ER neg,ER pos	MMP2	Sense (Fully Exonic)
736	ER pos	MMP2	Sense (Fully Exonic)
778	ER neg,ER pos	SPOCK1	Sense (Fully Exonic)
797	ER neg,ER pos	SPOCK1	Sense (Fully Exonic)
737	ER pos	COL1A1	Sense (Fully Exonic)
692	ER pos	SPARCL1	Sense (Fully Exonic)
857	ER neg	CD3D	Sense (Fully Exonic)
798	ER neg,ER pos	CD3E	Sense (Fully Exonic)
738	ER pos	COL3A1	Sense (Fully Exonic)
693	ER pos	CXCL14	Sense (Fully Exonic)
908	ER neg	ADAMTS4	Sense (Fully Exonic)
779	ER neg,ER pos	NCKAP1L	Sense (Fully Exonic)
694	ER pos	CALCOCO2	Sense (Fully Exonic)
909	ER neg	MMP11	Sense (Fully Exonic)
695	ER pos	TRIM22	Sense (Fully Exonic)
910	ER neg	OAS3	Sense (Fully Exonic)
739	ER pos	OAS3	Sense (Fully Exonic)
696	ER pos	POU2AF1	Sense (Fully Exonic)
780	ER neg,ER pos	GABBR1 /// UBD	Sense (Fully Exonic)
697	ER pos	FOSB	Sense (Fully Exonic)
781	ER neg,ER pos	IFI44L	Sense (Fully Exonic)
698	ER pos	PIM2	Sense (Fully Exonic)
858	ER neg	KLRK1	Sense (Fully Exonic)
859	ER neg	TAGLN	Sense (Fully Exonic)
860	ER neg	CD36	Sense (Fully Exonic)

861	ER neg	CYP27C1	Sense (Fully Exonic)
862	ER neg	PDLIM3	Sense (Fully Exonic)
911	ER neg	RNF144A	Sense (Fully Exonic)
782	ER neg,ER pos	SPOCK2	Sense (Fully Exonic)
699	ER pos	FAM19A5	Sense (Fully Exonic)
863	ER neg	ABI3BP	Sense (Fully Exonic)
912	ER neg	TMEM158	Sense (Fully Exonic)
700	ER pos	CCDC69	Sense (Fully Exonic)
913	ER neg	C1QA	Sense (Fully Exonic)
701	ER pos	ETV7	Sense (Fully Exonic)
914	ER neg	N/A	Sense (Fully Exonic)
702	ER pos	USP18	Sense (Fully Exonic)
915	ER neg	XAF1	Sense (Fully Exonic)
703	ER pos	PARP14	Sense (Fully Exonic)
916	ER neg	SAMD9	Sense (Fully Exonic)
740	ER pos	ASPN	Sense (Fully Exonic)
704	ER pos	FAM46C	Sense (Fully Exonic)
864	ER neg	PINX1	Sense (Fully Exonic)
865	ER neg	GIMAP4	Sense (Fully Exonic)
866	ER neg	N/A	Sense (includes Intronic)
917	ER neg	COL3A1	Sense (Fully Exonic)
741	ER pos	COL1A2	Sense (Fully Exonic)
799	ER neg,ER pos	IGHG1	AntiSense
742	ER pos	FN1	AntiSense
800	ER neg,ER pos	RAC2	Sense (Fully Exonic)
743	ER pos	GLRX	Sense (Fully Exonic)
744	ER pos	BIRC3	AntiSense
783	ER neg,ER pos	TAP1	Sense (Fully Exonic)
957	ER neg,ER pos	POSTN	Sense (Fully Exonic)
705	ER pos	THBS1	AntiSense
867	ER neg	TNFAIP8	AntiSense
958	ER neg,ER pos	N/A	Sense (Fully Exonic)
974	ER neg,ER pos	IGJ	Sense (Fully Exonic)

745	ER pos	ADAM12	Sense (Fully Exonic)
706	ER pos	COL14A1	Sense (Fully Exonic)
868	ER neg		No Transcript match
869	ER neg	CD53	Sense (Fully Exonic)
918	ER neg	MIR1245	Sense (Fully Exonic)
707	ER pos		No Genome match
919	ER neg		No Transcript match
708	ER pos	TNFSF10	Sense (Fully Exonic)
870	ER neg	NLRC5	Sense (Fully Exonic)
871	ER neg	WWC2	Sense (includes Intronic)
872	ER neg	RORA	AntiSense
873	ER neg	IGKV4-1	AntiSense
874	ER neg		No Genome match
975	ER neg,ER pos	MIAT	Sense (Fully Exonic)
746	ER pos	COL8A1	Sense (Fully Exonic)
747	ER neg,ER pos	IFI44	Sense (Fully Exonic)

In certain exemplary embodiments, all or a portion of the biomarkers recited in Table 1A, Table 1B and Table 1C, may be used in an expression signature. For example, expression signatures comprising the biomarkers in Table 1A, Table 1B, and Table 1C can be generated using the methods provided herein and can comprise between one, and all of the markers set forth in Tables 1A, 1B, 1C and each and every combination in between (e.g., four selected markers, 16 selected markers, 74 selected markers, etc.). In some embodiments, the expression signature comprises at least 5, 10, 20, 40, 60, 100, 150, 200, or 300 or more markers. In other embodiments, the predictive biomarker panel comprises no more than 5, 10, 20, 40, 60, 100, 150, 200, 300, 400, 500, 600 or 700 markers. In one exemplary embodiment, the expression signature includes a plurality of markers listed in Table 1A. In another exemplary embodiment, the expression signature includes a plurality of biomarkers listing in Table 1B. In another exemplary embodiment, the expression signature includes a plurality of biomarkers listed in Table 1C. In yet another exemplary embodiment, the expression signature includes a plurality of biomarkers listed in Table 1A, Table 1B, and Table 1C. In some embodiments the expression signature includes at least about 1%, about 5%, about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 95%, about 96%, about 97%, about 98%, or about 99% of the markers listed in Table 1A, Table 1B, Table 1C, or a combination thereof. Selected expression signatures can be assembled from the biomarkers

provided using methods described herein and analogous methods known in the art. In one embodiment, the expression signature contains all 250 genes or gene products in Table 1A. In another exemplary embodiment, the expression signature contains all 486 genes or gene products in Table 1B. In another exemplary embodiment, the expression signature contains all 343 genes or gene products in Table 1C.

4. Mathematical Models

[0076] The following methods may be used to derive expression signatures for distinguishing between subjects that are responsive or non-responsive to anti-angiogenic therapeutics, or as prognostic indicators of certain cancer types, including expression signatures derived from the biomarkers disclosed above. In certain other exemplary embodiments, the expression signature is derived using a decision tree (Hastie *et al.* The Elements of Statistical Learning, Springer, New York 2001), a random forest (Breiman, 2001 Random Forests, Machine Learning 45:5), a neural network (Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford 1995), discriminant analysis (Duda *et al.* Pattern Classification, 2nd ed., John Wiley, New York 2001), including, but not limited to linear, diagonal linear, quadratic and logistic discriminant analysis, a Prediction Analysis for Microarrays (PAM, (Tibshirani *et al.*, 2002, Proc. Natl. Acad. Sci. USA 99:6567-6572)) or a Soft Independent Modeling of Class Analogy analysis. (SIMCA, (Wold, 1976, Pattern Recogn. 8:127-139)).

[0077] Biomarker expression values may be defined in combination with corresponding scalar weights on the real scale with varying magnitude, which are further combined through linear or non-linear, algebraic, trigonometric or correlative means into a single scalar value via an algebraic, statistical learning, Bayesian, regression, or similar algorithms which together with a mathematically derived decision function on the scalar value provide a predictive model by which expression profiles from samples may be resolved into discrete classes of responder or non-responder, resistant or non-resistant, to a specified drug, drug class, or treatment regimen. Such predictive models, including biomarker membership, are developed by learning weights and the decision threshold, optimized for sensitivity, specificity, negative and positive predictive values, hazard ratio or any combination thereof, under cross-validation, bootstrapping or similar sampling techniques, from a set of representative expression profiles from historical patient samples with known drug response and/or resistance.

[0078] In one embodiment, the biomarkers are used to form a weighted sum of their signals, where individual weights can be positive or negative. The resulting sum ("expression score") is compared with a pre-determined reference point or value. The comparison with the reference point or value may be used to diagnose, or predict a clinical condition or outcome.

[0079] As described above, one of ordinary skill in the art will appreciate that the biomarkers included in the classifier provided in Tables 1A, 1B, and Table 1C will carry unequal weights in a classifier for responsiveness or resistance to a therapeutic agent. Therefore, while as few as one sequence may be used to diagnose or predict an outcome such as responsiveness to therapeutic agent, the specificity and sensitivity or diagnosis or prediction accuracy may increase using more sequences.

[0080] As used herein, the term “weight” refers to the relative importance of an item in a statistical calculation. The weight of each biomarker in a gene expression classifier may be determined on a data set of patient samples using analytical methods known in the art. As used herein the term “bias” or “offset” refers to a constant term derived using the mean expression of the signature genes in a training set and is used to mean-center the each gene analyzed in the test dataset.

[0081] In certain exemplary embodiments, the expression signature is defined by a decision function. A decision function is a set of weighted expression values derived using a linear classifier. All linear classifiers define the decision function using the following equation:

$$f(x) = w' \cdot x + b = \sum w_i \cdot x_i + b \quad (1)$$

[0082] All measurement values, such as the microarray gene expression intensities x_i , for a certain sample are collected in a vector x . Each intensity is then multiplied with a corresponding weight w_i to obtain the value of the decision function $f(x)$ after adding an offset term b . In deriving the decision function, the linear classifier will further define a threshold value that splits the gene expression data space into two disjoint halves. Exemplary linear classifiers include but are not limited to partial least squares (PLS), (Nguyen *et al.*, Bioinformatics 18 (2002) 39-50), support vector machines (SVM) (Schölkopf *et al.*, Learning with Kernels, MIT Press, Cambridge 2002), and shrinkage discriminant analysis (SDA) (Ahdesmäki *et al.*, Annals of applied statistics 4, 503-519 (2010)). In one exemplary embodiment, the linear classifier is a PLS linear classifier.

[0083] The decision function is empirically derived on a large set of training samples, for example from patients showing responsiveness or resistance to a therapeutic agent. The threshold separates a patient group based on different characteristics such as, but not limited to, responsiveness/non-responsiveness to treatment. The interpretation of this quantity, i.e. the cut-off threshold responsiveness or resistance to a therapeutic agent, is derived in the development phase (“training”) from a set of patients with known outcome. The corresponding weights and the responsiveness/resistance cut-off threshold for the decision score are fixed *a priori* from training data by methods known to those skilled in the art. In one exemplary embodiment, Partial Least Squares Discriminant Analysis (PLS-DA) is used for determining the weights. (L.

Ståhle, S. Wold, J. Chemom. 1 (1987) 185-196; D. V. Nguyen, D.M. Rocke, Bioinformatics 18 (2002) 39-50).

[0084] Effectively, this means that the data space, i.e. the set of all possible combinations of biomarker expression values, is split into two mutually exclusive groups corresponding to different clinical classifications or predictions, for example, one corresponding to responsiveness to a therapeutic agent and the other to non-responsiveness. In the context of the overall classifier, relative over-expression of a certain biomarker can either increase the decision score (positive weight) or reduce it (negative weight) and thus contribute to an overall decision of, for example, responsiveness or resistance to a therapeutic agent.

[0085] In certain exemplary embodiments of the invention, the data is transformed non-linearly before applying a weighted sum as described above. This non-linear transformation might include increasing the dimensionality of the data. The non-linear transformation and weighted summation might also be performed implicitly, for example, through the use of a kernel function. (Schölkopf *et al.* Learning with Kernels, MIT Press, Cambridge 2002).

[0086] In certain exemplary embodiments, the patient training set data is derived by isolated RNA from a corresponding cancer tissue sample set and determining expression values by hybridizing the isolated RNA to a microarray. In certain exemplary embodiments, the microarray used in deriving the expression signature is a transcriptome array. As used herein a “transcriptome array” refers to a microarray containing probe sets that are designed to hybridize to sequences that have been verified as expressed in the diseased tissue of interest. Given alternative splicing and variable poly-A tail processing between tissues and biological contexts, it is possible that probes designed against the same gene sequence derived from another tissue source or biological context will not effectively bind to transcripts expressed in the diseased tissue of interest, leading to a loss of potentially relevant biological information. Accordingly, it is beneficial to verify what sequences are expressed in the disease tissue of interest before deriving a microarray probe set. Verification of expressed sequences in a particular disease context may be done, for example, by isolating and sequencing total RNA from a diseased tissue sample set and cross-referencing the isolated sequences with known nucleic acid sequence databases to verify that the probe set on the transcriptome array is designed against the sequences actually expressed in the diseased tissue of interest. Methods for making transcriptome arrays are described in United States Patent Application Publication No. 2006/0134663, which is incorporated herein by reference. In certain exemplary embodiments, the probe set of the transcriptome array is designed to bind within 300 nucleotides of the 3' end of a transcript. Methods for designing transcriptome arrays with probe sets that bind within 300 nucleotides of the 3' end of target transcripts are disclosed in United States Patent Application

Publication No. 2009/0082218, which is incorporated by reference herein. In certain exemplary embodiments, the microarray used in deriving the gene expression profiles of the present invention is the Almac Ovarian Cancer DSA™ microarray (Almac Group, Craigavon, United Kingdom).

[0087] An optimal linear classifier can be selected by evaluating a linear classifier's performance using such diagnostics as "area under the curve" (AUC). AUC refers to the area under the curve of a receiver operating characteristic (ROC) curve, both of which are well known in the art. AUC measures are useful for comparing the accuracy of a classifier across the complete data range. Linear classifiers with a higher AUC have a greater capacity to classify unknowns correctly between two groups of interest (e.g., ovarian cancer samples and normal or control samples). ROC curves are useful for plotting the performance of a particular feature (e.g., any of the biomarkers described herein and/or any item of additional biomedical information) in distinguishing between two populations (e.g., individuals responding and not responding to a therapeutic agent). Typically, the feature data across the entire population (e.g., the cases and controls) are sorted in ascending order based on the value of a single feature. Then, for each value for that feature, the true positive and false positive rates for the data are calculated. The true positive rate is determined by counting the number of cases above the value for that feature and then dividing by the total number of cases. The false positive rate is determined by counting the number of controls above the value for that feature and then dividing by the total number of controls. Although this definition refers to scenarios in which a feature is elevated in cases compared to controls, this definition also applies to scenarios in which a feature is lower in cases compared to the controls (in such a scenario, samples below the value for that feature would be counted). ROC curves can be generated for a single feature as well as for other single outputs, for example, a combination of two or more features can be mathematically combined (e.g., added, subtracted, multiplied, etc.) to provide a single sum value, and this single sum value can be plotted in a ROC curve. Additionally, any combination of multiple features, in which the combination derives a single output value, can be plotted in a ROC curve. These combinations of features may comprise a test. The ROC curve is the plot of the true positive rate (sensitivity) of a test against the false positive rate (1-specificity) of the test.

[0088] In one exemplary embodiment an angiogenesis expression signature is directed to the 25 biomarkers detailed in Table 2A with corresponding ranks, weights and associated bias detailed in the table or alternative rankings, weightings and bias, depending, for example, on the disease setting. In another exemplary embodiment, an angiogenesis expression signature is directed to the 45 biomarkers detailed in Table 2B with corresponding ranks, weights and

associated bias detailed in the table or alternative rankings, weightings and bias, depending, for example, on the disease setting. In another exemplary embodiment, a non-angiogenesis expression signature is directed to the 63 biomarkers detailed in Table 2C with corresponding ranks, and weights and associated bias detailed in the table or alternative rankings, and weightings and bias, depending, for example, on the disease setting. Tables 2A, 2B and 2C rank the biomarkers in order of absolute decreasing weight, in an example classifier, in the compound decision score function.

Table 2A

25 Biomarker Signature			
Rank	Gene Symbol	Weight	Bias
1	CCDC80	0.0584	5.9594
2	INHBA	0.0508	3.8146
3	THBS2	0.0504	3.4342
4	SFRP2	0.0437	4.3293
5	MMP2	0.0367	5.3319
6	PLAU	-0.0323	8.3844
7	FAP	0.0300	3.4689
8	FN1	0.0277	10.0610
9	COL8A1	-0.0248	5.1984
10	RAB31	0.0244	5.6121
11	FAM38B	0.0242	5.0883
12	VCAN	0.0230	4.4460
13	GJB2	0.0223	5.8677
14	ITGA5	0.0216	5.4455
15	CRISPLD2	0.0192	6.5088
16	C17orf91	0.0167	6.4830
17	BGN	-0.0142	5.5396
18	TIMP3	0.0130	4.6739
19	ALPK2	0.0123	7.3709
20	LUM	0.0104	8.3279
21	NKD2	-0.0098	5.4448
22	LOX	-0.0082	8.2117
23	MIR1245	0.0059	4.9959
24	LOXL1	0.0052	2.9463
25	CXCL12	0.0048	9.7132

Table 2B

45 Biomarker Signature			
Rank	Gene Symbol	Weight	Bias
1	TMEM200A	0.0595	3.6813
2	GJB2	0.0560	4.4798
3	MMP13	0.0383	3.7241
4	GFPT2	0.0380	4.8602
5	POSTN	-0.0355	4.3599
6	BICC1	0.0304	3.6982
7	CDH11	0.0283	4.9968
8	MRVI1	0.0256	5.0761
9	PMP22	0.0240	5.5645
10	COL11A1	-0.0237	3.5002
11	IGFL2	0.0222	3.3104
12	LUM	-0.0220	8.3363
13	NTM	-0.0218	4.2302
14	BGN	0.0211	10.1524
15	COL3A1	-0.0210	8.3236
16	COL10A1	0.0197	6.3538
17	RAB31	0.0180	5.3171
18	ANGPTL2	0.0166	5.6396
19	PLAU	0.0166	5.8488
20	COL8A1	0.0164	6.4193
21	MIR1245	0.0153	5.4552
22	POLD2	0.0146	9.3878
23	NKD2	0.0145	7.3717
24	FZD1	0.0143	4.1519
25	COPZ2	0.0139	5.1039
26	ITGA5	0.0136	8.3663
27	VGLL3	0.0125	4.5019
28	INHBA	-0.0118	4.6843
29	MMP14	0.0110	10.0841
30	VCAN	0.0100	5.5518
31	THBS2	-0.0087	8.1309
32	RUNX2	0.0083	4.7345
33	TIMP3	0.0081	6.4983
34	SFRP2	-0.0079	5.6017
35	COL1A2	0.0078	6.0100
36	COL5A2	-0.0072	3.5671
37	SERPINF1	0.0068	10.8334
38	KIF26B	-0.0052	4.9782
39	TNFAIP6	0.0050	5.3618
40	MMP2	0.0040	5.3622
41	FN1	0.0031	4.9840
42	ALPK2	0.0024	3.5136
43	CTSK	0.0015	5.7322
44	LOXL1	-0.0014	9.5939
45	FAP	0.0000	5.2254

Table 2C

63 Biomarker Signature			
Rank	Gene Symbol	Weight	Bias
1	IGF2	-0.01737	9.8884
2	SOX11	-0.01457	4.5276
3	INS	-0.01409	7.0637
4	CXCL17	0.012568	4.8478
5	SLC5A1	0.012426	4.8920
6	TMEM45A	-0.0124	6.1307
7	CXCR2P1	0.011427	3.1478
8	MFAP2	-0.01039	9.0516
9	MATN3	-0.01028	3.7313
10	RTP4	0.010052	4.9852
11	COL3A1	-0.01002	8.4130
12	CDR1	-0.00916	8.1778
13	RARRES3	0.009056	6.8964
14	TNFSF10	0.008876	6.2325
15	NUAK1	-0.0087	6.6771
16	SNORD114-14	-0.00864	5.6385
17	SRPX	-0.00862	5.0850
18	SPARC	-0.00848	6.0135
19	GJB1	0.008445	5.8142
20	TIMP3	-0.00823	6.5937
21	ISLR	-0.0079	8.9876
22	TUBA1A	-0.00754	9.6540
23	DEXI	0.007271	5.5913
24	BASP1	-0.00724	8.4396
25	PXDN	-0.00724	7.7570
26	GBP4	0.007226	3.1119
27	SLC28A3	0.007201	4.2125
28	HLA-DRA	0.007197	8.3089
29	TAP2	0.007189	4.8464
30	ACSL5	0.007155	6.8703
31	CDH11	-0.00708	4.9925
32	PSMB9	0.006962	4.1122
33	MMP14	-0.00683	10.1689
34	CD74	0.006825	9.2707
35	LOXL1	-0.00676	9.6429
36	CIITA	0.006623	5.5396
37	ZNF697	-0.00658	7.0319
38	SH3RF2	0.006549	5.0029
39	MIR198	-0.00654	5.1935
40	COL1A2	-0.00645	6.0427
41	TNFRSF14	0.006421	9.0366
42	COL8A1	-0.00642	6.4565
43	C21orf63	0.006261	5.9811
44	TAP1	0.006215	8.6458
45	PDPN	-0.00612	5.3198
46	RHOBTB3	-0.00597	3.5609

47	BCL11A	0.005943	4.3818
48	HLA-DOB	0.005851	4.6075
49	XAF1	0.005742	7.9229
50	ARHGAP26	0.005632	4.3991
51	POLD2	-0.00558	9.4183
52	DPYSL2	-0.00533	8.3469
53	COL4A1	-0.0052	7.0317
54	ID3	-0.00516	7.5673
55	CFB	0.005077	5.7951
56	NID1	-0.00494	4.7186
57	FKBP7	-0.00489	2.9437
58	TIMP2	-0.00468	7.5253
59	RCBTB1	-0.00458	7.4491
60	ANGPTL2	-0.00448	5.6807
61	ENTPD7	-0.00442	7.3772
62	SHISA4	-0.00403	6.0601
63	HINT1	0.003651	6.0724

[0089] In one exemplary embodiment, an expression signature comprises all or a portion of the following biomarkers: CCDC80, INHBA, THBS2, SFRP2, MMP2, PLAUI, FAP, FN1, COL8A1, RAB31, FAM38B, VCAN, GJB2, ITGA5, CRISPLD2, C17, f91, BGN, TIMP3, ALPK2, LUM, NKD2, LOX, MIR1245, LOXL1, and CXCL12.

[0090] In another exemplary embodiment, an expression signature comprises CCDC80, INHBA, THBS2 and SFRP2 and at least N additional biomarkers selected from the list of biomarkers in Table 2A, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 21.

[0091] In another exemplary embodiment, the expression signature comprises CCDC80 and at least N additional biomarkers selected from the list of biomarkers in Table 2A, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24.

[0092] In another exemplary embodiment, the expression signature comprises INHBA and at least N additional biomarkers selected from the list of biomarkers in Table 2A, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24.

[0093] In another exemplary embodiment, the expression signature comprises THBS2 and at least N additional biomarkers selected from the list of biomarkers in Table 2A, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24.

[0094] In another exemplary embodiment, the expression signature comprises SFRP2 and at least N additional biomarkers selected from the list of biomarkers in Table 2A, wherein N equals 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24.

[0095] In another example embodiment, an example expression signatures comprises the biomarkers and the corresponding biomarker weighted values listed in Table 2A. In another

exemplary embodiment, an example expression signature consists of the biomarkers and the corresponding biomarker weighted values listed in Table 2A

[0096] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising CCDC80, INHBA, THBS2 and SFRP2 and at least N additional biomarkers selected from the list of biomarkers in Table 2A, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or 21.

[0097] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising CCDC80 and at least N additional biomarkers selected from the list of biomarkers in Table 2A, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24.

[0098] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising INHBA and at least N additional biomarkers selected from the list of biomarkers in Table 2A, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24.

[0099] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising THBS2 and at least N additional biomarkers selected from the list of biomarkers in Table 2A, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24.

[00100] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising SFRP2 and at least N additional biomarkers selected from the list of biomarkers in Table 2A, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 or 24.

[00101] In another exemplary embodiment, an expression signature comprises all or a portion of the following biomarkers; TMEM200A, GJB2, MMP13, GFPT2, POSTN, BICC1, CDH11, MRVI1, PMP22, COL11A1, IGFL2, LUM, NTM, BGN, COL3A1, COL10A1, RAB31, ANGPTL2, PLA2, COL8A1, MIR1245, POLD2, NKD2, FZD1, COPZ2, ITGA5, VGLL3, INHBA, MMP14, VCAN, THBS2, RUNX2, TIMP3, SFRP2, COL1A2, COL5A2, SERPINF1, KIF26B, TNFAIP6, MMP2, FN1, ALPK2, CTSK, LOXL1 and FAP.

[0100] In another exemplary embodiment, an expression signature comprises TMEM200A, GJB2, MMP13 and GFPT2 and at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, or 41.

[0101] In another exemplary embodiment, an expression signature comprises TMEM200A and at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44.

[0102] In another exemplary embodiment, an expression signature comprises GJB2 and at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44.

[0103] In another exemplary embodiment, an expression signature comprises MMP13 and at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44.

[0104] In another exemplary embodiment, an expression signature comprises GFPT2 and at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44.

[0105] In another exemplary embodiment, an example expression signature comprises the biomarkers and the corresponding biomarker weighted values listed in Table 2B. In another exemplary embodiment, the expression signature consists of the biomarkers and corresponding biomarker weighted values listed in Table 2B.

[0106] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising TMEM200A, GJB2, MMP13 and GFPT2 and at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40 or 41.

[0107] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising TMEM200A and at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 1, 2,

3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44.

[0108] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising GJB2 and at least N additional biomarkers selected from the list of biomarkers in Table 2a, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44.

[0109] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising MMP13 and at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44.

[0110] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising GFPT2 and at least N additional biomarkers selected from the list of biomarkers in Table 2B, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44.

[0111] In another exemplary embodiment, an expression signature comprises all or a portion of the following biomarkers; IGF2, SOX11, INS, CXCL17, SLC5A1, TMEM45A, CXCR2P1, MFAP2, MATN3, RTP4, COL3A1, CDR1, RARRES3, TNFSF10, NUA1, SNORD114-14, SRPX, SPARC, GJB1, TIMP3, ISLR, TUBA1A, DEXI, BASP1, PXDN, GBP4, SLC28A3, HLA-DRA, TAP2, ACSL5, CDH11, PSMB9, MMP14, CD74, LOXL1, CIITA, ZNF697, SH3RF2, MIR198, COL1A2, TNFRSF14, COL8A1, C21orf63, TAP1, PDPN, RHOBTB3, BCL11A, HLA-DOB, XAF1, ARHGAP26, POLD2, DPYSL2, COL4A1, ID3, CFB, NID1, FKBP7, TIMP2, RCBTB1, ANGPTL2, ENTPD7, SHISA4, and HINT1,

[0112] In another exemplary embodiment, an expression signature comprises IGF2, SOX11, INS, and CXCL17 and at least N additional biomarkers selected from the list of biomarkers in Table 2C, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, or 59.

[0113] In another exemplary embodiment, an expression signature comprises IGF2 and at least N additional biomarkers selected from the list of biomarkers in Table 2C, wherein N equals

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, or 62.

[0114] In another exemplary embodiment, an expression signature comprises SOX11 and at least N additional biomarkers selected from the list of biomarkers in Table 2C, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, or 62.

[0115] In another exemplary embodiment, an expression signature comprises INS and at least N additional biomarkers selected from the list of biomarkers in Table 2C, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, or 62.

[0116] In another exemplary embodiment, an expression signature comprises CXCL17 and at least N additional biomarkers selected from the list of biomarkers in Table 2C, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, or 62.

[0117] In another exemplary embodiment, an example expression signature comprises the biomarkers and corresponding biomarker weighted values listed in Table 2C. In another exemplary embodiment, an example expression signature consists of the biomarkers and corresponding biomarker weighted values listed in Table 2C.

[0118] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising all or a portion of the biomarkers listed in Table 2C.

[0119] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising IGF2, SOX11, INS, and CXCL17 and at least N additional biomarkers selected from the list of biomarkers in Table 2C, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, or 59.

[0120] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or

more biomarkers in a biomarker panel, the biomarker panel comprising IGF2 and at least N additional biomarkers selected from the list of biomarkers in Table 2C, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, or 62.

[0121] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising SOX11 and at least N additional biomarkers selected from the list of biomarkers in Table 2C, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, or 62.

[0122] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising INS and at least N additional biomarkers selected from the list of biomarkers in Table 2C, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, or 62.

[0123] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising CXCL17 and at least N additional biomarkers selected from the list of biomarkers in Table 2C, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, or 62.

[0124] In one exemplary embodiment, the expression signature comprises all or a portion of the following biomarkers; ALPK2, BGN, COL8A1, FAP, FN1, GJB2, INHBA, ITGA5, LOXL1, LUM, MIR1245, MMP2, NKD2, PLAU, RAB31, SFRP2, THBS2, TIMP3 and VCAN.

[0125] In another exemplary embodiment, the expression signature comprises ALPK2, BGN, COL8A1, FAP and at least N additional biomarkers selected from FN1, GJB2, INHBA, ITGA5, LOXL1, LUM, MIR1245, MMP2, NKD2, PLAU, RAB31, SFRP2, THBS2, TIMP3 and VCAN, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15.

[0126] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising ALPK2, BGN, COL8A1 and FAP at least N additional biomarkers selected from FN1, GJB2, INHBA, ITGA5, LOXL1, LUM, MIR1245, MMP2, NKD2, PLAU, RAB31, SFRP2, THBS2, TIMP3 and VCAN, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15.

[0127] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising ALPK2 and at least N additional biomarkers selected from the list of biomarkers BGN, COL8A1, FAP, FN1, GJB2, INHBA, ITGA5, LOXL1, LUM, MIR1245, MMP2, NKD2, PLAU, RAB31, SFRP2, THBS2, TIMP3 and VCAN wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15.

[0128] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising BGN and at least N additional biomarkers selected from the list of biomarkers ALPK2, COL8A1, FAP, FN1, GJB2, INHBA, ITGA5, LOXL1, LUM, MIR1245, MMP2, NKD2, PLAU, RAB31, SFRP2, THBS2, TIMP3 and VCAN, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15.

[0129] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising COL8A1 and at least N additional biomarkers selected from the list of biomarkers ALPK2, BGN, , FAP, FN1, GJB2, INHBA, ITGA5, LOXL1, LUM, MIR1245, MMP2, NKD2, PLAU, RAB31, SFRP2, THBS2, TIMP3 and VCAN, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15.

[0130] In a further aspect, the methods of the present invention comprise conducting an assay on a biological sample from an individual to determine the expression levels of one or more biomarkers in a biomarker panel, the biomarker panel comprising FAP and at least N additional biomarkers selected from the list of biomarkers ALPK2, BGN, COL8A1, FN1, GJB2, INHBA, ITGA5, LOXL1, LUM, MIR1245, MMP2, NKD2, PLAU, RAB31, SFRP2, THBS2, TIMP3 and VCAN, wherein N equals 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15.

[0131] In another exemplary embodiment, an expression signature comprises all or a portion of the following biomarkers; GJB2, INHBA, THBS2, SFRP2, PLAU and at least N additional biomarkers from Table 1A-1C, wherein N is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43

or 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, or 70.

[0132] In another exemplary embodiment, an expression signature comprises all GJB2 and at least N additional biomarkers from Table 1A-1C, wherein N is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, or 74.

[0133] In another exemplary embodiment, an expression signature comprises all INHBA and at least N additional biomarkers from Table 1A-1C, wherein N is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, or 74.

[0134] In another exemplary embodiment, an expression signature comprises all THBS2 and at least N additional biomarkers from Table 1A-1C, wherein N is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, or 74.

[0135] In another exemplary embodiment, an expression signature comprises all SFRP2 wherein N is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, or 74.

[0136] In another exemplary embodiment, an expression signature comprises all PLAU and at least N additional biomarkers from Table 1A-1C, wherein N is 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 29, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43 or 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, or 70.

[0137] In another exemplary embodiment, an expression signature comprises all GJB2, INHBA, THBS2, SFRP2, PLAU and at least about 1%, about 5%, about 10%, about 20%, about 30%, about 40%, about 50%, about 60%, about 70%, about 80%, about 90%, about 95%, or about 99% of the biomarkers listed in Table 1A and Table 1B or a combination thereof.

Classifying New Test Samples using an Expression Signature

[0138] To classify new test samples using an expression signature, such as those described above, the relative expression levels of biomarkers in a cancer tissue are measured to form a test sample expression profile. In certain exemplary embodiments, the test sample expression profile is summarized in the form of a compound decision score ("expression score") and

compared to a threshold score that is mathematically derived from a training set of patient data. The score threshold separates a patient group based on different characteristics such as, but not limited to, responsiveness/non-responsiveness to treatment. The patient training set data is preferably derived from cancer tissue samples having been characterized by prognosis, likelihood of recurrence, long term survival, clinical outcome, treatment response, diagnosis, cancer classification, or personalized genomics profile. Expression profiles, and corresponding decision scores from patient samples may be correlated with the characteristics of patient samples in the training set that are on the same side of the mathematically derived score decision threshold. The threshold of the linear classifier scalar output is optimized to maximize the sum of sensitivity and specificity under cross-validation as observed within the training dataset.

[0139] The overall expression data for a given sample is normalized using methods known to those skilled in the art in order to correct for differing amounts of starting material, varying efficiencies of the extraction and amplification reactions, etc.

[0140] In one embodiment, the biomarker expression profile of a patient tissue sample is evaluated by a linear classifier. As used herein, a linear classifier refers to a weighted sum of the individual biomarker intensities into a compound decision score (“decision function”). The decision score is then compared to a pre-defined cut-off score threshold, corresponding to a certain set-point in terms of sensitivity and specificity which indicates if a sample is above the score threshold (decision function positive) or below (decision function negative).

[0141] Using a linear classifier on the normalized data to make a diagnostic or prognostic call (e.g. responsiveness or resistance to therapeutic agent) effectively means to split the data space, i.e. all possible combinations of expression values for all genes in the classifier, into two disjoint halves by means of a separating hyperplane. This split is empirically derived on a large set of training examples, for example from patients showing responsiveness or resistance to a therapeutic agent. Without loss of generality, one can assume a certain fixed set of values for all but one biomarker, which would automatically define a threshold value for this remaining biomarker where the decision would change from, for example, responsiveness or resistance to a therapeutic agent. Expression values above this dynamic threshold would then either indicate resistance (for a biomarker with a negative weight) or responsiveness (for a biomarker with a positive weight) to a therapeutic agent. The precise value of this threshold depends on the actual measured expression profile of all other biomarkers within the classifier, but the general indication of certain biomarkers remains fixed, i.e. high values or “relative over-expression” always contributes to either a responsiveness (genes with a positive weight) or resistance (genes with a negative weights). Therefore, in the context of the overall gene expression classifier,

relative expression can indicate if either up- or down-regulation of a certain biomarker is indicative of responsiveness or resistance to a therapeutic agent.

[0142] There are a number of suitable methods for measuring expression profiles of test samples depending on the type of biomarker to be assayed. Measuring mRNA in a biological sample may be used as a surrogate for detection of the level of the corresponding protein in the biological sample. Thus, any of the biomarkers or biomarker panels described herein can also be detected by detecting the appropriate RNA. Methods of gene expression profiling include, but are not limited to, microarray, RT-PCT, qPCR, NGS, northern blots, SAGE, mass spectrometry.

[0143] mRNA expression levels are measured by reverse transcription quantitative polymerase chain reaction (RT-PCR followed with qPCR). RT-PCR is used to create a cDNA from the mRNA. The cDNA may be used in a qPCR assay to produce fluorescence as the DNA amplification process progresses. By comparison to a standard curve, qPCR can produce an absolute measurement such as number of copies of mRNA per cell. Northern blots, microarrays, Invader assays, and RT-PCR combined with capillary electrophoresis have all been used to measure expression levels of mRNA in a sample. See *Gene Expression Profiling: Methods and Protocols*, Richard A. Shimkets, editor, Humana Press, 2004.

[0144] miRNA molecules are small RNAs that are non-coding but may regulate gene expression. Any of the methods suited to the measurement of mRNA expression levels can also be used for the corresponding miRNA. Recently many laboratories have investigated the use of miRNAs as biomarkers for disease. Many diseases involve widespread transcriptional regulation, and it is not surprising that miRNAs might find a role as biomarkers. The connection between miRNA concentrations and disease is often even less clear than the connections between protein levels and disease, yet the value of miRNA biomarkers might be substantial. Of course, as with any RNA expressed differentially during disease, the problems facing the development of an in vitro diagnostic product will include the requirement that the miRNAs survive in the diseased cell and are easily extracted for analysis, or that the miRNAs are released into blood or other matrices where they must survive long enough to be measured. Protein biomarkers have similar requirements, although many potential protein biomarkers are secreted intentionally at the site of pathology and function, during disease, in a paracrine fashion. Many potential protein biomarkers are designed to function outside the cells within which those proteins are synthesized.

[0145] Gene expression may also be evaluated using mass spectrometry methods. A variety of configurations of mass spectrometers can be used to detect biomarker values. Several types of mass spectrometers are available or can be produced with various configurations. In general, a mass spectrometer has the following major components: a sample inlet, an ion source, a mass

analyzer, a detector, a vacuum system, and instrument-control system, and a data system. Difference in the sample inlet, ion source, and mass analyzer generally define the type of instrument and its capabilities. For example, an inlet can be a capillary-column liquid chromatography source or can be a direct probe or stage such as used in matrix-assisted laser desorption. Common ion sources are, for example, electrospray, including nanospray and microspray or matrix-assisted laser desorption. Common mass analyzers include a quadrupole mass filter, ion trap mass analyzer and time-of-flight mass analyzer. Additional mass spectrometry methods are well known in the art (see Burlingame *et al.*, Anal. Chem. 70:647 R-716R (1998); Kinter and Sherman, New York (2000)).

[0146] Protein biomarkers and biomarker values can be detected and measured by any of the following: electrospray ionization mass spectrometry (ESI-MS), ESI-MS/MS, ESI-MS/(MS)_n, matrix-assisted laser desorption ionization time-of-flight mass spectrometry (MALDI-TOF-MS), surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF-MS), desorption/ionization on silicon (DIOS), secondary ion mass spectrometry (SIMS), quadrupole time-of-flight (Q-TOF), tandem time-of-flight (TOF/TOF) technology, called ultraflex III TOF/TOF, atmospheric pressure chemical ionization mass spectrometry (APCI-MS), APCI-MS/MS, APCI-(MS).sup.N, atmospheric pressure photoionization mass spectrometry (APPI-MS), APPI-MS/MS, and APPI-(MS).sup.N, quadrupole mass spectrometry, Fourier transform mass spectrometry (FTMS), quantitative mass spectrometry, and ion trap mass spectrometry.

[0147] Sample preparation strategies are used to label and enrich samples before mass spectroscopic characterization of protein biomarkers and determination biomarker values. Labeling methods include but are not limited to isobaric tag for relative and absolute quantitation (iTRAQ) and stable isotope labeling with amino acids in cell culture (SILAC). Capture reagents used to selectively enrich samples for candidate biomarker proteins prior to mass spectroscopic analysis include but are not limited to aptamers, antibodies, nucleic acid probes, chimeras, small molecules, an F(ab')₂ fragment, a single chain antibody fragment, an Fv fragment, a single chain Fv fragment, a nucleic acid, a lectin, a ligand-binding receptor, affybodies, nanobodies, ankyrins, domain antibodies, alternative antibody scaffolds (e.g. diabodies etc) imprinted polymers, avimers, peptidomimetics, peptoids, peptide nucleic acids, threose nucleic acid, a hormone receptor, a cytokine receptor, and synthetic receptors, and modifications and fragments of these.

[0148] The foregoing assays enable the detection of biomarker values that are useful in methods for predicting responsiveness of a cancer therapeutic agent, where the methods comprise detecting, in a biological sample from an individual, at least N biomarker values that

each correspond to a biomarker selected from the group consisting of the biomarkers provided in Tables 1A, 1B, 2A, 2B or Groups I and II, wherein a classification, as described in detail below, using the biomarker values indicates whether the individual will be responsive to a therapeutic agent. While certain of the described predictive biomarkers are useful alone for predicting responsiveness to a therapeutic agent, methods are also described herein for the grouping of multiple subsets of the biomarkers that are each useful as a panel of two or more biomarkers. Thus, various embodiments of the instant application provide combinations comprising N biomarkers, wherein N is at least three biomarkers. It will be appreciated that N can be selected to be any number from any of the above-described ranges, as well as similar, but higher order, ranges. In accordance with any of the methods described herein, biomarker values can be detected and classified individually or they can be detected and classified collectively, as for example in a multiplex assay format.

b) Microarray methods

[0149] In one embodiment, the present invention makes use of "oligonucleotide arrays" (also called herein "microarrays"). Microarrays can be employed for analyzing the expression of biomarkers in a cell, and especially for measuring the expression of biomarkers of cancer tissues.

[0150] In one embodiment, biomarker arrays are produced by hybridizing detectably labeled polynucleotides representing the mRNA transcripts present in a cell (e.g., fluorescently-labeled cDNA synthesized from total cell mRNA or labeled cRNA) to a microarray. A microarray is a surface with an ordered array of binding (e.g., hybridization) sites for products of many of the genes in the genome of a cell or organism, preferably most or almost all of the genes. Microarrays can be made in a number of ways known in the art. However produced, microarrays share certain characteristics. The arrays are reproducible, allowing multiple copies of a given array to be produced and easily compared with each other. Preferably the microarrays are small, usually smaller than 5 cm², and they are made from materials that are stable under binding (e.g., nucleic acid hybridization) conditions. A given binding site or unique set of binding sites in the microarray will specifically bind the product of a single gene in the cell. In a specific embodiment, positionally addressable arrays containing affixed nucleic acids of known sequence at each location are used.

[0151] It will be appreciated that when cDNA complementary to the RNA of a cell is made and hybridized to a microarray under suitable hybridization conditions, the level of hybridization to the site in the array corresponding to any particular gene will reflect the prevalence in the cell of mRNA transcribed from that gene/biomarker. For example, when detectably labeled (e.g., with a fluorophore) cDNA or cRNA complementary to the total cellular

mRNA is hybridized to a microarray, the site on the array corresponding to a gene (i.e., capable of specifically binding the product of the gene) that is not transcribed in the cell will have little or no signal (e.g., fluorescent signal), and a gene for which the encoded mRNA is prevalent will have a relatively strong signal. Nucleic acid hybridization and wash conditions are chosen so that the probe "specifically binds" or "specifically hybridizes" to a specific array site, i.e., the probe hybridizes, duplexes or binds to a sequence array site with a complementary nucleic acid sequence but does not hybridize to a site with a non-complementary nucleic acid sequence. As used herein, one polynucleotide sequence is considered complementary to another when, if the shorter of the polynucleotides is less than or equal to 25 bases, there are no mismatches using standard base-pairing rules or, if the shorter of the polynucleotides is longer than 25 bases, there is no more than a 5% mismatch. Preferably, the polynucleotides are perfectly complementary (no mismatches). It can be demonstrated that specific hybridization conditions result in specific hybridization by carrying out a hybridization assay including negative controls using routine experimentation.

[0152] Optimal hybridization conditions will depend on the length (e.g., oligomer vs. polynucleotide greater than 200 bases) and type (e.g., RNA, DNA, PNA) of labeled probe and immobilized polynucleotide or oligonucleotide. General parameters for specific (i.e., stringent) hybridization conditions for nucleic acids are described in Sambrook *et al.*, supra, and in Ausubel *et al.*, "Current Protocols in Molecular Biology", Greene Publishing and Wiley-interscience, NY (1987), which is incorporated in its entirety for all purposes. When the cDNA microarrays are used, typical hybridization conditions are hybridization in 5xSSC plus 0.2% SDS at 65°C for 4 hours followed by washes at 25°C in low stringency wash buffer (1xSSC plus 0.2% SDS) followed by 10 minutes at 25°C in high stringency wash buffer (0.1SSC plus 0.2% SDS) (see Shena *et al.*, Proc. Natl. Acad. Sci. USA, Vol. 93, p. 10614 (1996)). Useful hybridization conditions are also provided in, e.g., Tijessen, Hybridization With Nucleic Acid Probes", Elsevier Science Publishers B.V. (1993) and Kricka, "Nonisotopic DNA Probe Techniques", Academic Press, San Diego, Calif. (1992).

c) Immunoassay methods

[0153] Immunoassay methods are based on the reaction of an antibody to its corresponding target or analyte and can detect the analyte in a sample depending on the specific assay format. To improve specificity and sensitivity of an assay method based on immunoreactivity, monoclonal antibodies are often used because of their specific epitope recognition. Polyclonal antibodies have also been successfully used in various immunoassays because of their increased affinity for the target as compared to monoclonal antibodies. Immunoassays have been designed

for use with a wide range of biological sample matrices Immunoassay formats have been designed to provide qualitative, semi-quantitative, and quantitative results.

[0154] Quantitative results may be generated through the use of a standard curve created with known concentrations of the specific analyte to be detected. The response or signal from an unknown sample is plotted onto the standard curve, and a quantity or value corresponding to the target in the unknown sample is established.

[0155] Numerous immunoassay formats have been designed. ELISA or EIA can be quantitative for the detection of an analyte/biomarker. This method relies on attachment of a label to either the analyte or the antibody and the label component includes, either directly or indirectly, an enzyme. ELISA tests may be formatted for direct, indirect, competitive, or sandwich detection of the analyte. Other methods rely on labels such as, for example, radioisotopes (I^{125}) or fluorescence. Additional techniques include, for example, agglutination, nephelometry, turbidimetry, Western blot, immunoprecipitation, immunocytochemistry, immunohistochemistry, flow cytometry, Luminex assay, and others (see *ImmunoAssay: A Practical Guide*, edited by Brian Law, published by Taylor & Francis, Ltd., 2005 edition).

[0156] Exemplary assay formats include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay, fluorescent, chemiluminescence, and fluorescence resonance energy transfer (FRET) or time resolved-FRET (TR-FRET) immunoassays. Examples of procedures for detecting biomarkers include biomarker immunoprecipitation followed by quantitative methods that allow size and peptide level discrimination, such as gel electrophoresis, capillary electrophoresis, planar electrochromatography, and the like.

[0157] Methods of detecting and/or quantifying a detectable label or signal generating material depend on the nature of the label. The products of reactions catalyzed by appropriate enzymes (where the detectable label is an enzyme; see above) can be, without limitation, fluorescent, luminescent, or radioactive or they may absorb visible or ultraviolet light. Examples of detectors suitable for detecting such detectable labels include, without limitation, x-ray film, radioactivity counters, scintillation counters, spectrophotometers, colorimeters, fluorometers, luminometers, and densitometers.

[0158] Any of the methods for detection can be performed in any format that allows for any suitable preparation, processing, and analysis of the reactions. This can be, for example, in multi-well assay plates (e.g., 96 wells or 384 wells) or using any suitable array or microarray. Stock solutions for various agents can be made manually or robotically, and all subsequent pipetting, diluting, mixing, distribution, washing, incubating, sample readout, data collection and analysis can be done robotically using commercially available analysis software, robotics, and detection instrumentation capable of detecting a detectable label.

Kits

[0159] Reagents, tools, and/or instructions for performing the methods described herein can be provided in a kit. For example, the kit can contain reagents, tools, and instructions for determining an appropriate therapy for a cancer patient. Such a kit can include reagents for collecting a tissue sample from a patient, such as by biopsy, and reagents for processing the tissue. The kit can also include one or more reagents for performing a gene or gene product expression analysis, such as reagents for performing nucleic acid amplification (e.g RT-PCR, qPCR), sequencing (e.g. next generation sequencing), northern blot, proteomic analysis, or immunohistochemistry to determine expression levels of gene or gene product markers in a sample of a patient. For example, primers for performing RT-PCR, probes for performing northern blot analyses, and/or antibodies for performing proteomic analysis such as Western blot, immunohistochemistry and ELISA analyses can be included in such kits. Appropriate buffers for the assays can also be included. Detection reagents required for any of these assays can also be included. The appropriate reagents and methods are described in further detail below. The kits may include suitable primers and/or probes to detect the expression levels of at least one (up to all) of the biomarkers of in Table 2A and/or 2B and/or 2C. Where expression is determined at the protein level the kit may contain binding reagents specific for the proteins of interest. The binding reagents may comprise antibodies to include all fragments and derivatives thereof. In the context of the various embodiments of the present invention the term “antibody” includes all immunoglobulins or immunoglobulin-like molecules with specific binding affinity for the relevant protein (including by way of example and without limitation, IgA, IgD, IgE, IgG and IgM, combinations thereof, and similar molecules produced during an immune response in any vertebrate, for example, in mammals such as humans, goats, rabbits and mice). Specific immunoglobulins useful in the various embodiments of the invention include IgG isotypes. The antibodies useful in the various embodiments of the invention may be monoclonal or polyclonal in origin, but are typically monoclonal antibodies. Antibodies may be human antibodies, non-human antibodies, or humanized versions of non-human antibodies, or chimeric antibodies. Various techniques for antibody humanization are well established and any suitable technique may be employed. The term “antibody” also refers to a polypeptide ligand comprising at least a light chain or heavy chain immunoglobulin variable region which specifically recognizes and binds an epitope of an antigen, and it extends to all antibody derivatives and fragments that retain the ability to specifically bind to the relevant protein. These derivatives and fragments may include Fab fragments, F(ab')₂ fragments, Fv fragments, single chain antibodies, single domain antibodies, Fc fragments etc. The term antibody encompasses antibodies comprised of both heavy and light chains, but also heavy chain (only) antibodies (which may be derived from

various species of cartilaginous fish or camelids). In specific embodiments, the antibodies may be engineered so as to be specific for more than protein, for example bi-specific to permit binding to two different target proteins as identified herein (see Tables 2A, 2B and 2C).

[0160] In some embodiments, the kits may also contain the specific anti-angiogenic therapeutic agent to be administered in the event that the test predicts responsiveness. This agent may be provided in a form, such as a dosage form, that is tailored to the specific treatment. The kit may be provided with suitable instructions for administration according to an appropriate treatment regimen.

[0161] The kits featured herein can also include an instruction sheet describing how to perform the assays for measuring gene or gene product expression. The instruction sheet can also include instructions for how to determine a reference cohort, including how to determine expression levels of gene or gene product markers in the reference cohort and how to assemble the expression data to establish a reference for comparison to a test patient. The instruction sheet can also include instructions for assaying gene or gene product expression in a test patient and for comparing the expression level with the expression in the reference cohort to subsequently determine the appropriate chemotherapy for the test patient. Methods for determining the appropriate chemotherapy are described above and can be described in detail in the instruction sheet.

[0162] Informational material included in the kits can be descriptive, instructional, marketing or other material that relates to the methods described herein and/or the use of the reagents for the methods described herein. For example, the informational material of the kit can contain contact information, e.g., a physical address, email address, website, or telephone number, where a user of the kit can obtain substantive information about performing a gene expression analysis and interpreting the results, particularly as they apply to a human's likelihood of having a positive response to a specific therapeutic agent.

[0163] The kits featured herein can also contain software necessary to infer a patient's likelihood of having a positive response to a specific therapeutic agent from the gene product marker expression.

Therapeutic agents

[0164] As described above, the methods described herein permit the classification of a patient as responsive or non-responsive to a therapeutic agent that targets angiogenic processes and signaling within tumors. Some current such therapeutics used to treat cancer include, but are not limited to, the following agents; VEGF pathway-targeted therapeutic agent, including multi-targeted pathway inhibitors (VEGF/PDGF/FGF/EGFT/FLT-3/c-KIT), Angiopoietin-TIE2 pathway inhibitors, endogenous angiogenic inhibitors, immunomodulatory Agents. VEGF

specific inhibitors include, but are not limited to, Bevacizumab (Avastin), Aflibercept (VEGF Trap), IMC-1121B (Ramucirumab). Multi-targeted pathway inhibitors include, but are not limited to, Imatinib (Gleevec), Sorafenib (Nexavar), Gefitinib (Iressa), Sunitinib (Sutent), Erlotinib, Tivozinib, Cediranib (Recentin), Pazopanib (Votrient), BIBF 1120 (Vargatef), Dovitinib, Semaxanib (Sugen), Axitinib (AG013736), Vandetanib (Zactima), Nilotinib (Tasigna), Dasatinib (Sprycel), Vatalanib, Motesanib, ABT-869, TKI-258. Angiopoietin-TIE2 pathway inhibitors include, but are not limited to, AMG-386, PF-4856884 CVX-060, CEP-11981, CE-245677, MEDI-3617, CVX-241, Trastuzumab (Herceptin). Endogenous angiogenic inhibitors include, but are not limited to, Thrombospondin, Endostatin, Tumstatin, Canstatin, Arrestin, Angiostatin, Vasostatin, Interferon alpha. Immunomodulatory Agents include, but are not limited to, Thalidomide and Lenalidomide.

[0165] The invention is further defined with reference to the following numbered clauses:

1. A method of predicting a subject's responsiveness to an anti-angiogenic therapeutic agent comprising:

measuring expression levels of one or more biomarkers in a test sample of diseased tissue obtained from the subject to determine a sample expression score, wherein the biomarkers are defined by an expression signature;

comparing the sample expression score to a threshold expression score; and

classifying the subject as responsive or non-responsive, based on whether the sample expression score is above or below the threshold expression score.

2. The method of clause 1, wherein the diseased tissue is cancer tissue.

3. The method of clause 2, wherein the cancer is ovarian cancer, glioblastoma, lung cancer, colon cancer, prostate cancer, or breast cancer.

4. The method of any one of clauses 1 to 3, wherein the sample expression value is determined by measuring an expression level for each biomarker and multiplying it by a corresponding weight, wherein the weight for each biomarker is determined by the expression signature.

5. The method of clause 4, wherein the expression signature is derived by a method comprising:

isolating total RNA from a sample set of diseased tissue;

hybridizing the isolated total RNA to a microarray to obtain a sample expression data set;

selecting those probes on the microarray with a variability above a defined significance threshold to form a preliminary biomarker set;

generating clusters of biomarkers within the preliminary biomarker set having a similar expression profile using a clustering algorithm;

identifying the biological processes or biological pathways for each cluster of biomarkers;

selecting the cluster corresponding to the biological process or biological pathway of interest; and

defining an expression signature by analysing the expression levels of the biomarkers in the selected cluster in a the sample set of diseased tissue using a supervised or unsupervised training algorithm.

6. The method of clause 5, wherein the microarray is a transcriptome array comprising probe sets that binds to RNA transcripts verified as expressed in a test sample set by isolating and sequencing RNA transcripts from the test sample set and cross-validating the isolated RNA transcript sequences to known RNA transcript sequences, wherein the test sample set comprises the same tissue as the patient test sample

7. The method of clause 6, wherein the probe set comprises probes that bind within 300 nucleotides of the 3' end of each RNA transcript.

8. The method of any one of clauses 5 to 7, wherein RNA transcripts comprising coding and non-coding transcripts that include messenger RNAs (mRNA), alternatively spliced mRNAs, ribosomal RNAs (rRNA), transfer RNAs (tRNA), small nuclear RNAs (snRNA), microRNAs (miRNAs) and antisense RNAs.

9. The method of any one of clauses 5 to 8, wherein the expression signature is defined using a PLS classifier, a SVM classifier, a SDA classifier, or a DSDA classifier.

10. The method of clause 9, wherein the expression signature is defined using a PLS classifier.

11. The method of any one of clauses 1 to 10, wherein the expression signature comprises two or more genes from Table 1A, Table 1B, or Table 2C.
12. The method of any one of clauses 1 to 10, wherein the expression signature comprises two or more genes from Table 2A, 2B, or 2C.
13. The method of any one of clauses 1 to 10, wherein the expression signature comprises ALPK2, BGN, COL8A1, FAP, FN1, GJB2, INHBA, ITGA5, LOXL1, LUM, MIR1245, MMP2, NKD2, PLAU, RAB31, SFRP2, THBS2, TIMP3, VCAN.
14. The method of any one of clauses 1 to 10, wherein the expression signature comprises CCDC80, INHBA, THBS2, SFRP2, and MMP2.
15. The method of any one of clauses 1 to 10, wherein the expression signature comprises TMEM200A, GJB2, MMP13, GFPT2, and POSTN.
16. The method of any one of clauses 1 to 10, wherein the expression signature comprises IGF2, SOX11, INS, CSCL17, and SLC5A1.
17. The method of any one of clauses 1 to 10, wherein the expression signatures comprises the genes listed in Table 2A.
18. The method of any one of clauses 1 to 10, wherein the expression signature comprises the genes listed in Table 2B.
19. The method of any one of clauses 1 to 10, wherein the expression signature comprises the genes listed in Table 2C.
20. The method of any one of clauses 1 to 10,, wherein the expression signature comprises GJB2, INHBA, THBS2, SFRP2, and PLAU.
21. The method of any one of clauses 1 to 20, wherein the anti-angiogenic therapeutic agent is a VEGF-pathway-targeted therapeutic agent, an angiopoietin-TIE2 pathway inhibitor, endogenous angiogenic inhibitors, and immunomodulatory agents.

22. The method of clause 21, wherein the VEGF pathway-targeted therapeutic agents include Bevacizumab (Avastin), Aflibercept (VEGF Trap), IMC-1121B (Ramucirumab), Imatinib (Gleevec), Sorafenib (Nexavar), Gefitinib (Iressa), Sunitinib (Sutent), Erlotinib, Tivozinib, Cediranib (Recentin), Pazopanib (Votrient), BIBF 1120 (Vargatef), Dovitinib, Semaxanib (Sugen), Axitinib (AG013736), Vandetanib (Zactima), Nilotinib (Tasigna), Dasatinib (Sprycel), Vatalanib, Motesanib, ABT-869, TKI-258 or a combination thereof.

23. The method of clause 21, wherein the angiopoietin-TIE2 pathway inhibitor includes AMG-386, PF-4856884 CVX-060, CEP-11981, CE-245677, MEDI-3617, CVX-241, Trastuzumab (Herceptin) or a combination thereof.

24. The method of clause 21, wherein the endogenous angiogenic inhibitors include Thrombospondin, Endostatin, Tumstatin, Canstatin, Arrestin, Angiostatin, Vasostatin, Interferon alpha or a combination thereof.

25. The method of clause 21, wherein the immunomodulatory agents include thalidomide and lenalidomide or a combination thereof.

26. A method of diagnosing a subject as having a cancer or being susceptible to developing the cancer response to anti-angiogenic therapeutics comprising:

measuring expression levels of biomarkers in a test sample of a diseased tissue obtained from the subject to determine obtain a sample expression score, wherein the biomarkers are defined by an expression signature;

comparing the sample expression value to a threshold score; and

classifying the subject as responsive or non-responsive based on whether the expression score is above or below the threshold score.

27. The method of clause 26, wherein the cancer is ovarian cancer, breast cancer, prostate cancer, lung cancer, colon cancer or a glioblastoma.

28. The method of any one of clauses 26 to 27, wherein the reference expression value is determined by calculating an expression value for each biomarker and multiplying it by a corresponding weight, wherein the weight for each biomarker is determined by the expression signature.

29. The method of clause 26, wherein the expression signature is derived by a method comprising:

isolating total RNA from a sample set of the diseased tissue;

hybridizing the isolated total RNA to a microarray to obtain a set of expression values;

selecting those probes on the microarray with the variability above a defined significance threshold to form a preliminary biomarker set;

generating clusters of biomarkers within the preliminary biomarker set having a similar gene expression profile using a clustering algorithm;

identifying the biological process or biological pathway for each cluster of biomarkers;

selecting the cluster corresponding to the biological processes or biological pathways of interest; and

deriving the expression signature by analysing the expression levels of the biomarkers in the selected cluster in a the sample set of cancer tissue using a supervised or unsupervised training algorithm, wherein the expression value defines a set of biomarkers, corresponding weights for each biomarker and the reference expression value.

30. The method of clause 29, wherein the microarray is a transcriptome array comprising a probe set that binds to RNA transcripts verified as expressed in a sample set of cancer tissue by isolating and sequencing RNA transcripts from the cancer tissue and cross-validating the isolated RNA transcript sequences to known RNA transcript sequences.
31. The method of clause 30, wherein the probe set comprises probes that bind within 300 nucleotides of the 3' end of each RNA transcript.
32. The method of any one of clauses 29 to 31, wherein RNA transcripts comprising coding and non-coding transcripts that include messenger RNAs (mRNA), alternatively spliced mRNAs, ribosomal RNAs (rRNA), transfer RNAs (tRNA), small nuclear RNAs (snRNA), microRNAs (miRNAs) and antisense RNA.
33. The method of any one of clauses 27 to 32, wherein the expression signature is defined using a PLS classifier, a SVM classifier, a SDA classifier, or a DSDA classifier.
34. The method of clause 33, wherein the expression signature is defined using a PLS classifier.
35. The method of any one of clauses 26 to 34, wherein the expression signature comprises two or more genes from Table 1A, Table 1B, or Table 1C.
36. The method of any one of clauses 26 to 34, wherein the expression signature comprises two or more genes from Table 2A, Table 2B, or Table 2C.
37. The method of any one of clauses 26 to 34, wherein the expression signature comprises ALPK2, BGN, COL8A1, FAP, FN1, GJB2, INHBA, ITGA5, LOXL1, LUM, MIR1245, MMP2, NKD2, PLAU, RAB31, SFRP2, THBS2, TIMP3, VCAN.
38. The method of any one of clauses 26 to 34, wherein the expression signature comprises CCDC80, INHBA, THBS2, SFRP2, MMP2.
39. The method of any one of clauses 26 to 34, wherein the expression signature comprises TMEM200A, GJB2, MMP13, GFPT2, and POSTN.

40. The method of any one of clauses 26 to 34, wherein the expression signature comprises IGF2, SOX11, INS, CSCL17, and SLC5A1.
41. The method of any one of clauses 26 to 34, wherein the expression signature comprises the genes listed in Table 2A.
42. The method of any one of clauses 26 to 34, wherein the expression signature comprises the genes listed in Table 2B.
43. The method of any one of clauses 26 to 34, wherein the expression signature comprises the genes listed in Table 2C.
44. The method of any one of clauses 26 to 34, wherein the expression signature comprises GJB2, INHBA, THBS2, SFRP2, and PLAUR.
45. The method of any one of clauses 26 to 44, wherein the anti-angiogenic agent is a VEGF-pathway-targeted therapeutic agent, an angiopoietin-TIE2 pathway inhibitor, endogenous angiogenic inhibitors, and immunomodulatory agents.
46. The method of clause 45, wherein the VEGF pathway-targeted therapeutic agents include Bevacizumab (Avastin), Aflibercept (VEGF Trap), IMC-1121B (Ramucirumab), Imatinib (Gleevec), Sorafenib (Nexavar), Gefitinib (Iressa), Sunitinib (Sutent), Erlotinib, Tivozinib, Cediranib (Recentin), Pazopanib (Votrient), BIBF 1120 (Vargatef), Dovitinib, Semaxanib (Sugen), Axitinib (AG013736), Vandetanib (Zactima), Nilotinib (Tasigna), Dasatinib (Sprycel), Vatalanib, Motesanib, ABT-869, and TKI-258.
47. The method of clause 45, wherein the angiopoietin-TIE2 pathway inhibitor includes AMG-386, PF-4856884, CVX-060, CEP-11981, CE-245677, MEDI-3617, CVX-241, Trastuzumab (Herceptin).
48. The method of clause 45, wherein the endogenous angiogenic inhibitors include Thrombospondin, Endostatin, Tumstatin, Canstatin, Arrestin, Angiostatin, Vasostatin, Interferon alpha.

49. The method of clause 45, wherein the immunomodulatory agents include thalidomide and lenalidomide.
50. A method of determining a prognosis of a subject with cancer comprising:
measuring expression levels of one or more biomarkers in a test sample of diseased tissue obtained from the subject to determine a sample expression score, wherein the biomarkers are defined by an expression signature;
comparing the sample expression score to a threshold expression score; and
classifying the subject as responsive or non-responsive, based on whether the sample expression score is above or below the threshold expression score.
47. The method of clause 46, wherein the diseased tissue is cancer tissue.
48. The method of clause 47, wherein the cancer is ovarian cancer, lung cancer, colon cancer, prostate cancer, glioblastoma, or breast cancer.
49. The method of any one clauses 46 to 48, wherein the sample expression value is determined by measuring an expression level for each biomarker and multiplying it by a corresponding weight, wherein each weight for each biomarker is determined by the expression signature.
50. The method of clause 49, wherein the expression signature is derived by a method comprising:
isolating total RNA from a sample set of diseased tissue;
hybridizing the isolated total RNA to a microarray to obtain a sample expression data set;
selecting those probes on the microarray with a variability above a defined significance threshold to form a preliminary biomarker set;
generating clusters of biomarkers within the preliminary biomarker set having a similar expression profile using a clustering algorithm;
identifying the biological processes or biological pathways for each cluster of biomarkers;
selecting the cluster corresponding to the biological process or biological pathway of interest; and

defining an expression signature by analysing the expression levels of the biomarkers in the selected cluster in a the sample set of diseased tissue using a supervised or unsupervised training algorithm.

51. The method of clause 50, wherein the microarray is a transcriptome array comprising probe sets that binds to RNA transcripts verified as expressed in a test sample set by isolating and sequencing RNA transcripts from the test sample set and cross-validating the isolated RNA transcript sequences to known RNA transcript sequences, wherein the test sample set comprises the same tissue as the patient test sample

52. The method of clause 51, wherein the probe set comprises probes that bind within 300 nucleotides of the 3' end of each RNA transcript.

53. The method of any one of clauses 50 to 57, wherein RNA transcripts comprising coding and non-coding transcripts that include messenger RNAs (mRNA), alternatively spliced mRNAs, ribosomal RNAs (rRNA), transfer RNAs (tRNA), small nuclear RNAs (snRNA), microRNAs (miRNAs) and antisense RNAs.

54. The method of any one of clauses 50 to 53, wherein the expression signature is defined using a PLS classifier, a SVM classifier, a SDA classifier, or a DSDA classifier.

55. The method of clause 46 to 54, wherein the expression signature is defined using a PLS classifier.

56. The method of any one of clauses 46 to 55, wherein the expression signature comprises two or more genes from Table 1A, Table 1B, or Table 1C.

57. The method of any one of clauses 46 to 55, wherein the expression signature comprises two or more genes from Table 2A, 2B, or 2C.

58. The method of any one of clauses 46 to 55, wherein the expression signature comprises ALPK2, BGN, COL8A1, FAP, FN1, GJB2, INHBA, ITGA5, LOXL1, LUM, MIR1245, MMP2, NKD2, PLAU, RAB31, SFRP2, THBS2, TIMP3, VCAN.

59. The method of any one of clauses 46 to 55, wherein the expression signature comprises CCDC80, INHBA, THBS2, SFRP2, and MMP2.

60. The method of any one of clauses 46 to 55, wherein the expression signature comprises TMEM200A, GJB2, MMP13, GFPT2, and POSTN.

61. The method of any one of clauses 46 to 55, wherein the expression signature comprises IGF2, SOX11, INS, CSCL17, and SLC5A1.

62. The method of any one of clauses 46 to 55, wherein the expression signatures comprises the genes listed in Table 2A.

63. The method of any one of clauses 46 to 55, wherein the expression signature comprises the genes listed in Table 2B.

64. The method of any one of clauses 46 to 55, wherein the expression signature comprises the genes listed in Table 2C.

65. The method of any one of clauses 46 to 55, wherein the expression signature comprises GJB2, INHBA, THBS2, SFRP2, and PLAU.

66. A method for determining a course of treatment for a subject with cancer comprising:
measuring expression levels of one or more biomarkers in a test sample of diseased tissue obtained from the subject to determine a sample expression score using a non-angiogenesis expression signature, wherein the biomarkers are defined by the expression signature

determining a prognosis for the subject by comparing the sample expression score from the non-angiogenesis signature to a threshold score, wherein a sample expression score greater than the threshold score indicates a good prognosis and a sample expression score below the the threshold scored indicates a poor prognosis;

measuring expression levels of one or more biomarkers in a test sample of diseased tissue obtained from the subject to determine a sample expression score using an angiogenesis expression signature, wherein the biormarkers are defined by the angiogenesis expression signature;

determining the subject's responsiveness to a therapeutic agent by comparing the sample expression score to a threshold score, wherein a sample expression score greater than the threshold value indicates responsiveness to the therapeutic agent and a sample expression score less than the threshold value indicates non-responsiveness to the therapeutic agent.

67. The method of clause 66, wherein the non-angiogenesis signature comprises two or more biomarkers listed in Table 2C

.

68. The method of clause 66, wherein the non-angiogenesis signature consists of the biomarkers listed in Table 2C.

69. The method of clause 66, wherein the angiogenesis signature comprises two or more biomarkers listed in Table 2A, Table 2B, or a combination thereof.

70.

71. The method of clause 66, wherein the angiogenesis signature consists of the biomarkers listed in Table 1A.

72. The method of clause 66, wherein the angiogenesis signature consists of the biomarkers listed in Table 1B.

73. The method of clause 66, wherein the cancer is ovarian cancer, breast cancer, colon cancer, prostate cancer, lung cancer, or glioblastoma.

74. The method of clause 66, wherein the therapeutic agent is an anti-angiogenic therapeutic agent.

75. The method of clause 74, wherein the anti-angiogenic therapeutic agent is a VEGF-pathway-targeted therapeutic agent, an angiopoietin-TIE2 pathway inhibitor, endogenous angiogenic inhibitors, and immunomodulatory agents.

76. The method of clause 75, wherein the VEGF pathway-targeted therapeutic agents include Bevacizumab (Avastin), Afibercept (VEGF Trap), IMC-1121B (Ramucirumab), Imatinib (Gleevec), Sorafenib (Nexavar), Gefitinib (Iressa), Sunitinib (Sutent), Erlotinib, Tivozinib, Cediranib (Recentin), Pazopanib (Votrient), BIBF 1120 (Vargatef), Dovitinib,

Semaxanib (Sugen), Axitinib (AG013736), Vandetanib (Zactima), Nilotinib (Tasigna), Dasatinib (Sprycel), Vatalanib, Motesanib, ABT-869, TKI-258 or a combination thereof.

77. The method of clause 75, wherein the angiopoietin-TIE2 pathway inhibitor includes AMG-386, PF-4856884 CVX-060, CEP-11981, CE-245677, MEDI-3617, CVX-241, Trastuzumab (Herceptin) or a combination thereof.

78. The method of clause 75, wherein the endogenous angiogenic inhibitors include Thrombospondin, Endostatin, Tumstatin, Canstatin, Arrestin, Angiostatin, Vasostatin, Interferon alpha or a combination thereof.

79. The method of clause 75, wherein the immunomodulatory agents include thalidomide and lenalidomide or a combination thereof.

[0166] For the avoidance of doubt, the scope of the invention is defined by the appended claims.

[0167] This invention is further illustrated by the following examples, which are not to be construed in any way as imposing limitations upon the scope thereof. On the contrary, it is to be clearly understood that resort may be had to various other embodiments, modifications, and equivalents thereof which, after reading the description herein, may suggest themselves to those skilled in the art without departing from the spirit of the present invention and/or scope of the appended claims.

EXAMPLES

Example 1: Tissue processing, hierarchical clustering, subtype identification and classifier development

Tumor Material.

[0168] Exemplary expression signatures were identified from gene expression analysis of a cohort of macrodissected epithelial serous ovarian tumor FFPE tissue samples sourced from the NHS Lothian and University of Edinburgh.

[0169] The protocol for histological classification of epithelial ovarian cancer to define serous, endometrioid, clear cell and mucinous histologies has recently been updated. One of the consequences of this is that many tumors that would previously been classified as endometrioid are now being classified as serous. (McCluggage, W.G. "Morphological subtypes of ovarian carcinoma: a review with emphasis on new developments and pathogenesis," *PATHOLOGY* 2011 Aug;43(5):420-32). Serous samples that were used in this study were among a larger set of epithelial ovarian cancer samples of all histologies that were harvested between 1984 and 2006.

Pathology to assign histology status was performed by the pathologists at each of the centers at the time of harvesting. During March and April 2012, 357 of these epithelial ovarian samples were reviewed for histology classification by two independent consultant ovarian cancer pathologists according to the revised protocol. This resulted in a reclassification of several of these samples as indicated Table 3.

Table 3: results of pathology review of 357 epithelial ovarian cancer samples.

(Original histological status is presented in rows, and updated histological status is presented in columns)

Original/Updated	CLEAR CELL	ENDOMETRIOID	MUCINOUS	SEROUS	TOTAL
CLEAR CELL	19	1	0	5	25
ENDOMETRIOID	2	33	0	38	73
MUCINOUS	0	1	8	1	10
OTHER MIXED	3	5	0	8	16
SEROUS	1	3	1	193	198
SEROUS/ENDO	0	2	0	25	27
UNCLASSIFIED	0	0	0	4	4
UNDIFFERENTIATED	1	0	0	3	4
TOTAL	26	45	9	277	357

The original three serous subtypes identified below, and consequently a 25 gene signature (Table 2A) described in the example below (Figure1) were identified from 199 samples that were classified as serous according to the original pathologist reports. Bioinformatic analysis was similarly performed on the 277 samples classified as stage III and IV high grade serous ovarian cancer using the updated pathology classification method. This analysis identified the updated serous subgroups detailed in Figure 6 and consequently used to define a 45 gene “angiogenesis” signature (Table 2B) and a 63 gene “non-angiogenesis” signature (Table 2C)

Gene Expression Profiling from FFPE

[0170] Total RNA was extracted from macrodissected FFPE tissue using the High Pure RNA Paraffin Kit (Roche Diagnostics GmbH, Mannheim, Germany). RNA was converted into

complementary deoxyribonucleic acid (cDNA), which was subsequently amplified and converted into single-stranded form using the SPIA® technology of the WT-Ovation™ FFPE RNA Amplification System V2 (NuGEN Technologies Inc., San Carlos, CA, USA). The amplified single-stranded cDNA was then fragmented and biotin labeled using the FL-Ovation™ cDNA Biotin Module V2 (NuGEN Technologies Inc.). The fragmented and labeled cDNA was then hybridized to the Almac Ovarian Cancer DSA™. Almac's Ovarian Cancer DSA™ research tool has been optimised for analysis of FFPE tissue samples, enabling the use of valuable archived tissue banks. The Almac Ovarian Cancer DSA™ research tool is an innovative microarray platform that represents the transcriptome in both normal and cancerous ovarian tissues. Consequently, the Ovarian Cancer DSA™ provides a comprehensive representation of the transcriptome within the ovarian disease and tissue setting, not available using generic microarray platforms. Arrays were scanned using the Affymetrix Genechip® Scanner 7G (Affymetrix Inc., Santa Clara, CA).

Data preparation

[0171] Quality Control (QC) of profiled samples was carried out using MAS5 pre-processing algorithm. Different technical aspects were addressed: average noise and background homogeneity, percentage of present call (array quality), signal quality, RNA quality and hybridization quality. Distributions and Median Absolute Deviation of corresponding parameters were analyzed and used to identify possible outliers.

[0172] Almac's Ovarian Cancer DSA™ contains probes that primarily target the area within 300 nucleotides from the 3' end. Therefore standard Affymetrix RNA quality measures were adapted – for housekeeping genes intensities of 3' end probe sets with ratios of 3' end probe set intensity to the average background intensity were used in addition to usual 3'/5' ratios. Hybridization controls were checked to ensure that their intensities and present calls conform to the requirements specified by Affymetrix.

[0173] Pre-processing of the raw data generated from expression profiling of the epithelial serous ovarian cancer training set was performed in Expression Console v1.1 with Robust Multi-array Analysis (RMA).

Hierarchical Clustering and Functional Analysis

a. Hierarchical clustering analysis

[0174] Hierarchical clustering techniques were applied to microarray data from the epithelial serous ovarian tumors analysed using the Ovarian Cancer DSA™ (disease specific array) platform. Raw expression data was preprocessed using the standard Robust Multichip Algorithm (RMA) procedure. Non-biological systematic variance in the data set was identified and

removed. Those probe sets whose expression levels varied significantly from tumor to tumor were identified. These probe sets formed the intrinsic list.

[0175] Two dimensional cluster analysis (tumor, probe set) was performed to establish tumor relationships based on the intrinsic list. Hierarchical agglomerative clustering was applied (Pearson correlation – Original analysis – or Euclidean distance – updated analysis – and Ward’s linkage). Optimal partition number was selected using the GAP index (Tibshirani *et al.*, 2002, J. R. Stat. Soc., 63:411-423). All probe sets available in the cluster subgroups were mapped to genes names.

b. Functional analysis of gene clusters

[0176] To establish the functional significance of the probe set clusters, probe sets were mapped to genes (Entrez gene ID) and an enrichment analysis was performed. Enrichment significance was calculated based on the hypergeometric function (False Discovery Rate applied (Benjamini and Hochberg, 1995, J. R. Stat. Soc. 57:289:300)). Over-representation of biological processes and pathways were analysed for each gene group generated by the hierarchical clustering for the epithelial serous ovarian cancer samples using Almac Diagnostics’ proprietary Functional Enrichment Tool (FET). Antisense probe sets were excluded from the analysis. Hypergeometric p-values were assessed for each enriched functional entity class. Functional entity classes with the highest p-values were selected as representative of the group and a general functional category representing these functional entities was assigned to the gene clusters based on significance of representation (i.e. p-value).

[0177] To generate an angiogenesis classifier using the original 199 epithelial serous ovarian tumors, genes in clusters enriched for angiogenesis, vasculature development and immune response general functional terms were grouped into a putative angiogenesis gene group and used for the signature generation. The sample clusters presenting high expression for the genes involved in angiogenesis, vasculature development and immune response general functional terms were selected for classification and labeled as ‘angiogenesis’. Those not showing high expression for the genes involved in these functional terms were labeled as ‘non-angiogenesis’.

[0178] To generate an angiogenesis classifier using the reclassified 265 epithelial serous ovarian tumors, genes in clusters enriched for angiogenesis and vasculature development general functional terms were grouped into a putative angiogenesis gene group and used for the signature generation. The sample clusters presenting high expression for the genes involved in angiogenesis and vasculature development general functional terms were selected for classification and labeled as ‘angiogenesis’. Those not showing high expression for the genes involved in these functional terms were labeled as ‘others ‘others’ (Figure 6 top label identifying “angiogenesis” samples (red) plus “others” (grey)).

[0179] To generate a “non-angiogenesis classifier” using the reclassified 265 epithelial serous ovarian tumors the sample clusters presenting low expression for the genes involved in angiogenesis and vasculature development general functional terms were selected for classification and labeled as “non-angiogenesis.” Those not showing low expression for the genes involved in these functional terms were labeled as ‘others’ (Figure 6 bottom label identifying “non-angiogenesis” samples plus “others”).

Classifier development at a gene level

[0180] To facilitate validation of the classifier across multiple array platforms, the angiogenesis classifier was generated at the gene level. The following steps outline the procedures that were taken for gene level signature development (each step performed over internal cross validation using 10 repeats of 5-fold cross-validation):

Gene level signature development

- Pre-processing:
 - RMA background correction.
 - Reference set of genes are those genes (sense probe sets only) unique to the ovarian DSA platform.
 - Gene level summarization was performed in two steps; first probes to probe set summarization was performed by calculating the median expression of the probes in a probe set; secondly, median expression of the (sense only) probe sets mapping to each gene in the reference distribution is calculated, yielding a “gene level” expression matrix.
 - Quantile normalization was performed on the full gene expression data matrix and a reference quantile derived from the training data was used to normalize the test samples within each round of cross validation.
- Feature selection: Filtering 75% of genes by variance, intensity and correlation to cDNA concentration, followed by either recursive feature elimination (RFE) or filter feature selection (FFS) based on CAT scores.
- Classification algorithms: Partial Least Squares (PLS), SDA (Shrinkage Discriminate Analysis) and DSDA (Diagonal SDA).

Model selection

[0181] The criteria used for model selection were AUC and/or hazard ratio (HR) over internal cross-validation and feature elimination. Functional enrichment of the signatures over cross validation using FET based on the gene ontologies, interim validation sets which included two sets of technical replicates for which standard deviation in signature scores for repeated

samples was evaluated over cross-validation and feature elimination and an assessment of the independence to clinical and technical factors over cross validation (factors listed in Table 4).

[0182] It should be noted, since the subgroup (i.e. class label) derivation was performed using microarray expression from the same sample cohort that was used for signature development, there was an expected positive bias in any performance estimates based on AUC. This highlights the importance of widening the criteria used for model selection, by including additional metrics such as the hazard ratio, functional enrichment and assessing the independence to clinical and technical factors.

Table 4 List of clinical and technical factors investigated

Clinical factors:	Technical factors:	Sample processing factors:
Debulking	Hospital	Post Amp operator
	Block age	
	RNA concentration	
	cDNA yield	

Calculating classifier scores for validation data sets

[0183] All datasets were background corrected using RMA. For each validation set, the probe sets that map to the classifier genes were determined, excluding anti-sense probe sets (if applicable). Annotation for Affymetrix Plus 2.0 and U133A arrays are available from the Affymetrix website. The median intensity over all probes mapping to each gene in the classifier was calculated and the quantile normalization model from the training set was applied to normalize the testing samples one at a time, resulting in a gene intensity matrix. The classifier was then applied to this data matrix to produce a classifier score/prediction for each sample.

Univariate and Multivariate analysis

[0184] Univariate and multivariate analysis may be carried out in relation to the glioblastoma dataset to assess respectively the association between the angiogenesis subtype classifier and survival, and to determine if the association, if any, was independent to known clinical predictors. The p-values for univariate analysis were calculated using logistic regression in MATLAB. For the multivariate analysis we used a likelihood ratio test from logistic regression was used, where the p-values represent the drop in the log-likelihood when comparing the model with the clinical covariates and the predictions to a reduced model with clinical covariates only. The likelihood ratio test measures the importance of the gene predictor in modeling survival, and highlights its independence as a predictor relative to the clinical predictors. In both univariate and multivariate analysis, a p-value < 0.05 was used as the

criterion for significance. Furthermore, samples with unknown clinical factors were excluded in this assessment.

Results

Identification of subgroups and generation of signature from original and updated histology classification

Hierarchical Clustering Analysis

[0185] Feature selection resulted in the selection of 1200 probe sets from the original epithelial serous ovarian cancer data set (199 samples) and 1400 PS from the reclassified epithelial serous ovarian cancer data set (265 samples). The GAP analysis revealed three sample clusters and three probe set cluster groups within both sample sets (Figure 1, Figure 6).

Classification of tumors into ‘angiogenesis’ or ‘non-angiogenesis’ sample groups

[0186] The classification of samples as ‘angiogenesis’ or ‘non-angiogenesis’ was based upon the results of the functional analysis of the epithelial serous ovarian cancer dataset (Figure 1 Figure 6). The objective of this study was to characterize at a transcriptomic level a set of genes that would be capable of determining responsiveness or resistance of a pathogenic cell to anti-angiogenic agents and potentially identify patients who could benefit from anti-angiogenic therapy. With this in mind, those samples within the epithelial serous ovarian cancer datasets that best represented this biology were to be selected and compared to the remaining samples for classifier generation (see next section). It was decided that the samples from the sample angiogenesis cluster within the original epithelial serous ovarian cancer sample set (199 samples) (Figure 1) were the most relevant samples for this selection as these samples demonstrated an up-regulation of genes involved in signaling related to angiogenic and immune response processes and pathways as defined by functional analysis (Figure 2A and 2B). It was decided that the samples from sample cluster three within the reclassified epithelial serous ovarian cancer sample set (265 samples) (Figure 6) were the most relevant samples for this selection as these samples demonstrated an up-regulation of genes involved in signaling related to angiogenic processes and pathways as defined by functional analysis (Figure 2A and 2B).

[0187] It was decided that the samples from sample cluster 1 within the reclassified epithelial serous ovarian cancer sample set (265 samples) (Figure 6) were the most relevant samples for selection to represent the “non-angiogenic” subgroup as these samples demonstrated a down-regulation of genes involved in signaling related to angiogenic processes and pathways as defined by functional analysis (Figure 2A and 2B).

[0188] An identical hierarchical clustering approach was applied to 105 breast cancer samples. The dominant biology in Breast cancer is ER status and therefore in order to identify the structure in the biology of the samples this cohort was divided into 2 populations for cluster

analysis. We identified angiogenesis and vasculature development subtypes (Figure 11A and 11B) demonstrating the expositing of an angiogenesis subtype from breast cancer samples.

Development and validation of the angiogenesis subtype classifier models

[0189] For ease of reference, the following steps are detailed in reference to expression signatures derived from Table 1A or Table 1B. Following the identification of a class of tumors, that form the putative ‘angiogenesis’ subgroup, computational classification of these tumors versus all others in the tumor cohort with reference to the functional ‘angiogenesis’ (angiogenesis, vasculature development, immune response) gene list (Table 1A or Table 1B) was performed to identify a refined gene classification model, which classifies the ‘angiogenesis’ subtype.

[0190] The classification pipeline was used to derive a model using the set of epithelial serous ovarian cancer samples. The classification pipeline has been developed in accordance with commonly accepted good practice (MAQC Consortium, *Nat Biotechnol* 2010). The process will, in parallel: 1) derive gene classification models from empirical data; and 2) assess the classification performance of the models, both under cross-validation. The performance and success of the classifier generation depends on a number of parameters that can be varied, for instance the choice of classification method or probe set filtering. Taking this into account, two feature sets were evaluated (i) the full feature list with 75% variance/intensity filtering (with forced inclusion of the angiogenesis gene list, Table 1A) and (ii) the angiogenesis gene list only; and three classification algorithms were evaluated, namely PLS; SDA and DSDA. Two feature elimination methods were adopted: (i) Recursive Feature Elimination (RFE) which is an iterative procedure removing a fraction of the lowest-ranked features at each iteration; stopping when only a minimum number of features remain; (ii) Filter Feature Selection (FFS) which pre-ranks the features based on correlation adjusted t-scores (CAT-scores) (ref) and removes the lowest ranking features at each round of feature elimination. The Area Under the Receiver Operator Characteristic Curve (AUC) was used to assess the classification performance, as this measure is independent of cut-off between groups and prevalence rates in the data. It is also one of the recognized measurements of choice for classification performance. As such, the best number of features for each model was chosen based on the average AUC under cross-validation.

[0191] From the analysis described above, the PLS FFS model was deemed to be the most suitable classifier model. Weights were calculated for each gene using PLS regression, resulting in the final gene classifier models (25-gene classifier model for the original approach, and a 45-gene classifier for samples reclassified reflecting recent changes to standard histology protocols) that may be used for validation on external data sets from different array platforms. The gene

signature development process was focused upon identification of the ontological processes and pathways relevant to angiogenesis to ensure biological relevance of any signature developed. As such, functional analysis was performed upon both signatures to qualify their relevance to angiogenesis and related processes. The significance processes in FIGURE 3 and FIGURE 8 are related to angiogenesis and vasculature development.

Example 2: *In silico* validation of the angiogenesis subtype and angiogenesis classifier models

[0192] The performance of both the 25-gene (original approach) and 45-gene (reclassification approach) angiogenesis classifier models were validated by the Area Under the ROC (Receiver Operator Characteristic) Curve (AUC) within the original Almac epithelial serous ovarian cancer dataset and two independent datasets. The AUC is a statistic calculated on the observed disease scale and is a measure of the efficacy of prediction of a phenotype using a classifier model (Wray *et al.*, PLoS Genetics Vol 6, 1-9). An AUC of 0.5 is typical of a random classifier, and an AUC of 1.0 would represent perfect separation of classes. Therefore, in order to determine if the angiogenesis subtype classifier model is capable of predicting response to, and selecting patients for anti-angiogenic ovarian cancer therapeutic drug classes either as single agent or in combination with standard of care therapies, the hypothesis is that the AUCs following application within these datasets should be above 0.5 with the lowest confidence interval also above 0.5.

Application of the classifier models to an independent prostate cancer cell-line dataset

[0193] To assess the predictive power of the 25-gene and 45-gene classifier models, they were applied to a dataset of 16 prostate cell-lines following treatment with Dasatanib. The cell-lines were defined as being either a ‘responder’ or ‘non-responder’ based upon cell-proliferation assays. This analysis revealed that the 25-gene classifier model is associated with response to Dasatanib, with an AUC of 0.8364 (CI = 0.5255 – 1.0000), indicating that the 25-gene classifier is predictive of response to Dasatanib. The analysis revealed that the 45-gene classifier model is associated with response to the same compound, with an AUC of 0.9455 (CI = 0.7949 – 1.0000) indicating that the 45-gene classifier is also predictive of response to Dasatanib.

Example 3: Identification of and *in silico* validation of a “non-angiogenic subgroup of ovarian cancer

Classification of tumors into ‘non -angiogenesis’ sample groups

[0194] The expression of angiogenesis genes in probe set cluster 2 is down regulated in all samples in sample Cluster 1 of hierarchial clustering of 265 samples newly classified as serous (Figure 6 bottom labelled Blue bar)). These samples in sample cluster 2 have a better prognosis than the rest of the serous samples in samples from cluster 1 and 3 combined together as

demonstrated in Figure 10. This indicated that this group is defined by down regulation of expression of the angiogenesis genes identified in Table 1B. Patients with downregulation of genes involved in angiogenesis and therefore this subgroup is termed a “non-angiogenesis” group. This phenotype has also been identified in ER+ and ER- breast cancer as can be see the middle sample group in Figure 11A and the second sample group in Figure 11B.

Development and validation of the non-angiogenesis subtype classifier models

[0195] Following the identification of a class of tumors, within the 265 samples newly classified as serous, that form the putative ‘non-angiogenesis’ subgroup, computational classification of these tumors versus all others in the tumor cohort with reference to the functional ‘angiogenesis’ (angiogenesis, vasculature development, immune response) gene list (Table 1A or Table 1B) was performed to identify a refined gene classification model, which classifies the ‘non-angiogenesis’ subtype.

[0196] The classification pipeline was used to derive a model using the set of epithelial serous ovarian cancer samples. The methods are analogous to those described above in Example 1.

[0197] Applying the analysis described in Example 1, the PLS FFS model was deemed to be the most suitable classifier model. Weights were calculated for each gene using PLS regression, resulting in the final gene classifier model (63-gene classifier model) that may be used for validation on external data sets from different array platforms.

Application of non-angiogenesis classifier model to the discovery data and an independent microarray dataset

[0198] The performance of the 63-gene non-angiogenesis classifier model was validated using hazard ratio (HR) within the original Almac epithelial serous ovarian cancer dataset and one independent dataset. The hazard ratio is the ratio of the hazard rates corresponding to the conditions described by two sets of explanatory variables. For example, a treated population may die at twice the rate per unit time as the control population. The hazard ratio would be 2, indicating higher hazard of death from the treatment. Therefore, in order to determine if the non-angiogenesis subtype classifier model is capable of predicting non-response to, and selecting patients that should not be treated with anti-angiogenic cancer therapeutic drug classes, either as single agent or in combination with standard of care therapies, it is expected that the HRs should be below 1.0 with a p-value less than 0.05.

[0199] To assess the prognostic power of the 63-gene classifier models, it was applied first to the discovery data of 265 epithelial serous ovarian samples and then to a dataset of 275 epithelial serous and endometrioid ovarian samples (14). Within both datasets the samples were taken at the time of initial surgical resection from patients without prior therapy.

[0200] This analysis revealed that the 63-gene classifier model was independently associated with prognosis in both ovarian cancer datasets. In the discovery set the non-angiogenesis group was associated with significantly better survival compared to the other samples (FIGURE 12); Hazard Ratio = 0.4842, $p = 1.307 \times 10^{-5}$). Similarly in the independent microarray dataset the non-angiogenesis group demonstrated significantly better survival compared to the other samples (FIGURE 13) Hazard Ratio = 0.62 [0.41-0.95], $p = 0.029$ and OS (multivariable analysis; HR=0.32 [0.19-0.54], $p=0.00001$). This validates the fact that the 63-gene classifier is an independent prognostic biomarker of survival in ovarian serous and endometrioid cancer samples.

[0201] The International Collaboration on Ovarian Neoplasms 7 (ICON7) trial is a Gynecologic Cancer Intergroup phase 3 trial that assessed the effects of adding bevacizumab, concurrently and as a continuation, to standard chemotherapy with carboplatin and paclitaxel in patients with primary peritoneal carcinoma, fallopian tube carcinoma, and epithelial ovarian carcinoma (Perren TJ, Swart AM, Pfisterer J, Ledermann JA, Pujade-Lauraine E, Kristensen G, et al. A phase 3 trial of bevacizumab in ovarian cancer. *N Engl J Med.* 365(26): 2484-96, Aghajanian C, Blank SV, Goff BA, Judson PL, Teneriello MG, Husain A, et al. OCEANS: A randomized, double-blind, placebo-controlled phase III trial of chemotherapy with or without bevacizumab in patients with platinum-sensitive recurrent epithelial ovarian, primary peritoneal, or fallopian tube cancer. *Journal of Clinical Oncology.* 2012; 30(17): 2039-45).

[0202] Patient characteristics, progression-free survival, toxicity, and preliminary overall survival data and a summary of quality-of-life (QoL) data have been reported from ICON7. In the standard chemotherapy group, 696 (91%) of 764 women received 18 weeks of chemotherapy by protocol. In the bevacizumab group, 719 (94%) of 764 women received 18 weeks of chemotherapy and bevacizumab and 472 (62%) continued bevacizumab to protocol completion at 54 weeks. The hazard ratio for progression-free survival with standard chemotherapy and bevacizumab was 0.81 (95% CI 0.70–0.94, $p=0.004$). In patients at high risk of progression, defined as International Federation of Gynecology and Obstetrics (FIGO) stage IV disease or stage III disease with greater than 1.0 cm of residual disease after debulking surgery, the hazard ratio for death in the bevacizumab group was 0.64 (95% CI 0.48–0.85; $p=0.002$).

[0203] Access was obtained to the ICON7 trial samples via the Medical Research Council (MRC). An honest broker holds the associated clinical data from the MRC. A randomization strategy for profiling the samples has been performed based on clinical factors. All reagents, arrays, and reference samples were previously tested and passed qualification criteria.

[0204] Total RNA was extracted from macrodissected FFPE tissue using the High Pure RNA Paraffin Kit (Roche Diagnostics GmbH, Mannheim, Germany). RNA is converted into

complementary deoxyribonucleic acid (cDNA), which is subsequently amplified and converted into single-stranded form using the SPIA® technology of the WT-Ovation™ FFPE RNA Amplification System V2 (NuGEN Technologies Inc., San Carlos, CA, USA). The amplified single-stranded cDNA is then fragmented and biotin labeled using the FL-Ovation™ cDNA Biotin Module V2 (NuGEN Technologies Inc.). The fragmented and labeled cDNA was then hybridized to the Almac Ovarian Cancer DSA, on this the signature was developed. Arrays are scanned using the Affymetrix Genechip® Scanner 7G (Affymetrix Inc., Santa Clara, CA).

[0205] Control UHR and a pooled clinical reference sample are processed in each post RNA extraction batch. QC is performed at several steps of the process, from RNA Extraction QC to array QC.

[0206] Quality Control (QC) of profiled samples is carried out using the RMA pre-processing algorithm. Different technical aspects are assessed: average noise and background homogeneity, percentage of present call (array quality), signal quality, RNA quality and hybridization quality. Distributions and Median Absolute Deviation of corresponding parameters are analyzed and used to identify possible outliers.

[0207] Almac's Ovarian Cancer DSA™ contains probes that primarily target the area within 300 nucleotides from the 3' end. Therefore standard Affymetrix RNA quality measures were adapted – for housekeeping genes intensities of 3' end probe sets with ratios of 3' end probe set intensity to the average background intensity were used in addition to usual 3'/5' ratios. Hybridization controls were checked to ensure that their intensities and present calls conform to the requirements specified by Affymetrix.

[0208] Samples that pass QC metrics will be deemed suitable for inclusion in subsequent analysis, and signature scores will be calculated (per sample) using the following:

- Background correction
- RefRMA model to pre-process the data, one sample at a time
- The signature score is calculated as a weighted sum of the expression of the

genes in the signature:

$$SignatureScore = \sum_i w_i \times (ge_i - b_i) + k$$

where w_i is a weight for each gene, b_i is a gene-specific bias (Supplementary Table S5), ge_i is the observed gene expression level after pre-processing, and $k = 0.2953$ is a constant offset.

[0209] Sample IDs and corresponding signature scores will be sent to the honest broker to evaluate the predictive performance of the Ovarian Immune signature to predict resistance to Avastin in the treatment arm of the study.

Application of angiogenesis classifier model to the discovery data and an independent microarray dataset

[0210] Class labels were assigned to samples based upon expression of the angiogenesis-related genes and a 25 and 45 gene expression signatures were developed to identify the pro-angiogenic molecular subgroup of high grade serous (HGS) carcinoma.

[0211] Angiogenesis is central to the process of cancer growth and metastasis and has a role in the progression and prognosis of ovarian cancer. VEGF is an important promoter of angiogenesis produced by normal and neoplastic cells. Bevacizumab (Avastin, Riche) is a recombinant humanised version of a murine anti-human VEGF monoclonal antibody and has been studied in the management of many tumours. The International Collaboration on Ovarian Neoplasms 7 (ICON7) trial is a Gynecologic Cancer Intergroup phase 3 trial that assessed the effects of adding bevacizumab, concurrently and as a continuation, to standard chemotherapy with carboplatin and paclitaxel in patients with primary peritoneal carcinoma, fallopian tube carcinoma, and epithelial ovarian carcinoma (Perren TJ, Swart AM, Pfisterer J, Ledermann JA, Pujade-Lauraine E, Kristensen G, et al. A phase 3 trial of bevacizumab in ovarian cancer. *N Engl J Med.* 365(26): 2484-96, Aghajanian C, Blank SV, Goff BA, Judson PL, Teneriello MG, Husain A, et al. OCEANS: A randomized, double-blind, placebo-controlled phase III trial of chemotherapy with or without bevacizumab in patients with platinum-sensitive recurrent epithelial ovarian, primary peritoneal, or fallopian tube cancer. *Journal of Clinical Oncology.* 2012; 30(17): 2039-45).

[0212] Patient characteristics, progression-free survival, toxicity, and preliminary overall survival data and a summary of quality-of-life (QoL) data have been reported from ICON7. In the standard chemotherapy group, 696 (91%) of 764 women received 18 weeks of chemotherapy by protocol. In the bevacizumab group, 719 (94%) of 764 women received 18 weeks of chemotherapy and bevacizumab and 472 (62%) continued bevacizumab to protocol completion at 54 weeks. The hazard ratio for progression-free survival with standard chemotherapy and bevacizumab was 0.81 (95% CI 0.70–0.94, $p=0.004$). In patients at high risk of progression, defined as International Federation of Gynecology and Obstetrics (FIGO) stage IV disease or stage III disease with greater than 1.0 cm of residual disease after debulking surgery, the hazard ratio for death in the bevacizumab group was 0.64 (95% CI 0.48–0.85; $p=0.002$).

[0213] HGS patients with the angiogenic subgroup may benefit from anti-angiogenic targeted therapies. Access to the ICON7 trial samples was obtained via the MRC, and are being profiled on Ovarian DSAs in order to validate this hypothesis.

[0214] H & E Sections of FFPE material from ICON7 trial samples have been pathology reviewed, marked for tumour content and the tumour material macro dissected.

[0215] An honest broker obtained and holds the associated clinical data from the MRC. They have also performed the randomisation strategy, based on clinical factors, for profiling the samples. All reagents, arrays and reference samples were previously tested and passed qualification criteria.

[0216] Total RNA was extracted from macrodissected FFPE tissue using the High Pure RNA Paraffin Kit (Roche Diagnostics GmbH, Mannheim, Germany). RNA is converted into complementary deoxyribonucleic acid (cDNA), which is subsequently amplified and converted into single-stranded form using the SPIA® technology of the WT-Ovation™ FFPE RNA Amplification System V2 (NuGEN Technologies Inc., San Carlos, CA, USA). The amplified single-stranded cDNA is then fragmented and biotin labeled using the FL-Ovation™ cDNA Biotin Module V2 (NuGEN Technologies Inc.). The fragmented and labeled cDNA was then hybridized to the Almac Ovarian Cancer DSA, on this the signature was developed. Arrays are scanned using the Affymetrix Genechip® Scanner 7G (Affymetrix Inc., Santa Clara, CA).

[0217] Control UHR and a pooled clinical reference sample are processed in each post RNA extraction batch. QC is performed at several steps of the process, from RNA Extraction QC to array QC.

[0218] Quality Control (QC) of profiled samples is carried out using the RMA pre-processing algorithm. Different technical aspects are assessed: average noise and background homogeneity, percentage of present call (array quality), signal quality, RNA quality and hybridization quality. Distributions and Median Absolute Deviation of corresponding parameters are analyzed and used to identify possible outliers.

[0219] Almac's Ovarian Cancer DSATM contains probes that primarily target the area within 300 nucleotides from the 3' end. Therefore standard Affymetrix RNA quality measures were adapted – for housekeeping genes intensities of 3' end probe sets with ratios of 3' end probe set intensity to the average background intensity were used in addition to usual 3'/5' ratios. Hybridization controls were checked to ensure that their intensities and present calls conform to the requirements specified by Affymetrix.

[0220] Samples that pass QC metrics will be deemed suitable for inclusion in subsequent analysis, and signature scores will be calculated (per sample) using the following:

Background correction

[0221] RefRMA model to pre-process the data, one sample at a time

[0222] The signature score is calculated as a weighted sum of the expression of the genes in the signature:

$$SignatureScore = \sum_i w_i \times (ge_i - b_i) + k$$

[0223]

[0224] where w_i is a weight for each gene, b_i is a gene-specific bias (Supplementary Table S5), ge_i is the observed gene expression level after pre-processing, and $k = 0.2953$ is a constant offset.

[0225] Sample IDs and corresponding signature scores will be sent to the honest broker to evaluate the predictive performance of the pro-angiogenic signature to predict benefit from Avastin in the treatment arm of the study.

Application of classifier model combination to identify three molecular subgroups

[0226] As a secondary endpoint, analysis of the combined signature approach (see Figure 17) will be evaluated. This approach utilizes the 63 gene signature to identify 2 groups, one with a better survival i.e. the group with reduced activation of angiogenic processes, and the other with poorer survival. The 45 gene signature will be applied to the group with poorer survival outcome to identify a further 2 groups, one with elevated activation of angiogenic processes and the other with indifferent angiogenic activation. It is hypothesised that the group identified with elevated angiogenic activity should respond to Avastin in the treatment arm of the study.

Example 4: Predictive utility of non-angiogenesis signature in colorectal cancer.

[0227] A public array data set obtained from the Gene Expression Omnibus database for a cohort of recurrent or metastatic colorectal cancer responders and non-responders to Bevacizumab on plus 2 arrays (E-GEOD-19862) was obtained and analyzed using the example 63 gene signature of Table 2C. The 63 gene ovarian immune signatures predicts response to bevacizumab with an AUC: 0.86 (0.60-1.00). See Figure 14.

Example 5: Predictive utility of the 25 and 45 angiogenesis gene signatures in colorectal cancer

[0228] In a public array data set obtained from the Gene Expression Omnibus database from a cohort of recurrent or metastatic colorectal cancer responders and non-responders to Bevacizumab on plus 2 arrays (E-GEOD-19862), the 25 gene and 45 gene angiogenic signatures both predict response to bevacizumab – see Figures 18 and 19 respectively.

References

1. Friedman HS, Prados MD, Wen PY, *et al.* Bevacizumab alone and in combination with irinotecan in recurrent glioblastoma. *J Clin Oncol*; **27**:4733–40 (2009).
2. Hurwitz H, Fehrenbacher L, Novotny W, *et al.* Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N Engl J Med*; **350**:2335–42 (2004).
3. Rini BI, Halabi S, Rosenberg JE, *et al.* Bevacizumab plus interferon alfa compared with interferon alfa monotherapy in patients with metastatic renal cell carcinoma: CALGB 90206. *J Clin Oncol*; **26**:5422–8 (2008).
4. Sandler A, Gray R, Perry MC, *et al.* Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. *N Engl J Med*; **355**: 2542–50 (2006).
5. Wolmark N, Yothers G, O'Connell MJ, *et al.* A phase III trial comparing mFOLFOX6 to mFOLFOX6 plus bevacizumab in stage II or III carcinoma of the colon: results of NSABP protocol C-08. *J Clin Oncol*; **27**:LBA4 (2009).
6. Yang JC, Haworth L, Sherry RM, *et al.*, A randomized trial of bevacizumab, an anti-vascular endothelial growth factor antibody, for metastatic renal cancer, *N Engl J Med* **349** 427–434 (2003).
7. Willett CG, Boucher Y, di Tomaso E, *et al.*, Direct evidence that the VEGF-specific antibody bevacizumab has antivascular effects in human rectal cancer, *Nat. Med.* **10**, 145–147 (2004).
8. Miller K, Wang M, Gralow J, *et al.*, Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer, *N Engl J Med* **357** 2666–2676 (2007).
9. Miller KD, Chap LI, Holmes FA, *et al.*, Randomized phase III trial of capecitabine compared with bevacizumab plus capecitabine in patients with previously treated metastatic breast cancer, *J Clin Oncol* **23** 792–799 (2005).
10. O'Shaughnessy J, Miles D, Gray RJ, *et al.*, A meta-analysis of overall survival data from three randomized trials of bevacizumab (BV) and first-line chemotherapy as treatment for patients with metastatic breast cancer (MBC), *J Clin Oncol* **28** (suppl) (abstr 1005) (2010).
11. Reck M, von Pawel J, Zatloukal P, *et al.*, Phase III trial of cisplatin plus gemcitabine with either placebo or bevacizumab as first-line therapy for nonsquamous non-small-cell lung cancer: AVAIL, *J Clin Oncol* **27**, 1227–1234 (2009).
12. Escudier B, Bellmunt J, Negrier S *et al.*, Phase III trial of bevacizumab plus interferon alfa-2a in patients with metastatic renal cell carcinoma (AVOREN): final analysis of overall survival, *J Clin Oncol* **28**, 2144–2150 (2010)
13. Burger RA, Sill MW, Monk BJ, *et al.* Phase II trial of bevacizumab in persistent or recurrent epithelial ovarian cancer or primary peritoneal cancer: a Gynecologic Oncology Group Study. *J Clin Oncol*; **20**;25(33):5165-71 (2007).

- 14 Tothill RW *et al.* Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. *Clin Cancer Res.* **14**(16), 5198-208 (2008).

CLAIMS

We claim:

1. A method for predicting a subject's responsiveness to an anti-angiogenic therapeutic agent comprising:
 - obtaining a test sample from a subject;
 - measuring expression levels of a biomarker panel from the test sample obtained from the subject, wherein the biomarker panel comprises one or more biomarkers selected from Table 2C;
 - determining a sample expression score for the biomarker panel;
 - comparing the sample expression score to a threshold score; and
 - classifying the subject as responsive or non-responsive to the anti-angiogenic therapeutic agent based on whether the sample expression score is above or below the threshold expression score.
2. The method of claim 1, wherein the biomarker panel comprises one or more of IGF2, SOX11, INS, CXCL17, SLC5A1, TMEM45A, CXCR2P1, MFAP2, MATN3, or RTP4.
3. The method of claim 1, wherein the biomarker panel comprises the biomarkers listed in Table 2C.
4. The method of any one of claims 1 to 3, wherein the sample expression score is determined by measuring an expression level for each biomarker and multiplying it by a corresponding weight, wherein an expression signature defines the weight for each biomarker.
5. The method of any one of claims 1 to 4, wherein the expression level of each biomarker in the biomarker panel is measured using a microarray, quantitative PCR, or an immunoassay.

6. The method of claim 4, wherein the expression signature is derived by a method comprising:

isolating total RNA from a sample set of diseased tissue;

hybridizing the isolated total RNA to a microarray to obtain a sample expression data set;

selecting those probes on the microarray with a variability above a defined significance threshold to form a preliminary biomarker set;

generating clusters of biomarkers within the preliminary biomarker set having a similar expression profile using a clustering algorithm;

identifying the biological processes or biological pathways for each cluster of biomarkers;

selecting the cluster corresponding to the biological process or biological pathway of interest; and

defining an expression signature by analyzing the expression levels of the biomarkers in the selected cluster in a the sample set of diseased tissue using a supervised or unsupervised training algorithm.

7. The method of claim 6, wherein the expression signature is defined using a PLS classifier, a SVM classifier, a SDA classifier, or a DSDA classifier.

8. The method of any one of claims 1 to 7, wherein the diseased tissue is cancer tissue.

9. The method of claim 8, wherein the cancer is ovarian cancer, prostate, breast cancer, colon cancer, or colorectal cancer.

10. The method of any one of claims 1 to 8, wherein the anti-angiogenic therapeutic agent is a VEGF-pathway-targeted therapeutic agent, an angiopoietin-TIE2 pathway inhibitor, an endogenous angiogenic inhibitor, or an immunomodulatory agent.

11. The method of claim 10, wherein the VEGF pathway-targeted therapeutic agents include Bevacizumab (Avastin), Afibercept (VEGF Trap), IMC-1121B (Ramucirumab), Imatinib (Gleevec), Sorafenib (Nexavar), Gefitinib (Iressa), Sunitinib (Sutent), Erlotinib, Tivozinib, Cediranib (Recentin), Pazopanib (Votrient), BIBF 1120 (Vargatef), Dovitinib, Semaxanib (Sugen), Axitinib (AG013736), Vandetanib (Zactima), Nilotinib (Tasigna), Dasatinib (Sprycel), Vatalanib, Motesanib, ABT-869, TKI-258 or a combination thereof.

12. The method of claim 10, wherein the angiopoietin-TIE2 pathway inhibitor includes AMG-386, PF-4856884 CVX-060, CEP-11981, CE-245677, MEDI-3617, CVX-241, Trastuzumab (Herceptin) or a combination thereof.

13. The method of claim 10, wherein the endogenous angiogenic inhibitors include Thrombospondin, Endostatin, Tumstatin, Canstatin, Arrestin, Angiostatin, Vasostatin, Interferon alpha or a combination thereof.

14. The method of claim 10, wherein the immunomodulatory agents include thalidomide and lenalidomide.

15. The method of any one of claims 1 to 14, further comprising determining the subject's prognosis, wherein a classification as responsive to the anti-angiogenic therapeutic agent indicates a good prognosis, the good prognosis indicating at least an increased expected survival time as compared to a classification as non-responsive to the anti-angiogenic therapeutic agent.

16. A method for predicting subject responsiveness to anti-angiogenic therapeutic agents and providing treatment prognosis, the method comprising:

obtaining a test sample from a subject with a disease or disorder;

measuring expression levels of a first biomarker panel from the test sample of diseased tissue obtained from the subject, wherein the biomarker panel comprises one or more biomarkers selected from Table 2C;

determining a first sample expression score for the first biomarker panel;

comparing the first sample expression score to a first threshold score

classifying the subject as having a good prognosis or bad prognosis based on whether the first sample expression score is above or below the first threshold score, wherein a good prognosis indicates at least an increased expected survival time as compared to a bad prognosis and wherein, in the event of a bad prognosis; measuring expression levels of a second biomarker panel from the test sample of diseased tissue obtained from the subject, wherein the second biomarker panel comprises one or more biomarkers selected from Table 2A or Table 2B;

determining a second sample expression score for the second biomarker panel;

comparing the second sample expression score to a second threshold score; and

classifying the subject as responsive or non-responsive to an anti-angiogenic therapeutic agent based on whether the second sample expression score is above or below the second threshold expression score.

17. The method of claim 16, wherein the first biomarker panel comprises one or more of IGF2, SOX11, INS, CXCL17, SLC5A1, TMEM45A, CXCR2P1, MFAP2, MATN3, or RTP4, and the second biomarker panel comprises one or more of CCDC80, INHBA, THBS2, SFRP2, MMP2, PLAU, FAP, FN1, COL8A1, and RAB31.

18. The method of claim 16, wherein the first biomarker panel comprises one or more of IGF2, SOX11, INS, CXCL17, SLC5A1, TMEM45A, CXCR2P1, MFAP2, MATN3, or RTP4, and the second biomarker panel comprises one or more of TMEM200A, GJB2, MMP13, GFPT2, POSTN, BICC1, CDH11, MRVI1, PMP22, and COL11A1.

19. The method of claim 16, wherein the first biomarker panel consists of the biomarkers in Table 2C, and the second biomarker panel consists of the biomarkers in Table 2A or 2B.

20. The method of any one of claims 16 to 19, wherein the expression level of each biomarker in the biomarker panel is measured using a microarray, quantitative PCR, or an immunoassay.

21. A method for diagnosing a subject as having a cancer, or being susceptible to developing a cancer, that is responsive to an anti-angiogenic therapeutic, the method comprising;

obtaining a test sample from a subject having or suspected of having a cancer;

measuring expression levels of a biomarker panel from the test sample obtained from the subject, wherein the biomarker panel comprises one or more biomarkers selected from Table 2C;

determining a sample expression score for the biomarker panel;

comparing the sample expression score to a threshold score; and

diagnosing the subject as having a cancer, or being susceptible to developing a cancer that is responsive to an anti-angiogenic therapeutic based on whether the sample expression score is above or below the threshold expression score.

22. The method of claim 21, wherein the biomarker panel comprises one or more of IGF2, SOX11, INS, CXCL17, SLC5A1, TMEM45A, CXCR2P1, MFAP2, MATN3, or RTP4.

23. The method of claim 21, wherein the biomarker panel comprises the biomarkers listed in Table 2C.

24. The method of any one of claims 21 to 23, wherein the sample expression score is determined by measuring an expression level for each biomarker and multiplying it by a corresponding weight, wherein an expression signature defines the weight for each biomarker.

25. The method of claim 24, wherein the expression signature is defined using a PLS classifier, a SVM classifier, a SDA classifier, or a DSDA classifier.

26. The method of any one of claims 21 to 25, wherein the cancer is ovarian cancer, glioblastoma, breast cancer, prostate cancer, colorectal cancer or colon cancer.

27. The method of any one of claims 21 to 26, wherein the expression level of each biomarker in the biomarker panel is measured using a microarray, quantitative PCR, or an immunoassay.