(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 August 2015 (27.08.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/127101 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2015/016650

(22) International Filing Date:
19 February 2015 (19.02.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/966,271 19 February 2014 (19.02.2014) US

(71) Applicant: **THE TRUSTEES OF COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK** [US/US]; 412 Low Memorial Library, 535 West 116th Street, New York, NY 10027 (US).

(72) Inventors: **ABATE-SHEN, Corinne**; 737 Saint Nicholas Ave., New York, NY 10031 (US). **CALIFANO, Andrea**; 251 West 19th Street, Apt. 8C, New York, NY 10011 (US). **SHEN, Michael**; 737 Saint Nicholas Ave., New York, NY 10031 (US).

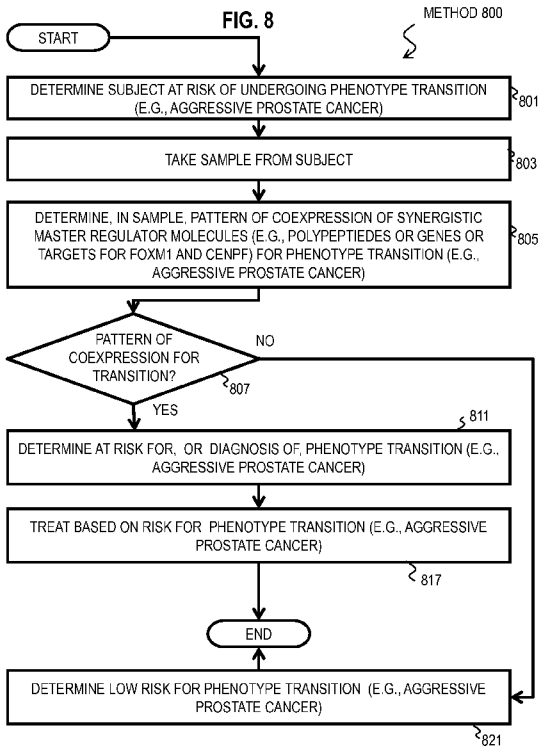
(74) Agent: **DAVITZ, Michael, MD, JD**; Ascenda Law Group PC, 333 W. San Carlos ST, Ste 200, San Jose, CA 95110 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK,

[Continued on next page]

(54) Title: METHOD AND COMPOSITION FOR DIAGNOSIS OF AGGRESSIVE PROSTATE CANCER



(57) Abstract: Techniques for diagnosis of aggressive prostate cancer include determining a level of expression of each of the genes encoding (FOXM1) Forkhead box protein M1 and Centromere protein F (CENPF) in a test sample. If the level of expression of each of the FOXM1 and CENPF genes in the test sample is at least 35% higher than the corresponding level in a control sample, then it is determined that the subject has an aggressive form of prostate cancer or has a high risk of prostate cancer progressing to an aggressive form. Alternatively, if at least 50% of prostate cancer cells in the sample express both FOXM1 protein and CENPF protein at a composite score of at least 100 for each, then the above diagnosis is made. Composite score is calculated by multiplying a percent staining value by a staining intensity value.

WO 2015/127101 A1

SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

METHOD AND COMPOSITION FOR DIAGNOSIS OF AGGRESSIVE PROSTATE CANCER

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims benefit of Provisional Appln. 61/966,271, filed February 19, 2014 under 35 U.S.C. §119(e).

STATEMENT OF GOVERNMENTAL INTEREST

[0002] This invention was made with Government support under CA084294, U54 CA121852 and CA154293 awarded by the National Institutes of Health. The Government has certain rights in the invention.

BACKGROUND

[0003] Cancer is not a single entity but rather a highly individualized spectrum of diseases characterized by a number of genetic and genomic alterations (Hanahan and Weinberg, 2011). Distinguishing molecular alterations that constitute true drivers of cancer phenotypes from the multitude that are simply de-regulated has proven to be a daunting task, which is further exacerbated by the complexity of elucidating how such drivers interact synergistically to elicit cancer phenotypes. Prostate cancer is particularly challenging because its notorious heterogeneity, combined with a relative paucity of recurrent gene mutations, has made prostate cancer especially difficult to identify molecularly distinct subtypes with known clinical outcomes (Baca et al., 2013; Schoenborn et al., 2013; Shen and Abate-Shen, 2010). Additionally, while early-stage prostate tumors are readily treatable (Cooperberg et al., 2007), advanced prostate cancer frequently progresses to castration-resistant disease, which is often metastatic and nearly always fatal (Ryan and Tindall, 2011; Scher and Sawyers, 2005).

[0004] It should be noted that several factors, including an increase in the aging population and widespread screening for prostate specific antigen (PSA), have contributed to a substantial rise in diagnoses of prostate cancer. The primary means of determining the appropriate treatment course for men diagnosed with prostate cancer still relies on Gleason

grading, a histopathological evaluation that lacks a precise molecular correlate. While patients with high Gleason score (Gleason ≥8) tumors are recommended to undergo immediate treatment, the appropriate treatment for those with low (Gleason 6) or intermediate (Gleason 7) Gleason score tumors remains more ambiguous. Indeed, although the majority of Gleason grade 6 tumors, as well as many Gleason grade 7 tumors, are likely to remain indolent (i.e., low-risk, non-aggressive or non-invasive), a minority (~ 10%) will progress to aggressive disease.

[0005] Indeed, the current lack of reliable and reproducible assays to identify tumors destined to remain indolent versus those that are aggressive, has resulted in substantial overtreatment of patients that would not die of the disease if left untreated. Consequently, "active surveillance" has emerged as an alternative for monitoring men with indolent prostate cancer, with the goal of avoiding treatment unless there is evidence of disease progression. The obvious advantage of active surveillance is that it avoids overtreatment; however, the potential concern is that it may miss the opportunity for early intervention for patients with aggressive tumors. Therefore, better methods with a molecular correlate for diagnosing aggressive prostate cancer have great value.

SUMMARY

[0006] Applicants have determined that there is a need to identify molecular determinants of cancers, including but not limited to, aggressive prostate cancer subtypes, a need to identify other prognostic biomarkers of disease outcome, and a need to treat such cancers. The subject matter disclosed herein addresses this need.

[0007] In a first set of embodiments, a method includes obtaining a test prostate cancer sample from a subject having prostate cancer and determining a level of expression of each of the genes encoding (FOXM1) Forkhead box protein M1 and Centromere protein F (CENPF) in the test sample and a control sample. The method also includes comparing the level of expression of each of the FOXM1 and CENPF genes in the test sample to the corresponding level in the control sample. The method further includes determining that the subject has an aggressive form of prostate cancer or has a high risk of prostate cancer progressing to an aggressive form, if the level of expression of each of the FOXM1 and CENPF genes in the test sample is at least 35% higher than the corresponding level in the control sample.

[0008] In a second set of embodiments, a method includes obtaining a prostate cancer sample from a subject having prostate cancer, and determining a level of expression of FOXM1 protein and CENPF protein in the prostate cancer sample by immunostaining with a first antibody that specifically binds to FOXM1 and a second antibody that specifically binds to CENPF. The method further includes determining that the subject has an aggressive form of prostate cancer or has a high risk of prostate cancer progressing to an aggressive form, if at least 50% of prostate cancer cells in the prostate cancer sample express both FOXM1 protein and CENPF protein at a composite score of at least 100 for each protein. The composite score is calculated by multiplying a percent staining value by a staining intensity value.

[0009] In a third set of embodiments, a method includes obtaining a prostate cancer sample from a subject having prostate cancer (or at risk of developing prostate cancer). The method also includes applying a first antibody that specifically binds to FOXM1 protein in the sample, wherein presence of FOXM1 creates an antibody-FOXM1 complex; and applying a second antibody that specifically binds to CENPF in the sample, wherein presence of the

CENPF creates an antibody-CENPF complex. The method further includes applying a first detection agent that detects the antibody-FOXM1 complex; and a second detection agent that detects the antibody-CENPF complex. The method still further includes then determining that the subject has an aggressive form of prostate cancer or has a high risk of prostate cancer progressing to an aggressive form, if at least 50% of prostate cancer cells in the sample express both FOXM1 protein and CENPF protein at a composite score of at least 100 for each protein.

[0010] In a fourth set of embodiments, a diagnostic kit for detecting an expression level of an mRNA or a protein encoding FOXM1 or CENPF or both in a biological sample inlcudes oligonucleotides that specifically hybridize to each of the respective mRNAs or one or more agents that specifically bind to each of the respective proteins, or both.

[0011] In a fifth set of embodiments, a microarray includes a plurality of oligonucleotides that specifically hybridize to an mRNA encoded by each of the FOXM1 or CENPF genes, which oligonucleotides are fixed on the microarray.

[0012] In a sixth set of embodiments, a microarray includes a plurality of antibodies or antibody fragments that specifically bind to either or both of FOXM1 protein or CENPF protein or biologically active fragment thereof, which antibodies or antibody fragments are fixed on the microarray.

[0013] Still other aspects, features, and advantages of the invention are readily apparent from the following detailed description, simply by illustrating a number of particular embodiments and implementations, including the best mode contemplated for carrying out the invention. The invention is also capable of other and different embodiments, and its several details can be modified in various obvious respects, all without departing from the spirit and scope of the invention. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] The present invention is illustrated by way of example, and not by way of limitation, in the figures of the accompanying drawings and in which like reference numerals refer to similar elements and in which:

[0015] FIG. 1A is a block diagram and graph that illustrate example gene profiling data for multiple human subjects with various stages of prostate cancer, according to an embodiment;

[0016] FIG. 1B is a block diagram and graph that illustrate example gene profiling data for multiple mouse models for prostate cancer, according to an embodiment;

[0017] FIG. 1C is a block diagram and graph that illustrate example effects on gene profiling data for mouse models in response to various perturbagens, according to an embodiment;

[0018] FIG. 2A is a block diagram and graph that illustrate example interactomes for human and mouse models with prostate cancer, according to an embodiment;

[0019] FIG. 2B is a graph that illustrates example percentage of the interactomes that are conserved between human and mouse models with prostate cancer, according to an embodiment;

[0020] FIG. 3A is a Venn diagram and table that illustrate example selection of a subset of master regulators from a full set determined by available automated computer processes, according to an embodiment;

[0021] FIG. 3B is a diagram that illustrates example ranking of master regulators for their impacts on prostate cancer, according to an embodiment;

[0022] FIG. 3C is a table that illustrates example ranking of master regulators for their impacts on prostate cancer by various available algorithms, according to an embodiment;

[0023] FIG. 4 is a table that illustrates example predicted synergy of FOXM1 And CENPF among the subset of master regulators using available algorithms, according to an embodiment;

[0024] FIG. 5 is a table that illustrates example clinical datasets used to determine whether synergistic master regulators FOXM1 and CENPF are prognostic biomarkers of prostate cancer outcomes, according to an embodiment;

[0025] FIG. 6A is an image that illustrates example micrographs of FOXM1 and CENPF stained tissues showing enhanced concentrations of both in aggressive prostate cancer tumors compared to other prostate tumors, according to an embodiment;

[0026] FIG. 6B is an image that illustrates example micrographs of FOXM1 and CENPF stained tissues showing enhanced concentrations of both in metastasized lung and liver tumors, according to an embodiment;

[0027] FIG. 6C through FIG. 6E are graphs that illustrate example Kaplan-Meier survival analysis based on protein expression levels of FOXM1 and CENPF with respect to time to biochemical recurrence, time to prostate cancer-specific death, or time to metastatic progression, respectively, according to an embodiment;

[0028] FIG. 6F and FIG. 6G are graphs that illustrate example Kaplan-Meier survival analysis based on the interactome-inferred activity levels of FOXM1 and CENPF with respect to time to biochemical recurrence, or time to prostate cancer-specific death, respectively, according to an embodiment;

[0029] FIG. 7 is a table that illustrates example prognostic power of co-expression of protein levels of FOXM1 and CENPF, with death due to prostate cancer and time to metastasis as evaluation endpoints, according to an embodiment;

[0030] FIG. 8 is a flow chart that illustrates an example diagnostic method for determining whether a subjects is at risk based on coexpression of the synergistic master regulators FOXM1 and CENPF, according to an embodiment;

[0031] FIG. 9A is a graph that illustrates example resulting relative mRNA expression levels for the shared targets of FOXM1 and CENPF in the indicated cell lines following individual or co-silencing of FOXM1 and CENPF, according to an embodiment;

[0032] FIG. 9B is a graph that illustrates example enrichment of FOXM1 binding normalized to input with and without silencing of CENPF, according to an embodiment;

[0033] FIG. 9C is an image of micrographs that illustrate example changes in subcellular localization of FOXM1 and CENPF proteins in prostate cancer cells after silencing either, according to an embodiment; and

[0034] FIG. 10 is a flow chart that illustrates an example method for determining coexpression of the synergistic master regulators FOXM1 and CENPF, according to an embodiment.

DEFINITIONS

[0035] 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 invention belongs. The practice of the present invention will employ, unless indicated specifically to the contrary, conventional methods of molecular biology and recombinant DNA techniques within the skill of the art, many of which are described below for the purpose of illustration. Such techniques are fully explained in the literature. See, e.g., Singleton et al., Dictionary of Microbiology and Molecular Biology 3rd.sup.ed., J. Wiley & Sons (2001); March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 5th.sup.ed., J. Wiley & Sons (2001); Sambrook & Russell, eds., Molecular Cloning: A Laboratory Manual 3rd ed., Cold Spring Harbor Laboratory Press (2001); Glover, ed., DNA Cloning: A Practical Approach, vol. I & II (2002); Gait, ed., Oligonucleotide Synthesis: A practical approach, Oxford University Press (1984); Herdewijn, ed., Oligonucleotide Synthesis: Methods and Applications, Humana Press (2005); Hames & Higgins, eds., Nucleic Acid Hybridisation: A Practical Approach, IRL Press (1985); Buzdin & Lukyanov, eds., Nucleic Acid Hybridization: Modern Applications, Springer (2007); Hames & Higgins, eds., Transcription and Translation: A Practical Approach, IRL Press (1984); Freshney, ed., Animal Cell Culture, Oxford UP (1986); Freshney, Culture of Animal Cells: A Manual of Basic Technique and Specialized Applications, 6th ed., John Wiley & Sons (2010); Perbal, A Practical Guide to Molecular Cloning, 3rd ed., Wiley-Liss (2014); Farrell, RNA Methodologies: A Laboratory Guide for Isolation and Characterization, 3rd ed., Elsevier/Focal Press (2005); Lilley & Dahlberg, eds., Methods in Enzymology: DNA Structures, Part A: Synthesis and Physical Analysis of DNA, Academic Press (1992); Harlow & Lane, Using Antibodies: A Laboratory Manual: Portable Protocol no. 1, Cold Spring Harbor Laboratory Press (1999); Harlow & Lane, eds., Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1988); Seethala & Fernandes, eds., Handbook of Drug Screening, Marcel Dekker (2001); and Roskams & Rodgers, eds., Lab Ref: A Handbook of Recipes, Reagents, and Other Reference Tools for Use at the Bench, Cold Spring Harbor

Laboratory (2002) provide one skilled in the art with a general guide to many of the terms used in the present application.

[0036] One skilled in the art will recognize many methods and materials similar or equivalent to those described herein, which could be used in the practice of the present invention. Other features and advantages of the invention will become apparent from the following detailed description, taken in conjunction with the accompanying drawings, which illustrate, by way of example, various features of embodiments of the invention. Indeed, the present invention is in no way limited to the methods and materials described. For convenience, certain terms employed herein in the specification, examples and appended claims are collected here.

[0037] Unless stated otherwise, or implicit from context, the following terms and phrases include the meanings provided below. Unless explicitly stated otherwise, or apparent from context, the terms and phrases below do not exclude the meaning that the term or phrase has acquired in the art to which it pertains. The definitions are provided to aid in describing particular embodiments, and are not intended to limit the claimed invention, because the scope of the invention is limited only by the claims. Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

[0038] The term "nucleic acid" as used herein refers to any natural and synthetic linear and sequential arrays of nucleotides and nucleosides, cDNA, genomic DNA, mRNA, oligonucleotides and derivatives thereof. The term "nucleic acid" further includes modified or derivatized nucleotides.

[0039] An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid molecule, namely cancerous or noncancerous biological samples. An "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[0040] As used herein "oligonucleotide" refers to an oligomer or polymer of ribonucleic acid (RNA) or deoxyribonucleic acid (DNA) or mimetics thereof. This term includes oligonucleotides composed of naturally-occurring nucleobases, sugars and covalent internucleoside (backbone) linkages as well as oligonucleotides having non-naturally-occurring portions which function similarly. Such modified or substituted oligonucleotides are often preferred over native forms because of desirable properties such as, for example, enhanced cellular uptake, enhanced affinity for nucleic acid target and increased stability in the presence of nucleases.

[0041] As used herein an "inhibitory oligonucleotide" includes antisense, siRNA, shRNA, ribozymes and MIRs or other oligonucleotide that reduces the expression of a targeted FOXM1 or CENPF gene or protein.

[0042] "Biological sample" refers to a sample of prostate cells. The sample can be prostate cancer cells, for example, those taken from a prostate biopsy from a subject having prostate cancer, or of normal prostate cells either taken from a normal control subject or in some embodiments from a noncancerous area of the prostate of the subject having prostate cancer. In other embodiments, the biological sample comprises circulating prostate cancer cells isolated from the blood, cerebrospinal fluid (CSF) or serum of a subject having prostate cancer or exosomes derived from prostate cancer cells.

[0043] "Indolent prostate cancer" means low-risk, non-aggressive or non-invasive prostate cancers which would not lead to subject death if left untreated.

[0044] "Aggressive prostate cancer" means prostate cancer that leads to a shortened life expectancy of the subject or an increased occurrence of metastasis to other tissue cancers.

[0045] "At high risk of progressing to aggressive prostate cancer" means that the subject has prostate cancer that, more likely than not, is or will become aggressive prostate cancer.

.

[0046] A "subject" is a mammal, typically a human, but optionally a mammalian animal of veterinary importance, including but not limited to horses, cattle, sheep, dogs, and cats. In some embodiments a "subject" refers to either one who has been previously diagnosed with or identified as suffering from prostate cancer or to one who does not have prostate cancer, i.e., a normal or control subject.

[0047] As used herein, the term "diagnosis" includes the detection, typing, monitoring, dosing, and comparison at various stages of prostate cancer in a subject. Diagnosis includes the assessment of a predisposition or risk of developing an aggressive form of prostate cancer.

[0048] As used herein, the terms "treat," "treatment," "treating," or "amelioration" when used in reference to prostate cancer refer to therapeutic treatments for the prostate cancer, wherein the object is to reverse, alleviate, ameliorate, inhibit, slow down or stop the progression of the prostate cancer to an aggressive form, or reduce the severity of a symptom or condition. The term "treating" includes reducing or alleviating at least one adverse effect or symptom of a condition. Treatment is generally "effective" if one or more symptoms or clinical markers such as prostate-specific antigen (PSA) are reduced.

[0049] "FOXM1" as used herein refers to Forkhead box protein M1 is a protein that in humans is encoded by theFOXM1 gene. The protein encoded by this gene is a member of the FOX family of transcription factors. FOXM1 is also referred to as FKHL16; FOXM1B; HFH-11; HFH11; HNF-3; INS-1; MPHOSPH2; MPP-2; MPP2; PIG29; TGT3; TRIDENT. The human and mouse reference mRNA sequences are NM_001243088 (SEQ ID NO: 49) and NM_008021 (SEQ ID NO: 51), respectively. The human and mouse protein sequences are NP_001230017 (SEQ ID NO: 50) and NP_032047 (SEQ ID NO: 52), respectively. For the purpose of the methods and compositions of the invention, "FOXM1 protein" includes orthologs (analogs in different species).

[0050] "CENPF" as used herein refers to centromere protein F, a protein that in humans is encoded by the CENPF gene. The CENPF protein associates with the centromere-kinetochore complex. The protein is a component of the nuclear matrix during the G2 phase of interphase. CENPF is also referred to as CENF; PRO1779; hcp-1. The human and mouse

reference mRNA sequences are NM_016343 (SEQ ID NO: 53) and NM_001081363 (SEQ ID NO: 55) respectively; and the human and mouse protein sequences are NP_057427 (SEQ ID NO: 54) and NP_001074832 (SEQ ID NO: 56), respectively. For the purpose of the methods and compositions of the invention, "CENPF protein" includes orthologs (analogs in different species).

[0051] "Protein expression" refers to expression of protein as measured quantitatively by methods including without limitation Western blot, 2-dimensional SDS-PAGE and mass spectrometry.

[0052] "mRNA expression" refers to the expression of mRNA that can be measured quantitatively by methods including but not limited to nuclease protection assays, northern blots, real time quantitative PCR, and in-situ hybridization.

[0053] "Control level" and "normal level of expression" as used herein refer to a level or range of levels of FOXM1 or CENPF expressed in normal prostate tissue or indolent prostate cancer tumors.

[0054] "Threshold" or "threshold level" as used herein refers to a level or range of levels that separate normal level of expression of FOXM1 and CENPF from a pattern, level or ranges of levels of expression of FOXM1 and CENPF that indicate a high risk of aggressive prostate cancer. When the levels of expression of FOXM1 and CENPF are equal to or greater than the threshold level then it is determined that the subject is at high risk of developing aggressive prostate cancer or has aggressive prostate cancer, and vice versa.

[0055] "Protein" as used herein is a generic term referring to and used interchangeably with biologically active native protein, fragments, peptides, or analogs thereof.

[0056] "Subcellular localization of FOXM1 and CENPF" as used herein refers to the presence of FOXM1 and CENPF inside a cell. "Colocalization" means that both FOXM1 and CENPF are present in the same cell or if so designated, in the same subcellular compartment, for example colocalization in the nucleus or cytoplasm.

[0057] "Master regulator" as used herein refers to a protein that acts to drive any intermediary proteins in a key signaling pathway for a phenotype transition, such as a transition from indolent prostate cancer cell to an aggressive prostate cancer cell.

[0058] "Synergistic master regulator" as used herein refers to multiple master regulators that together have a measured effect greater than a predicted sum of their individual measured effects.

[0059] "Cross-species computational analysis" as used herein refers to automatically searching molecular interaction networks ("interactomes") of two or more species, such as human and mouse models for human cells, using a computer system to discover interactions present ("conserved") in both species.

[0060] The term "probe" refers to any molecule which is capable of selectively binding to a specifically intended target molecule, for example, an oligonucleotide probe that specifically hybridizes to a prognostic biomarker mRNA such as CENPF or FOXM1, or an antibody that specifically binds CENPF or FOXM1. Probes can be either synthesized by one skilled in the art, or derived from appropriate biological preparations. For purposes of detection of the target molecule, probes may be specifically designed to be labeled, as described herein. Examples of molecules that can be utilized as probes include, but are not limited to RNA, DNA, RNA/DNA chimeras, proteins, antibodies, and organic molecules.

[0061] Unless otherwise specified, the terms "antibody" and "antibodies" broadly encompass naturally-occurring forms of antibodies (e.g., IgG, IgA, IgM, IgE) and recombinant antibodies such as single-chain antibodies, chimeric and humanized antibodies and multi-specific antibodies, as well as fragments and derivatives of all of the foregoing, which fragments and derivatives have at least an antigenic binding site. Antibody derivatives may comprise a protein or chemical moiety conjugated to an antibody moiety.

DETAILED DESCRIPTION

[0062] A method, composition of matter, article of manufacture and apparatus are described for discovery of synergistic master regulators and the diagnosis of aggressive prostate cancer. In the following description, for the purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the present invention. It will be apparent, however, to one skilled in the art that the present invention may be practiced without these specific details. In other instances, well-known structures and devices are shown in block diagram form in order to avoid unnecessarily obscuring the present invention.

[0063] It has been discovered that the genes encoding FOXM1 and CENPF are prognostic biomarkers that are synergistic master regulators of aggressive prostate cancer in humans. Significantly elevated co-expression of both FOXM1 and CENPF genes in a prostate cancer sample at levels at least 35% above control levels is diagnostic of aggressive prostate cancer or of a high risk of developing aggressive prostate cancer, as is described in sample embodiments. Gene expression can be determined by mRNA or protein expression or combinations thereof. Regulatory drivers of prostate cancer malignancy were identified by assembling genome-wide regulatory networks (interactomes) for both human and mouse prostate cancer from expression profiling datasets of human tumors and genetically engineered mouse models, respectively. Cross-species computational analysis of these interactomes identified FOXM1 and CENPF as synergistic master regulators of prostate cancer malignancy that promote tumor growth by coordinated regulation of target gene expression and activation of key signaling pathways associated with prostate cancer malignancy. Thus, co-expression of FOXM1 and CENPF was identified for the first time as a robust prognostic indicator of aggressive prostate cancer with poor survival and metastasis.

[0064] Based on the data described herein, certain embodiments of the invention are directed to methods for diagnosing aggressive prostate cancer or a high risk of prostate cancer progressing to an aggressive form if the level of mRNA or protein expression for each of FOXM1 and CENPF in a prostate cancer sample from a subject is at least 35% higher than the corresponding level in a control prostate sample . In another embodiment aggressive

prostate cancer is diagnosed if at least 50% of the cells in the prostate cancer sample from a subject express elevated levels of both FOXM1 protein and CENPF protein.

1. Overview

[0065] While both FOXM1 and CENPF have been implicated in various cancers, the current work has uncovered a novel synergistic interaction that had not been previously anticipated. FOXM1 encodes a Forkhead domain transcription factor that is frequently over-expressed in many different types of cancer, including prostate, see (Alvarez-Fernandez and Medema, 2013; Halasi and Gartel, 2013a; Kalin et al., 2011; Koo et al., 2012), for reviews. Many previous studies have established a role for FOXM1 expression and activity in the regulation of cellular proliferation, DNA damage, genomic stability, drug resistance, and metastasis, and have shown that FOXM1 interacts with other key regulators such as β–Catenin and MYB (Lefebvre *et al.,* 2010; Zhang et al., 2011). In particular, the relevance of FOXM1 for prostate cancer has been shown by its gain- or loss-of-function *in vivo*, which elicit modest effects on tumor growth (Cai et al., 2013; Kalin et al., 2006).

[0066] CENPF (also known as mitosin or LEK1 in mouse), a known target of FOXM1, has also been implicated in various cancers, although not previously in prostate, and in some cases has been shown to undergo gene amplification and be associated with disease outcome (see Ma et al., 2006; Varis et al., 2006 for reviews). However, the actual functional role of CENPF has been more elusive and difficult to reconcile. In particular, while CENPF is named for its association with the centromere-kinetochore protein complex, such association is only transient. In fact, CENPF has been shown to have other functions, including regulation of mitosis and cellular proliferation (Bomont et al., 2005; Feng et al., 2006; Holt et al., 2005), which are mediated in part by protein interactions, including with members of the Retinoblastoma gene family as well as with the ATF transcription factor (see Ma et al., 2006; Varis et al., 2006 for reviews).

2. Cross Species Discovery of Regulatory Genes for Aggressive Prostate Cancer

[0067] To assemble a human prostate cancer interactome, gene expression profile data reported in (Taylor et al., 2010) was analyzed, which is ideally suited because: (i) it is

relatively large (n = 185) and diverse, including gene expression profiles from primary prostate tumors, adjacent normal prostate tissue, metastases, and cell lines; (ii) its primary tumors encompass the full range of pathological Gleason scores and have well-annotated clinical outcome data; and (iii) it displays extensive genetic diversity and tumor heterogeneity, as shown by t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis. Several characteristics of this dataset are described below with reference to FIG. 5. Notably, interactomes assembled from three alternative human prostate cancer datasets, also characterized in FIG. 5, were neither as complete nor as extensive. FIG. 1A is a block diagram and graph that illustrate example gene profiling data for multiple human subjects with various stages of prostate cancer, according to an embodiment. Details are set forth in Example 2.

[0068] Analysis of genetically engineered mouse models (GEMMs) can circumvent challenges associated with the inherent complexity of the more heterogeneous human cancer phenotypes. Investigations of mouse models of prostate cancer have contributed to characterization of disease-specific pathways, led to the identification of biomarkers of disease progression, and provided useful preclinical models for prevention and therapy (Irshad and Abate-Shen, 2013; Ittmann et al., 2013). Following the description of the first transgenic model of prostate cancer nearly 20 years ago, there are now numerous GEMMs that collectively model key molecular pathways de-regulated in human prostate cancer, and recapitulate the various stages of disease progression including pre-invasive lesions (prostate intraepithelial neoplasia, PIN), adenocarcinoma, castration-resistance, and metastasis (Irshad and Abate-Shen, 2013; Ittmann et al., 2013).

[0069] Inherent species differences often hinder direct comparative analyses of mouse models and human cancer. As described herein, a novel combination of computational approaches were applied to enable accurate cross-species integration of regulatory information from mouse to man in the context of prostate cancer. Recent advances in systems biology have led to the reverse engineering of regulatory networks (interactomes) that integrate large-scale datasets encompassing gene expression profiles, protein-protein interactions, genomic alterations, and epigenetic changes associated with cancer and other

diseases (see Lefebvre et al., 2012 for a review). While individual analyses of human and murine interactomes led to relevant biological discoveries, cross-species interactome-based interrogation strategies have not been systematically implemented until now.

[0070] The results described here are based on an approach for accurate cross-species analysis of conserved cancer pathways based on reverse engineering and interrogation of genome-wide regulatory networks (i.e., interactomes) representing both human and mouse prostate cancer. To accomplish this, the first regulatory network obtained from *in vivo* perturbation of a repertoire of mouse cancer models was introduced, as well as its comparative analysis with a complementary regulatory network generated from human prostate cancer samples. Cross-species computational interrogation of these paired interactomes followed by experimental validation thus elucidated the synergistic interaction of FOXM1 and CENPF as a driver of aggressive prostate cancer malignancy.

[0071] To assemble a corresponding mouse prostate cancer interactome, it was first necessary to generate an appropriately sized gene expression profile dataset representing sufficient expression variability. To address this challenge, 13 distinct GEMMs were first selected, which together represent the full spectrum of prostate cancer phenotypes, including normal epithelium (wild- type), low-grade PIN (Nkx3.1 and APT), high-grade PIN and adenocarcinoma (APT-P, APC, Hi- Myc, NP, Erg-R, and NP53), castration-resistant prostate cancer (NP-AI), and metastatic prostate cancer (NPB, NPK, and TRAMP). FIG. 1B is a block diagram and graph that illustrate example gene profiling data for multiple mouse models for prostate cancer, according to an embodiment. The diagram groups the mouse models by phenotype listed above. The graph plots the t-SNE analysis showing relative distribution of the GEMMs. More detail is set forth in Example 2.

[0072] To generate a sufficient number of samples, while further increasing the variability of the corresponding expression profiles, a controlled set of exogenous perturbations was introduced by *in vivo* administration of 13 different small-molecule perturbagens to each GEMM. Perturbagens were selected for their clinical relevance and/or ability to modulate key prostate cancer pathways, including: hormone signaling (testosterone, calcitriol, or enzalutamide); PI3 kinase activity (MK2206, LY294002, and rapamycin); MAP kinase

activity (PD035901); tyrosine kinase activity (imatinib, dasatinib, and sorafenib); NFκB signaling (BAY 11-7082); JAK/STAT activity (WP1066); and chemotherapy (docetaxel). Following pilot studies to define the appropriate dose and schedule that produced the broadest range of gene expression changes, a universal schedule was adopted of 1 treatment per day for 5 days with dosage determined independently for each perturbagen, as described below in an experimental procedures section.

[0073] The resulting dataset contained 384 gene expression profiles, corresponding to the 13 GEMMs each treated with the 13 perturbagens or vehicles. The t-SNE analysis revealed that the resulting mouse dataset represented an extensive range of gene expression variability, as requisite for ARACNe. Specifically, while expression profiles from the same GEMMs and perturbagens clustered together, suggesting their effect was highly replicable, the diverse GEMMs and perturbagens provided independent and highly effective axes of expression heterogeneity. FIG. 1C is a block diagram and graph that illustrate example effects on gene profiling data for mouse models in response to various perturbagens, according to an embodiment. The schematic diagram depicts perturbagens used to treat the GEMMs. The graph plots the t-SNE analysis showing the relative distribution of perturbagens for a representative GEMM (i.e., the NP model).

[0074] Regulatory networks (interactomes) for human and mouse prostate cancer were generated using the Algorithm for the Reconstruction of Accurate Cellular Networks. FIG. 2A is a block diagram and graph that illustrates example interactomes for human and mouse models with prostate cancer, according to an embodiment. The suitability of these mouse and human interactomes for cross-species interrogation was next evaluated by developing a novel computational approach to assess the global conservation of their transcriptional programs described in detail in Example 2. Notably, conserved transcriptional regulators included many genes known to play important roles in prostate cancer, such as AR, ETS1, ETV4, ETV5, STAT3, MYC, BRCA1, and NKX3.1. In particular, AR displayed extensive correlation of its transcriptional activity between the human and mouse interactomes, consistent with its known role as a key regulator of prostate development and prostate tumorigenesis

[0075] The Master Regulator Inference algorithm (MARINa) (Carro et al., 2010; Lefebvre et al., 2010) was used to infer candidate master regulators (MRs) that act individually or synergistically to drive malignant prostate cancer in the conserved interactomes. MARINa estimates differential activity (DA) based on enrichment (differential expression, DE) of their activated and repressed targets in the malignancy signature. More specifically, MARINa identified candidate MRs based on the concerted differential expression of their ARACNe-inferred targets (i.e., their differential activity, DA). Specifically, "activated" MRs have positively-regulated and repressed targets significantly enriched among upregulated and downregulated genes, respectively, while "repressed" MRs have the converse. To interrogate the human prostate cancer interactome, a gene signature was used representing prostate cancer malignancy from the Taylor dataset, which compares aggressive prostate tumors (Gleason score ≥ 8 with rapid biochemical recurrence; sample size n =10) versus indolent ones (Gleason score 6 tumors with no biochemical recurrence; sample size n = 39). The resulting independent lists of human and mouse MRs were then integrated to produce a ranked list of 20 conserved MRs, including 7 activated and 13 repressed (joint p-value: $p \leq 0.0074$ by Stouffer's method). Notably, these conserved MRs were more likely to be associated with disease outcome than the non-conserved ones, and were also more likely to be differentially expressed in aggressive prostate tumors (metastatic versus non-metastatic; 100% versus 60%). FIG. 3C is a table that illustrates example ranking of master regulators for their impact on prostate cancer by various available algorithms. Using the ARACNe method to analyze all possible pairs among the conserved activated MRs, the only pair that was found to be statistically significant was FOXM1 and CENPF. Both FOXM1 and CENPF were differentially co-expressed at significantly elevated levels in aggressive prostate tumors and were predicted to be significantly associated with disease outcome. Thus, subsequent analyses were focused on this pair of cross-species conserved, synergistic MRs.

3. Method of Diagnosis: FOXM1 and CENPF are prognostic biomarkers of aggressive prostate cancer

[0076] Analysis of high-density tissue microarrays (TMAs) revealed that the co-expression of FOXM1 and CENPF constituted a highly informative biomarker of poor disease outcome. FIG. 5 is a table that illustrates example clinical datasets used to determine whether synergistic master regulators FOXM1 and CENPF are prognostic biomarkers of prostate cancer outcomes, according to an embodiment. The datasets are listed along the top row, with their use in this study grouped by primary dataset (Taylor et al., 2010); secondary datasets; RNA gene expression datasets (Sboner et al., 2010; Glinsky et al., 2004); and protein immunohistochemistry tissue microarray (TMA) datasets (Outcome TMA from MSKCC, and Metastasis TMA from Michigan). The categories of data in each dataset are given by the rows, as applicable. One row breaks down the number of samples for each cell type; one row gives the median age of the subjects. The next rows give the Pathology T stage; the clinical T stage; the Pathology N stage; the Pathology Gleason score; the biopsy Gleason score; the survival index (SVI); the extracapsular extension percentage; the biochemical recurrence (BCR) median time in months; the median overall survival in months; and the median time to metastasis in months.

[0077] Analysis of protein expression of FOXM1 and CENPF was performed using high-density tissue primary tumor microarray (TMAs) (Donovan et al., 2008) and a metastasis TMA (Shah et al., 2004). Available clinico-pathological features of these cohorts as well as independent human datasets used for clinical validation are summarized in the Table of FIG. 5.

[0078] In particular, a high-density TMA containing primary tumors from a large cohort of subjects (sample size n = 916) that had undergone prostatectomy at Memorial Sloan-Kettering Cancer Center from 1985 to 2003 (Donovan *et al.*, 2008) was analyzed. These cases have extensive clinical follow-up data for up to 20 years, including time to biochemical recurrence, prostate-cancer specific survival, and time to metastasis. A second TMA was evaluated from the rapid autopsy program at the University of Michigan containing prostate cancer metastases (sample size n = 60), including 6 lung, 11 liver, 22 lymph node, and 14

other sites (Shah et al., 2004). Immunostaining for FOXM1 or CENPF was performed on adjacent sections of each TMA slide and staining intensity was evaluated (see experimental procedures section, below).

[0079] FIG. 6A is an image that illustrates example micrographs of FOXM1 and CENPF stained tissues showing enhanced concentrations of both in aggressive prostate cancer tumors compared to other prostate tumors, according to an embodiment. These micrographs are based on the MSKCC prostatectomy TMA; and, analysis revealed that FOXM1 and CENPF were over-expressed in 33% and 37% of all cases, respectively (sample size n = 821 informative cases), with a trend toward increased expression in tumors with higher Gleason scores. Spearman rank correlation coefficient of FOXM1 and CENPF protein expression levels was 0.57 with p value $< 2.2 \times 10^{-16}$, indicating the coexpression relationship is highly significant.

[0080] FIG. 6B is an image that illustrates example micrographs of FOXM1- and CENPF-stained tissues showing enhanced concentrations of both in prostate cancer that metastasized to lung and liver tumors. These micrographs are based on the Michigan metastasis TMA; and, analysis revealed that FOXM1 and CENPF were coexpressed in most of the prostate cancer metastases (88% and 90%, respectively, sample size n = 53 informative cases) at significantly elevated levels. Spearman rank correlation coefficient of FOXM1 and CENPF protein expression levels was 0.43 with p value < 0.001, indicating the coexpression is significant.

[0081] Thus, co-expression of FOXM1 and CENPF at above-threshold levels, particularly their nuclear colocalization, as described in more detail below, was well correlated in both the MSKCC prostatectomy TMA and the Michigan metastasis TMA. Additionally, both FOXM1 and CENPF were overexpressed at the mRNA level and their co-expression was well-correlated in advanced prostate cancer and metastases from independent cohorts of human prostate cancer.

[0082] To determine whether expression of FOXM1 and/or CENPF is associated with disease outcome on the MSKCC TMA, 4 groups of subjects were defined based on their expression levels: (i) low/normal expression of both FOXM1 and CENPF (sample size n =

418); (ii) high expression of FOXM1 and low/normal expression of CENPF (sample size n = 97); (iii) high expression of CENPF and low/normal expression of FOXM1 (sample size n = 133); and (iv) high expression of both FOXM1 and CENPF (sample size n = 173). FIG. 6C through FIG. 6E are graphs that illustrate example Kaplan-Meier survival analysis based on protein expression levels of FOXM1 and CENPF with respect to time to biochemical recurrence, time to prostate cancer-specific death, or time to metastatic progression, respectively, according to an embodiment.

[0083] Kaplan-Meier survival analysis of these subject groups revealed that those having elevated expression of both FOXM1 and CENPF were associated with the worst outcome with high significance (low values of p) for three independent clinical endpoints, namely, time to biochemical-free recurrence ($p \leq 4.4\times10^{-6}$), death due to prostate cancer ($p \leq 5.9\times 10^{-9}$), and time to metastasis ($p \leq 1.0\times10^{-16}$). The p-values correspond to a log-rank test and indicate the statistical significance of the association with outcome for each indicated branch compared to control (i.e., subjects with low protein expression of both FOXM1 and CENPF). Notably, co-subcellular localization of FOXM1 and CENPF in prostate tumors was also associated with the worst outcome for all three independent clinical endpoints, as described in more detail below. In contrast, elevated expression of only FOXM1 or CENPF was either not significant or marginally significant for biochemical recurrence and prostate-specific survival ($p \leq 0.053$ and $p \leq 0.011$ for FOXM1, respectively; $p \leq 0.078$ and $p \leq 0.402$ for CENPF, respectively), and was 10 to 13 orders of magnitude less significant, respectively, than co-expression for time to metastasis ($p \leq 0.001$ for FOXM1 and $p \leq 3.1\times10^{-6}$ for CENPF, respectively).

[0084] Association of FOXM1 and CENPF with disease outcome was independently corroborated in two independent human prostate cancer datasets that had not been used for training purposes elsewhere in this study; namely, the Glinsky dataset, in which biochemical recurrence is the clinical endpoint (Glinsky et al., 2004), and the Sboner dataset, in which the clinical endpoint is prostate cancer-specific overall survival (Sboner et al., 2010). Using these independent cohorts, the mRNA expression levels of FOXM1 and CENPF was evaluated as

well as their MARINa-inferred activity. Kaplan-Meier survival analysis was then performed on 4 subject groups: (i) those with low inferred activity or expression for FOXM1 and CENPF; (ii) those with high inferred activity or expression only for FOXM1; (iii) those with high inferred activity or expression only for CENPF; and (iv) those with high inferred activity or expression for both FOXM1 and CENPF. FIG. 6F and FIG. 6G are graphs that illustrate example Kaplan-Meier survival analysis based on the interactome-inferred activity levels of FOXM1 and CENPF with respect to time to biochemical recurrence, or time to prostate cancer-specific death, respectively, according to an embodiment.

[0085] Similar to the analysis of protein expression on the TMA, subjects with high inferred activity or mRNA expression for both CENPF and FOXM1 were associated with the worst outcome in both cohorts, as measured by biochemical recurrence ($p \leq 6.5\times10^{-5}$) and prostate cancer-specific survival ($p \leq 4.0\times10^{-5}$). The ARACNe-inferred activities levels were assessed for each subject sample in both cohorts. The p-values correspond to a log-rank test and indicate the statistical significance of the association with outcome for each indicated branch compared to control (i.e., subjects with low activity levels of both FOXM1 and CENPF). Notably, these findings reveal that their ARACNe-inferred activities are well-correlated with the actual expression of FOXM1 and CENPF proteins on the TMA, and further demonstrate the striking association of their co- expression/co-activity with poor disease outcome.

[0086] FIG. 7 is a table that illustrates example prognostic power of co-expression of protein levels of FOXM1 and CENPF, with death due to prostate cancer and time to metastasis as evaluation endpoints, according to an embodiment. C-statistics give the proportion of pairs in which the predicted event probability (e.g., probability of survival from prostate cancer) is higher for the subject who experienced the event of interest (e.g., coexpression of FOXM1 and CENPF) than that of the subject who did not experience the event. Analysis of co-expression of FOXM1 and CENPF on the MSKCC prostatectomy TMA using C-statistics revealed their robust prognostic value for disease-specific survival (C = 0.71; confidence interval = 0.59-0.84, $p \leq 2.4\times10^{-4}$), as well as time to metastasis (C = 0.77; confidence interval = 0.71-0.83, $p \leq 3.0\times10^{-19}$). Notably, co-expression of FOXM1 and CENPF proteins

as diagnostic markers of aggressive prostate cancer dramatically improved the prognostic value compared to Gleason score alone, for both disease-specific survival (C = 0.86; confidence interval =. 0.80-0.93, $p \leq 1.0\times10^{-30}$; p value for improvement, $p \leq 2.0\times10^{-4}$) and time to metastasis (C = 0.86; confidence interval = 0.81-0.89, $p \leq 6.5\times10^{-58}$; p value for improvement, $p \leq 5.3\times10^{-13}$). In certain embodiments of the invention, diagnosis of aggressive prostate cancer further includes determining, in addition to elevated coexpression of both FOXM1 and CENPF, high Gleason scores of score ≥ 8.

[0087] Taken together, these analyses of independent clinical cohorts using distinct statistical models demonstrate that elevated levels of co-expression of FOXM1 and CENPF is a good predictor of disease outcome. In an embodiment FOXM1 and CENPF are prognostic for aggressive prostate cancer or of prostate cancer progressing to an aggressive form when they are coexpressed at elevated levels of at least 35% compared to the levels expressed in control prostate tissue. Based on the results, certain embodiments are directed to a method for diagnosing aggressive prostate cancer or of identifying subjects with prostate cancer that is at high risk of progressing to an aggressive form by a) obtaining a test prostate cancer sample from a subject having prostate cancer, and a control prostate tissue sample,

b) determining a level of expression of the prognostic genes (FOXM1) Forkhead box protein M1 and Centromere protein F (CENPF) in the test and control samples, c) comparing the level of expression of prognostic genes FOXM1 and CENPF in the test sample to the corresponding level in the control sample, and d) if the level of expression of both of the prognostic genes FOXM1 and CENPF in the test sample is at least 35% higher than the corresponding level in the control sample, then determining that the subject has an aggressive form of prostate cancer or is at high risk of developing an aggressive form of prostate cancer.

[0088] In certain embodiments the level of expression of FOXM1 and CENPF is determined by the level of mRNA encoding FOXM1 and CENPF, respectively; or by the level of FOXM1 protein and CENPF protein in the sample or combinations thereof. In another embodiment, a diagnosis of aggressive prostate cancer is reached by a) obtaining a prostate cancer sample from a subject having prostate cancer, b) determining a level of expression of

FOXM1 protein and CENPF protein expression in the cancer cells in the sample by immunostaining with a first antibody that specifically binds to FOXM1 and a second antibody that specifically binds to CENPF, and c) diagnosing aggressive prostate cancer if at least 50% of the cells in the test sample express both FOXM1 protein and CENPF protein at a composite score of at least 100 for each protein, wherein the composite score is calculated by multiplying the percent staining value by the staining intensity value. Any other method for determining that at least 50% of the cells in a prostate cancer sample coexpress both FOXM1 and CENPF proteins at levels of at least 35% above levels in control prostate tissue can be used, these include flow cytometry with differential florescent labeling of both proteins.

[0089] FIG. 8 is a flow chart that illustrates an example diagnostic method 800 for determining whether a subject has or is at high risk of undergoing a specific phenotypic transition based on coexpression of at least two synergistic master regulators, such as FOXM1 and CENPF for aggressive prostate cancer, according to an embodiment. Although steps are depicted in FIG. 8, and in subsequent flowchart FIG. 10, as integral steps in a particular order for purposes of illustration, in other embodiments, one or more steps, or portions thereof, are performed in a different order, or overlapping in time, in series or in parallel, or are omitted, or one or more additional steps are added, or the method is changed in some combination of ways.

[0090] In step 810, a subject is identified who has or is at high risk of producing the phenotype transition of interest, here development of aggressive prostate cancer from a prostate tumor or nodule detected in a subject. In step 803 a sample is taken from the identified subject, such as a biopsy of the prostate tumor or nodule. Biological samples for use in the present embodiments also include circulating prostate cancer cells or prostate tumor cells which can be detected using a variety of methods known in the art that select cells based on surface markers for example using antibodies against the surface markers, or expression of other prostate cancer markers. In some embodiments the prostate cancer cells can be selected by size. Biological samples for use in the present embodiments also include exosomes derived from prostate cancer cells, which are cell-derived vesicles that are present

in many and perhaps all biological fluids, including blood. The reported diameter of exosomes is between 30 and 100 nm.

[0091] In step 805, the pattern of coexpression of synergistic master regulators, such as FOXM1 and CENPF, for the phenotype transition, such as to aggressive prostate cancer, is determined. For example, the pattern of certain genes' expression is determined by determining the relative level of coexpression of mRNA coding for the master regulators, or the intensity of immunostaining of polypeptides encoded by the master regulators. Step 805 is described in more detail below with reference to FIG. 10.

[0092] In step 807, it is determined whether the synergistic master regulators are coexpressed at significantly elevated (or reduced) levels above (below) some threshold level, or otherwise have different patterns than, determined in control prostate tissue. For diagnosis of aggressive prostate cancer or the risk of a tumor progressing to an aggressive form, it is determined whether FOXM1 and CENPF levels are both above corresponding threshold control levels, which are the levels seen in normal prostate tissue or indolent prostate cancer tumors. The threshold or pattern depends on the measurement type as described in more detail below with reference to FIG. 10.

[0093] If it is determined in step 807, that the synergistic master regulators are coexpressed at some elevated (or reduced) level or other different pattern, then control passes to step 811. Otherwise control passes to step 821.

[0094] In step 811, it is determined that the subject is at risk for developing the phenotype transition, or in fact is undergoing, or has undergone, the phenotype transition. In some embodiments the phenotype transition is to aggressive prostate cancer. Control then passes to step 817 to treat the subject based on this risk or diagnosis.

[0095] In step 821, it is determined that the subject does not have or is at low risk or no risk for developing the phenotype transition, e.g., aggressive prostate cancer; and, the process ends, or is repeated with the same subject at a later time or with another subject.

[0096] More information on the secondary effects of elevated co-expression of FOXM1 and CENPF was obtained by experiments that were performed to validate synergistic interactions of master regulators and elucidate underlying mechanisms, as well as to evaluate their

relevance for clinical outcome. FIG. 9A is a graph shows relative mRNA expression levels for the indicated genes in the indicated cell lines following individual or co-silencing (i.e., silencing both) of FOXM1 and CENPF. The p-values (indicated by one or two *) show the significance of the predicted additive effect versus actual observed effect on gene expression (* = $p<0.01$; ** = $p<0.001$). Silencing was performed using lentivirus vectors for shRNA for each or both of the two genes FOXM1 and CENPF, as described in more detail below. The ARACNe-inferred common target genes include BRCA1, BUB1, KI67, CYCLIN A, TIMELESS, CDC25, TRIP13, PLK1, HHMR, MYBL2, BIRC5, AURKA, AURKB.

[0097] Although target gene expression was somewhat reduced by their individual silencing, as shown in FIG. 9A, co-silencing of FOXM1 and CENPF produced a significantly greater reduction for the majority of targets, consistent with the synergistic regulation of target gene expression by FOXM1 and CENPF. Notably, these findings were observed in each cell line that express both FOXM1 and CENPF, but not in LNCaP cells which do not express CENPF.

[0098] In addition, analyses of genomic binding of FOXM1 to its known target sites using chromatin immunoprecipitation (ChIP) was followed by quantitative PCR analyses. FIG. 9B is a graph that illustrates example enrichment of FOXM1 binding normalized to input with and without silencing of CENPF, according to an embodiment. Cells were infected with a lentivirus expressing a V5-tagged FOXM1 plus shRNA CENPF (or a control) and ChIP was done using an anti-V5 antibody. Data are expressed as fold of enrichment of FOXM1 binding normalized to input. This revealed that FOXM1 binding to its targets was abrogated by silencing CENPF, therefore suggesting that CENPF is required for appropriate genomic binding by FOXM1.

[0099] Interestingly, although a direct protein-protein interaction of FOXM1 and CENPF in co-immunoprecipitation assays was not observed, it was observed that FOXM1 and CENPF were co- localized in the nucleus of prostate cancer cells and that their subcellular colocalization was mutually dependent. FIG. 9C is an image of micrographs that illustrate example changes in subcellular localization of FOXM1 and CENPF proteins in prostate cancer cells after silencing either, according to an embodiment. Shown are microphotographs of immunofluorescence staining for FOXM1 or CENPF in the control or silenced cells as

indicated. Arrows indicate subcellular localization or the shift in localization following silencing.

[0100] In particular, silencing of CENPF resulted in the redistribution of FOXM1 to the cytoplasm as well the nucleus, and conversely silencing of FOXM1 resulted in the accumulation of CENPF at the nuclear periphery. Notably, subcellular co-localization of FOXM1 and CENPF was also observed in human prostate tumors and associated with disease outcome. Taken together, these findings show that FOXM1 and CENPF synergistically regulate expression of mutual target genes, which mediated in part through their subcellular colocalization in prostate cancer cells.

Determining co-expression of the synergistic master regulators FOXM1 and CENPF

[0101] FIG. 10 is a flow chart that illustrates an example method 1000 for determining coexpression of the synergistic master regulators such as FOXM1 and CENPF, according to an embodiment. Method 1000 is a particular embodiment of step 805, described above. In step 1051 it is determined whether to use expression of the genes for the synergistic master regulators, or expression of the master regulator proteins (e.g., polypeptides included), or expression of the genes or polypeptides of one or more the targets in the signaling pathways of the master regulators, or combinations thereof.

[0102] In step 1053, it is determined whether the gene expression or polypeptide expression is to be evaluated. If the gene expression is to be determined, control passes to step 1061. Otherwise control passes to step 1071.

[0103] In step 1061, the normalized level of mRNA is determined for FOXM1 and CENPF in the prostate sample. Certain methods and primers for determining the relative or normalized levels of mRNA for FOXM1 and CENPF compared to other genes are also described in the experimental procedures section, below, with primer sequences listed in Table 1 in that later section. In an example, total RNA was isolated from a subject sample of prostate tissues/tumors using a MagMAX-96 total RNA isolation kit and biotin-labeled using the Illumina TotalPrep RNA Amplification Kit (Life Technologies). Slides were scanned using an iScan (Illumina) and the resulting files were uploaded and background-corrected in

BeadStudio 3.1.3.0 (Illumina, Inc.). Expression profiling data were normalized using standard variance stabilizing transformation (VST) and robust spline normalization (RSN) with lumiT and lumiN functions from lumi library, in R-system v2.14.0 (The R Foundation for Statistical Computing, ISBN 3-900051-07-0).

[0104] In step 1063, the threshold for significantly elevated coexpression is determined, e.g., retrieved from data storage. An mRNA level in the subject prostate cancer sample that is at least 35% higher than the level expressed in normal prostate tissue for each gene is considered elevated. In other embodiments other thresholds are used. In some embodiments, the threshold for each to be considered elevated is selected in a range from about 35% to about 100% or more. In some embodiments the threshold is at least 50%; and in other embodiments the threshold is at least 75%. Control then passes to step 807 of FIG. 8, described above, to determine if the measured level exceeds the threshold.

[0105] In step 1071, the level of immunostaining is determined for FOXM1 and CENPF proteins. Certain antibodies for immunostaining FOXM1 and CENPF are identified in the experimental procedures section below and listed in Table 2, and procedures for quantifying the intensity levels are also described. Any antibody that selectively binds to either FOXM1 or CENPF can be used. Protein levels were determined by percent of staining (e.g., from 0 to 100%) and intensity level of staining (e.g., 0, 1, 2, or 3) in each tumor sample. A composite protein level is determined by multiplying percent of staining and its intensity level for each tumor sample, for FOXM1 or CENPF. In some embodiments, step 1071 includes determining the relative amounts of FOXM1 and CENPF inside the membrane of the nucleus of the cells where the staining is observed in the top row of FIG. 9C when both FOXM1 and CENPF are expressed. In various embodiments, this determination of the relative amount within the nucleus is done in addition to, or instead of, determining the composite protein level.

[0106] In step 1073, the threshold for elevated coexpression is determined, e.g., retrieved from data storage. Composite protein level exceeding 100 for each protein was considered elevated. Thus, in some embodiments, FOXM1 and CENPF are considered to be co-

expressed if the determined composite protein level in the subject prostate tumor sample for each is above about 100.

[0107] In some embodiments, the pattern of coexpression is determined by the relative amount of the total FOXM1 and CENPF that is inside the same cell, or in a particular subcellular compartment, such as inside of the nuclear membrane. For example, if at least 50% of the cells in the test sample express both FOXM1 protein and CENPF protein at a composite score of at least 100 for each protein, then determining that coexpression is elevated.

[0108] Control then passes to step 807 of FIG. 8, described above, to determine if the measured level exceeds the threshold.

Method for detecting mRNA expression

[0109] In some embodiments, the methods described herein comprise detecting the presence of FOXM1 or CENPF RNA expression (e.g., mRNA expression), including detecting of the absolute or relative quantity of the RNA, the half-life of the RNA, a splicing or processing of the RNA, the nuclear export of the RNA or the sub-cellular location of the RNA. Such detection can be by various techniques known in the art, including by sequencing all or part of the FOXM1 or CENPF RNA or by selective hybridization or selective amplification of all or part of the FOXM1 or CENPF RNA. As described herein, there exist many suitable methods for detecting the presence and level of a nucleic acid encoding a FOXM1 polypeptide or a CENPF polypeptide, including, but not limited to genotyping a sample, for example via gene sequencing, selective hybridization, amplification, gene expression analysis (e.g. microarray analysis), oligonucleotide ligation assay, a confirmation based assay, a hybridization assay, a sequencing assay, an allele-specific amplification assay, a microsequencing assay, a melting curve analysis, a denaturing high performance liquid chromatography (DHPLC) assay (for example, see Jones et al., 2000), or a combination thereof. Sequencing can be carried out using techniques well known in the art, using automatic sequencers. The sequencing can be performed on the complete gene or on specific domains thereof, such as those known or suspected to carry deleterious mutations or other

alterations. Other suitable methods include allele-specific oligonucleotide (ASO), oligonucleotide ligation, allele-specific amplification, Southern blot (for DNAs), Northern blot (for RNAs), single-stranded conformation analysis (SSCA), PFGE, fluorescent in situ hybridization (FISH), gel migration, clamped denaturing gel electrophoresis, denaturing HLPC, melting curve analysis, heteroduplex analysis, RNase protection, chemical or enzymatic mismatch cleavage, ELISA, radio-immunoassays (RIA) and immuno-enzymatic assays (IEMA). Some other approaches are based on specific hybridization between nucleic acids from the subject and a probe specific for wild type gene or RNA. The probe can be in suspension or immobilized on a substrate. The probe can be labeled to facilitate detection of hybrids. Some of these approaches are suited for assessing a polypeptide sequence or expression level, such as Northern blot, ELISA and RIA. These latter require the use of a ligand-specific for the polypeptide, for example, the use of a specific antibody.

[0110] In certain embodiments, detection or quantification of a nucleic acid encoding a nucleic acid encoding a FOXM1 polypeptide or a CENPF polypeptide (or a fragment thereof) can be by hybridization based methods. In certain embodiments, hybridization-based detection methods can employ a step of forming specific hybrids between complementary nucleic acid sequences that serve to detect nucleic acid sequences. Microarrays are a suitable hybridization based detection technique that can be used in connection with the methods described herein. Microarrays employ nucleic acid probes specific for wild type gene or RNA and can be used to investigate the expression of a nucleic acid encoding a FOXM1 polypeptide or a CENPF polypeptide in samples from patients in a diagnostic context. In general, microarrays comprise a two dimensional arrangement of nucleic acid or polypeptide probes which comprises an intentionally created collection of nucleic acid or polypeptide probes of any length spotted onto a substrate/solid support. The array itself can have different formats, e.g. libraries of soluble probes or libraries of probes tethered to resin beads, silica chips, or other solid supports. The process of microarray fabrication is well-known to the person skilled in the art. The process can comprise preparing a glass (or other) slide (e.g., chemical treatment of the glass to enhance binding of the nucleic acid probes to the glass surface), obtaining DNA sequences representing genes of a genome of interest, and spotting

sequences these sequences of interest onto glass slide. Sequences of interest can be obtained via creating a cDNA library from an mRNA source or by using publicly available databases, such as GeneBank, to annotate the sequence information of custom cDNA libraries or to identify cDNA clones from previously prepared libraries. Generally, it is recommendable to amplify obtained sequences by PCR in order to have sufficient amounts of DNA to print on the array. The liquid containing the amplified probes can be deposited on the array by using a set of microspotting pins. Ideally, the amount deposited should be uniform. The process can further include UV-crosslinking in order to enhance immobilization of the probes on the array. Microarray chips suitable for use with the methods described herein are well known to those of skill in the art (see, e.g., U.S. Pat. Nos. 6,308,170; 6,183,698; 6,306,643; 6,297,018; 6,287,850; 6,291,183, each incorporated herein by reference). These are exemplary patents that disclose nucleic acid microarrays and those of skill in the art are aware of numerous other methods and compositions for producing microarrays. A microarray composition of the present invention can be employed for the diagnosis and treatment of any condition or disease in which the expression of FOXM1 and/or CENPF is implicated. The microarray-based methods can be used for large scale genetic or gene expression analysis of a large number of target sequences, including nucleic acids encoding a FOXM1 polypeptide or nucleic acids encoding a CENPF polypeptide. The microarray can also be used in the diagnosis of diseases and in the monitoring of treatments. Further, microarrays can also be employed to investigate an individual's predisposition to a disease. Furthermore, the microarrays can be employed to investigate cellular responses to infection, drug treatment, and the like.

[0111] When microarrays are used in connection with the methods described herein, the formation of a plurality of detectable complexes between probes and target nucleic acid sequences can be assessed. The expression profiles can show unique expression patterns that are characteristic of the presence or absence of a disease or condition, such as a malignant prostate cancer. In certain embodiments where expression profiles are examined using microarray technology, complexes can be formed by hybridization of one or more probes having complementarity to a nucleic acid encoding a FOXM1 polypeptide or a nucleic acid

encoding a CENPF polypeptide. Such a microarray can be employed in several applications including diagnostics, prognostics and treatment regimens, drug discovery and development, toxicological and carcinogenicity studies, forensics, pharmacogenomics, and the like. The probe can be in suspension or immobilized on a substrate or support (for example, as in nucleic acid array or chips technologies). For example, a sample from the subject can be contacted with a nucleic acid probe specific for a nucleic acid encoding a FOXM1 polypeptide or a nucleic acid encoding CENPF polypeptide.

[0112] In certain embodiments, the expression profile can be used to a nucleic acid encoding a FOXM1 polypeptide or a nucleic acid encoding CENPF polypeptide infer changes in the expression of target genes implicated in disease wherein the expression of such target genes is upregulated or downregulated by FOXM1, CENPF, or by the concerted action of FOXM1 and CENPF. Example probes and primers useful for obtaining gene expression profiles in normal and malignant cells, and comparing the gene expression in malignant and corresponding normal cells are known in the art (Okabe et al., 2001; Kitahara et al., 2001; Lin et al., 2002; Hasegawa et al., 2002).

[0113] In certain embodiments, microarray-based detection and/or quantification of a nucleic acid encoding FOXM1 polypeptide and/or a nucleic acid encoding or a CENPF polypeptide can comprise steps of providing a biological sample from a person suspected of having a cancer (e.g. a malignant prostate cancer), and determining the level of expression of a nucleic acid encoding FOXM1 polypeptide and/or a nucleic acid encoding or a CENPF polypeptide in the cells of the biological sample. In particular, such embodiments of the methods described herein can comprises comprising the following steps: (a) contacting a cell sample nucleic acid with a microarray under conditions suitable for hybridization; (b) providing hybridization conditions suitable for hybrid formation between the cell sample nucleic acid and a polynucleotide of the microarray; (c) detecting the hybridization; and (d) diagnosing the disorder condition based on the results of detecting the hybridization.

[0114] For example, methods of purification of nucleic acids are described in Tijssen Laboratory Techniques in Biochemistry and Molecular Biology: Hybridization With Nucleic Acid Probes, Part I. Theory and Nucleic Acid Preparation, Elsevier, New York, 1993. In one

case, total RNA is isolated using the TRIZOL reagent (Life Technologies, Gaithersburg Md.), and mRNA is isolated using oligo d (T) column chromatography or glass beads. Alternatively, when target polynucleotides are derived from an mRNA, the target polynucleotides can be a cDNA reverse transcribed from an mRNA, an RNA transcribed from that cDNA, a DNA amplified from that cDNA, an RNA transcribed from the amplified DNA, and the like. When the target polynucleotide is derived from DNA, the target polynucleotide can be DNA amplified from DNA or RNA reverse transcribed from DNA. In yet another alternative, the targets are target polynucleotides prepared by more than one method.

[0115] When target polynucleotides are amplified, it is desirable to amplify the nucleic acid sample and maintain the relative abundances of the original sample, including low abundance transcripts. Total mRNA can be amplified by reverse transcription using a reverse transcriptase and a primer consisting of oligo d(T) and a sequence encoding the phage T7 promoter to provide a single stranded DNA template. The second DNA strand is polymerized using a DNA polymerase and a RNAse which assists in breaking up the DNA/RNA hybrid. After synthesis of the double stranded DNA, T7 RNA polymerase can be added, and RNA transcribed from the second DNA strand template (Van Gelder et al. U.S. Pat. No. 5,545,522). RNA can be amplified in vitro, in situ or in vivo (see Eberwine, U.S. Pat. No. 5,514,545).

[0116] The sequence of the probes and primers suitable for use with hybridization or amplification based detection methods described herein can be derived from the sequences of a nucleic acid encoding a FOXM1 polypeptide or a CENPF polypeptide. According to the invention, a probe can be a polynucleotide sequence which is complementary to and specifically hybridizes with a, or a target portion of a nucleic acid encoding a FOXM1 polypeptide or a CENPF polypeptide, such as a DNA or RNA molecule encoding such polypeptides. Probes and primers suitable for use with the methods described herein include those that are complementary to a nucleic acid encoding a FOXM1 polypeptide or a CENPF polypeptide, can comprise single-stranded nucleic acids of between 8 to 1000 nucleotides in length, for instance between 10 and 800, between 15 and 700, or between 20 and 500.

Exemplary probes and primers may be 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or more 100 nucleotides in length. In one embodiment, a useful probe or primers of the invention is a single stranded nucleic acid molecule of between 8 to 500 nucleotides in length, which can specifically hybridize to a region of a nucleic acid encoding a FOXM1 polypeptide or a CENPF polypeptide. Longer polynucleotides encoding 250, 500, or 1000 bases and longer are contemplated as well. Such oligonucleotides will find use, for example, as probes in Southern and Northern blots and as primers in amplification reactions.

[0117] Conditions can be selected for hybridization where an exactly complementary target and probes can hybridize, i.e., each base pair must interact with its complementary base pair. Alternatively, conditions can be selected where a target and probes have mismatches but are still able to hybridize. Suitable conditions can be selected, for example, by varying the concentrations of salt in the prehybridization, hybridization and wash solutions, by varying the hybridization and wash temperatures, or by varying the polarity of the prehybridization, hybridization or wash solutions.

[0118] Suitable hybridization conditions for the diagnostic methods are those conditions that allow the detection of gene expression from identifiable expression units such as genes. Exemplary stringent hybridization conditions include but are not limited to hybridization at 42°C. in a solution (i.e., a hybridization solution) comprising 50% formamide, 1% SDS, 1 M NaCl, 10% dextran sulfate, and washing twice for 30 minutes at 60°C. in a wash solution comprising 0.1xSSC and 1% SDS. Hybridization can be performed at low stringency with buffers, such as 6xSSPE with 0.005% Triton X-100 at 37°C., which permits hybridization between target and probes that contain some mismatches to form target polynucleotide/probe complexes. Subsequent washes are performed at higher stringency with buffers, such as 0.5xSSPE with 0.005% Triton X-100 at 50°C., to retain hybridization of only those target/probe complexes that contain exactly complementary sequences. Alternatively, hybridization can be performed with buffers, such as 5xSSC/0.2% SDS at 60°C. and washes are performed in 2xSSC/0.2% SDS and then in 0.1xSSC. Background signals can be reduced by the use of detergent, such as sodium dodecyl sulfate, Sarcosyl or Triton X-100, or a

blocking agent, such as salmon sperm DNA. It is understood in the art that conditions of stringency can be achieved through variation of temperature and buffer, or salt concentration, as described in Ausubel et al., eds., Protocols in Molecular Biology, John Wiley & Sons (1994), pp. 6.0.3 to 6.4.10. After hybridization, the microarray can be washed to remove nonhybridized nucleic acids, and complex formation between the hybridizable array elements and the target polynucleotides is detected. Modifications in hybridization conditions can be empirically determined or precisely calculated based on the length and the percentage of guanosine/cytosine (GC) base pairing of the probe. The hybridization conditions can be calculated as described in Sambrook et al., eds., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y. (1989), pp. 9.47 to 9.51.

[0119] Detection of hybridization can be achieved by labeling probes or target polynucleotides (e.g. a nucleic acid encoding a FOXM1 polypeptide or a nucleic acid encoding a CENPF polypeptide with one or more labeling moieties. In one embodiment, the target polynucleotides are labeled with a fluorescent label, and measurement of levels and patterns of fluorescence indicative of complex formation is accomplished by fluorescence microscopy (e.g. confocal fluorescence microscopy). The labeling moieties can include compositions that can be detected by spectroscopic, photochemical, biochemical, bioelectronic, immunochemical, electrical, optical or chemical means. The labeling moieties include radioisotopes, such as 3H, 14C, 32P, 33P or 35S, chemiluminescent compounds, labeled binding proteins, heavy metal atoms, spectroscopic markers, such as fluorescent markers and dyes, magnetic labels, linked enzymes, mass spectrometry tags, spin labels, electron transfer donors and acceptors, and the like. Exemplary dyes include quinoline dyes, triarylmethane dyes, phthaleins, azo dyes, cyanine dyes, and the like. Fluorescent markers that emit light at wavelengths at least greater than 10 nm above the wavelength of the light absorbed can be used in some embodiments. Exemplary fluorescent markers include, but are not limited to, fluorescein, phycoerythrin, rhodamine, lissamine, and C3 and C5 available from Amersham Pharmacia Biotech (Piscataway N.J.). Labeling can also be carried out during an amplification reaction, such as polymerase chain reactions and in vitro transcription reactions, or by nick translation or 5' or 3'-end-labeling reactions. When the

label may be incorporated after or without an amplification step, the label is incorporated by using terminal transferase or by phosphorylating the 5' end of the target polynucleotide using, e.g., a kinase and then incubating overnight with a labeled oligonucleotide in the presence of T4 RNA ligase. Alternatively, the labeling moiety can be incorporated after hybridization once a probe/target complex has formed. Nucleotide substitutions can be performed, as well as chemical modifications of the probe. Such chemical modifications can be accomplished to increase the stability of hybrids (e.g., intercalating groups) or to label the probe. Some examples of labels include, without limitation, radioactivity, fluorescence, luminescence, and enzymatic labeling.

[0120] In embodiments where amplification used to detect the presence of a nucleic acid encoding a FOXM1 polypeptide or nucleic acid encoding a CENPF polypeptide, such methods can be based on the formation of specific hybrids between primers nucleic acid sequences having complete or partial complementarity to portions of a nucleic acid encoding a FOXM1 polypeptide or to portions of a nucleic acid encoding a CENPF polypeptide, wherein the primer sequences serve to initiate nucleic acid reproduction though, for example, PCR based methodologies. Numerous nucleic acid amplification techniques known in the art, including traditional polymerase chain reaction (PCR), quantitative PCR (qPCR), ligase chain reaction (LCR), strand displacement amplification (SDA) and nucleic acid sequence based amplification (NASBA). These techniques can be performed using commercially available reagents and protocols. Useful techniques in the art encompass real-time PCR, allele-specific PCR, or PCR-SSCP. Nucleic acid primers useful for amplifying a nucleic acid encoding a FOXM1 polypeptide or nucleic acid encoding a CENPF polypeptide include, but are not limited to primers that specifically hybridize with a DNA encoding a FOXM1 polypeptide or nucleic acid encoding a CENPF polypeptide, or an RNA encoding a FOXM1 polypeptide or nucleic acid encoding a CENPF polypeptide.

[0121] In some embodiments, the detection is performed by sequencing all or part of a nucleic acid encoding a FOXM1 polypeptide or a CENPF polypeptide or by selective hybridization or amplification of all or part of a nucleic acid encoding a FOXM1 polypeptide

or a CENPF polypeptide. In one embodiment, the sample can comprise prostate tissue sample from a subject.

[0122] Thus, in certain aspects, the diagnostic methods described herein comprise the use of a nucleic acid primer, wherein the primer can be complementary to and hybridize specifically to a portion of a coding sequence (e.g., gene or RNA) of a nucleic acid encoding FOXM1 or CENPF present in a sample form a subject having or at risk of developing a cancer, such as a prostate cancer, or a malignant prostate cancer. Primers suitable for use with the methods described herein include those that are specific for a nucleic acid encoding FOXM1 or CENPF. By using such primers, the detection of an amplification product indicates the presence of a nucleic acid encoding FOXM1 or CENPF or the absence of such. The use of such primers can also be employed to quantify the relative or absolute amount of a nucleic acid encoding FOXM1 or CENPF in a sample.

[0123] Primers suitable for use with the methods described herein, include, but are not limited to those having the sequence of SEQ ID NOs: 5, 6, 19 and 20. In certain embodiments, amplification of a FOXM1 nucleic acid sequence can be performed using a primer pair of SEQ ID NO: 5 and 19. In certain embodiments, amplification of a CENPF nucleic acid sequence can be performed using a primer pair of SEQ ID NO: 6 and 20. One of skill in the art will readily be able to design and synthesize primers suitable for amplifying FOXM1 or CENPF nucleic acid sequences.

[0124] Examples of primers of this invention can be single-stranded nucleic acid molecules of about 5 to 100 nucleotides in length, or about 8 to about 25 nucleotides in length. Exemplary primers may be 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 or more contiguous base pairs from the above sequences will be used, although others are contemplated. Primers suitable for use with the methods described herein can be labelled according to any method known in the art, including those described for use in labeling the probes and oligonucleotides suitable for use with the methods described herein. Labeling of primers can also be limited to labeling methods that do not interfere with the ability of the primer to be used for amplification purposes.

[0125] The sequence of a primer suitable for use with the methods described herein can be derived directly from a nucleic acid encoding FOXM1 or CENPF. Perfect complementarity is useful, to ensure high specificity. However, certain mismatch can be tolerated. For example, a nucleic acid primer or a pair of nucleic acid primers as described herein can be used in a method for detecting the presence of or a predisposition to prostate cancer in a subject.

[0126] Amplification methods include, e.g., polymerase chain reaction, PCR (PCR Protocols: A Guide to Methods and Applications, Innis, ed., Academic Press, N.Y., 1990 and PCR Strategies, Innis, ed., Academic Press, Inc., N.Y., 1995; ligase chain reaction (LCR) see, e.g., Wu and Wallace, 1989; Landegren et al., 1988; Barringer et al., 1990); transcription amplification (see, e.g., Kwoh et al., 1989); and self-sustained sequence replication (see, e.g., Guatelli et al., 1990); Q Beta replicase amplification (see, e.g., Smith et al., 1997), automated Q-beta replicase amplification assay (see, e.g., Burg et al., 1996) and other RNA polymerase mediated techniques (e.g., NASBA, Cangene, Mississauga, Ontario); see also Berger et al., 1987; Sambrook; Ausubel; U.S. Pat. Nos. 4,683,195 and 4,683,202; Sooknanan and Malek, 1995. All the references stated above are incorporated by reference in their entireties.

Methods for determining protein expression

[0127] According to the methods described herein, the coexpression of FOXM1 and CENPF protein is defined as being elevated to a diagnostic level if the amount of FOXM1 polypeptide and CENPF polypeptide expressed or present in a sample exceeds a defined composite score threshold. For example, in certain embodiments, a sample can be deemed to have elevated expression of FOXM1 and CENPF by investigating immunochemical staining (e.g. immunochemical staining using antibodies specific to FOXM1 or CENPF) of a sample from a subject, determining the percentage of the sample that is stained and assigning a percent staining value for the sample between 0% and 100%, determining an intensity for the staining and assigning a staining intensity value for the sample on a scale of 0, 1, 2, or 3, and calculating a score by multiplying the percent staining value by the staining intensity value, wherein a score exceeding 100% indicates that the FOXM1 polypeptide and CENPF

polypeptide present or expressed in the sample is at an elevated level. Thus, in certain aspects, the invention described herein related to the finding that a composite score based on (a) the percentage of a sample that is stained with immunochemical (e.g. an antibody, or a composition comprising an antibody), and (b) the intensity of the stain, can be used to diagnose a subject as having an aggressive or malignant prostate cancer, having a risk of dying from a prostate cancer, and having a risk of a prostate cancer undergoing metastasis. One of skill in the art will readily appreciate that the scoring scales described herein need not be limited to integers and may include fractional values. One of skill in the art will also understand that many variants of the composite scoring scale can be envisioned. For example, staining intensity can be ranked on a scale of 0 to 7 while retaining the fidelity of the method. Similarly, staining intensity can be ranked on a scale of 0 to 10 while retaining the fidelity of the system.

[0128] Detection of a polypeptide in accordance with the methods described herein can comprise detecting the presence of FOXM1 or CENPF polypeptide sequences in samples. In certain embodiments, detection of a polypeptide can comprise assaying for the presence of an elevated quantity FOXM1 or CENPF polypeptide in a subject prostate cancer sample as compared to a control (noncancerous or indolent cancer) sample. In certain embodiments, detection of a polypeptide can comprise detecting the subcellular localization of a quantity FOXM1 or CENPF polypeptide, and/or detection of colocalization of FOXM1 and CENPF polypeptide within a cell.

[0129] A variety of methods may be used to measure FOXM1 or CENPF protein levels including, but not limited to, immunologically based methods such as standard ELISA, immuno-polymerase chain reaction (immuno-PCR) (Sano et al., 1992), immunodetection amplified by T7 RNA polymerase (IDAT) (Zhang et al., 2001), radioimmunoassay, immunoblotting, etc. Other approaches include two-dimensional gel electrophoresis, mass spectrometry, and proximity ligation (Fredriksson et al., 2002).

[0130] In embodiments where detection of FOXM1 and CENPF is at the level of polypeptide expression, different types of ligands can be used, such as antibodies that specifically recognize FOXM1 or CENPF polypeptides. Thus, in certain embodiments where the

methods described herein involve detection of a FOXM1 or a CENPF polypeptide, a test sample can be contacted with an antibody specific for a FOXM1 or a CENPF polypeptide and the formation of an immune complex can be subsequently determined to determine the presence or location of the polypeptide. Various methods for detecting an immune complex can be used, such as ELISA, radioimmunoassays (RIA) and immuno-enzymatic assays (IEMA).

[0131] Antibodies suitable for use with the methods described herein can be polyclonal antibodies, a monoclonal antibodies, as well as fragments or derivatives thereof having substantially the same antigen specificity. Fragments of antibodies that are suitable for use with the methods described herein include Fab, Fab'2, or CDR regions. Derivatives of antibodies that are suitable for use with the methods described herein include single-chain antibodies, humanized antibodies, or poly-functional antibodies. An antibody specific for a FOXM1 polypeptide or a CENPF polypeptide can be an antibody that selectively binds to FOXM1 or CENPF, namely, an antibody raised against FOXM1 or CENPF polypeptide or an epitope-containing fragment of either polypeptide. Although non-specific binding towards other antigens can occur, binding to the target polypeptide occurs with a higher affinity and can be reliably discriminated from non-specific binding. One of skill in the art will appreciate that many methods exist for labeling antibodies for microscopic detection in samples. Exemplary labeling methods include, but are not limited fluorescent labeling, radioactive labeling, and quantum dots.

[0132] The diagnostic methods described herein can be performed on any suitable sample which contains nucleic acids or polypeptides, including in vitro, ex vivo, or in vivo samples. Examples of samples suitable for use with the methods described herein include prostate tissue samples, especially samples of prostate tumor or cancerous prostate cells from tissue biopsies taken from a subject having prostate cancer or at risk of developing it. In one embodiment, the sample comprises a tumor tissue. In one embodiment, the sample comprises prostate tissue. In another embodiment, the sample is an isolated population of prostate stem cells. The sample can be collected according to conventional techniques and used directly for diagnosis or stored. The sample can be treated prior to performing the method, in order to

render or improve availability of nucleic acids or polypeptides for testing. Treatments include, for instance, lysis (e.g., mechanical, physical, or chemical), and centrifugation. Also, the nucleic acids and/or polypeptides can be pre-purified or enriched by conventional techniques, and/or reduced in complexity. Nucleic acids and polypeptides can also be treated with enzymes or other chemical or physical treatments to produce fragments thereof. In one embodiment, the sample is contacted with reagents, such as probes, primers, or ligands, in order to assess the presence of FOXM1 or CENPF polypeptides or nucleic acids. Contacting can be performed in any suitable device, such as a plate, tube, well, or glass. In specific embodiments, the contacting is performed on a substrate coated with the reagent, such as a nucleic acid array or a specific ligand array. The substrate can be a solid or semi-solid substrate such as any support comprising glass, plastic, nylon, paper, metal, or polymers. The substrate can be of various forms and sizes, such as a slide, a membrane, a bead, a column, or a gel. The contacting can be made under any condition suitable for a complex to be formed between the reagent and the nucleic acids or polypeptides of the sample.

Diagnostic kits

[0133] The invention also provides for diagnostic kits comprising products and reagents for detecting in a sample from a subject the presence of a FOXM1 or CENPF polypeptides or nucleic acids or FOXM1 or CENPF activity. The kits can be useful for determining whether a sample from a subject expresses significantly elevated levels of FOXM1 or CENF compared to the level expressed in normal prostate tissue. For example, the diagnostic kit according to the present invention comprises any primer, any pair of primers, any nucleic acid probe and/or any ligand, suitable for use with the methods described herein. The diagnostic kits according to the present invention can further comprise reagents and/or protocols for performing a hybridization, amplification or antigen-antibody immune reaction. In certain embodiments, the kits can comprise nucleic acid primers that specifically hybridize to and can prime a polymerase reaction from a nucleic acid encoding FOXM1 or a nucleic acid encoding CENPF. In some kits nucleic acids that specifically hybridize to a nucleic acid

encoding FOXM1 or a nucleic acid encoding CENPF wherein in an embodiment the nucleic acid is affixed to a microarray support.

[0134] Some kits include anti-FOXM1 and/or anti-CENPF antibodies or fragments thereof, including monoclonal and polyclonal antibodies, and secondary antibodies that are labeled for easy detection for example with a fluororphore or horseradish peroxidase enzyme. The labeling moieties can include compositions that can be detected by spectroscopic, photochemical, biochemical, bioelectronic, immunochemical, electrical, optical or chemical means as described herein. For example, in certain embodiments, elevated nuclear colocalization of FOXM1 and CENPF can be microscopic immunofluorescent colocalization, wherein the extent of colocalization of FOXM1 and CENPF is determined using a Pearson colocalization co-efficient, a Spearman colocalization coefficient, or the like. In certain embodiments, an amount of nuclear colocalization yielding a Spearman colocalization coefficient P value of less than about 1.3 x10-11 indicates that the sample is from a subject having a prostate cancer that has undergone, or is at risk of undergoing metastasis. In certain embodiments, an amount of nuclear colocalization yielding a Spearman colocalization coefficient P value of less than about 6.2 x10-10 indicates that the sample is from a subject having a prostate cancer that has undergone, or is at risk of undergoing metastasis. In certain embodiments, an amount of nuclear colocalization yielding a Spearman colocalization coefficient P value of less than about 2.2 x10-6 indicates that the sample is from a subject at risk of dying from a prostate cancer. In certain embodiments, an amount of nuclear colocalization yielding a Spearman colocalization coefficient P value of less than about 3.5 x10-5 indicates that the sample is from a subject at risk of dying from a prostate cancer

4. Examples

4.1 Example 1. Experimental Procedures

[0135] Pilate perturbagen studies were performed to evaluate optimal dosage and scheduling. As an example, rapamycin treatment of NP mice involved treating mice for 1, 2, or 5 days and concentrations varied from 10, 25 and 50 mg/kg. Following treatment, expression profiling was done on prostate tumors to evaluate the dose and schedule that produced the

optimal range of gene expression changes. The number of differentially expressed genes at different p-value thresholds (0.01 or 0.05) with or without a 1.5 fold change (FC) cut-off was determined. The perturbagen studies were used for assembly of the mouse prostate cancer interactome.

[0136] The primary gene expression profile dataset used for ARACNe-based reverse engineering was Taylor et al (GSE21034), which consists of primary human prostatectomy samples, adjacent normal tissue, and metastases arrayed on a Affymetrix human Exon 1.0 ST microarray platform (Taylor et al., 2010). Additional expression profile datasets used were: (i) Yu et al (GSE6919): primary human prostatectomy samples and adjacent normal tissue arrayed on a Affymetrix U95a, U95b and U95c microarray platforms; (ii) Wang et al (a) (GSE17951): primary human prostatectomy samples, prostate biopsies, and normal prostate arrayed on a Affymetrix U133Plus2.0 platform; (iii) Wang et al (b) (GSE8218): primary human prostatectomy samples and normal prostate on a Affymetrix U133A platform; and (iv) Balk (GSE32269): biopsies of primary tumors from subjects with hormone-naïve prostate cancer and of bone marrow with confirmed tumor content from subjects with metastatic castration resistant prostate cancer (CRPC) on Affymetrix U133A platform. Available clinico-pathological information is provided in the table of FIG. 5.

[0137] For expression profiling analyses, prostate tumors were macrodissected, and the content of tumor/cellular atypia was verified by H&E analyses. Total RNA was isolated from prostate tissues/tumors using a MagMAX-96 total RNA isolation kit and biotin- labeled using the Illumina TotalPrep RNA Amplification Kit (Life Technologies). The resulting cRNA was hybridized on mouseWG-6 v2 BeadArrays (Illumina). Slides were scanned using an iScan (Illumina) and the resulting files were uploaded and background-corrected in BeadStudio 3.1.3.0 (Illumina, Inc.). Expression profiling data were normalized using standard variance stabilizing transformation (VST) and robust spline normalization (RSN) with lumiT and lumiN functions from lumi library, in R-system v2.14.0 (The R Foundation for Statistical Computing, ISBN 3-900051-07-0). The raw and normalized data files are deposited in Gene Expression Omnibus (GEO), accession number GSE53202

[0138] Immunostaining was performed as described in Irshad et al., 2013. Immunostaining for FOXM1 or CENPF was performed on adjacent sections of each TMA slide. For immunofluorescent staining on cells in culture, $1x10^5$ cells infected with the experimental or control shRNA were seeded in triplicate and grown in culture slides (BD Biosciences) for three days in the presence of 0.5 μg/ml of doxycycline. Cells were washed with PBS, fixed in ice cold acetone and permeabilized in 0.25% Triton X-100 in PBS and stained with antibodies for FOXM1 and CENPF (see Table 2, below). Images of the cellular localization of FOXM1 and CENPF were obtained using a Leica TCS SP5 spectral confocal microscope. Protein levels were determined by percent of staining (i.e. from 0 to 100%) and intensity level of staining (i.e., 0, 1, 2, or 3) in each tumor sample. We defined a composite protein level by multiplying percent of staining and its intensity level for each tumor sample, for FOXM1 or CENPF. Composite protein level exceeding 100 were considered elevated.

[0139] Statistical analysis was performed with survcomp package using R v2.14.0. Cox proportional hazard model was estimated with the surv and coxph functions. Kaplan-Meier survival analysis was performed using surv, survfit, and survdiff functions. Concordance indexes (c-index) were estimated and compared using coxp and concordance.index (counting ties) and cindex.comp functions.

[0140] Predicting additive effects by extrapolating individual effects of silencing FOXM1 or CENPF was evaluated as follows. . To quantitatively evaluate synergy versus additivity of the tumor growth rate, an estimate of an "additive" effect was projected using a log-linear regression model, which assumes that the silencing of either master regulator individually induces a fractional reduction in tumor growth from that of control mice. The difference between the projected "additive" model versus the actual observed consequence of co-silencing was calculated using a one sample t-test

[0141] MARINa was used to estimate the activity levels of FOXM1 and CENPF, based on their ARACNe-inferred transcriptional targets, for each sample (i.e., each subject) in the Sboner and Glinsky human prostate cancer datasets (FIG. 5) (Glinsky et al., 2004; Sboner et al., 2010). The activity was defined as elevated if activated targets were positively enriched in the sample signature (i.e., positive NES) and at the same time repressed targets were

negatively enriched in the sample signature (i.e., negative NES) and these enrichment scores fell into the upper/lower 35% percentile of NES distribution. Subjects were then divided into four groups: (i) those with non-elevated inferred activity for FOXM1 and CENPF; (ii) those with elevated inferred activity only for FOXM1; (iii) those with elevated inferred activity only for CENPF; and (iv) those with elevated inferred activity for both FOXM1 and CENPF. For these and all subsequent analyses, association with disease outcome was evaluated using Kaplan-Meier survival analysis calculated along with the log-rank p value using Surv, survfit, and survdiff functions from survcomp package in R v 2.14.0.

[0142] Gene silencing of FOXM1 and CENPF as well as forced expression of FOXM1 were done using lentiviral shRNAs or expression vectors (Open Biosystems and CCSB Human ORFeome Library, respectively). Functional studies were done in four independent human cancer cell lines, which were obtained from ATCC. All experiments using animals were performed according to protocols approved by the Institutional Animal Care and Use Committee (IACUC) at Columbia University Medical Center.

[0143] Silencing was performed using the pTRIPZ lentiviral vector (Open Biosystems), which express an shRNAmir (microRNA-adapted shRNA, hereafter referred to as shRNA) and, for functional analysis, a tRFP fluorescent reporter under the control of a tetracycline responsive element (TRE) promoter such that expression of the shRNA can be induced by addition of doxycycline (0.5 μg/ml). For two-color fluorescence analyses, which were used for selection of cells expressing two different shRNA, the pTRIPZ vector was engineered to express eGFP using the AgeI and ClaI sites to replace the tRFP cassette. Following induction with doxycycline, cells infected with the pTRIPZ-RFP virus are detected by RFP expression (red), those infected with the pTRIPZ-GFP virus by GFP expression (green), and those infected with both pTRIPZ- RFP/pTRIPZ-GFP virus by expression of both tRFP and the eGFP (yellow). The shRNAs used to silence FOXM1 and CENPF were purchased from Open Biosystems; sequences are provided in Table 1. Unless otherwise indicated, analyses were done using two alternative shRNA and co-silencing was done using each combination of the experimental or control shRNA lentivirus.

Table 1: Sequences of shRNA and Primers used for this study				
Purpose and name		**Sequence**		
shRNA		**Clone ID**	**SEQ ID**	**Mature antisense**
FOXM1 shRNA#1		V3THS_283849	1	ATAATTAGAGGATAATTTG
FOXM1 shRNA#2		V3THS_396941	2	TGATGGTCATGTTCCGGCG
CENPFshRNA#1		V2THS_115502	3	ATCTGATTCACTCAGTCTG
CENPF shRNA#2		V2THS_115504	4	TTTCTTCCAACAGTAACTG
Scramble shRNA		RHS4743		N/A
Real Time qPCR	**SEQ ID**	**Forward**	**SEQ ID**	**Reverse**
FOXM1	5	CGTCGGCCACTGATTCTCAAA	19	GGCAGGGGATCTCTTAGGTTC
CENPF	6	CTCTCCCGTCAACAGCGTTC	20	GTTGTGCATATTCTTGGCTTGC
BRCA1	7	GCTCGTGGAAGATTTCGGTGT	21	TCATCAATCACGGACGTATCATC
BUB1	8	AAATGACCCTCTGGATGTTTGG	22	GCATAAACGCCCTAATTTAAGCC
KI67	9	GGGCCAATCCTGTCGCTTAAT	23	GTTATGCGCTTGCGAACCT
CYCLIN	10	CGCTGGCGGTACTGAAGTC	24	GAGGAACGGTGACATGCTCAT
TIMELESS	11	TCTGATCCGCTATTTGAGGCA	25	GGCAGAAGGTCGCTCTGTAG
CDC25	12	ACGCACCTATCCCTGTCTC	26	CTGGAAGCGTCTGATGGCAA
TRIP13	13	ACTGTTGCACTTCACATTTTCC	27	TCGAGGAGATGGGATTTGACT
PLK1	14	AAAGAGATCCCGGAGGTCCTA	28	GGCTGCGGTGAATGGATATTTC
HMMR	15	ATGATGGCTAAGCAAGAAGGC	29	TTTCCCTTGAGACTCTTCGAGA
MYBL2	16	CCGGAGCAGAGGGATAGCA	30	CAGTGCGGTTAGGGAAGTGG
ACTIN	17	GTCTGCCTTGGTAGTGGATAATG	31	TCGAGGACGCCCTATCATGG
GAPDH	18	TGTGGGCATCAATGGATTTGG	32	ACACCATGTATTCCGGGTCAAT
ChIP				
FOXM1	33	CCGGAGCTTTCAGTTTGTTC	41	CGGAATGCCGAGACAAGG
CENPF	34	CACCTCCAGTAGAGGGGCTTG	42	TACCTCCACGCCTATTGGTC
AURKA	35	AGGACAAGGGCCTTCTTAGG	43	TAGTGGGTGGGGAGACAGAC
AURKB	36	GGGGTCCAAGGCACTGCTAC	44	GGGGCGGGAGATTTGAAAAG
BIRC5	37	CCATTAACCGCCAGATTTGA	45	TGTAGAGATGCGGTGGTCCT
CDC25	38	AAGAGCCCATCAGTTCCGCTTG	46	CCCATTTTACAGACCTGGACGC
PLK1	39	CCAGAGGGAGAAGATGTCCA	47	GTCGTTGTCCTCGAAAAAGC
CYCLIN B2	40	TCCTTTGCCGAAAGCTAGAG	48	GCAACTGCCAATCTGAAAAAG

[0144] Lentiviral particles were made using the 2nd generation packaging vectors, psPAX2 and pMD2.G (Addgene) in HEK-293T cells (ATCC), and concentrated using the Lenti-X Concentrator reagent (Clonetech). Human prostate cancer cells used in this study were DU145, PC3, LNCaP, and 22Rv1 (ATCC). Following infection with the lentiviruses, cells were selected using 4 μg/ml of puromycin for three days, following which shRNA expression was induced by addition of 0.5 μg/ml of doxycycline. The optimal time-point for silencing was determined to be 72h following induction and used for all analyses, unless otherwise indicated. For enrichment of shRNA-expression, single-cell suspensions of the induced cells were FACS sorted on a BD-FACSAria cell sorter (BD biosciences) using the FITC (emission wavelength 525nm, GFP positive) and/or PE-A (627-702nm emission wavelength, RFP positive) channels and cells having the 20% highest-level expression were collected and used for analyses. Silencing of FOXM1 and CENPF RNA and protein were confirmed by qPCR and western blot analyses, respectively. Sequences of primers used for real time PCR are provided above in Table 1; antibodies are described in Table 2, with antibodies for immunohistochemistry indicated by the initials IHC.

Table 2: Antibodies used in this study				
Description	**Source**	**Type**	**Dilution**	**Use**
FOXM1 (human)	Abcam, Ab550066	Mouse monoclonal	1:1000	Western blot, IF
FOXM1 (human)	Abcam, Ab550066	Mouse monoclonal	1:400	IHC
CENPF (human)	Abcam, Ab5	Rabbit polyclonal	1:200	Western blot, IF
CENPF (human)	Abcam, Ab90	Mouse monoclonal	1:400	IHC
pAKT	Cell Signaling #9271	Rabbit polyclonal	1:1000	Western blot
pERK	Cell Signaling #9101	Rabbit polyclonal	1:500	Western blot
pS6	Cell Signaling #2211	Rabbit polyclonal	1:1000	Western blot
Actin	Cell Signaling 4970	Rabbit polyclonal	1:2000	Western blot
PARP	Cell Signaling #9542	Rabbit polyclonal	1:1000	Western blot
V5	Invitrogen #R96025	Mouse monoclonal	1:5000	Western blot, ChIP
V5	Sigma #A7345	Mouse monoclonal	0.2 µg	IP

[0145] Determining mRNA expression of FOXM1 or CENPF or both, total RNA was isolated from lentiviral-infected cells using TRIZOL and hybridized on Illumina Human HT-12 v4 Expression BeadChip Arrays. Hybridization and expression data processing were done as described above. Differential gene expression analysis was estimated with student t-test using $p < 0.05$ as significant.

4.2 Example 2. Method of discovery

Gene profiling

[0146] The t-Distributed Stochastic Neighbor Embedding (t-SNE) is a machine learning algorithm for dimensionality reduction. It is a nonlinear dimensionality reduction technique that is particularly well suited for embedding high-dimensional data into a space of two or three dimensions, which can then be visualized in a scatter plot. Specifically, it models each high-dimensional object by a two- or three-dimensional point in such a way that similar

objects are modeled by nearby points and dissimilar objects are modeled by distant points. In the illustrated embodiment, the gene expression data is reduced to the two dimensions V1 and V2. The gene expression data was divided into 6 classes associated with normal cells adjacent to a tumor (AdjN), four Gleason scores (G6, G7, G8, and G9 or more), and metastasized (met). This analysis was done to evaluate the relative heterogeneity of human and mouse datasets used to assemble the prostate cancer interactomes. FIG. 1A depicts t-SNE analysis of the Taylor dataset relative to Gleason score. Each point is the two dimensional representation of the relative expression of many genes. The 26,445 genes are considered in the t-SNE analysis. Each point is the two dimensional representation of the similarity and divergence between the data sample (i.e., gene expression profile) and all other data samples.

Interactomes

[0147] Regulatory networks (interactomes) for human and mouse prostate cancer were generated using the Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNe) (Basso et al., 2005; Margolin et al., 2006b).

ARACNe is an unbiased algorithm that infers transcriptional interactions by computing the mutual information between each transcriptional regulator (transcription factors and co-factors) and its potential targets, and then by removing indirect interactions using the Data Processing Inequality (DPI). For optimal analysis, ARACNe requires large datasets of gene expression profiles ($\geq$ 100) having significant endogenous (i.e., genetic) and/or exogenous (i.e., perturbation-induced) heterogeneity. Thus ARACNe analysis was performed on the Taylor data set.

[0148] ARACNe was run independently on the human and mouse datasets using a conservative mutual information threshold ($p \leq 1.0\times10^{-9}$, e.g., $p \leq 0.05$, Bonferroni corrected for all candidate interactions). This resulted in highly robust regulatory networks in which the human interactome represented 249,896 interactions between 2,681 transcriptional regulators and their inferred target genes, while the mouse interactome represented 222,787 interactions for 2,072 transcriptional regulators.

[0149] FIG. 2A is a block diagram and graph that illustrates example interactomes for human and mouse models with prostate cancer, according to an embodiment. ARACNE sub-networks from the human and the mouse prostate cancer interactomes highlight selected conserved transcriptional regulators. The scaled size of the transcriptional regulator nodes (filled circles) indicates the level of conservation while the relative distance between them approximates the strength of their association.

[0150] The suitability of these mouse and human interactomes for cross-species interrogation was next evaluated by developing a novel computational approach to assess the global conservation of their transcriptional programs.

[0151] A quantitative metric was developed to compare conservation of the human and mouse interactomes. In particular, a modification of the MARINa algorithm was developed that allows for single- sample analysis to infer the differential activity of 2028 transcriptional regulators represented in both interactomes on a sample-by-sample basis, from the expression of their interactome-specific targets. The analysis was performed on 1009 expression profiles representing 4 human datasets listed in the Table of FIG. 5, as well as across the mouse datasets, to determine whether the activity of each regulator, inferred either from the expression of its human interactome targets or its murine interactome targets, was significantly correlated ($p \leq 0.05$), indicating that the murine and human regulatory programs were therefore conserved. The accuracy of this metric was demonstrated by comparing two equivalent same-species interactomes from the human and mouse datasets (i.e., positive control), in which virtually all transcriptional regulators were conserved (>90%), contrasting with randomized interactomes (i.e., negative control) that had virtually no conservation. Histogram (density plots) showed the distribution of the correlation coefficients of activity profiles of transcriptional regulators for randomized interactomes (negative control) and the positive control interactomes for human and mouse. The degree of correlations was measured by the Z-score, and the Spearman correlation coefficient. The Z-score, also called the standard score, is the (signed) number of standard deviations an observation or datum is above the mean; and, is useful in comparing different populations. The Spearman's rank correlation coefficient, also called Spearman's rho, is a nonparametric measure of statistical

dependence between two variables. It assesses how well the relationship between two variables can be described using a monotonic function. If there are no repeated data values, a perfect Spearman correlation of +1 or −1 occurs when each of the variables is a perfect monotone function of the other.

[0152] Using these metrics, it was found that 70% of the transcriptional regulators in the human and mouse prostate cancer interactomes regulate statistically conserved programs ($p \leq 0.05$). FIG. 2B is a graph that illustrates example percentage of the interactomes that are conserved between human and mouse models with prostate cancer, according to an embodiment. This histogram shows the distribution of the Z-scores for conservation of activity profiles between the human and mouse interactomes at $p \leq 0.05$. Comparison of the androgen receptor (AR) activity levels in each sample from Taylor *et al* and the mouse dataset was performed using the Spearman correlation coefficient.

[0153] Notably, conserved transcriptional regulators included many genes known to play important roles in prostate cancer, such as AR, ETS1, ETV4, ETV5, STAT3, MYC, BRCA1, and NKX3.1. In particular, AR displayed extensive correlation of its transcriptional activity between the human and mouse interactomes, consistent with its known role as a key regulator of prostate development and prostate tumorigenesis.

Master regulators

[0154] The Master Regulator Inference algorithm (MARINa) (Carro et al., 2010; Lefebvre et al., 2010) was then used to infer candidate master regulators (MRs) that act individually or synergistically to drive malignant prostate cancer in the conserved interactomes. MARINa estimates differential activity (DA) based on enrichment (differential expression, DE) of their activated and repressed targets in the malignancy signature. More specifically, MARINa identified candidate MRs based on the concerted differential expression of their ARACNe-inferred targets (i.e., their differential activity, DA). Specifically, "activated" MRs have positively-regulated and repressed targets significantly enriched among upregulated and downregulated genes, respectively, while "repressed" MRs have the converse.

[0155] To interrogate the human prostate cancer interactome, a gene signature was used representing prostate cancer malignancy from the Taylor dataset, which compares aggressive prostate tumors (Gleason score ≥ 8 with rapid biochemical recurrence; sample size n =10) versus indolent ones (Gleason score 6 tumors with no biochemical recurrence; sample size n = 39). These analyses identified 175 candidate MRs, including 49 activated and 126 repressed ($p \leq 0.05$).

[0156] To investigate the robustness of these MRs, MARINa was performed using a second, independent malignancy signature from the Balk dataset (see the table of FIG. 5), which compares lethal CRPC (sample size n=29) with indolent, hormone-naïve prostate cancer (sample size n=22). These independent MR analyses significantly overlapped with those identified from the Taylor malignancy signature (36 MRs in common; Fisher exact test $p < 0.0001$). The Fisher exact test was used to compare two populations with the same number of members and determine the probability p that deviations from the null hypothesis, here that the two distributions are the same could be explained by random events. Furthermore, MARINa analyses of 15 independent interactomes using the Taylor human prostate cancer malignancy signature showed that the inferred MRs were highly overlapping with those inferred from two additional independent prostate cancer interactomes ($p < 7 \times 10^{-9}$ and $p < 8 \times 10^{-20}$, Fisher exact test) but not with MRs inferred from non-prostate cancer specific interactomes (13 orders of magnitude different in significance). Thus, inference of master regulators of human prostate cancer malignancy required a prostate cancer-specific interactome but was independent of the specific dataset used for its interrogation.

[0157] To identify a corresponding mouse malignancy signature, MARINa was performed on four independent GEMM signatures, which are associated with prostate cancer malignancy and represent the diverse range of prostate cancer phenotypes represented among the GEMMs, including the NPK, NPB, NP, NP-AI, Myc, and NP53 mouse models. Meta-analyses of independent MR lists from these four independent GEMM signatures led to the identification of 229 candidate mouse MRs, including 110 activated and 119 repressed MRs ($p \leq 0.001$).

Conserved MRs were more likely to be associated with disease outcome than the non-conserved ones

[0158] The resulting independent lists of human and mouse MRs were then integrated to produce a ranked list of 20 conserved MRs, including 7 activated and 13 repressed (joint p-value: $p \leq 0.0074$ by Stouffer's method). FIG. 3A is a Venn diagram and table that illustrates example selection of a subset of master regulators from a full set determined by available automated computer processes, according to an embodiment. Notably, these conserved MRs were more likely to be associated with disease outcome than the non-conserved ones, as assessed by a univariate COX proportional hazard regression model (43% versus 21%; $p \leq 0.05$), and were also more likely to be differentially expressed in aggressive prostate tumors (metastatic versus non-metastatic; 100% versus 60%).

[0159] Subsequent analysis focused on the subset of activated conserved MRs, each of which has been associated with cancer-related biological processes: CHAF1A (chromatin activity); TRIB3 (regulation of cell signaling in transcriptional control); FOXM1 (cell cycle progression); CENPF (mitosis); PSRC1 (growth control); TSFM (translational elongation); and ASF1B (regulation of nucleosome assembly). FIG. 3B is a diagram that illustrates example ranking of activated master regulators for their impacts on prostate cancer, according to an embodiment. Conserved activated MRs are shown for the human (left) and mouse (right) malignancy signatures, depicting the different positive (activated; upper bars) and negative (repressed; lower bars) targets. The ranks of differential activity (DA) and differential expression (DE) are shown by the shaded boxes; the numbers indicate the rank of the DE in the signature. Differential expression is defined by comparing expression levels of a gene between two groups of samples (here, aggressive and indolent prostate cancer samples) using the t-test. Genes ranked (i.e., sorted from the most over-expressed to the most under-expressed) by their differential expression define a signature. For example, 411 represents a higher position in the signature and thus a stronger differential expression, compared to 13323.

[0160] FIG. 3C is a table that illustrates example ranking of master regulators for their impact on prostate cancer by various available algorithms, according to an embodiment. In

this summary of conserved MRs are shown: joint p-value from human and mouse MARINa analysis, calculated using Stouffer's method; p-value for COX proportional hazard regression model applied to mRNA expression levels and predicted MR activity; and average p-values for differential expression of MRs in metastatic versus non-metastatic primary tumors. Smaller p values means that the deviations from the null hypothesis, that the regulator is not important, are less likely due to chance and thus the corresponding regulator is more significant contributors. FOXM1 and CENPF are significant ($p < 0.05$) for all measures.

Synergistic master regulators FOXM1 and CENPF are differentially expressed in aggressive prostate tumors

[0161] These MRs were further prioritized by computationally evaluating their potential synergistic interactions. By these criteria, any pair of MRs was considered "synergistic" if their co-regulated ARACNe-inferred targets were significantly more enriched in the malignancy signature than their individual targets ($p \leq 0.001$) (Carro et al., 2010; Lefebvre et al., 2010). Using this computational approach to analyze all 21 possible pairs among the conserved activated MRs, the only pair that was found to be statistically significant was FOXM1 and CENPF.

[0162] FIG. 4 is a table that illustrates example predicted synergy of FOXM1 and CENPF among other pairs in the subset of master regulators using available algorithms, according to an embodiment. Shown are synergy p-values (i.e., enrichment of shared versus non-shared targets in the malignancy signature) for conserved MRs, inferred by MARINa. Clearly, the synergy of FOXM1 and CENPF is least likely to be random ($p < 0.001$), and thus most significant.

5. Alternatives and extensions

[0163] In the foregoing specification, the invention has been described with reference to specific embodiments thereof. It will, however, be evident that various modifications and changes may be made thereto without departing from the broader spirit and scope of the invention. The specification and drawings are, accordingly, to be regarded in an illustrative rather than a restrictive sense. Throughout this specification and the claims, unless the context requires otherwise, the word "comprise" and its variations, such as "comprises" and "comprising," will be understood to imply the inclusion of a stated item, element or step or group of items, elements or steps but not the exclusion of any other item, element or step or group of items, elements or steps. Furthermore, the indefinite article "a" or "an" is meant to indicate one or more of the item, element or step modified by the article.

6. References.

[0164] Reference to the following publications can be found herein.

Akavia, U. D., Litvin, O., Kim, J., Sanchez-Garcia, F., Kotliar, D., Causton, H. C., Pochanard, P., Mozes, E., Garraway, L. A., and Pe'er, D. (2010). An integrated approach to uncover drivers of cancer. Cell 143, 1005-1017.

Alvarez-Fernandez, M., and Medema, R. H. (2013). Novel functions of FoxM1: from molecular mechanisms to cancer therapy. Frontiers in oncology 3, 30.

Aytes, A., Mitrofanova, A., Kinkade, C. W., Lefebvre, C., Lei, M., Phelan, V., LeKaye, H. C., Koutcher, J. A., Cardiff, R. D., and Califano, A. (2013). ETV4 promotes metastasis in response to activation of PI3-kinase and Ras signaling in a mouse model of advanced prostate cancer. Proc Natl Acad Sci USA 110, E3506-3515.

Baca, S. C., Prandi, D., Lawrence, M. S., Mosquera, J. M., Romanel, A., Drier, Y., Park, K., Kitabayashi, N., MacDonald, T. Y., and Ghandi, M. (2013). Punctuated evolution of prostate cancer genomes. Cell 153, 666-677.

Barringer, K., Orgel, L., Wahl, G., and Gingeras, T. R. (1990). Blunt-end and single-strand ligations by Escherichia coli ligase: influence on an in vitro amplication scheme. Gene 89:117-122.

Basso, K., Margolin, A. A., Stolovitzky, G., Klein, U., Dalla-Favera, R., and Califano, A. (2005). Reverse engineering of regulatory networks in human B cells. Nat Genet 37, 382-390.

Berger, S., Kimmel, A., Abelson, J., and Simon, M. (1987). Methods in Enzymology 152:307-316.

Bomont, P., Maddox, P., Shah, J. V., Desai, A. B., and Cleveland, D. W. (2005). Unstable microtubule capture at kinetochores depleted of the centromere-associated protein CENP-F. EMBO J 24, 3927-3939.

Burg, L. J., Juffras, A. M., Wu, Y., Blomquist, C. L., and Du, Y. (1996). Single molecule detection of RNA reporter probes by amplification with Qβ replicase. Mol. Cell. Probes 10:257-271.

Cai, Y., Balli, D., Ustiyan, V., Fulford, L. A., Hiller, A., Misetic, V., Zhang, Y., Paluch, A. M., Waltz, S. E., Kasper, S., and Kalin, T. V. (2013). Foxm1 Expression in Prostate Epithelial Cells is Essential for Prostate Carcinogenesis. J Biol Chem 288, 22527-22541.

Carro, M. S., Lim, W. K., Alvarez, M. J., Bollo, R. J., Zhao, X., Snyder, E. Y., Sulman, E. P., Anne, S. L., Doetsch, F., and Colman, H. (2010). The transcriptional network for mesenchymal transformation of brain tumours. Nature 463, 318-325.

Donovan, M. J., Hamann, S., Clayton, M., Khan, F. M., Sapir, M., Bayer-Zubek, V., Fernandez, G., Mesa-Tejada, R., Teverovskiy, M., and Reuter, V. E. (2008). Systems pathology approach for the prediction of prostate cancer progression after radical prostatectomy. J Clin Oncol 26, 3923-3929.

Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J. J., and Gardner, T. S. (2007). Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol 5, e8.

Feng, J., Huang, H., and Yen, T. J. (2006). CENP-F is a novel microtubule-binding protein that is essential for kinetochore attachments and affects the duration of the mitotic checkpoint delay. Chromosoma 115, 320-329.

Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gustafsdottir, S. M., Ostman, A., and Landegren, U. (2002). Protein detection using proximity-dependent DNA ligation assays. Nat. Biotechnol. 20, 473-477.

Glinsky, G. V., Glinskii, A. B., Stephenson, A. J., Hoffman, R. M., and Gerald, W. L. (2004). Gene expression profiling predicts clinical outcome of prostate cancer. J Clin Invest 113, 913-923.

Guatelli, J. C., Whitfield, K. M., Kwoh, D. Y., Barringer, K. J., Richman, D. D., and Gingeras, T. R. (1990). Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. Proc. Natl. Acad. Sci. USA 87:1874-78.

Halasi, M., and Gartel, A. L. (2013a). FOX(M1) news--it is cancer. Mol Cancer Ther 12, 245- 254.

Halasi, M., and Gartel, A. L. (2013b). Targeting FOXM1 in cancer. Biochemical pharmacology 85, 644-652.

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. Cell 144, 646-67.

Hasegawa, S., Furukawa, Y., Li, M., Satoh, S., Kato, T., Watanabe, T., Katagiri, T., Tsunoda, T., Yamaoka, Y., and Nakamura, Y. (2002). Genome-Wide Analysis of Gene Expression in Intestinal-Type Gastric Cancers Using a Complementary DNA Microarray Representing 23,040 Genes. Cancer Res 62:7012-7.

Holt, S. V., Vergnolle, M. A., Hussein, D., Wozniak, M. J., Allan, V. J., and Taylor, S. S. (2005). Silencing Cenp-F weakens centromeric cohesion, prevents chromosome alignment and activates the spindle checkpoint. J Cell Sci 118, 4889-4900.

Irshad, S., and Abate-Shen, C. (2013). Modeling prostate cancer in mice: something old, something new, something premalignant, something metastatic. Cancer metastasis reviews 32, 109-122.

Ittmann, M., Huang, J., Radaelli, E., Martin, P., Signoretti, S., Sullivan, R., Simons, B. W., Ward, J. M., Robinson, B. D., and Chu, G. C. (2013). Animal models of human prostate

cancer: the consensus report of the New York meeting of the Mouse Models of Human Cancers Consortium Prostate Pathology Committee. Cancer Res 73, 2718-2736.

Jones, A. C., Sampson, J. R., Hoogendoorn, B., Cohen, D., and Cheadle, J. P. (2000). Application and evaluation of denaturing HPLC for molecular genetic analysis in tuberous sclerosis. Hum Genet., 106(6):663-8.

Kalin, T. V., Ustiyan, V., and Kalinichenko, V. V. (2011). Multiple faces of FoxM1 transcription factor: lessons from transgenic mouse models. Cell Cycle 10, 396-405.

Kalin, T. V., Wang, I. C., Ackerson, T. J., Major, M. L., Detrisac, C. J., Kalinichenko, V. V., Lyubimov, A., and Costa, R. H. (2006). Increased levels of the FoxM1 transcription factor accelerate development and progression of prostate carcinomas in both TRAMP and LADY transgenic mice. Cancer Res 66, 1712-1720.

Kitahara, O., Furukawa, Y., Tanaka, T., Kihara, C., Ono, K., Yanagawa, R., Nita, M., Takagi, T., Nakamura, Y., and Tsunoda, T. (2001). Alterations of Gene Expression during Colorectal Carcinogenesis Revealed by cDNA Microarrays after Laser-Capture Microdissection of Tumor Tissues and Normal Epithelia. Cancer Res 61: 3544-9.

Koo, C. Y., Muir, K. W., and Lam, E. W. (2012). FOXM1: From cancer initiation to progression and treatment. Biochim Biophys Acta 1819, 28-37.

Kwoh, D. Y., Davis, G. R., Whitfield, K. M., Chappelle, H. L., DiMichele, L. J., and Gingeras, T. R. (1989). Proc. Natl. Acad. Sci. USA 86:1173-77.

Landegren, U., Kaiser, R., Sanders, J., and Hood, L. (1988). A ligase-mediated gene detection technique. Science 241:1077-80.

Lefebvre, C., Rajbhandari, P., Alvarez, M. J., Bandaru, P., Lim, W. K., Sato, M., Wang, K., Sumazin, P., Kustagi, M., and Bisikirska, B. C. (2010). A human B-cell interactome identifies MYB and FOXM1 as master regulators of proliferation in germinal centers. Mol Syst Biol 6, 377.

Lefebvre, C., Rieckhof, G., and Califano, A. (2012). Reverse-engineering human regulatory networks. Wiley interdisciplinary reviews Systems biology and medicine 4, 311-325.

Lin, Y-M., Furukawa, Y., Tsunoda, T., Yue, C.-T., Yang, K.-C., and Nakamura, Y. (2002). Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. Oncogene 21:4120-8.

Ma, L., Zhao, X., and Zhu, X. (2006). Mitosin/CENP-F in mitosis, transcriptional control, and differentiation. Journal of biomedical science 13, 205-213.

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Dalla Favera, R., and Califano, A. (2006a). ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics 7 Suppl 1, S7.

Margolin, A. A., Wang, K., Lim, W. K., Kustagi, M., Nemenman, I., and Califano, A. (2006b). Reverse engineering cellular networks. Nat Protoc 1, 662-671.

Okabe, H., Satoh, S., Kato, T., Kitahara, O., Yanagawa, R., Yamaoka, Y., Tsunoda, T., Furukawa, Y., and Nakamura, Y. (2001). Genome-wide Analysis of Gene Expression in Human Hepatocellular Carcinomas Using cDNA Microarray: Identification of Genes Involved in Viral Carcinogenesis and Tumor Progression. Cancer Res 61:2129-37.

Pan, J., and Yeung, S. C. (2005). Recent advances in understanding the antineoplastic mechanisms of farnesyltransferase inhibitors. Cancer Res 65, 9109-9112.

Radhakrishnan, S. K., Bhat, U. G., Hughes, D. E., Wang, I. C., Costa, R. H., and Gartel, A. L. (2006). Identification of a chemical inhibitor of the oncogenic transcription factor forkhead box M1. Cancer Res 66, 9731-9735.

Ryan, C. J., and Tindall, D. J. (2011). Androgen receptor rediscovered: the new biology and targeting the androgen receptor therapeutically. J Clin Oncol 29, 3651-3658.

Sano, T., Smith, C. L., and Cantor, C. R. (1992). Immuno-PCR: Very sensitive antigen detection by means of specific antibody-DNA conjugates. Science 258, 120-122.

Sboner, A., Demichelis, F., Calza, S., Pawitan, Y., Setlur, S. R., Hoshida, Y., Perner, S., Adami, H. O., Fall, K., and Mucci, L. A. (2010). Molecular sampling of prostate cancer: a dilemma for predicting disease progression. BMC Med Genomics 3, 8.

Scher, H. I., and Sawyers, C. L. (2005). Biology of progressive, castration-resistant prostate cancer: directed therapies targeting the androgen-receptor signaling axis. J Clin Oncol 23, 8253-8261.

Schoenborn, J. R., Nelson, P., and Fang, M. (2013). Genomic profiling defines subtypes of prostate cancer with the potential for therapeutic stratification. Clin Cancer Res 19, 4058-4066.

Shah, R. B., Mehra, R., Chinnaiyan, A. M., Shen, R., Ghosh, D., Zhou, M., Macvicar, G. R., Varambally, S., Harwood, J., and Bismar, T. A. (2004). Androgen-independent prostate cancer is a heterogeneous group of diseases: lessons from a rapid autopsy program. Cancer Res 64, 9209-9216.

Shen, M. M., and Abate-Shen, C. (2010). Molecular genetics of prostate cancer: new prospects for old challenges. Genes Dev 24, 1967-2000.

Smith, J. H., Radcliffe, D., Rigmy, S., Mahan, D., Lane, D. J., and Klinger, J. D. (1997). Performance of an automated Q-beta replicase amplification assay for Mycobacterium tuberculosis in a clinical trial. Clin. Microbiol. 35:1477-1491.

Sooknanan, R., and Malek, L. T. (1995). A detection and amplification system uniquely suited for RNA. Nature Biotechnology 13:563-564.

Taylor, B. S., Schultz, N., Hieronymus, H., Gopalan, A., Xiao, Y., Carver, B. S., Arora, V. K., Kaushik, P., Cerami, E., and Reva, B. (2010). Integrative genomic profiling of human prostate cancer. Cancer Cell 18, 11-22.

Varis, A., Salmela, A. L., and Kallio, M. J. (2006). Cenp-F (mitosin) is more than a mitotic marker. Chromosoma 115, 288-295.

Wu, D. Y., and Wallace, R. B. (1989). The ligation amplification reaction (LAR)—amplification of specific DNA sequences using sequential rounds of template-dependent ligation. Genomics 4:560-569.

Zhang, H. T., Kacharmina, J. E., Miyashiro, K., Greene, M. I., and Eberwine, J. (2001). Protein quantification from complex protein mixtures using a proteomics methodology with single-cell resolution. J. Proc. Natl. Acad. Sci. USA 98, 5497-5502.

Zhang, N., Wei, P., Gong, A., Chiu, W. T., Lee, H. T., Colman, H., Huang, H., Xue, J., Liu, M., and Wang, Y. (2011). FoxM1 promotes beta-catenin nuclear localization and controls Wnt target-gene expression and glioma tumorigenesis. Cancer Cell 20, 427-442.

Zhang, Q. C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C. A., Bisikirska, B., Lefebvre, C., Accili, D., and Hunter, T. (2012). Structure-based prediction of protein-protein interactions on a genome-wide scale. Nature 490, 556-560.

CLAIMS

What is claimed is:

1. A method comprising:

a) obtaining a test prostate cancer sample from a subject having prostate cancer;

b) determining a level of expression of each of the genes encoding (FOXM1) Forkhead box protein M1 and Centromere protein F (CENPF) in the test sample and a control sample;

c) comparing the level of expression of each of the FOXM1 and CENPF genes in the test sample to the corresponding level in the control sample; and

d) if the level of expression of each of the FOXM1 and CENPF genes in the test sample is at least 35% higher than the corresponding level in the control sample, then determining that the subject has an aggressive form of prostate cancer or has a high risk of prostate cancer progressing to an aggressive form.

2. The method of claim 1, wherein determining the level of expression of the prognostic genes FOXM1 and CENPF comprises determining a level of mRNA encoding FOXM1 and CENPF in the sample, respectively, using a method selected from the group consisting of nuclease protection assays, northern blots, real time quantitative PCR, and in-situ hybridization.

3. The method of claim 1, wherein determining the level of expression of the prognostic genes FOXM1 and CENPF comprises determining a level of FOXM1 protein or CENPF protein in the sample, respectively, using a method selected from the group consisting western blots, 2-dimensional SDS-PAGE, and mass spectrometry.

4 A method comprising:

a) obtaining a prostate cancer sample from a subject having prostate cancer;

b) determining a level of expression of FOXM1 protein and CENPF protein in the prostate cancer sample by immunostaining with a first antibody that specifically binds to FOXM1 and a second antibody that specifically binds to CENPF; and

c) if at least 50% of prostate cancer cells in the prostate cancer sample express both FOXM1 protein and CENPF protein at a composite score of at least 100 for each protein, wherein the composite score is calculated by multiplying a percent staining value by a staining intensity value, then determining that the subject has an aggressive form of prostate cancer or has a high risk of prostate cancer progressing to an aggressive form.

5. The method of claim 4, wherein both FOXM1 protein and CENPF protein are colocalized in the nucleus of at least 50% of prostate cancer cells in the sample.

6. The method of one of claims 1 and 4, wherein the prostate cancer sample comprises circulating prostate cancer cells that have been isolated.

7. A method comprising:

a) obtaining a prostate cancer sample from a subject having prostate cancer (or at risk of developing prostate cancer),

b) applying a first antibody that specifically binds to FOXM1 protein in the sample, wherein presence of FOXM1 creates an antibody-FOXM1 complex; and applying a second antibody that specifically binds to CENPF in the sample, wherein presence of the CENPF creates an antibody-CENPF complex,

c) applying a first detection agent that detects the antibody-FOXM1 complex; and a second detection agent that detects the antibody-CENPF complex, and

d) if at least 50% of prostate cancer cells in the sample express both FOXM1 protein and CENPF protein at a composite score of at least 100 for each protein, wherein the composite score is calculated by multiplying a percent staining value by a staining intensity value, then determining that the subject has an aggressive form of prostate cancer or has a high risk of prostate cancer progressing to an aggressive form.

8. The method as in claim 1, 4 or 7, further comprising treating the subject for aggressive prostate cancer if a determination is made that the cancer is aggressive prostate cancer.

9. The method as in claim 1, 4 or 7, wherein the control prostate tissue sample comes from a normal subject that does not have cancer or from a noncancerous area of the subject's prostate.

10. A diagnostic kit for detecting an expression level of an mRNA or a protein encoding FOXM1 or CENPF or both in a biological sample, the kit comprising oligonucleotides that specifically hybridize to each of the respective mRNAs or one or more agents that specifically bind to each of the respective proteins, or both.

11. The diagnostic kit of claim 10, further comprising a forward primer and a reverse primer specific for each mRNA encoding FOXM1 or CENPF for use in a qRT-PCR assay to specifically quantify the expression level of each mRNA.

12. The diagnostic kit of claim 10, wherein the agents comprise one or more antibodies or antibody fragments that specifically bind to each of the respective FOXM1 or CENPF protein.

13. A microarray comprising a plurality of oligonucleotides that specifically hybridize to an mRNA encoded by each of the FOXM1 or CENPF genes, which oligonucleotides are fixed on the microarray.

14. The microarray of claim 13, wherein the oligonucleotides are labeled to facilitate detection of hybridization to the mRNAs.

15. The microarray of claim 14, wherein the oligonucleotides are radio-labeled, or biotin-labeled.

16. The microarray of claim 13, wherein the oligonucleotides are cDNAs.

17. A microarray comprising a plurality of antibodies or antibody fragments that specifically bind to either or both of FOXM1 protein or CENPF protein or biologically active fragment thereof, which antibodies or antibody fragments are fixed on the microarray.

18. The microarray of claim 17, wherein the antibodies or antibody fragments are labeled to facilitate detection of binding to the protein.

19. The microarray of claim 18, wherein the antibodies or antibody fragments are radio-labeled, biotin-labeled, chromophore-labeled, fluorophore-labeled, or enzyme-labeled.

FIG. 1A
HUMAN DATASET

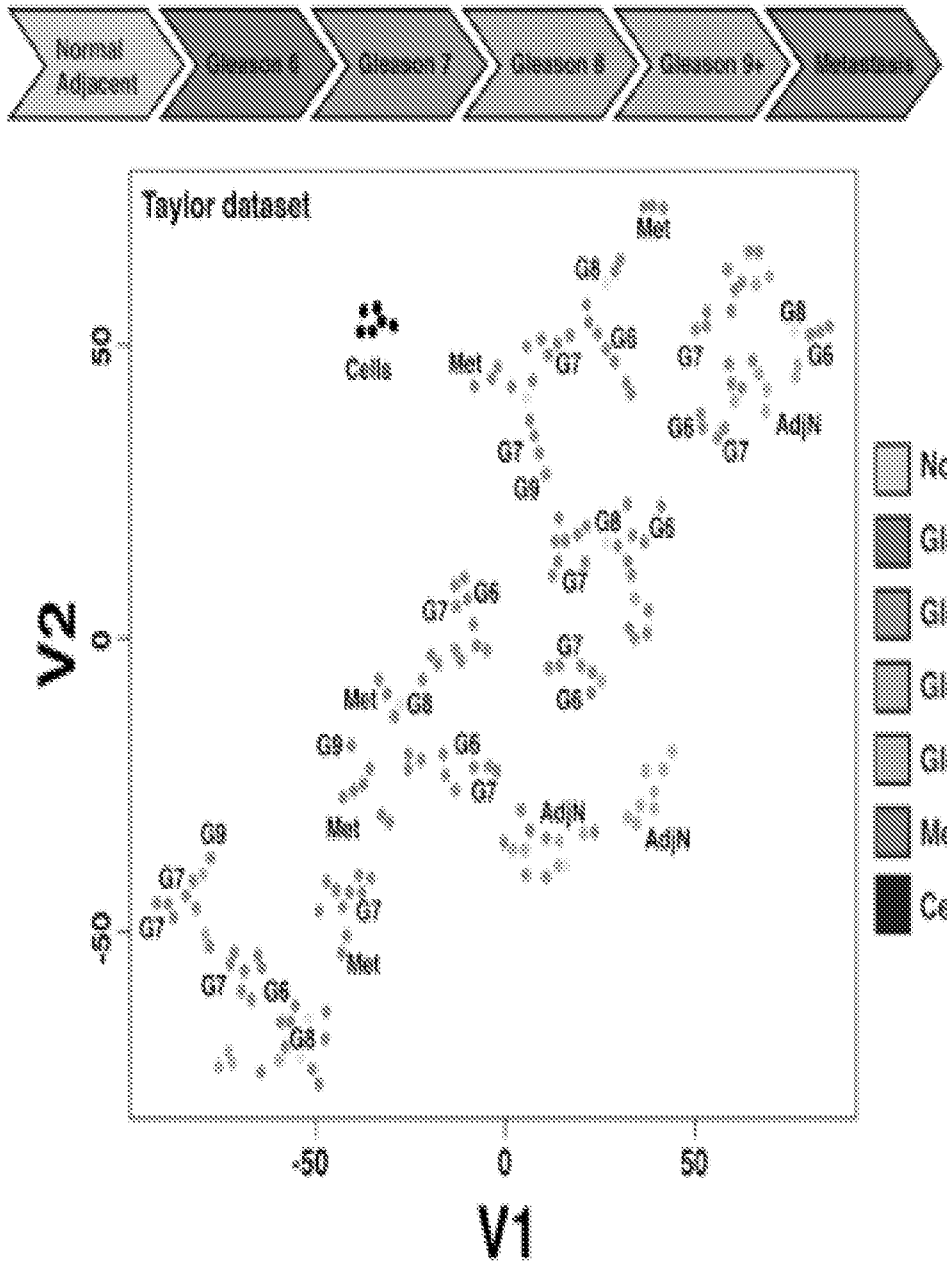


FIG. 1B
MOUSE MODELS

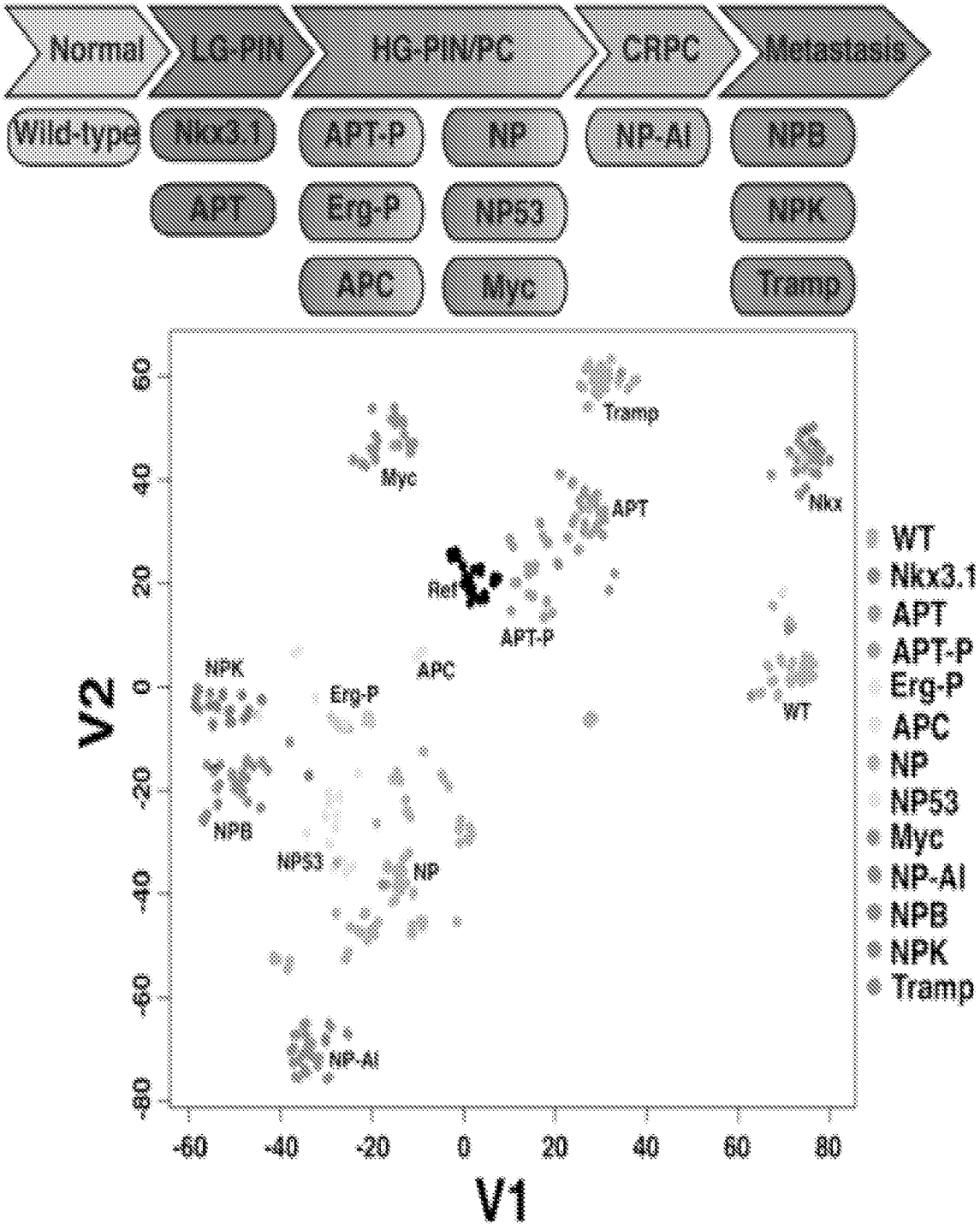


FIG. 1C

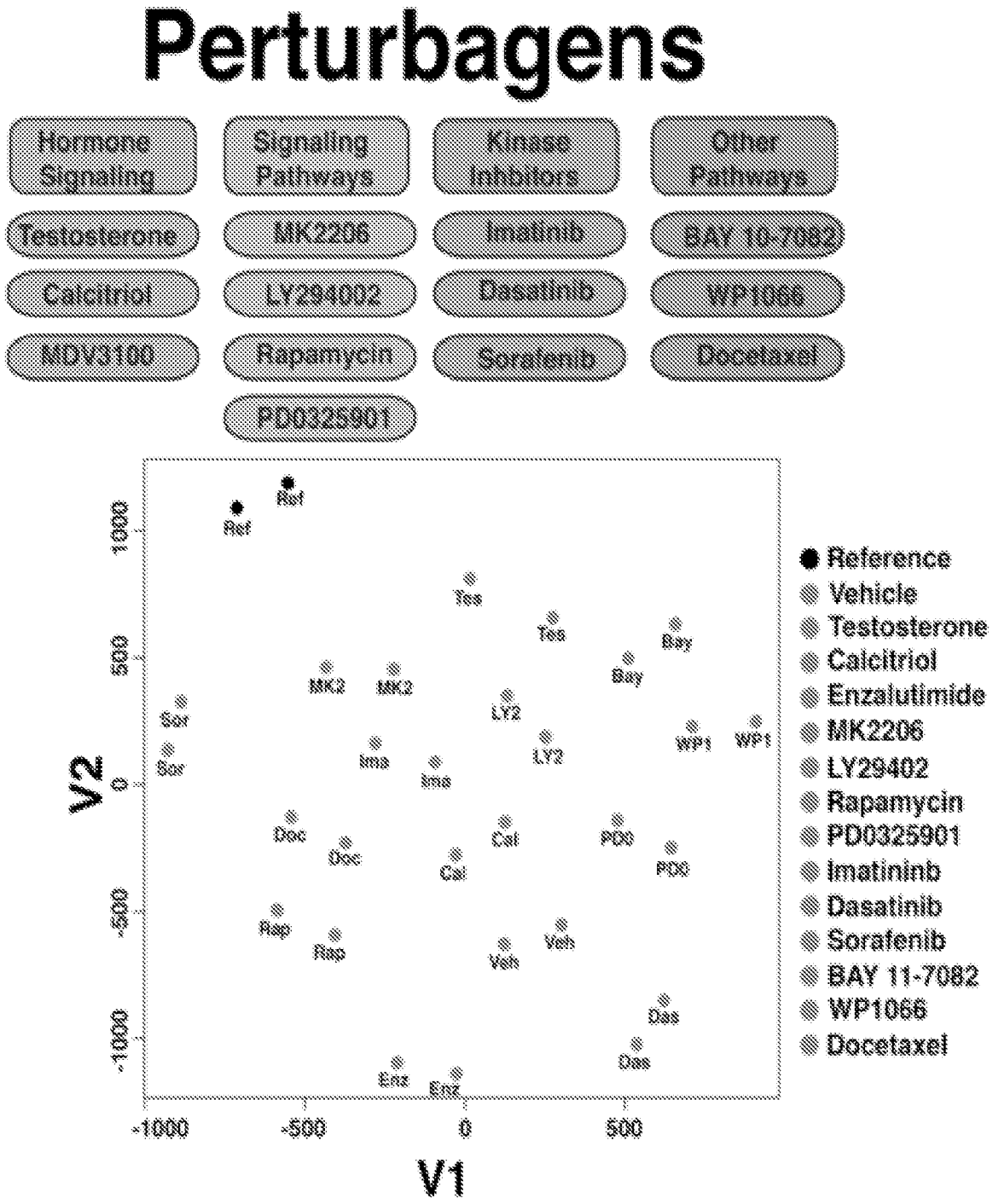

Perturbagens
Hormone Signaling
Signaling Pathways
Kinase Inhibitors
Other Pathways
Testosterone
MK2206
Imatinib
BAY 10-7082
Calcitriol
LY294002
Dasatinib
WP1066
MDV3100
Rapamycin
Sorafenib
Docetaxel
PD0325901
Reference
Vehicle
Testosterone
Calcitriol
Enzalutimide
MK2206
LY29402
Rapamycin
PD0325901
Imatininb
Dasatinib
Sorafenib
BAY 11-7082
WP1066
Docetaxel
V1
V2
-1000
-500
0
500
1000

FIG. 2A

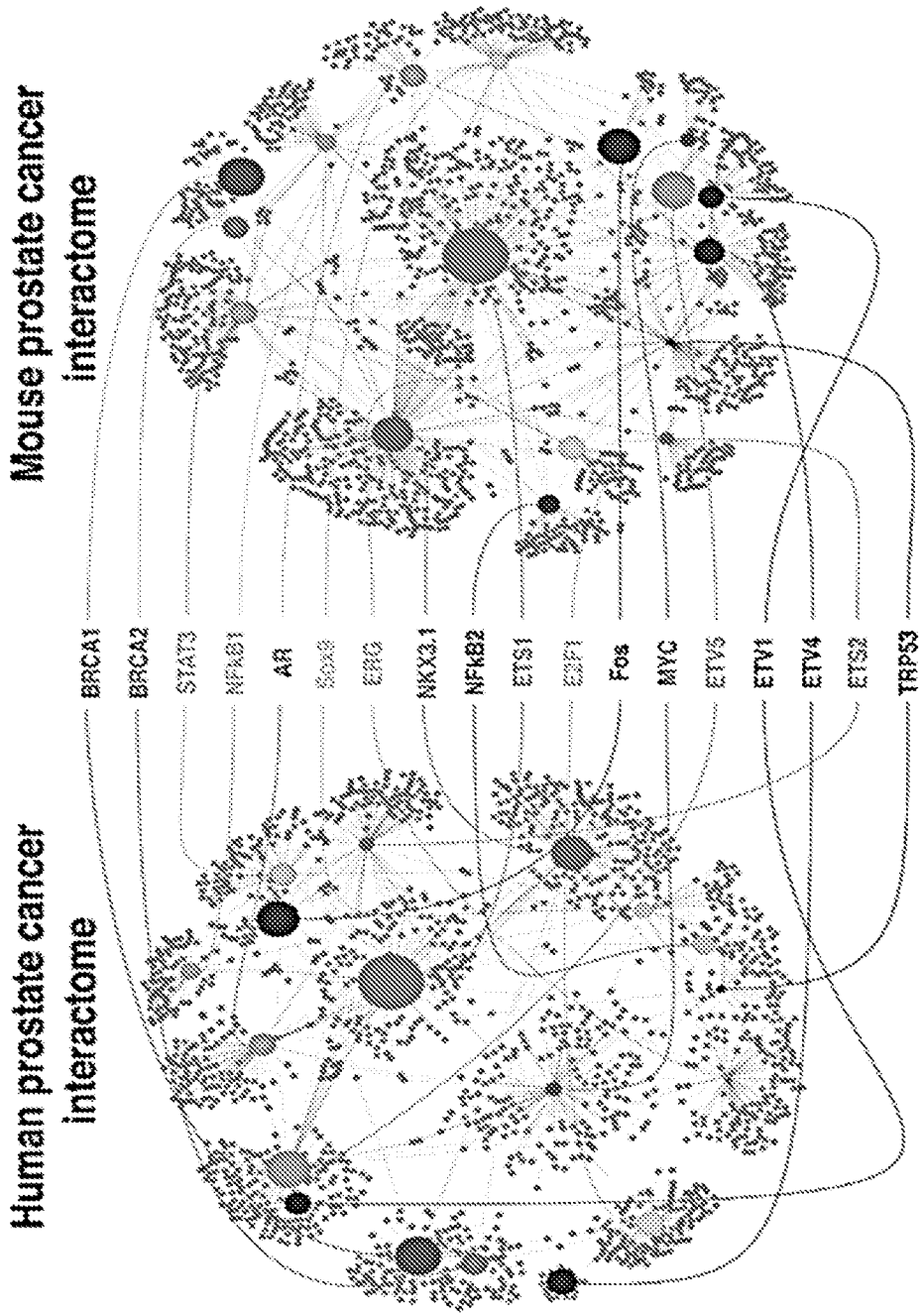

Mouse prostate cancer interactome
Human prostate cancer interactome
BRCA1
BRCA2
STAT3
NFkB1
AR
SOX9
ERG
NKX3.1
NFkB2
ETS1
EGR1
Fos
MYC
ETV5
ETV1
ETV4
ETS2
TRP53

FIG. 2B

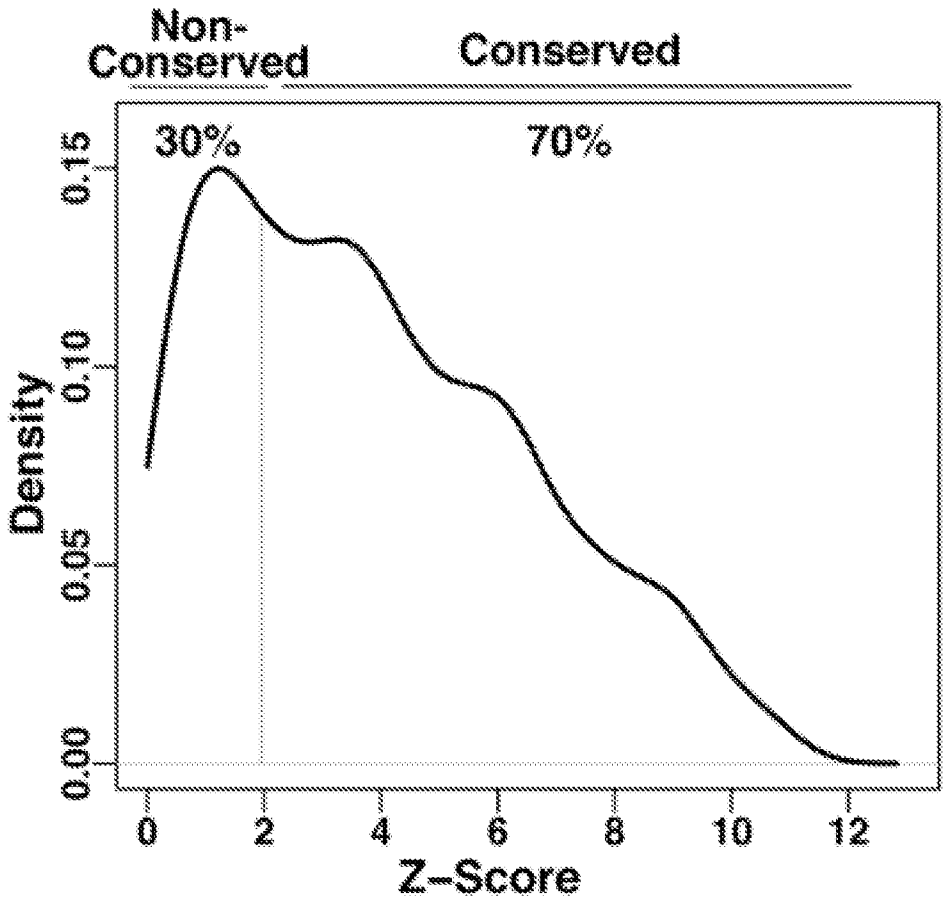
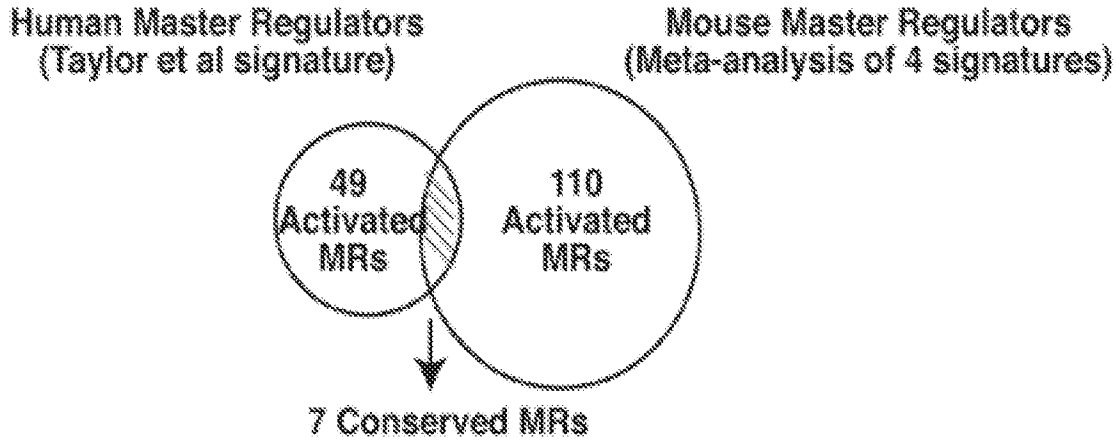

Non-
Conserved
Conserved
30%
70%
Density
0,15
0,10
0,05
0,00
0
2
4
6
8
10
12
Z-Score

FIG. 3A



Clinical features	Non-Selected MRs	Conserved MRs
Associated w/ disease outcome	21%	43%
Differentially expressed in advanced prostate cancer	60%	100%

FIG. 3B

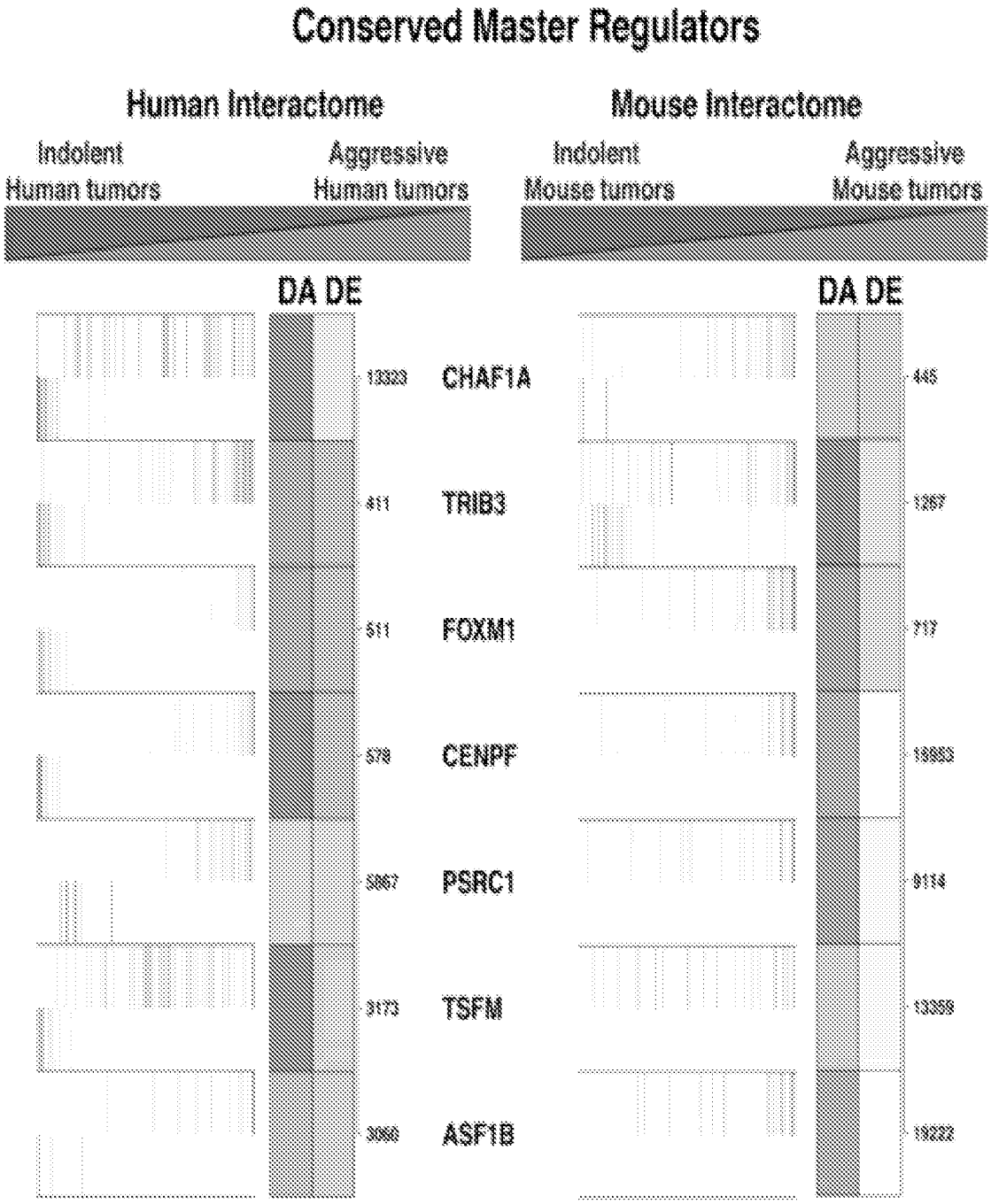

Conserved Master Regulators
Human Interactome
Indolent
Human tumors
Aggressive
Human tumors
Mouse Interactome
Indolent
Mouse tumors
Aggressive
Mouse tumors
DA DE
DA DE
13323
CHAF1A
411
TRIB3
511
FOXM1
578
CENPF
3867
PSRC1
3173
TSFM
3060
ASF1B
445
1267
717
15853
9114
13359
19222

FIG. 3C

Conserved MR	MARINa Joint p value	COX Expression p value	COX Activity p value	Oncomine expression p value
CHAF1A	2.3 x 10-5	0.9190	0.2662	0.0030
TRIB3	1.8 x 10-4	0.1174	0.1893	0.00004
FOXM1	3.0 x 10-4	0.0100	0.0094	0.0045
CENPF	8.3 x 10-4	0.0010	0.0019	0.0015
PSRC1	8.4 x 10-4	0.2249	0.1394	0.021
TSFM	1.0 x 10-3	0.0650	0.5167	0.251
ASF1B	7.0 x 10-3	0.1725	0.0408	0.044

FIG. 4

Predicted Synergy (p value)

	CENPF	FOXM1	TRIB3	CHAF1A	PSRC1	ASF1B
CENPF	---	< 0.001	1	1	1	1
FOXM1		---	1	1	1	0.110
TRIB3			---	1	1	1
CHAF1A				---	1	1
PSRC1					---	1
ASF1B						---
TSFM						

FIG. 5

Description		Primary dataset	Secondary datasets				RNA (gene expression)			Protein (immunohistochemistry)	
		Taylor	Yu	Wang (a)	Wang (b)	Balk	Sboner	Glinsky		Outcome TMA (MSKCC)	Metastasis TMA (Mich)
GEO accession #		GSE21034	GSE6919	GSE17951	GSE8218	GSE32269	GSE16560	NA			
Patients/samples		131 Primary 29 Ad. Normal 19 Metastases 6 cell lines	83 Primary 17 Normal 59 Ad. Normal	109 Primary 22 Biopsies 13 Normal	144 Primary 4 Normal	22 Primary 29 Bone metastases 4 normal bone	281 TURP	79 Primary		316 Primary (321 informative cases)	60 Metastases (53 informative)
Median Age (Years ± MAD)		58±4.4	NA	NA	NA	NA	74±5	60±6		61.6±7.5	NA
Pathology Stage T	T2 T3 T4	85 39 7	25 36 2	NA	NA	NA	NA	NA		555 240 26	NA
Clinical T- Stage	T1 T2 T3	75 51 5	NA	NA	NA	NA	NA	NA		404 368 16 1 (NA)	NA
Pathology Stage N	N0 N1 Nx	102 6 23	NA	NA	NA	NA	NA	79 3 (N1 or Nx)		NA	NA
Pathology Gleason score	≤5 6 7 ≥8	0 41 54 (3+4) 20 (4+3) 15	3 15 27 18	NA	NA	NA	NA	2 15 30 (3+4) 14 (4+3) 18		27 240 347 (3+4) 103 (4+3) 82	NA
Biopsy Gleason score	≤5 6 7 ≥8	1 77 29 (3+4) 13 (4+3) 11	NA	NA	NA	NA	0 63 79 (3+4) 38 (4+3) 81	13 24 30 (3+4) 12 (4+3) 10		82 423 156 (3+4) 72 (4+3) 58 32 (NA)	NA
SVI %	Negative Positive	89.3% 10.8%	NA	NA	NA	NA	NA	87.3% 12.7%		95.9% 4.1%	NA
Extracapsular Extension %	Present Absent	67.7% 32.3%	NA	NA	NA	NA	NA	79.4% 21.6		80.5% 19.5%	NA
BCR Median (Months ± MAD)		17.8±10.6 (N = 27)	NA	NA	NA	NA	NA	51.5±30.4 (N=37)		21.4±25.5 (N=266)	NA
Median overall survival (Months ± MAD)		NA	NA	NA	NA	NA	75±40	NA		87.4±39.3 (N=247)	NA
Median time to metastasis (Months ± MAD)		NA	NA	NA	NA	NA	NA	NA		54.2±44.1 (N=79)	NA

FIG. 6A

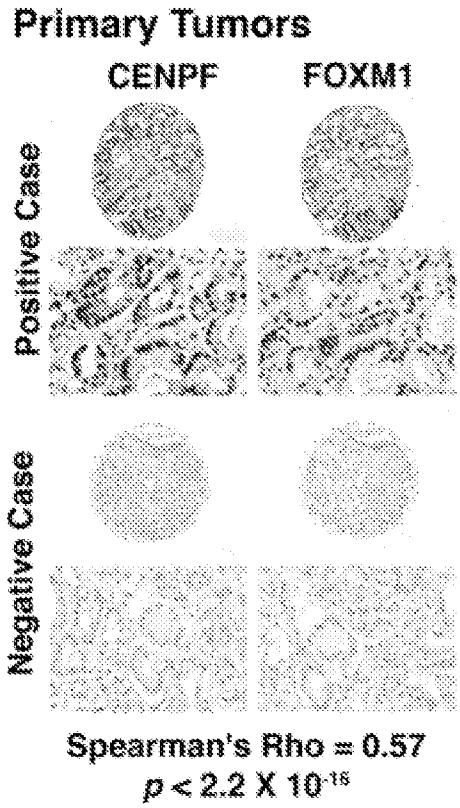

Primary Tumors
CENPF
FOXM1
Positive Case
Negative Case
Spearman's Rho = 0.57
p < 2.2 X 10-16

FIG. 6B

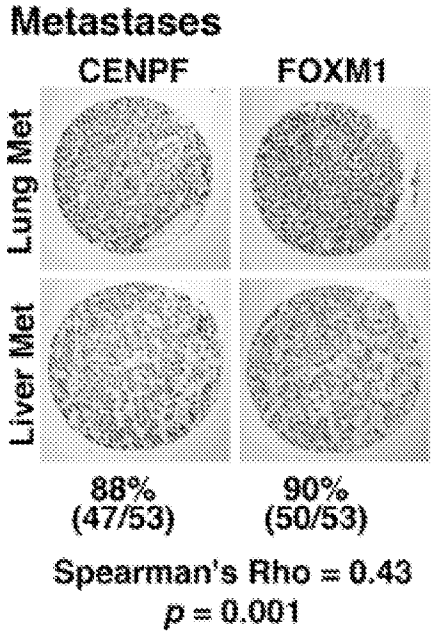

Metastases
CENPF
FOXM1
Lung Met
Liver Met
88%
(47/53)
90%
(50/53)
Spearman's Rho = 0.43
p = 0.001

FIG. 6C

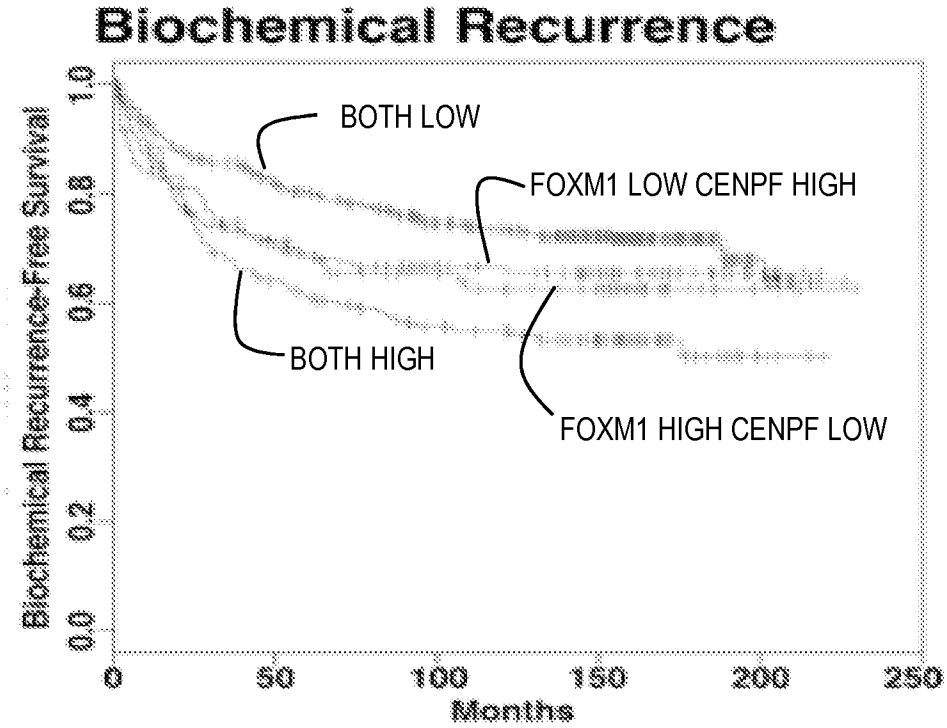

Biochemical Recurrence
Biochemical Recurrence-Free Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
Months
BOTH LOW
FOXM1 LOW CENPF HIGH
BOTH HIGH
FOXM1 HIGH CENPF LOW

FIG. 6D

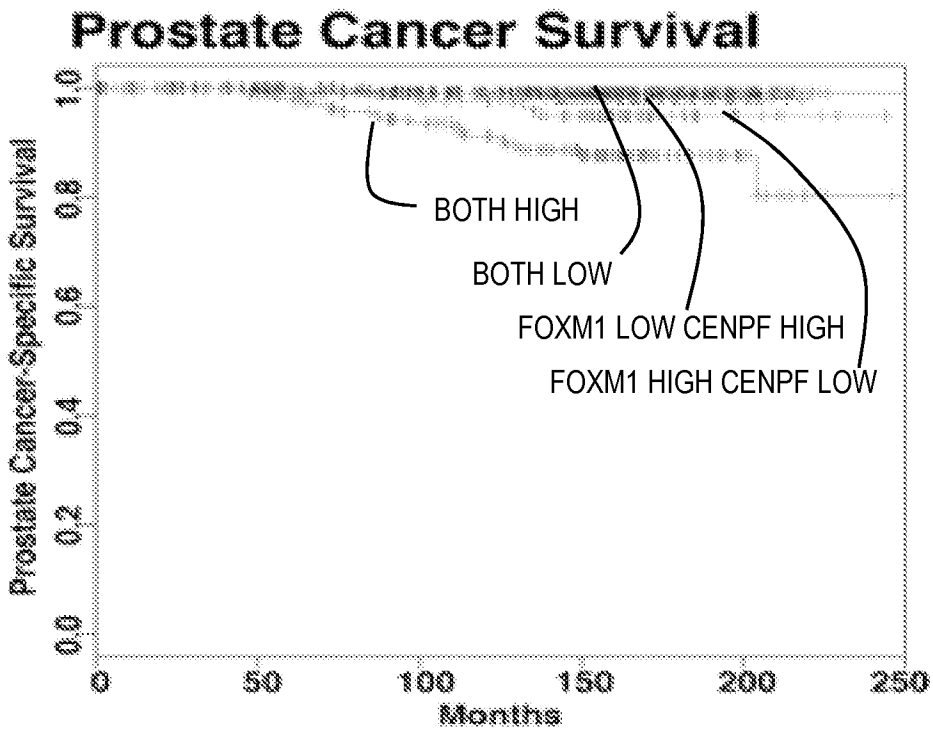

Prostate Cancer Survival
Prostate Cancer-Specific Survival
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
Months
BOTH HIGH
BOTH LOW
FOXM1 LOW CENPF HIGH
FOXM1 HIGH CENPF LOW

FIG. 6E

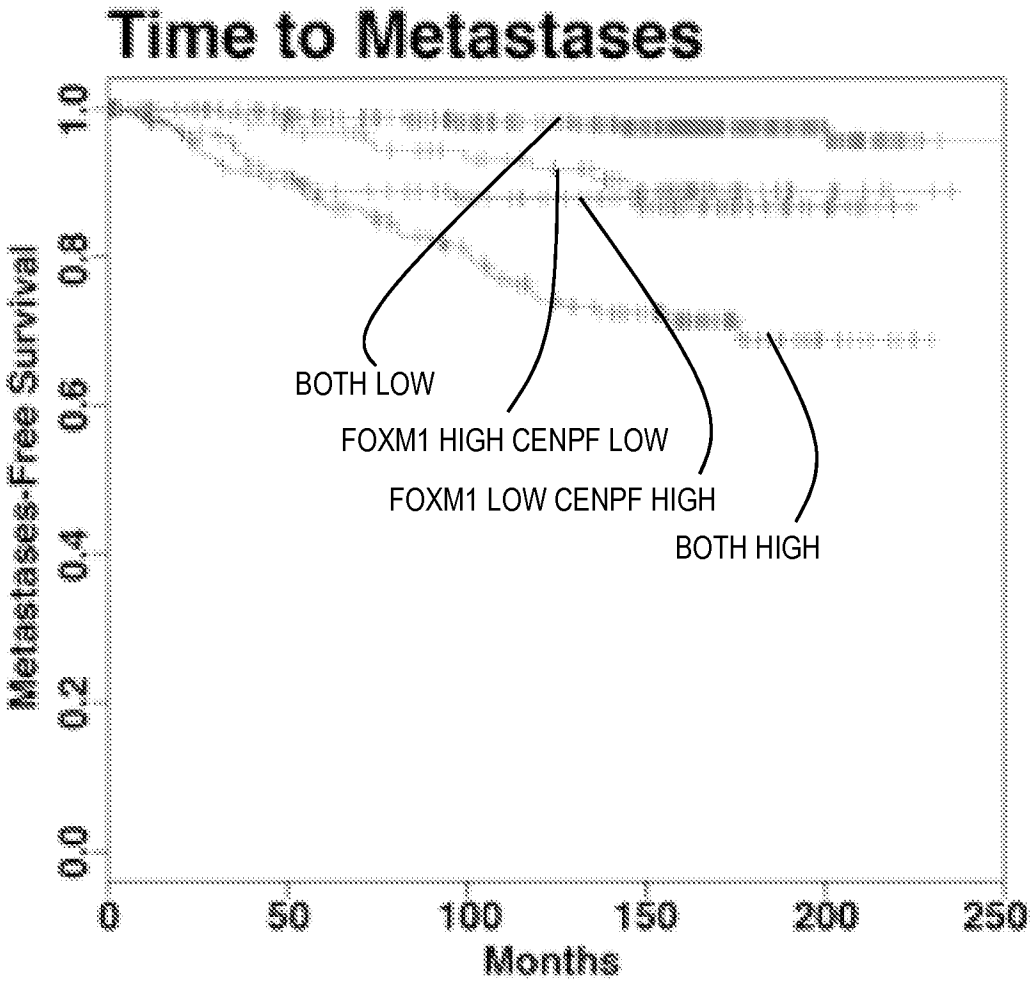

Time to Metastases
Metastases-Free Survival
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
250
Months
BOTH LOW
FOXM1 HIGH CENPF LOW
FOXM1 LOW CENPF HIGH
BOTH HIGH

FIG. 6F

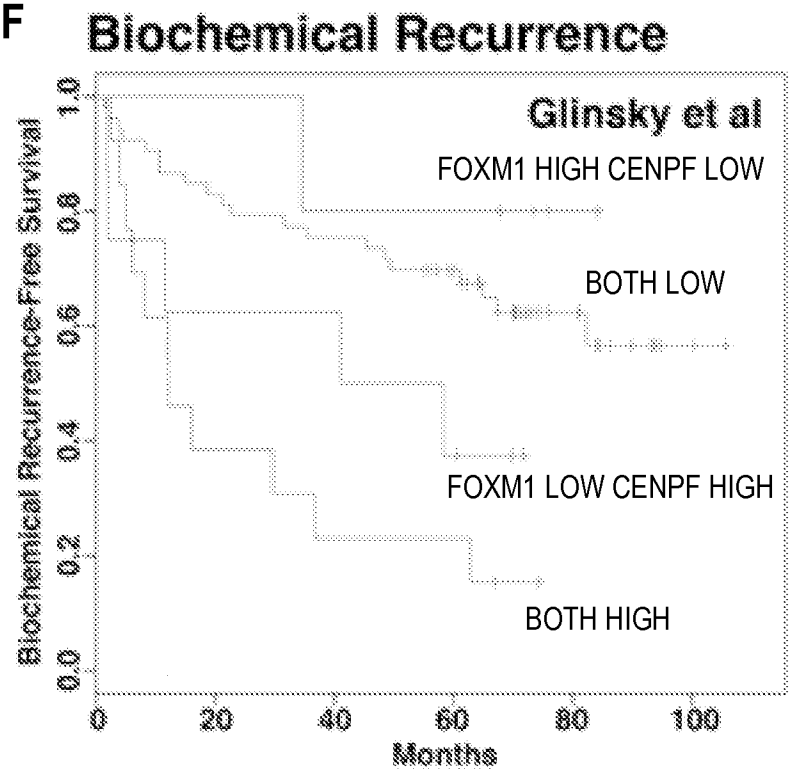

Biochemical Recurrence
Glinsky et al
FOXM1 HIGH CENPF LOW
BOTH LOW
FOXM1 LOW CENPF HIGH
BOTH HIGH
Biochemical Recurrence-Free Survival
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Months

FIG. 6G

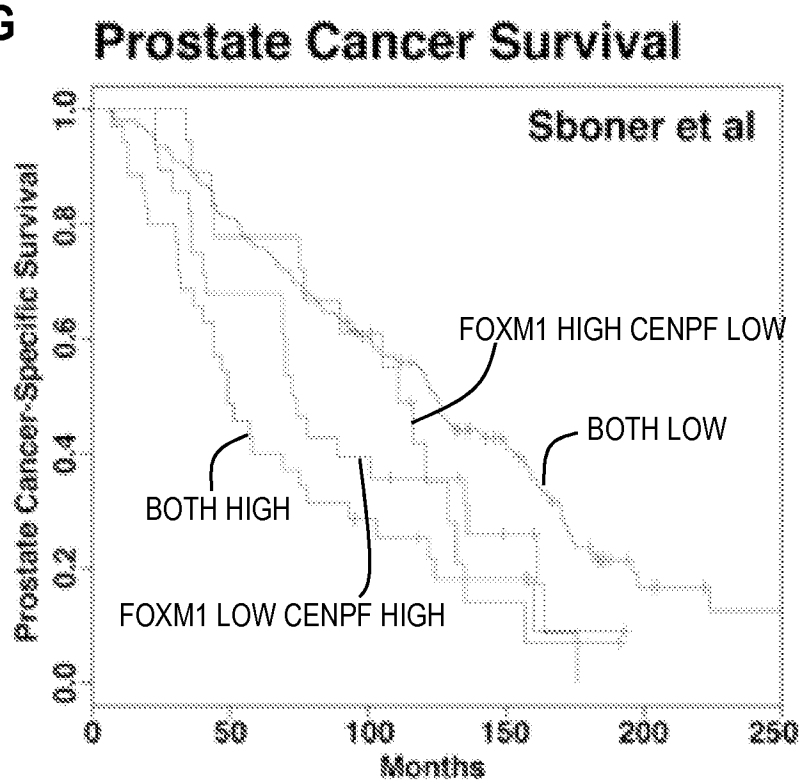

Prostate Cancer Survival
Sboner et al
FOXM1 HIGH CENPF LOW
BOTH LOW
BOTH HIGH
FOXM1 LOW CENPF HIGH
Prostate Cancer-Specific Survival
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
200
250
Months

FIG. 7

C-Statistics Model - MSKCC TMA

Prostate Cancer-Specific Survival

	C index	Confidence interval	p value
Gleason	0.65	0.48 - 0.81	0.038
FOXM1 & CENPF	0.71	0.59 - 0.84	2.4 x 10-4
Gleason + FOXM1 & CENPF	0.86	0.80 - 0.93	1.0 x 10-30

p value for improvement 2.7 X 10-4

Time to Metastases

	C index	Confidence interval	p value
Gleason	0.61	0.54 - 0.70	0.002
FOXM1 & CENPF	0.77	0.71 - 0.83	3.1 x 10-19
n + FOXM1 & CENPF	0.86	0.81 - 0.89	6.5 x 10-58

p value for improvement 5.3 X 10-13

FIG. 8

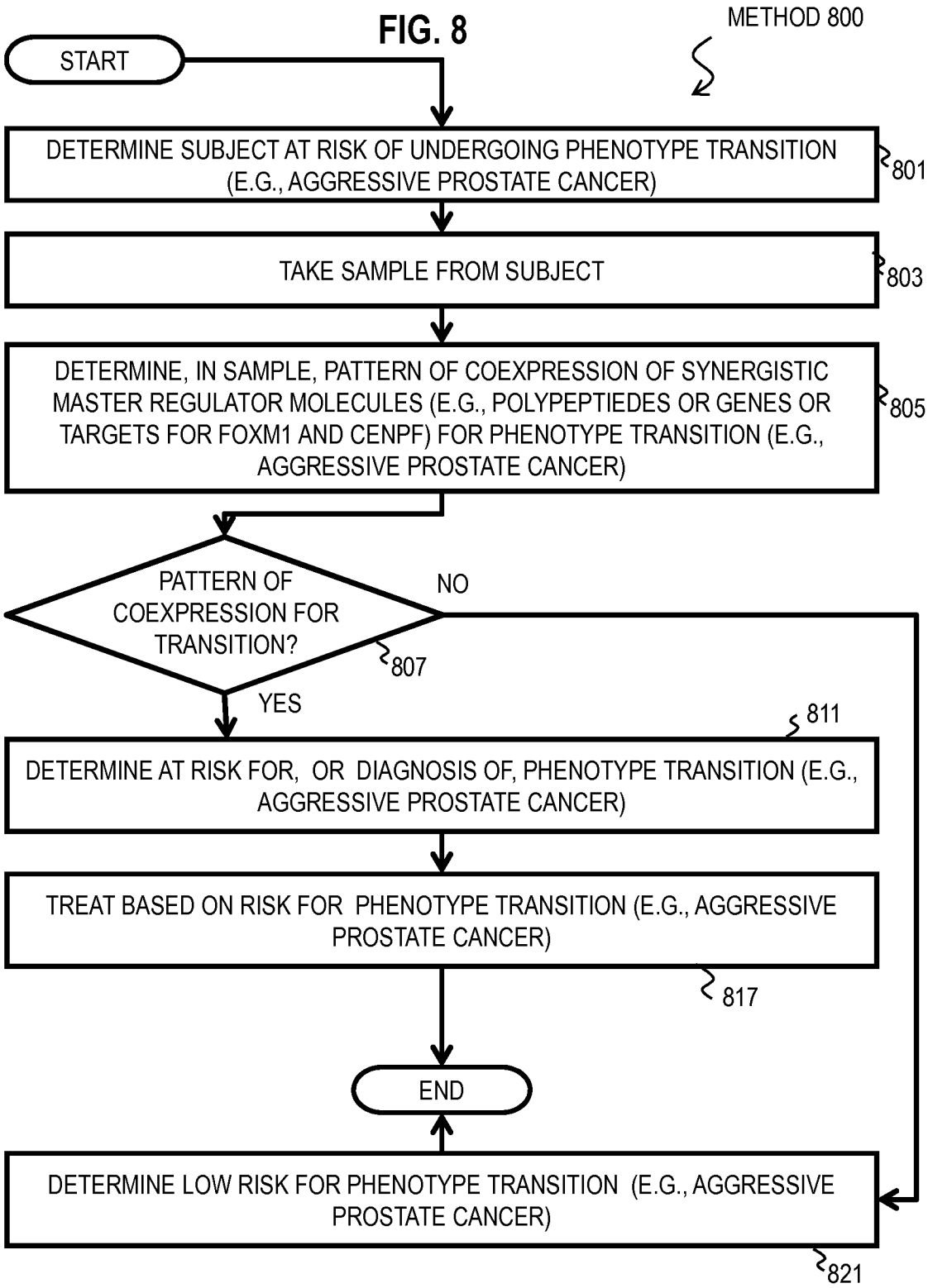

METHOD 800
START
DETERMINE SUBJECT AT RISK OF UNDERGOING PHENOTYPE TRANSITION (E.G., AGGRESSIVE PROSTATE CANCER)
801
TAKE SAMPLE FROM SUBJECT
803
DETERMINE, IN SAMPLE, PATTERN OF COEXPRESSION OF SYNERGISTIC MASTER REGULATOR MOLECULES (E.G., POLYPEPTIEDES OR GENES OR TARGETS FOR FOXM1 AND CENPF) FOR PHENOTYPE TRANSITION (E.G., AGGRESSIVE PROSTATE CANCER)
805
PATTERN OF COEXPRESSION FOR TRANSITION?
807
NO
YES
811
DETERMINE AT RISK FOR, OR DIAGNOSIS OF, PHENOTYPE TRANSITION (E.G., AGGRESSIVE PROSTATE CANCER)
TREAT BASED ON RISK FOR PHENOTYPE TRANSITION (E.G., AGGRESSIVE PROSTATE CANCER)
817
END
DETERMINE LOW RISK FOR PHENOTYPE TRANSITION (E.G., AGGRESSIVE PROSTATE CANCER)
821

FIG. 9A

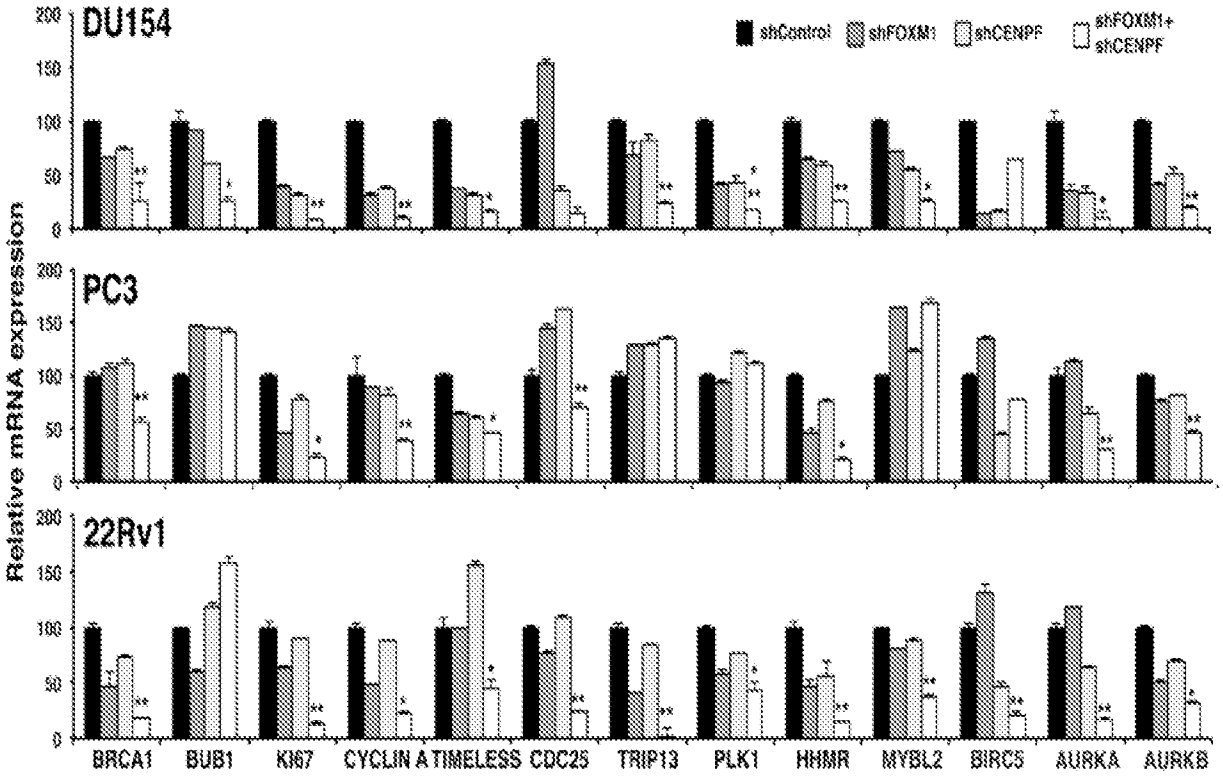

DU154
PC3
22Rv1
shControl
shFOXM1
shCENPF
shFOXM1+ shCENPF
Relative mRNA expression
BRCA1
BUB1
KI67
CYCLIN A
TIMELESS
CDC25
TRIP13
PLK1
HHMR
MYBL2
BIRC5
AURKA
AURKB

FIG. 9B

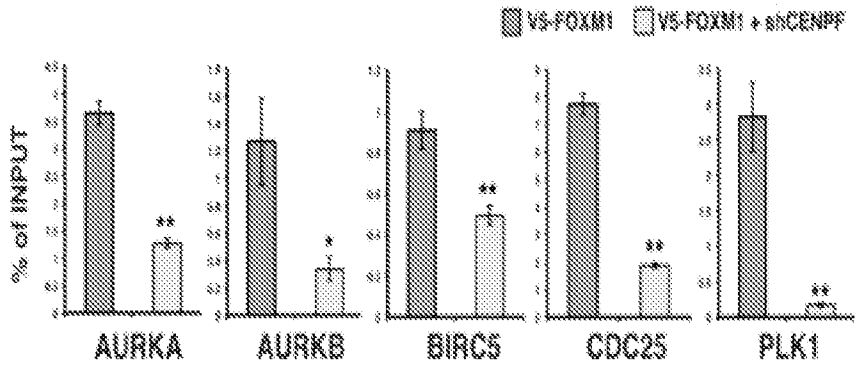

V5-FOXM1
V5-FOXM1 + shCENPF
% of INPUT
AURKA
AURKB
BIRC5
CDC25
PLK1

FIG. 9C

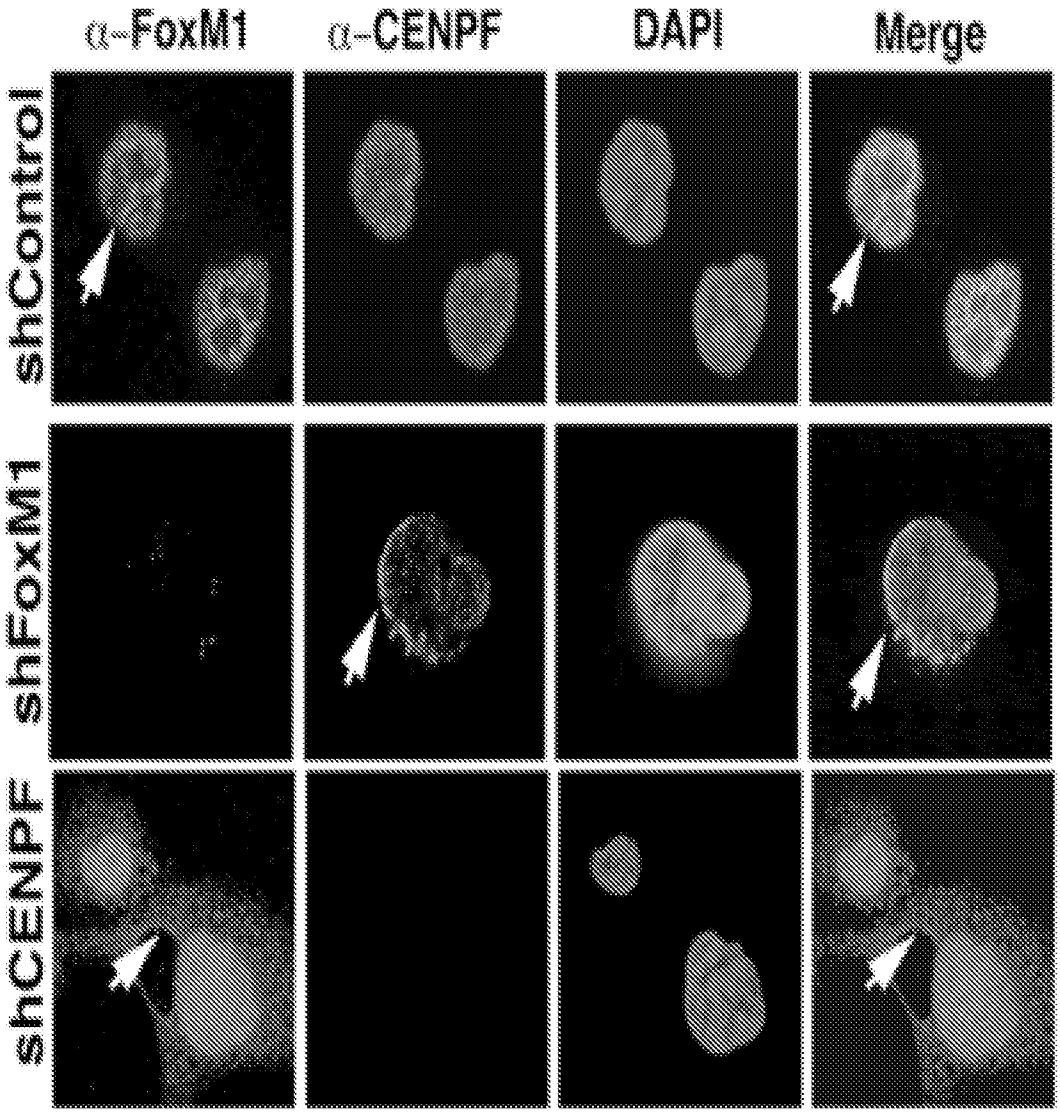
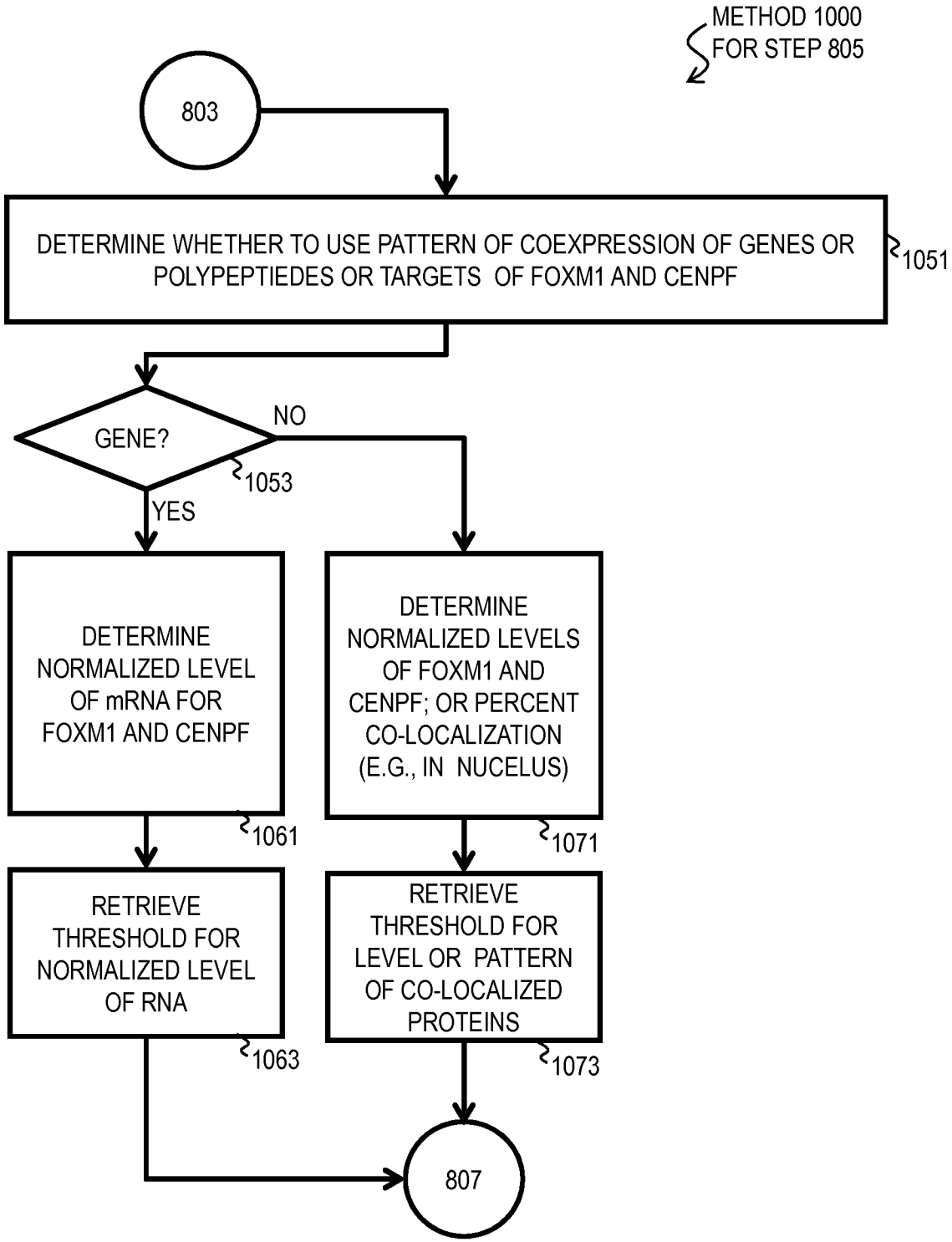


FIG. 10


METHOD 1000
FOR STEP 805
803
DETERMINE WHETHER TO USE PATTERN OF COEXPRESSION OF GENES OR POLYPEPTIEDES OR TARGETS OF FOXM1 AND CENPF
1051
GENE?
1053
NO
YES
DETERMINE NORMALIZED LEVEL OF mRNA FOR FOXM1 AND CENPF
1061
DETERMINE NORMALIZED LEVELS OF FOXM1 AND CENPF; OR PERCENT CO-LOCALIZATION (E.G., IN NUCELUS)
1071
RETRIEVE THRESHOLD FOR NORMALIZED LEVEL OF RNA
1063
RETRIEVE THRESHOLD FOR LEVEL OR PATTERN OF CO-LOCALIZED PROTEINS
1073
807

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/016650

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12Q 1/68 (2015.01)
CPC - G01N 33/57434 (2015.04)
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C12Q 1/00, 1/68; G01N 33/00, 33/48, 33/50 (2015.01)
CPC - G01N 33/50, 33/53, 33/574, 33/57407, 33/57434 (2015.04)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
CPC - G01N 33/50, 33/53, 33/574, 33/57407, 33/57434 (2015.04) (keyword delimited)
USPC - 435/4, 6.1, 6.14; 702/1, 19

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

Orbit, Google Patents, PubMed, Google
Search terms used: prostate cancer sample determine expression level foxm1 cenpf antibody complex staining value composite score

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	US 2012/0053253 A1 (STONE et al) 01 March 2012 (01.03.2012) entire document	1-3, 8-12 ---------- 6, 13-19
Y	CA 2 432 361 A1 (THE UNIVERSITY OF TEXAS SYSTEM) 19 August 1999 (19.08.1999) entire document	6
Y	US 2007/0298419 A1 (PARK et al) 27 December 2007 (27.12.2007) entire document	13-16
Y	US 2013/0053273 A1 (JUNCKER et al) 28 February 2013 (28.02.2013) entire document	17-19
A	US 2012/0315216 A1 (CLARKE et al) 13 December 2012 (13.12.2012) entire document	1-19

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 April 2015	15 MAY 2015
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774