



US 20230184776A1

(19) **United States**

(12) **Patent Application Publication**

Rao et al.

(10) **Pub. No.: US 2023/0184776 A1**

(43) **Pub. Date: Jun. 15, 2023**

(54) **METHOD OF ISOLATING CIRCULATING TUMOR CELLS**

(71) Applicants: **Duke University**, Durham, NC (US); **Menarini Silicon Biosystems S.p.A.**, Bologna (IT)

(72) Inventors: **Galla Chandra Rao**, Princeton Junction, NJ (US); **Mark C. Connelly**, Doylestown, PA (US); **Mariano A. Garcia-Blanco**, Hillsborough, NC (US); **Andrew J. Armstrong**, Chapel Hill, NC (US); **Rhonda L. Bitting**, Hillsborough, NC (US)

(21) Appl. No.: **18/063,678**

(22) Filed: **Dec. 8, 2022**

Related U.S. Application Data

(62) Division of application No. 16/206,932, filed on Nov. 30, 2018, now Pat. No. 11,567,081, which is a division of application No. 14/765,071, filed on Jul. 31, 2015, now Pat. No. 10,161,939, filed as application No. PCT/US13/43745 on May 31, 2013.

(60) Provisional application No. 61/806,358, filed on Mar. 28, 2013, provisional application No. 61/760,042, filed on Feb. 2, 2013.

Publication Classification

(51) **Int. Cl.**
G01N 33/574 (2006.01)
G01N 33/543 (2006.01)
G01N 33/569 (2006.01)

(52) **U.S. Cl.**
CPC . G01N 33/57484 (2013.01); **G01N 33/54326** (2013.01); **G01N 33/56966** (2013.01); **G01N 33/57415** (2013.01); **G01N 33/57434** (2013.01); **G01N 33/57492** (2013.01); **G01N 2333/70589** (2013.01); **G01N 2333/47** (2013.01); **G01N 2333/4703** (2013.01); **G01N 2333/4742** (2013.01); **G01N 2333/705** (2013.01); **G01N 2333/70503** (2013.01); **G01N 2333/70596** (2013.01)

(57) **ABSTRACT**

Provided are methods for detecting or isolating circulating tumor cells (CTCs) in a subject. The methods may include detecting the expression of at least one epithelial mesenchymal transition (EMT) biomarker. Further provided are kits for detecting or isolating CTCs. The kits may include antibodies to at least one EMT biomarker. Further provided are methods of predicting the responsiveness of a subject to a cancer drug, methods of targeting delivery of a cancer drug in a subject, methods of providing a cancer prognosis to a subject, and methods for following the progress of cancer in a subject.

Specification includes a Sequence Listing.

FIG. 1A

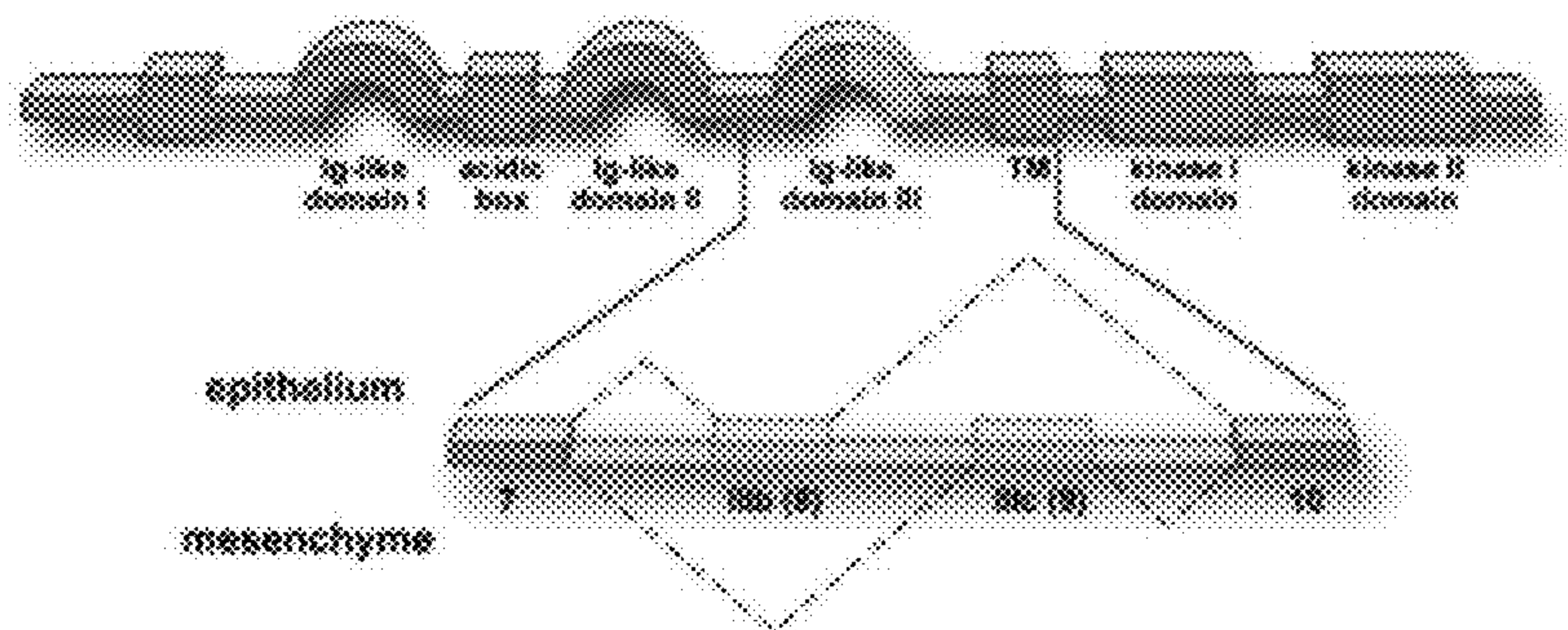


FIG. 1B



FIG. 1C

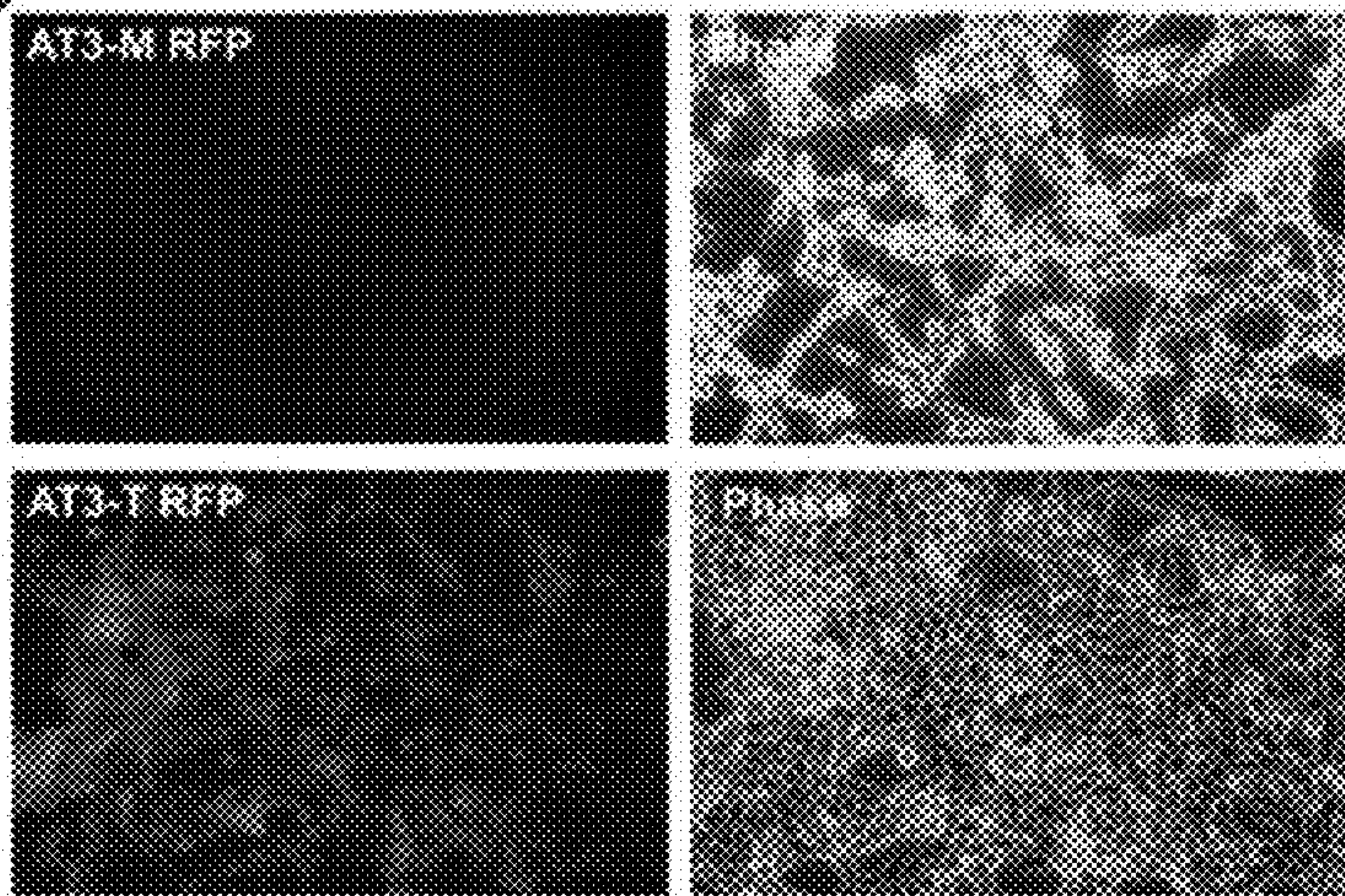
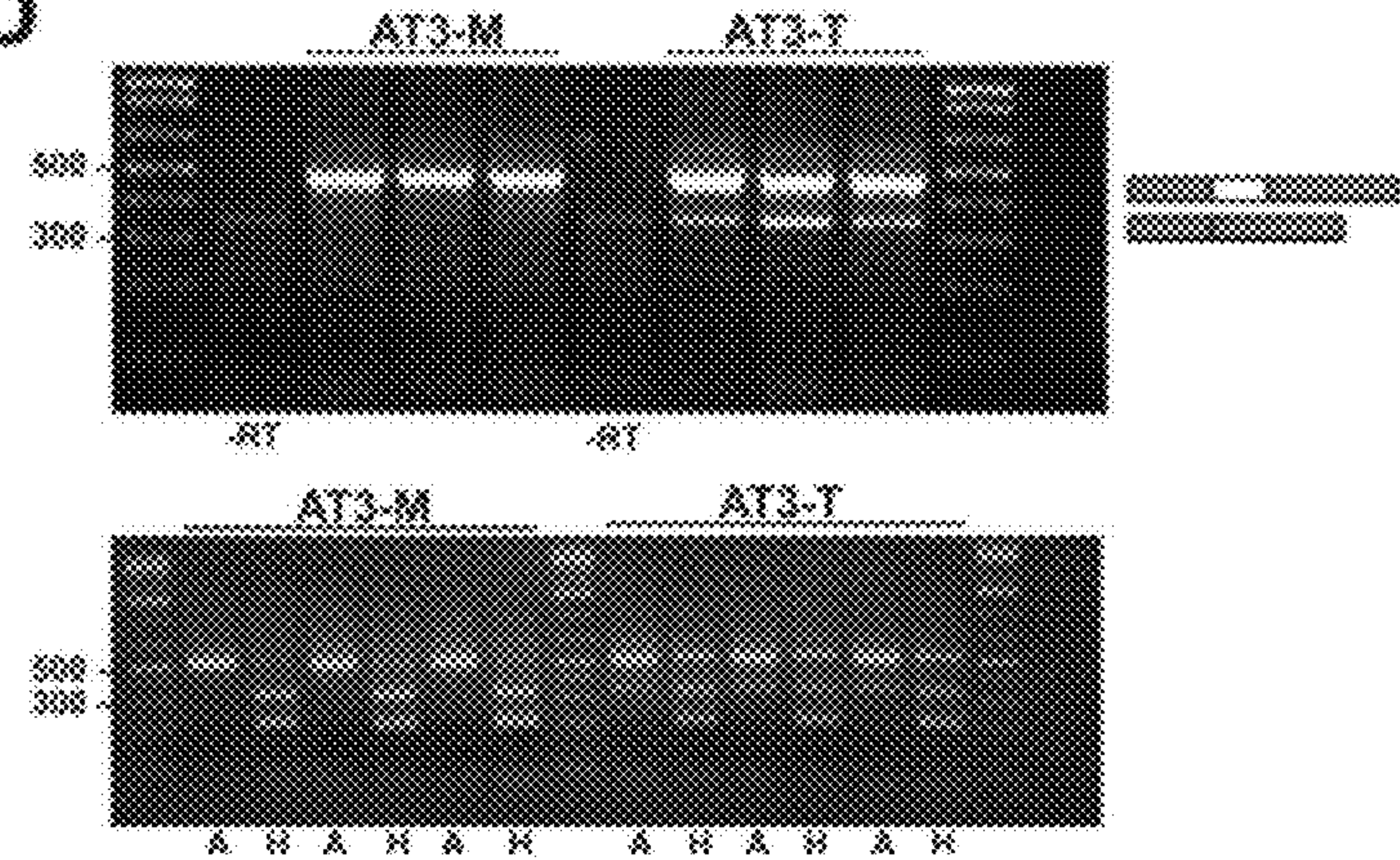


FIG. 1D



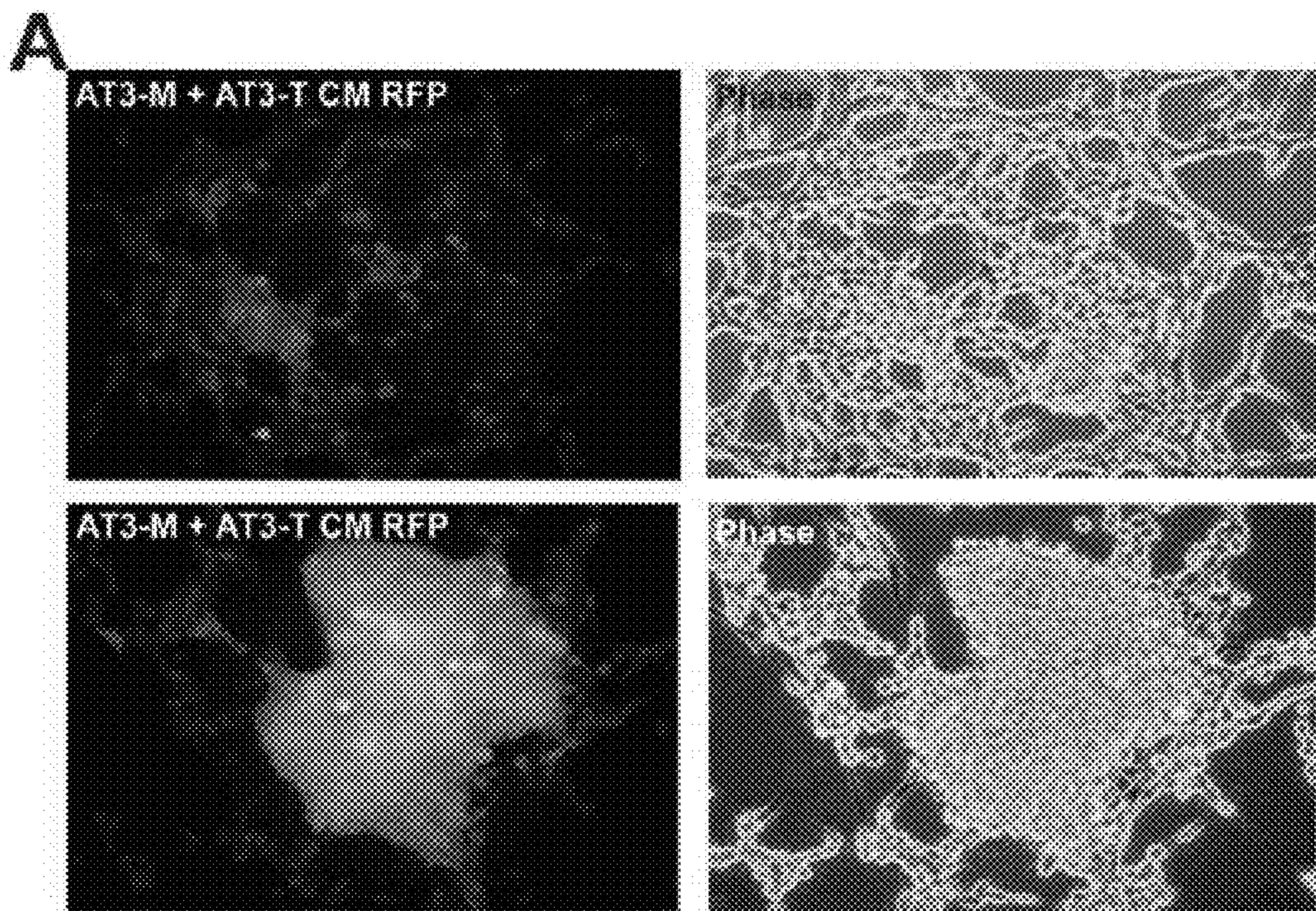


FIG. 2A

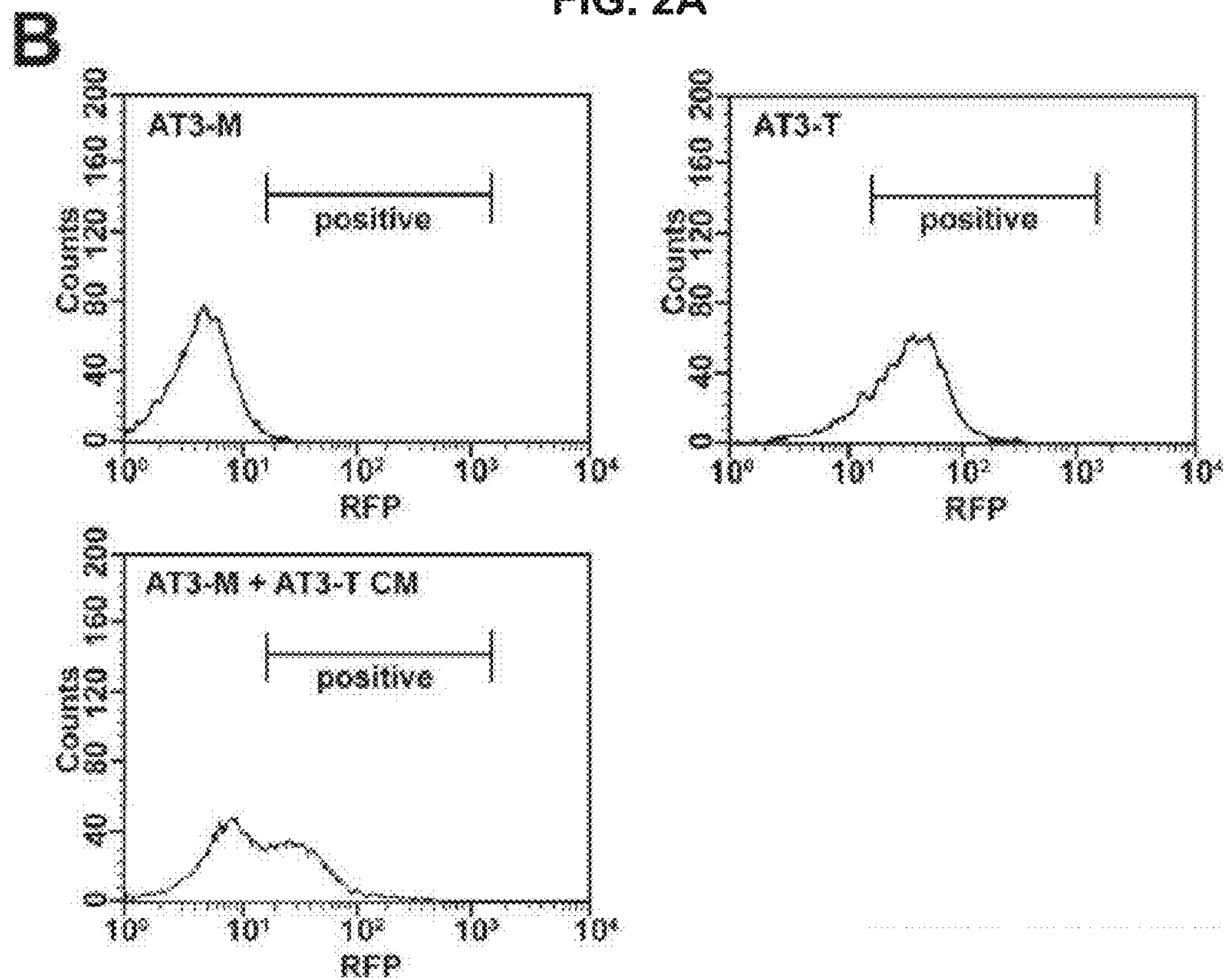


FIG. 2B

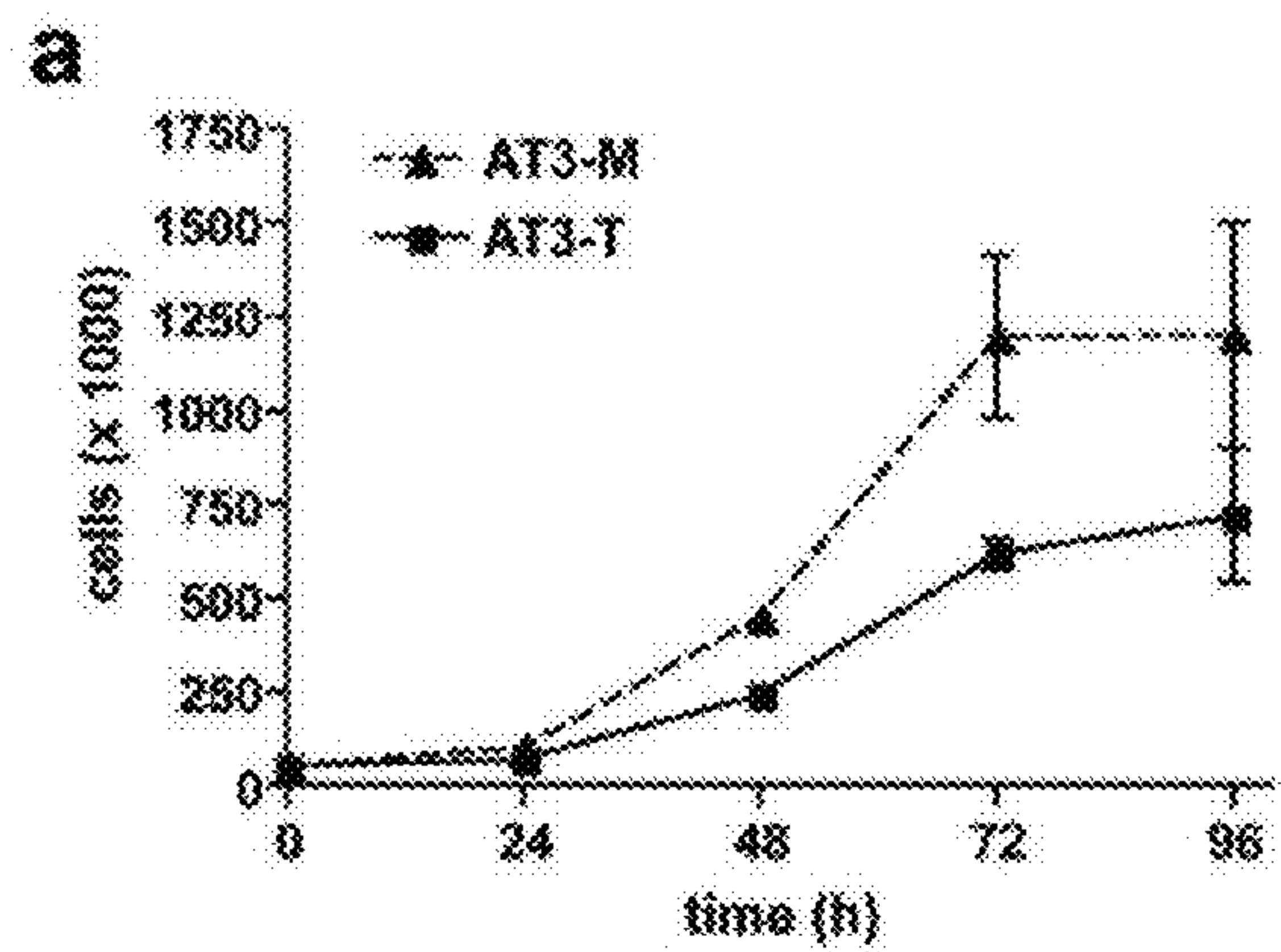


FIG. 3A

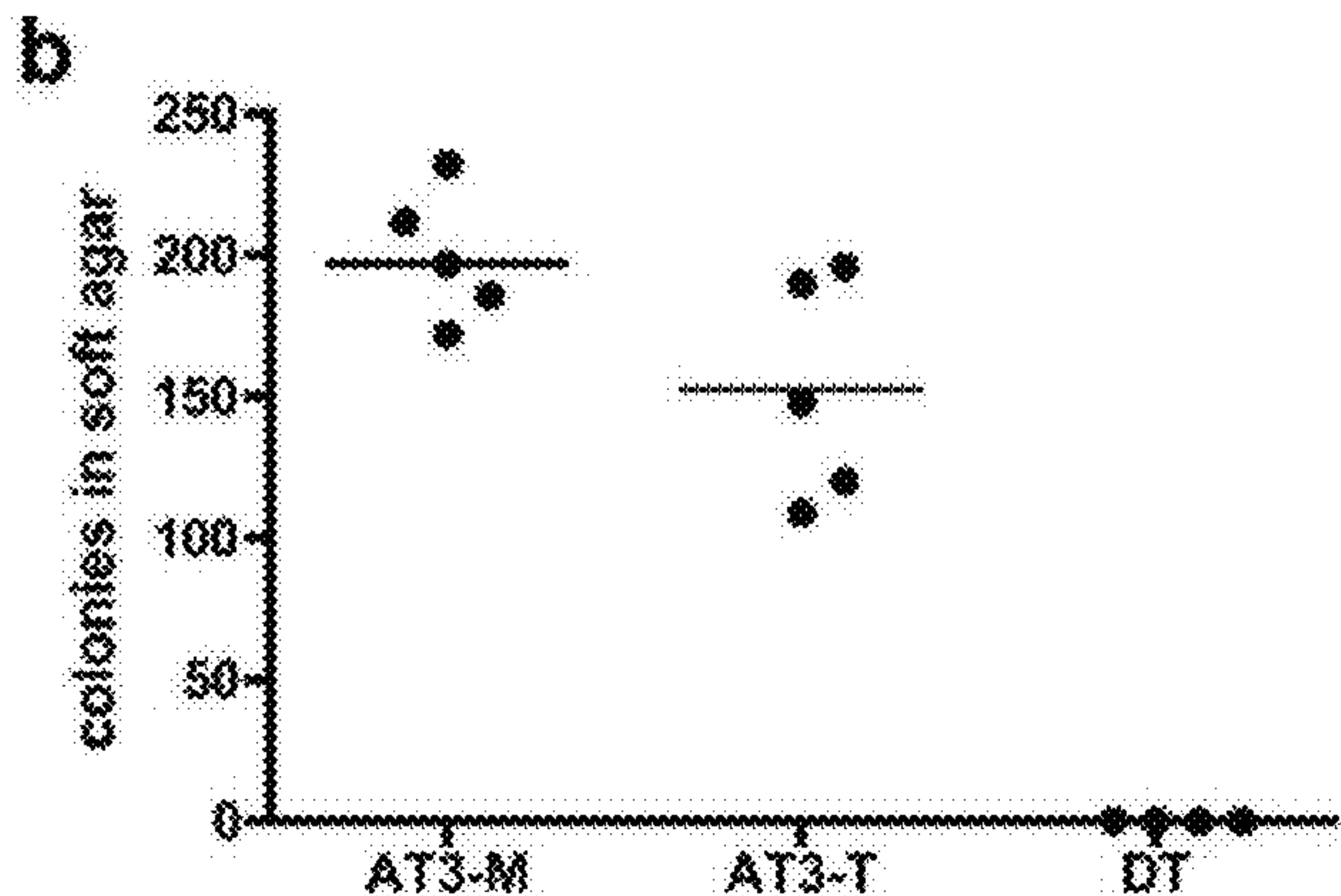


FIG. 3B

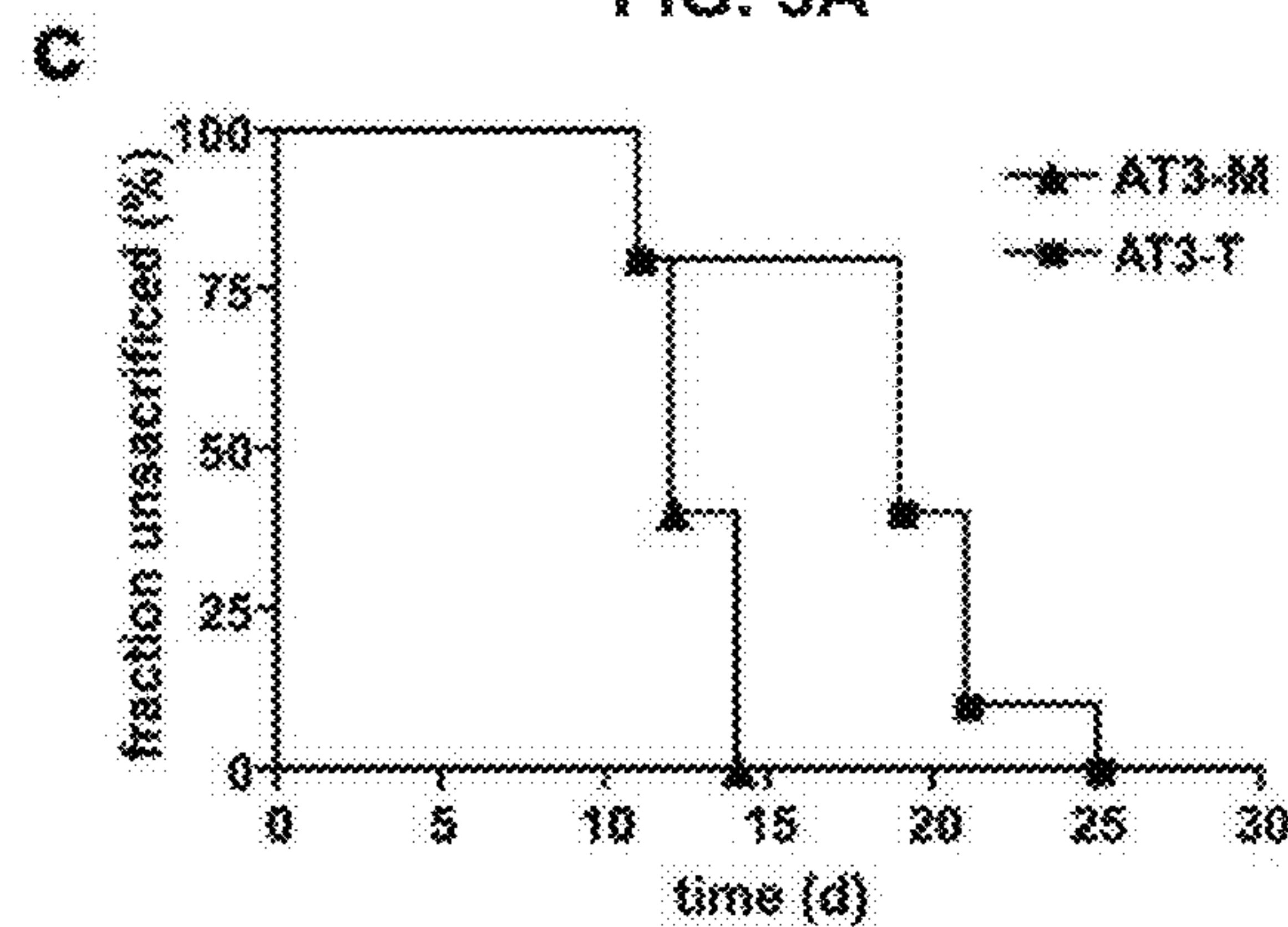


FIG. 3C

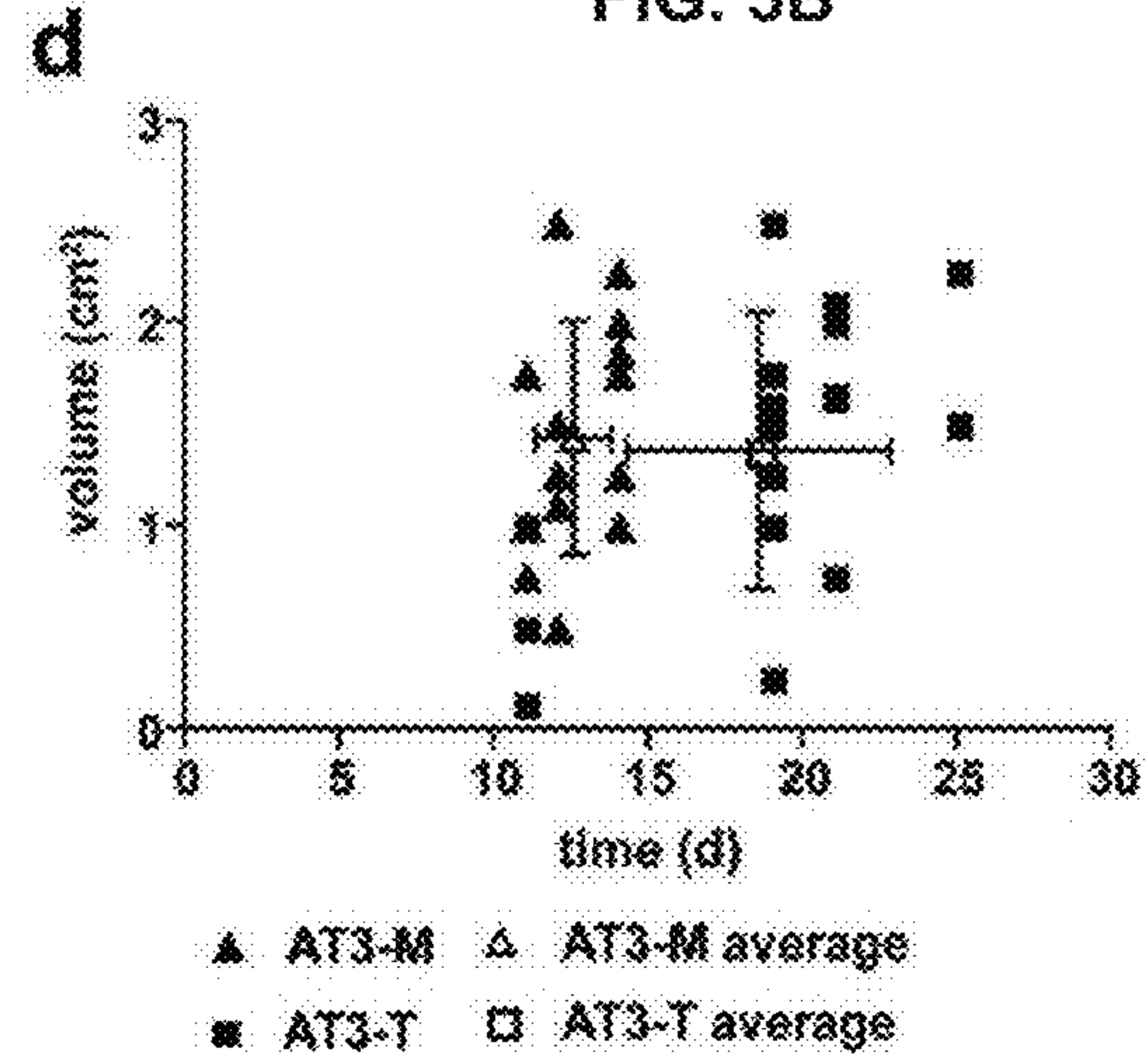


FIG. 3D

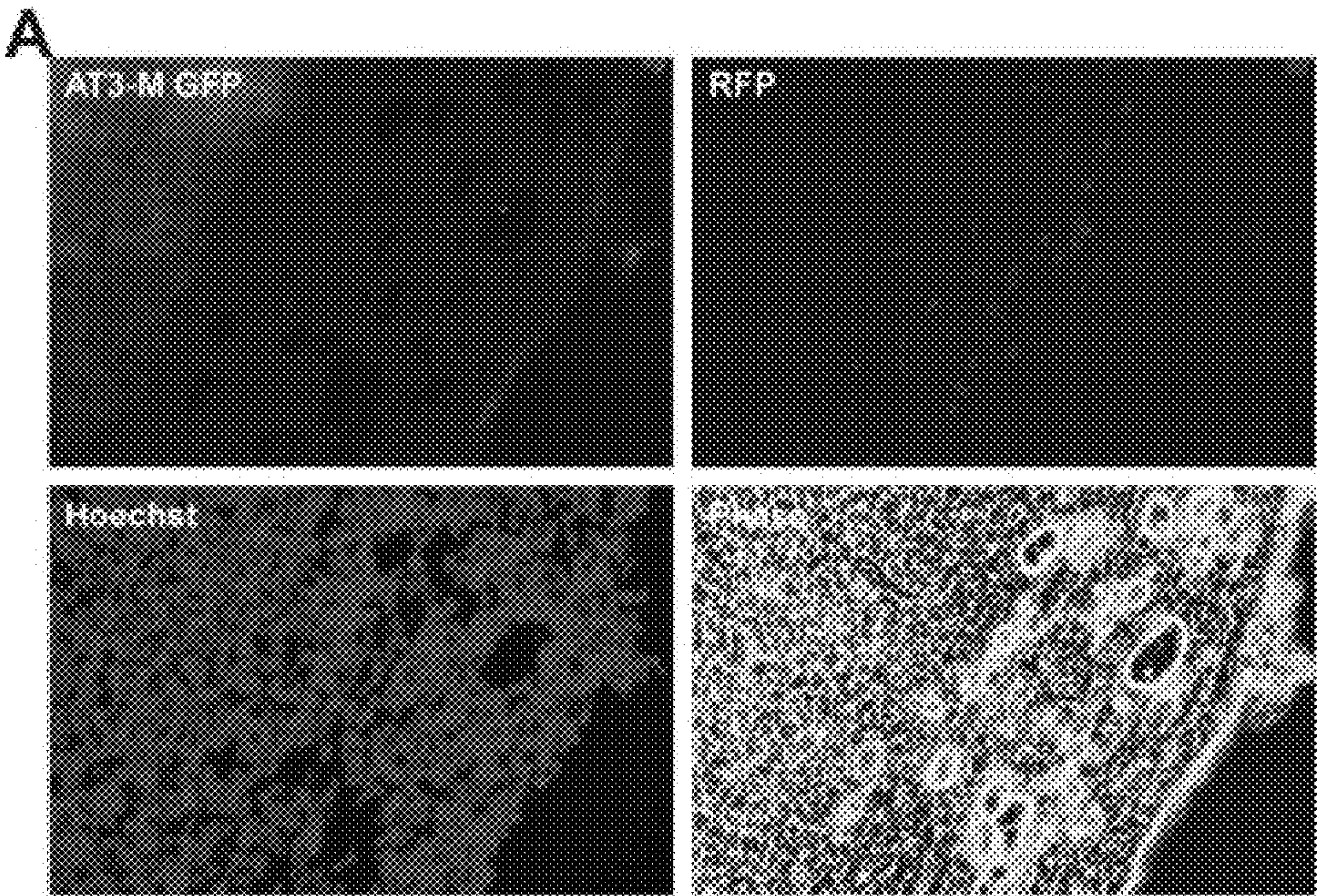


FIG. 4A

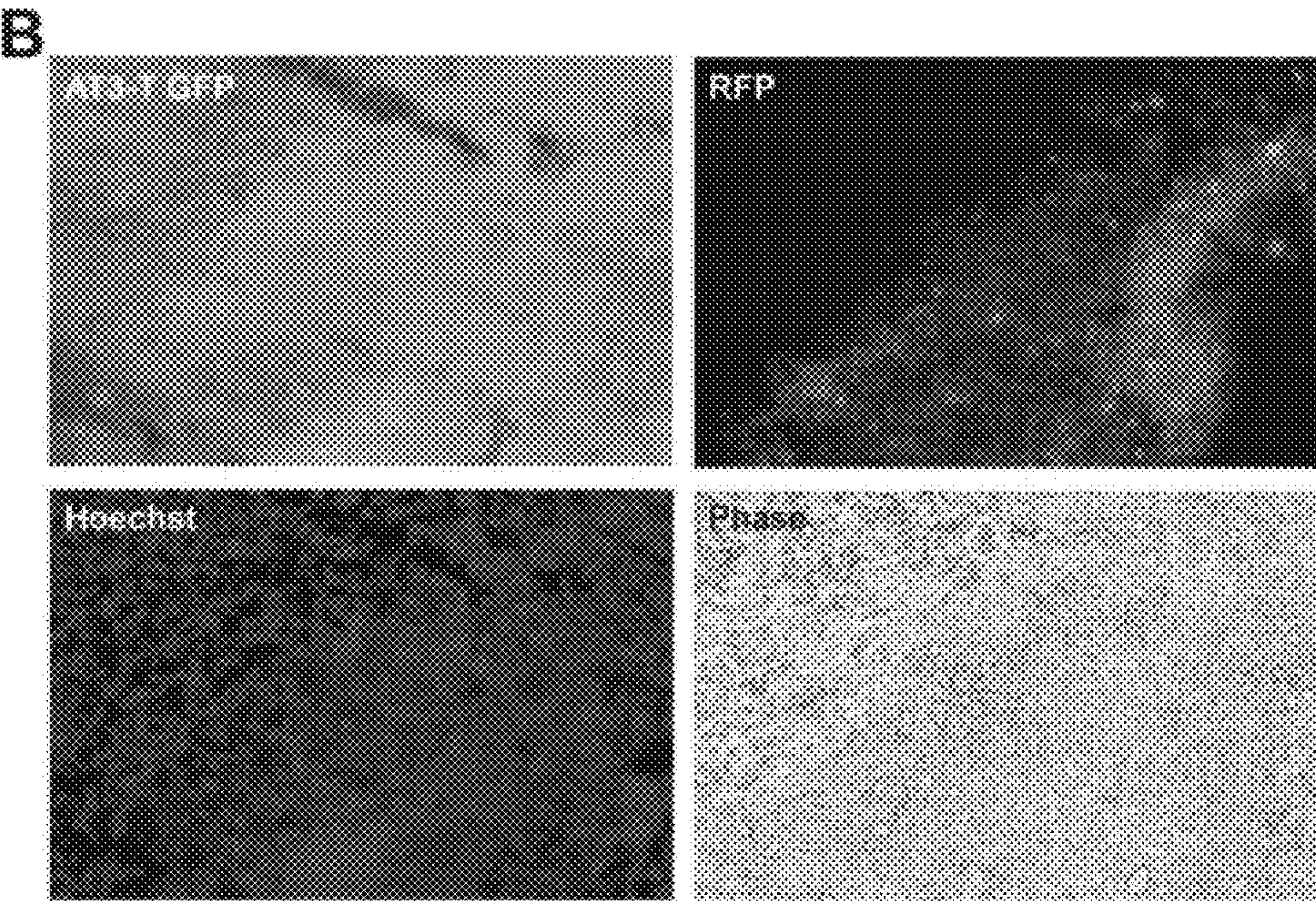


FIG. 4B

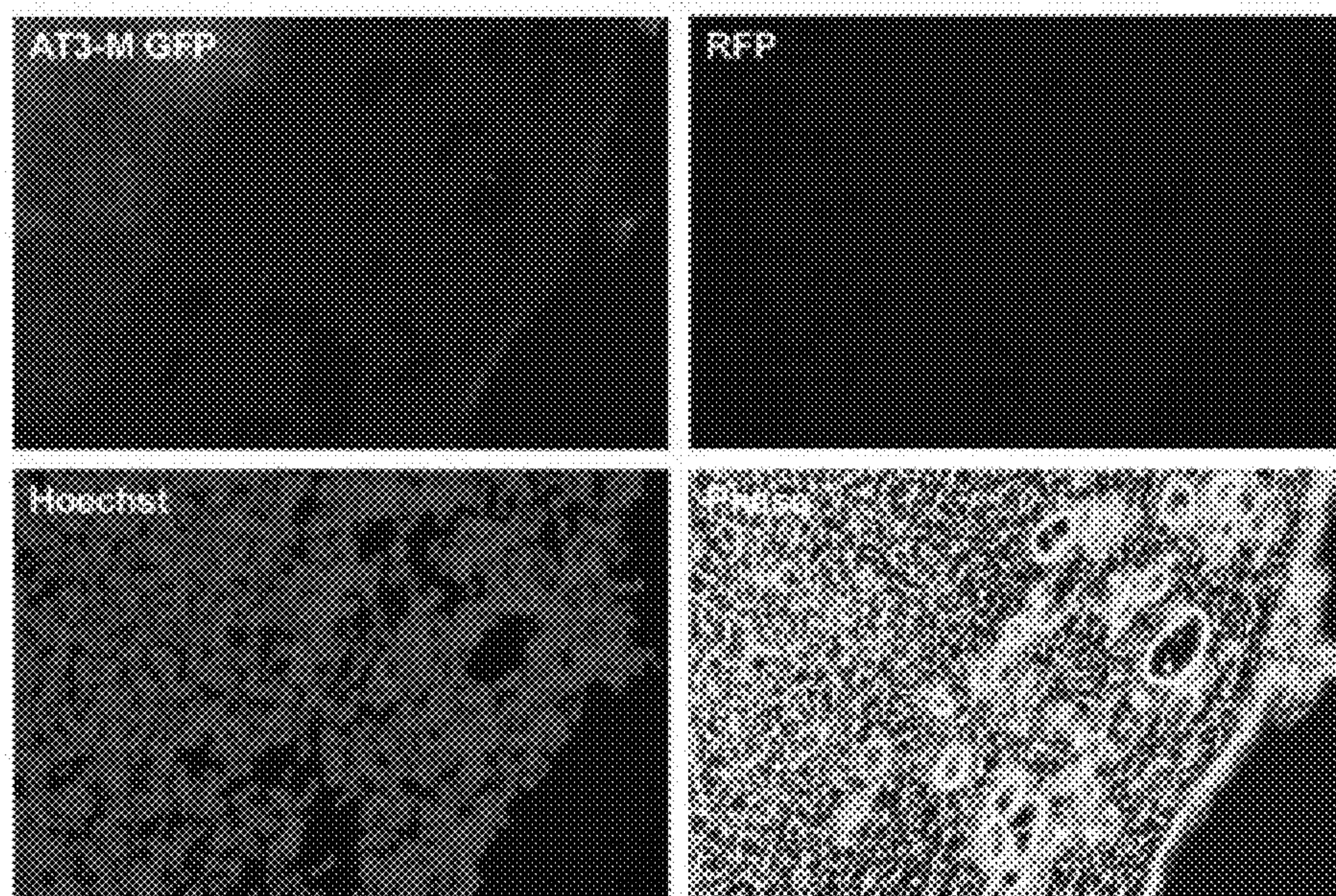


FIG. 5

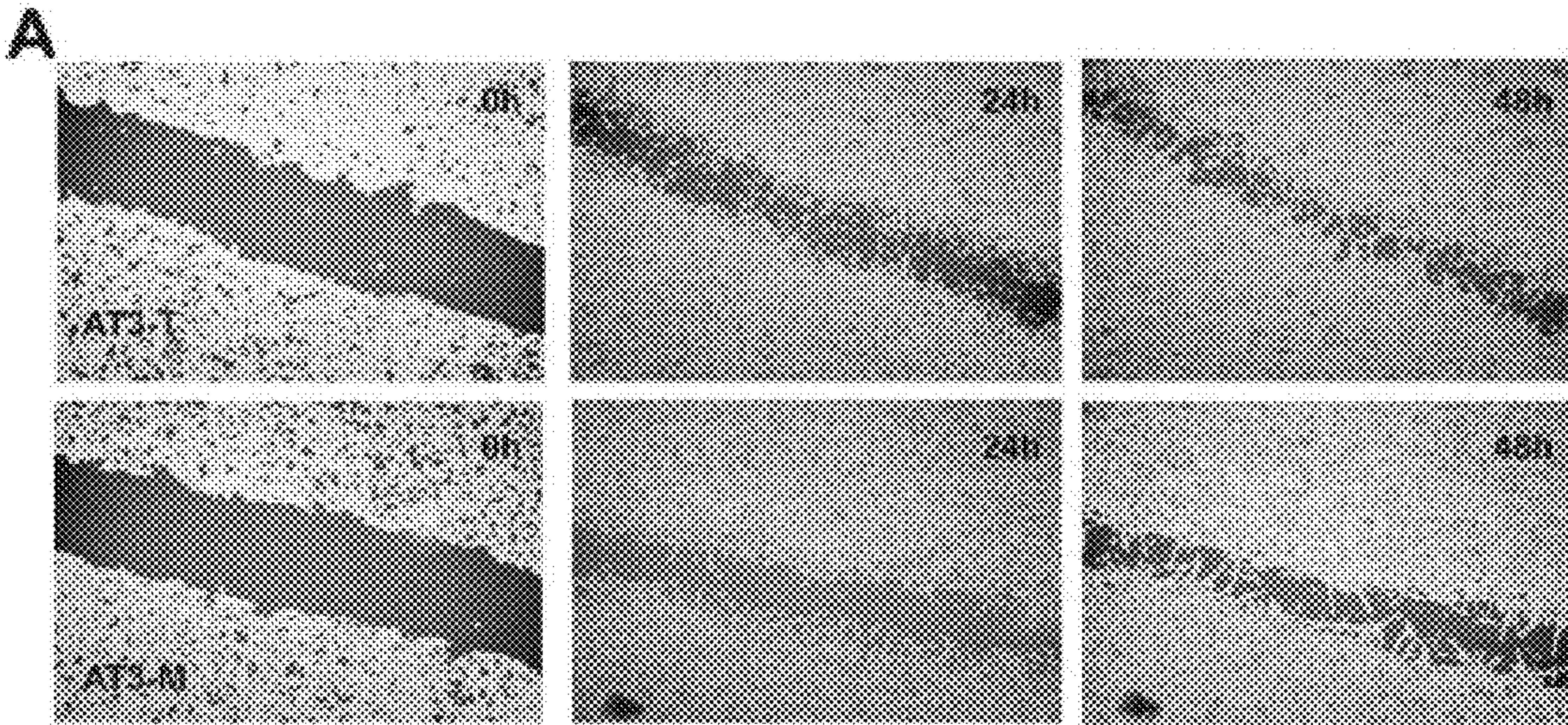


FIG. 6A

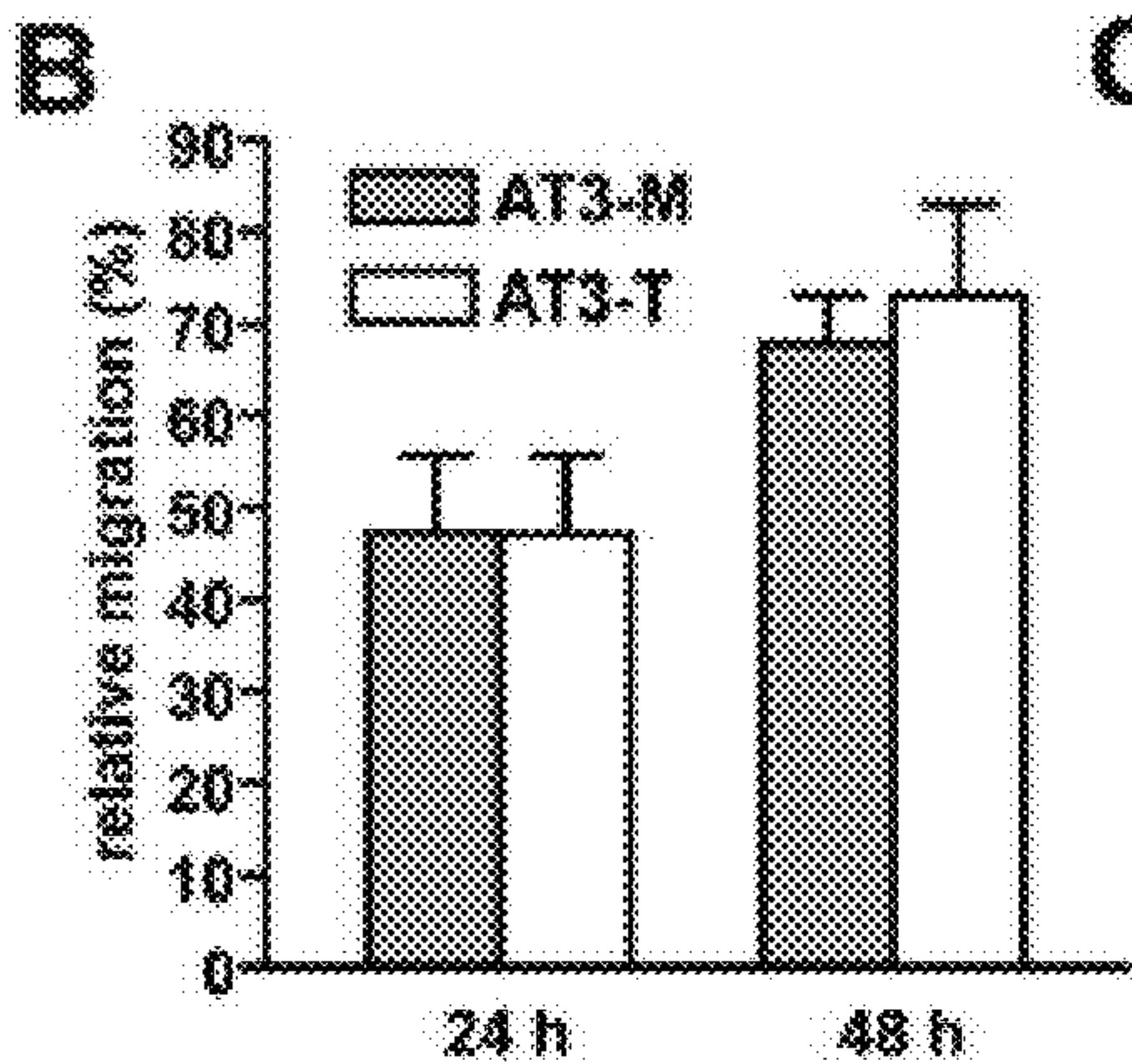


FIG. 6B

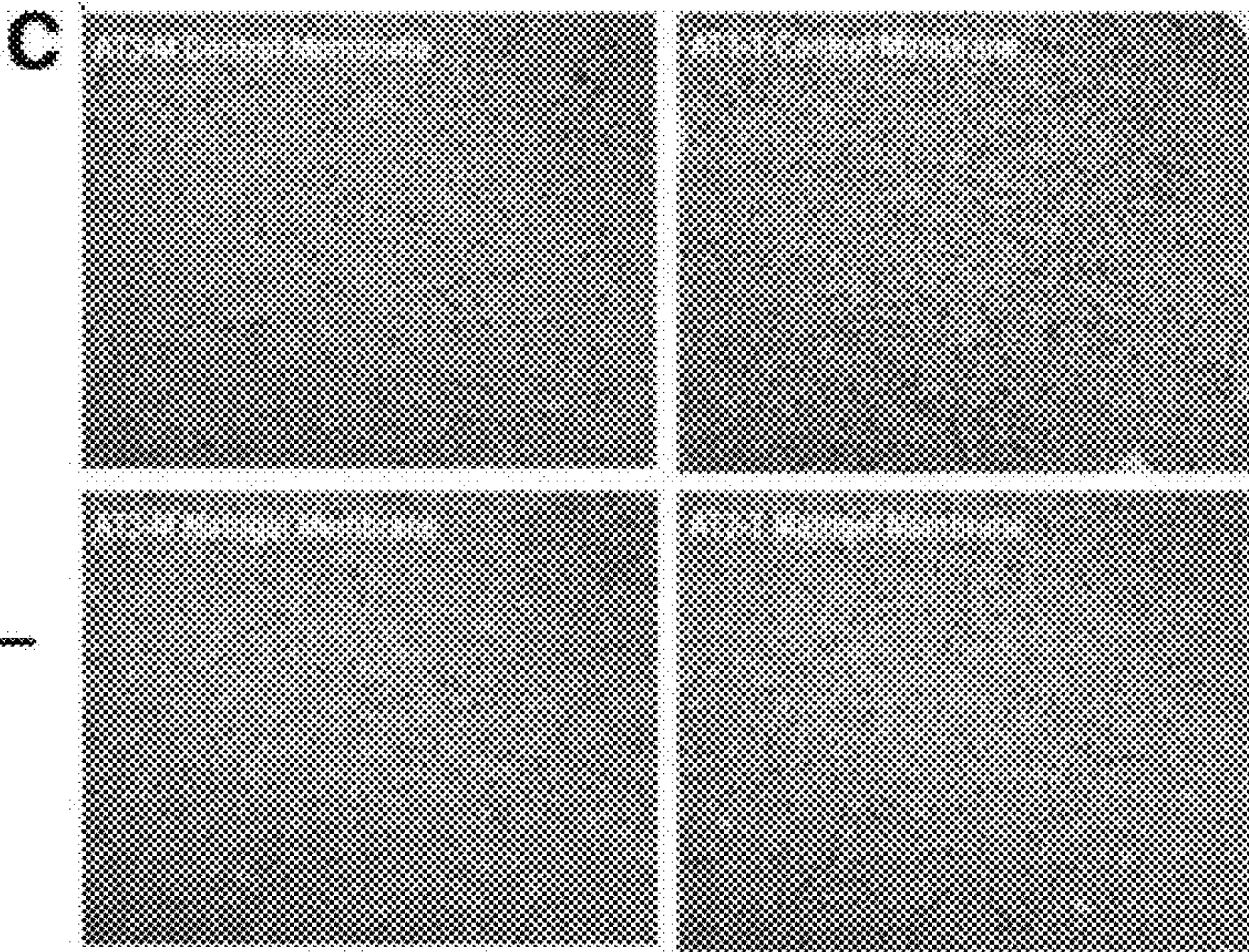


FIG. 6C

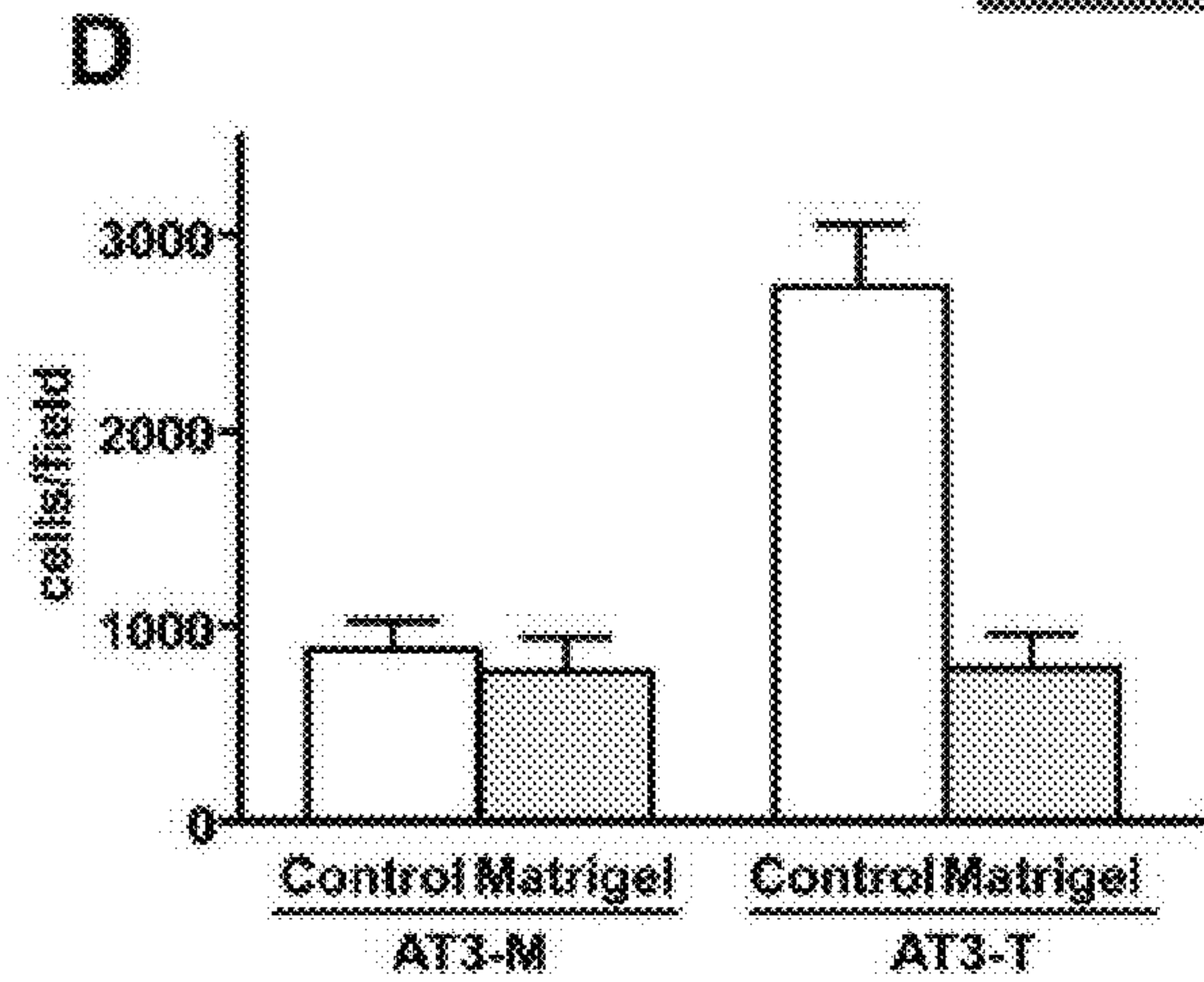
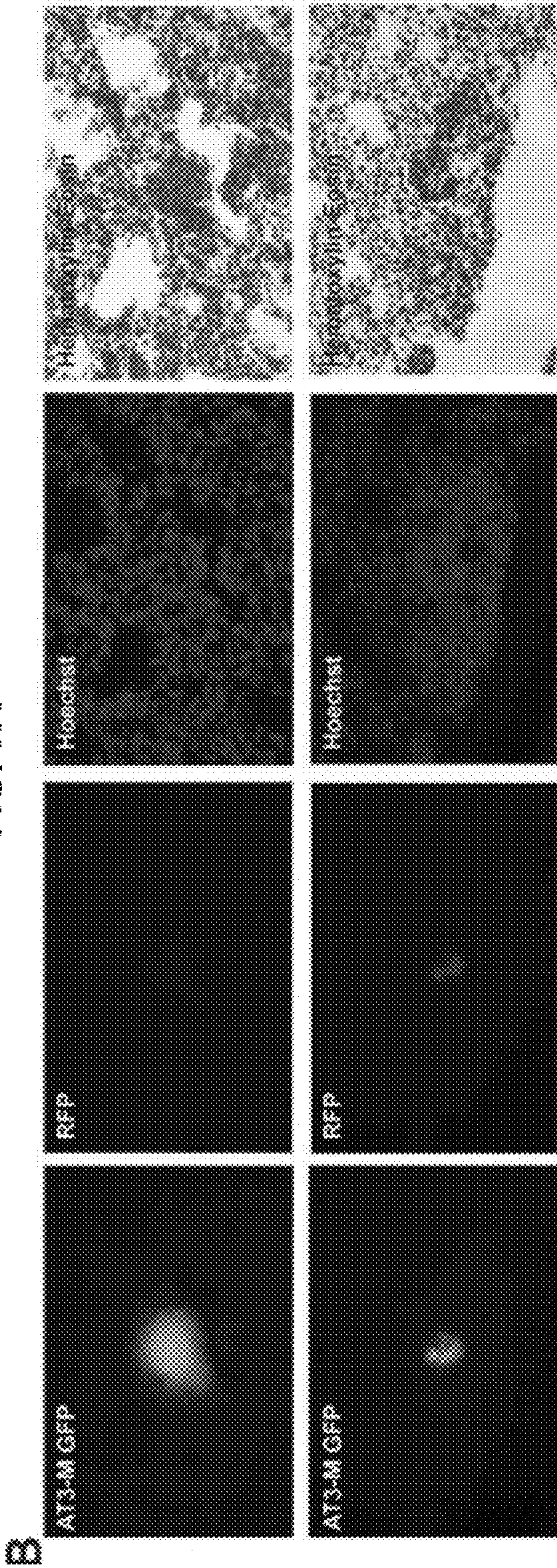
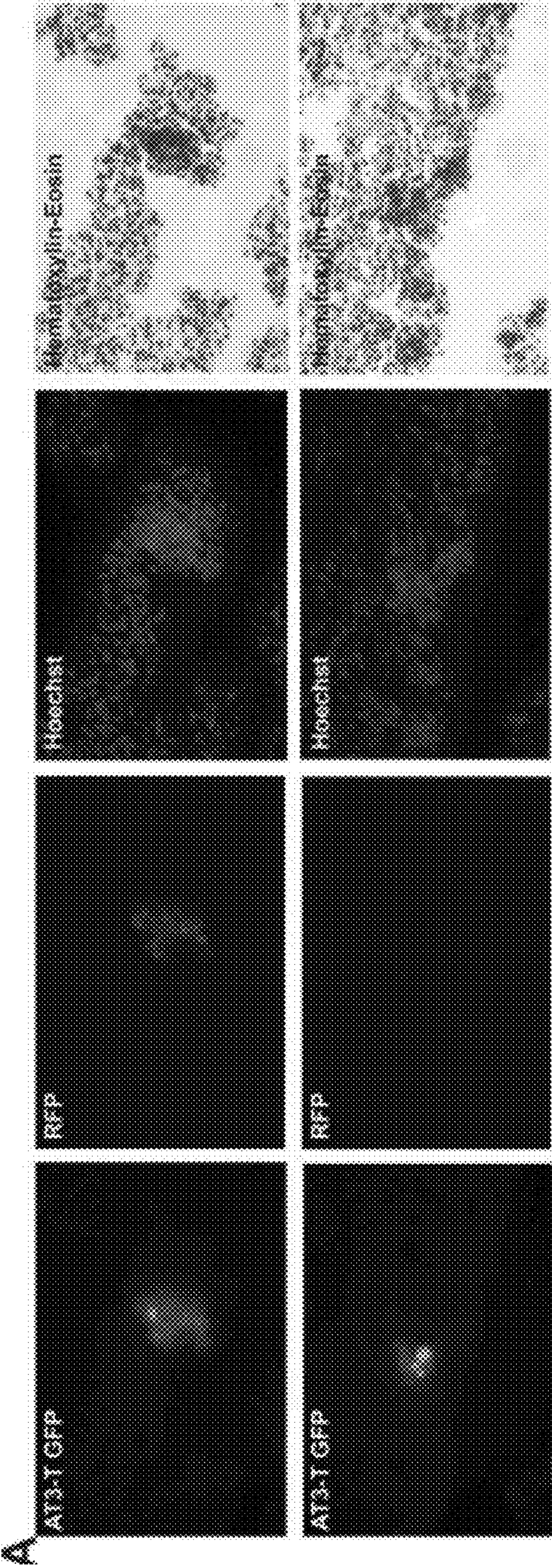


FIG. 6D



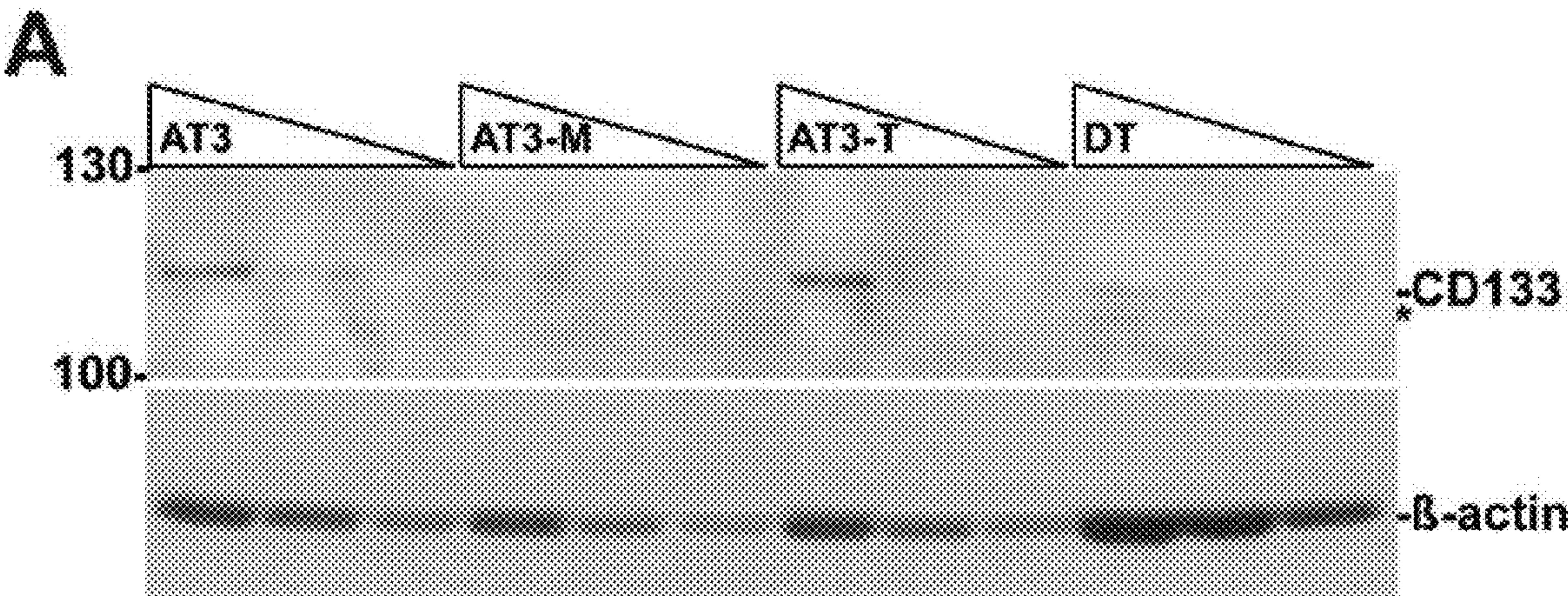


FIG. 8A

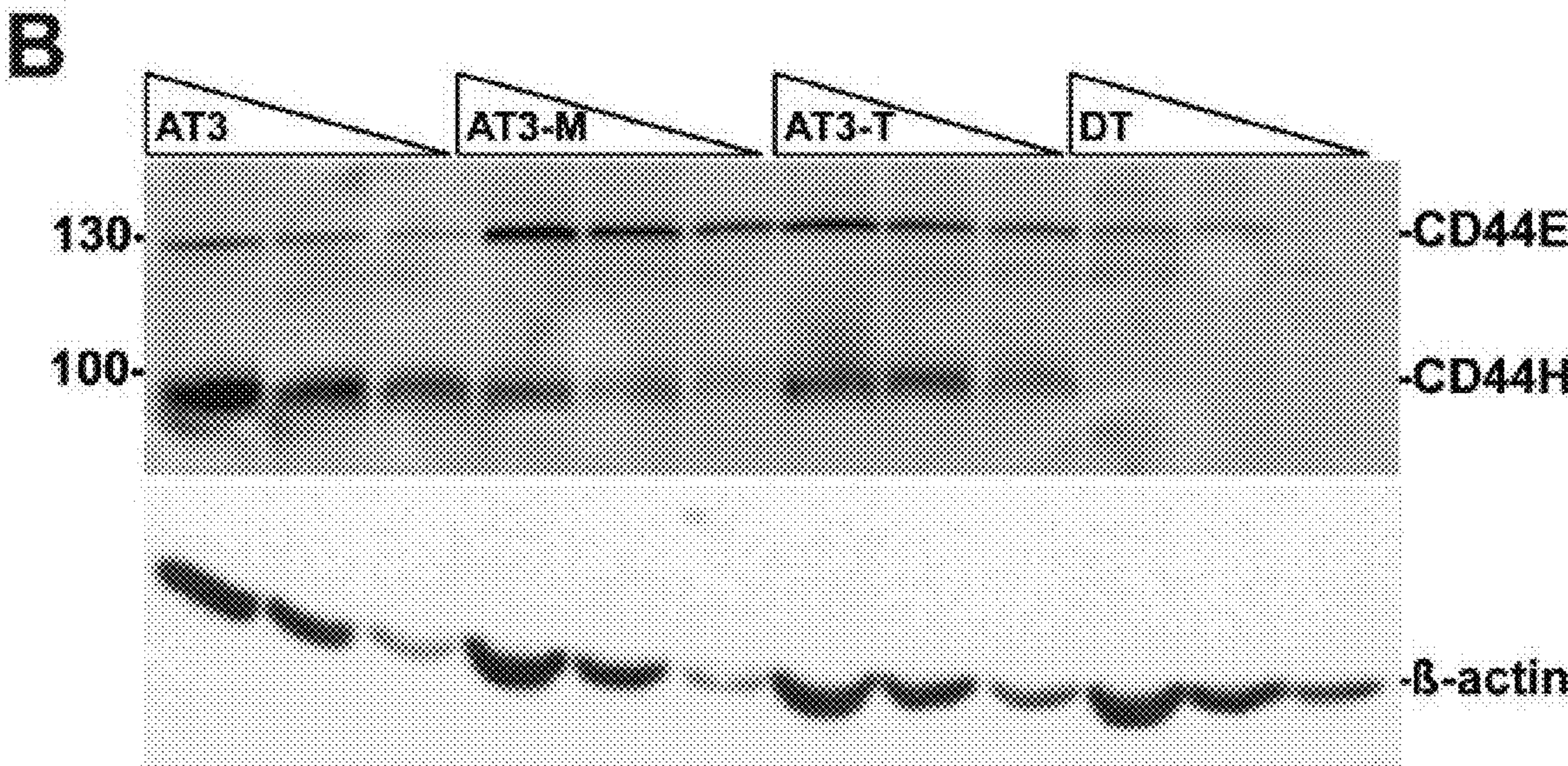


FIG. 8B

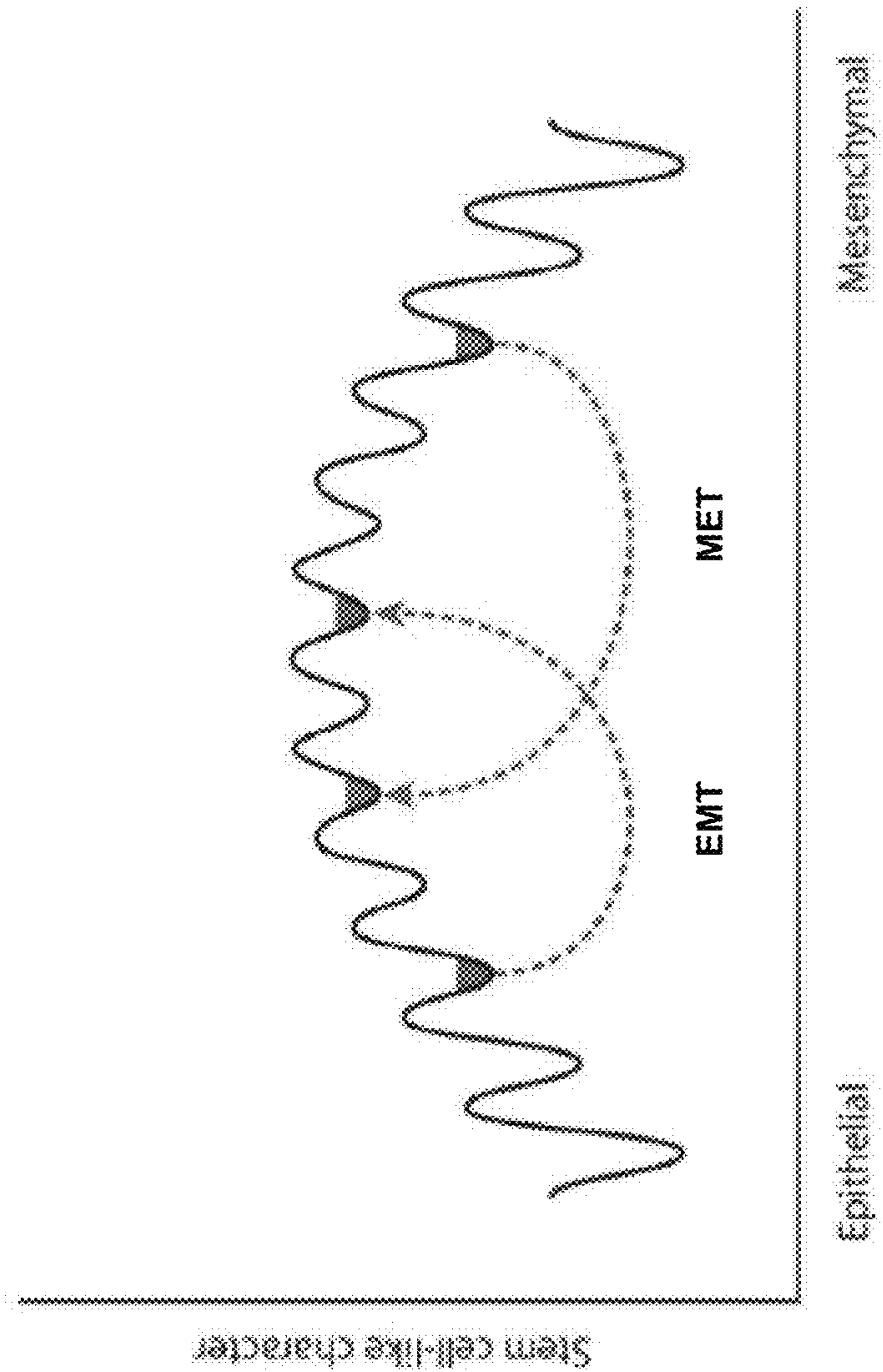


FIG. 9

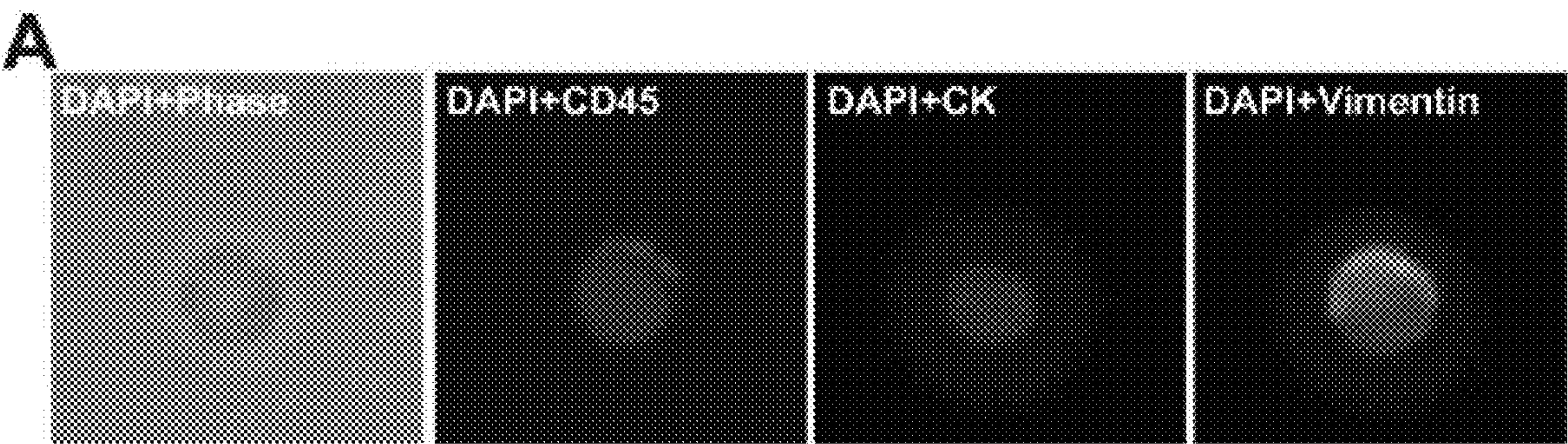


FIG. 10A

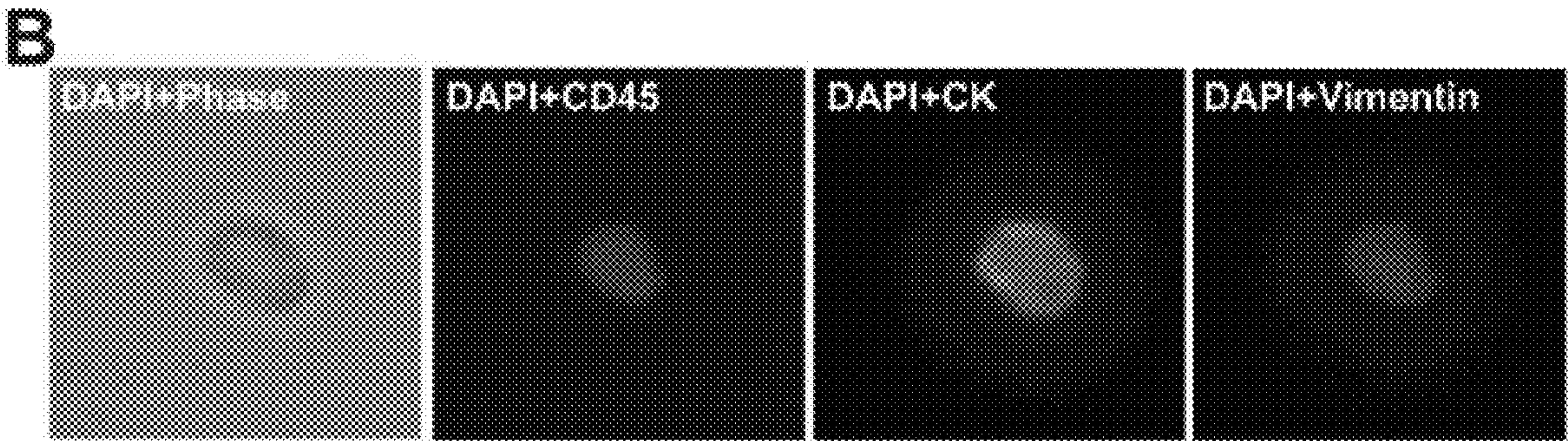


FIG. 10B

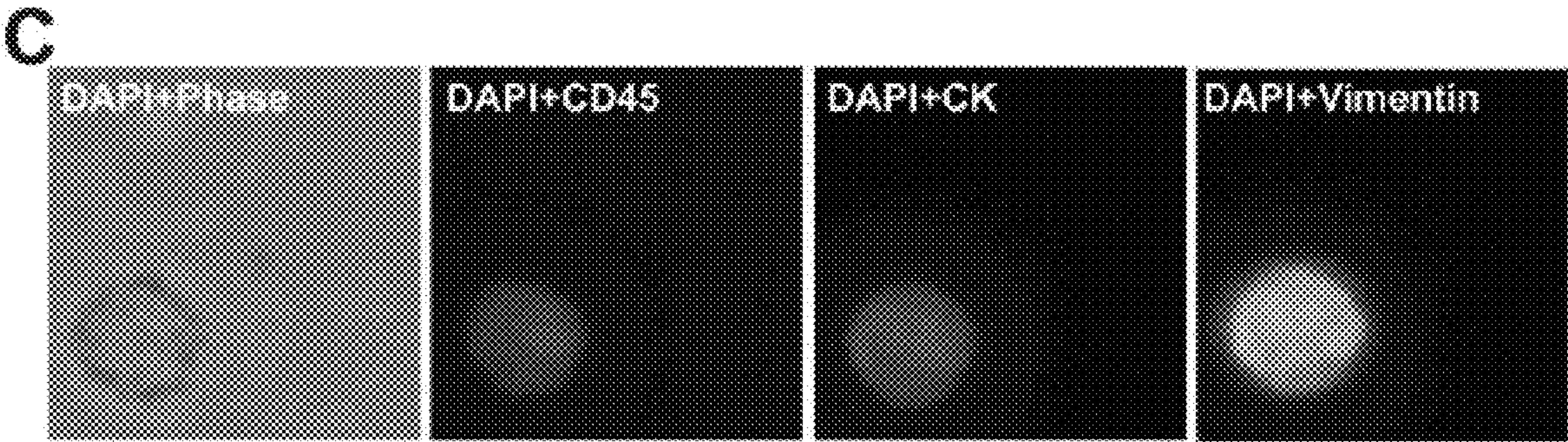


FIG. 10C

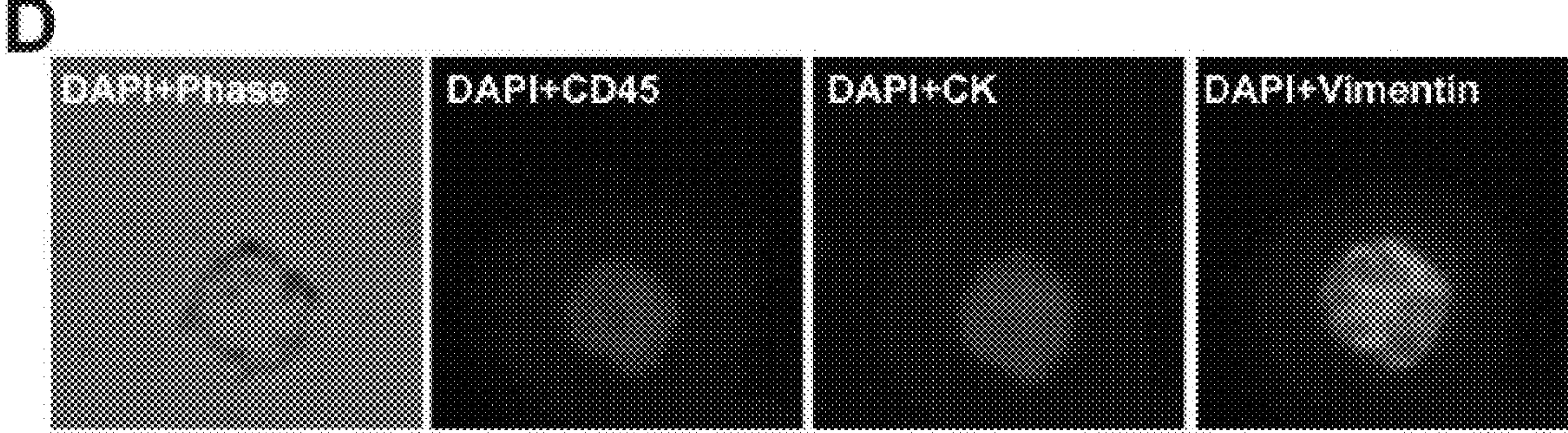


FIG. 10D

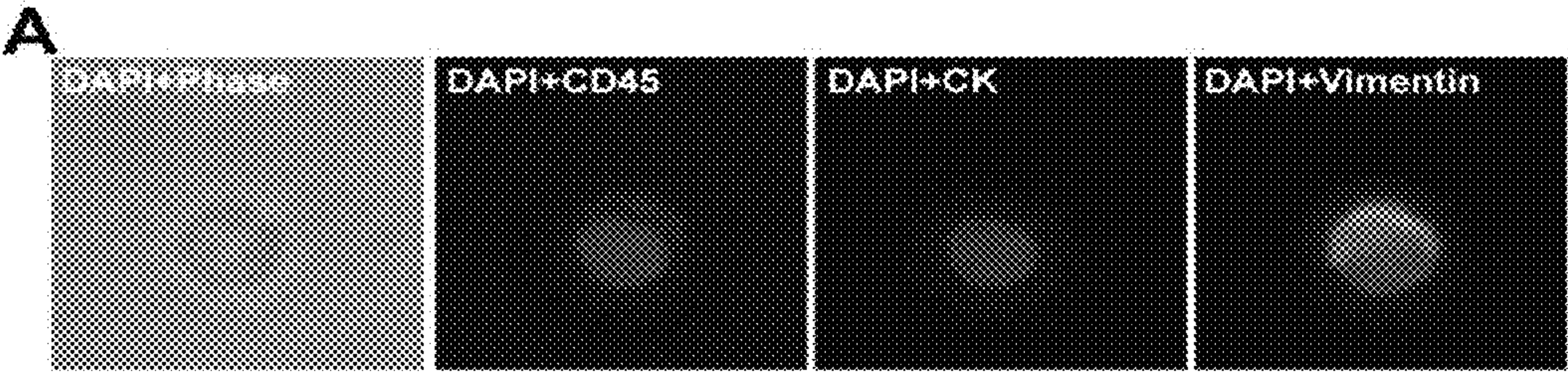


FIG. 11A

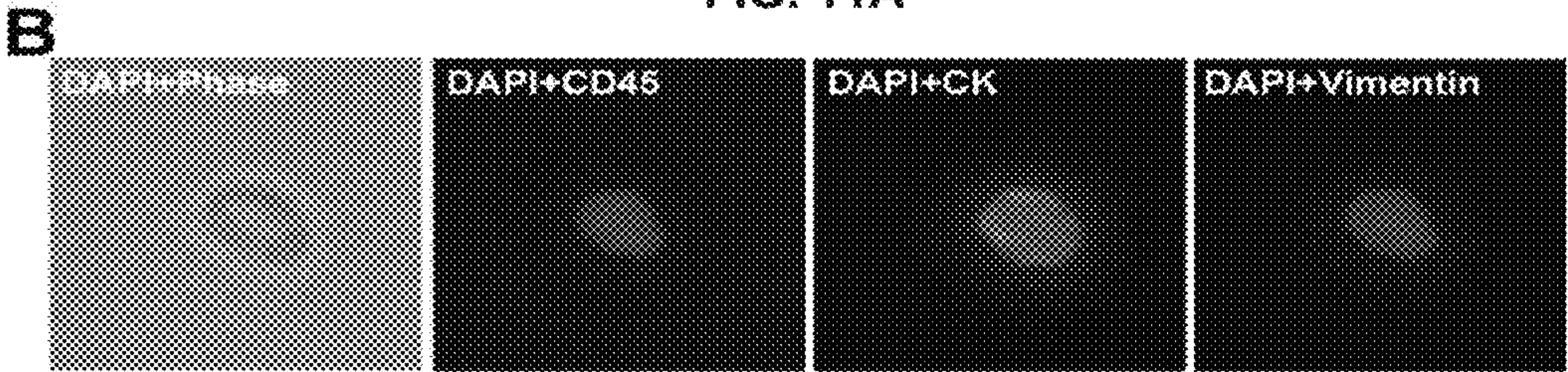


FIG. 11B

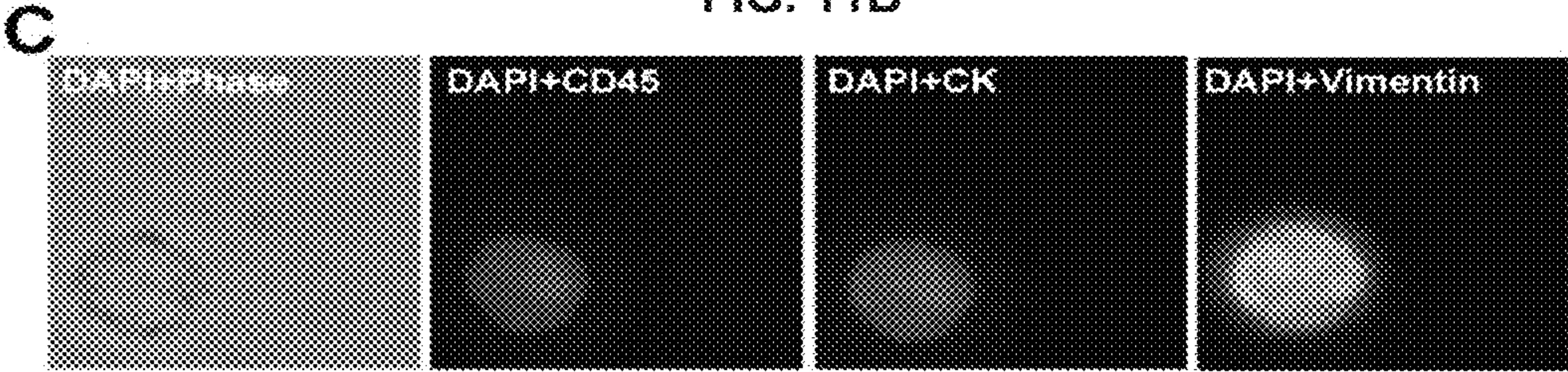


FIG. 11C

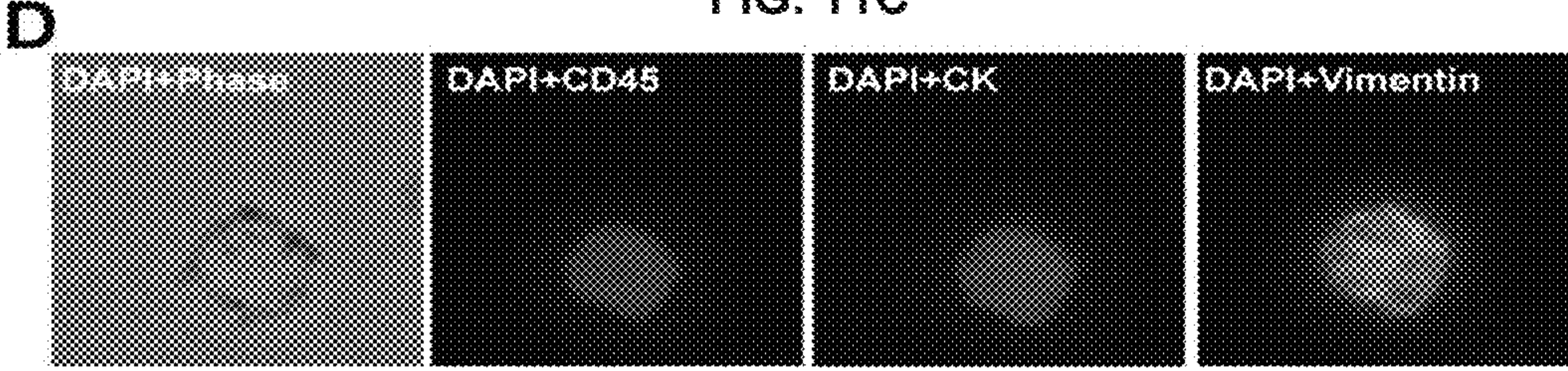
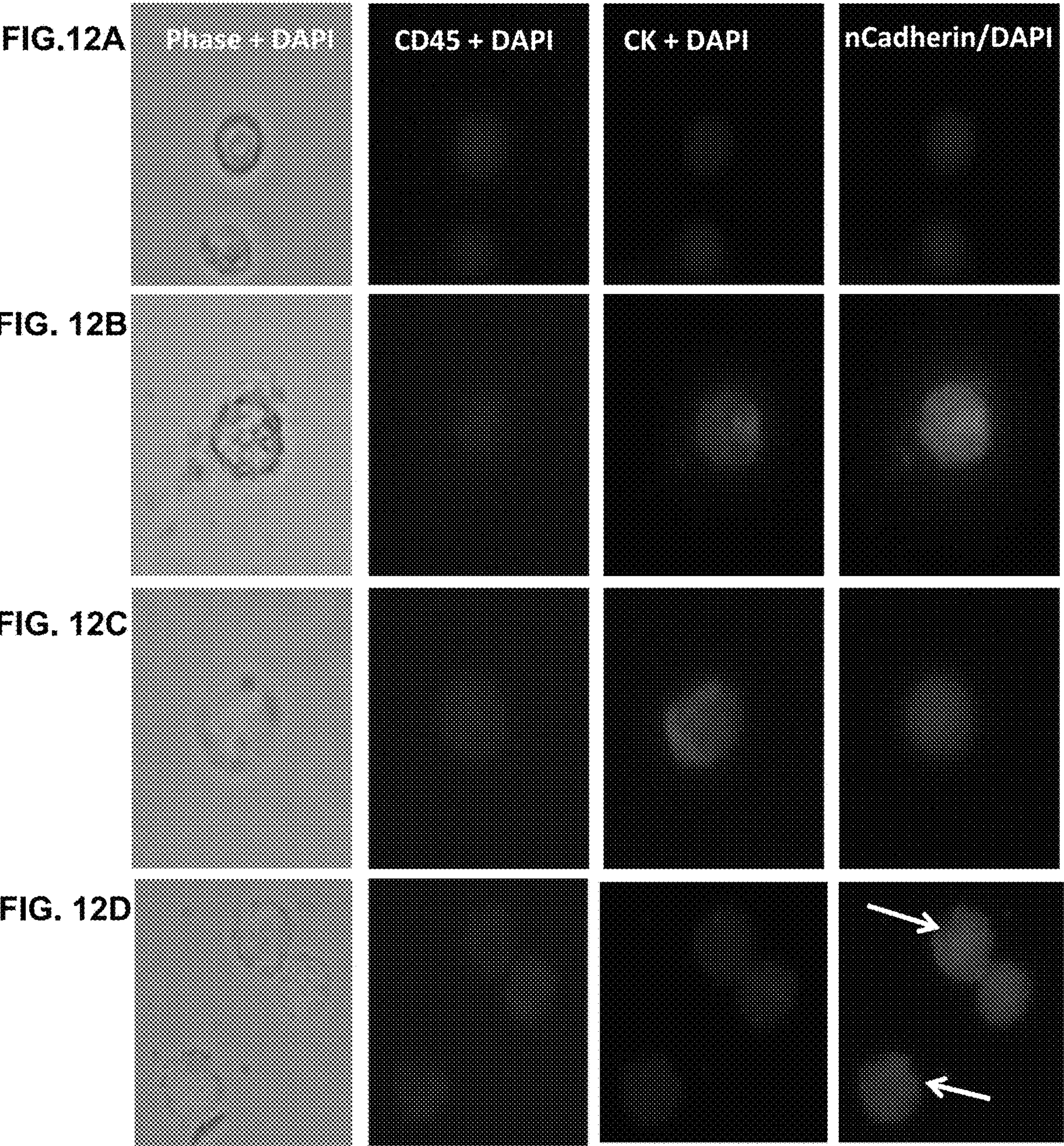
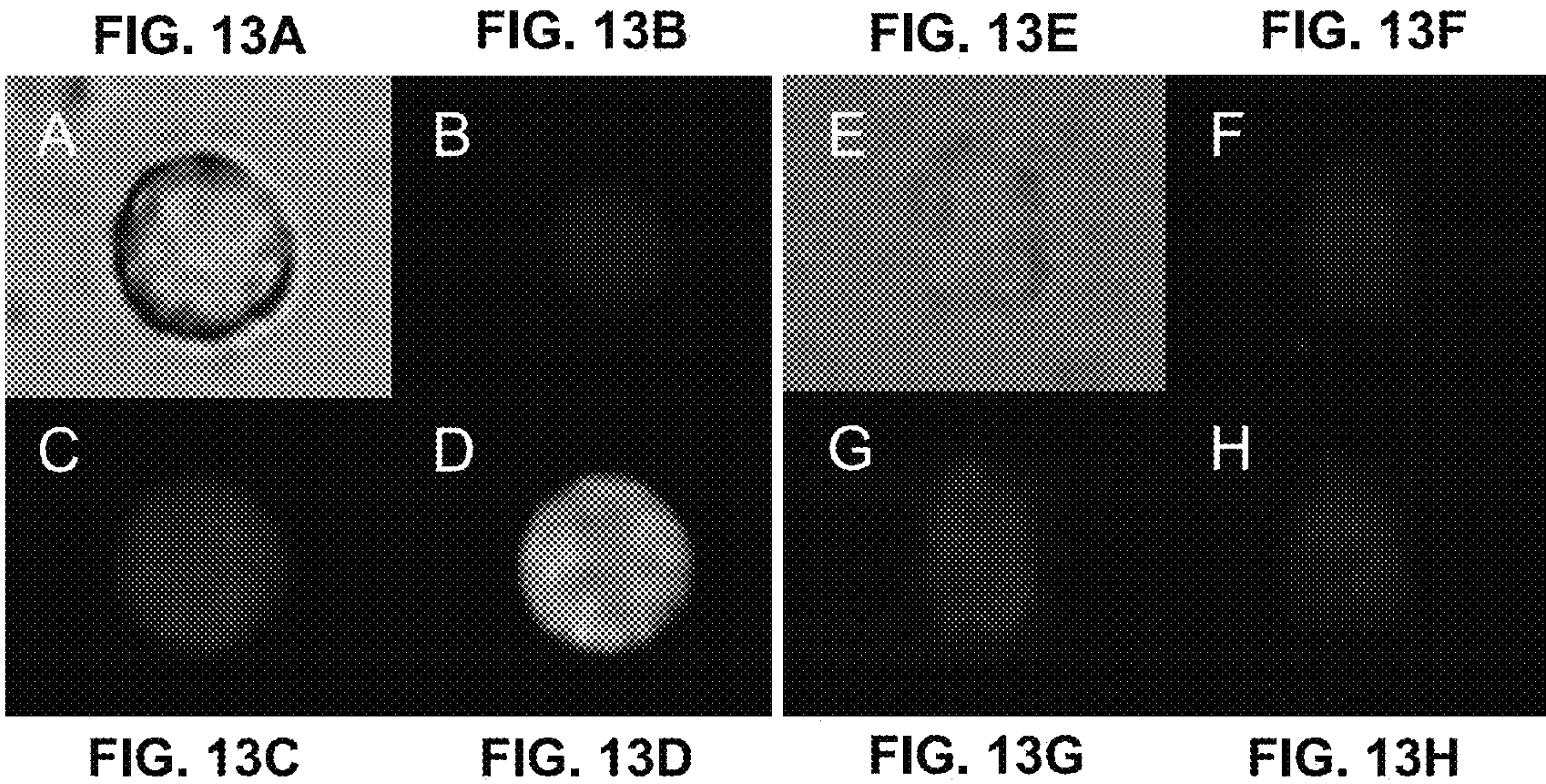
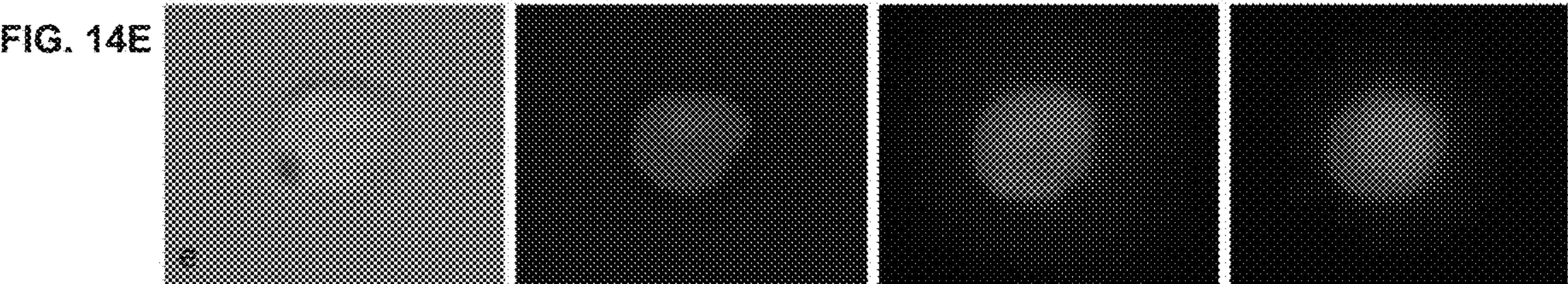
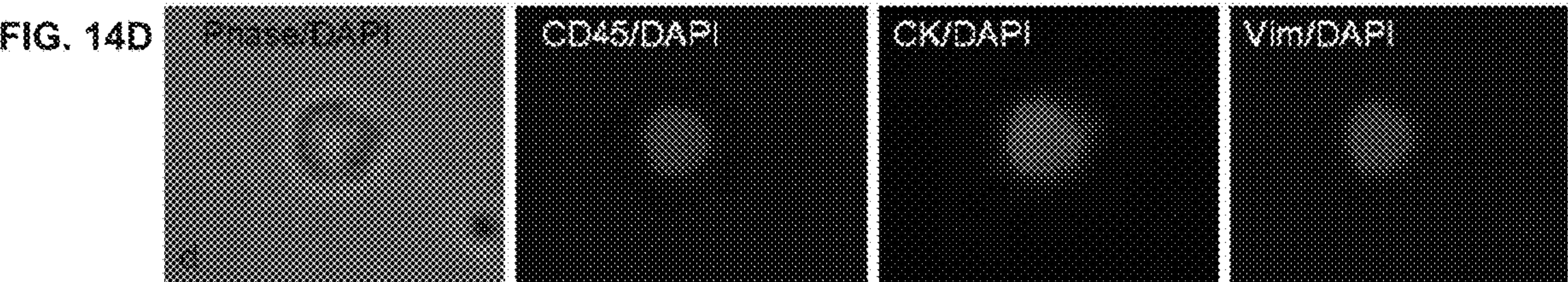
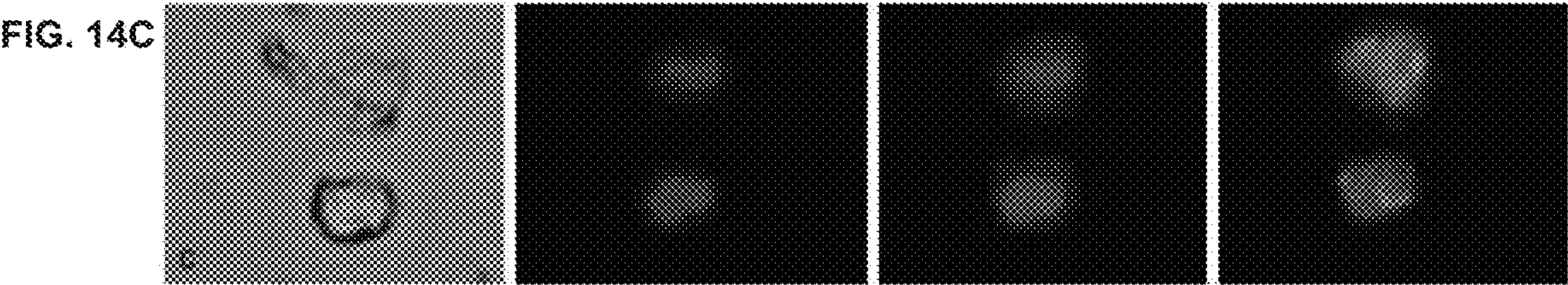
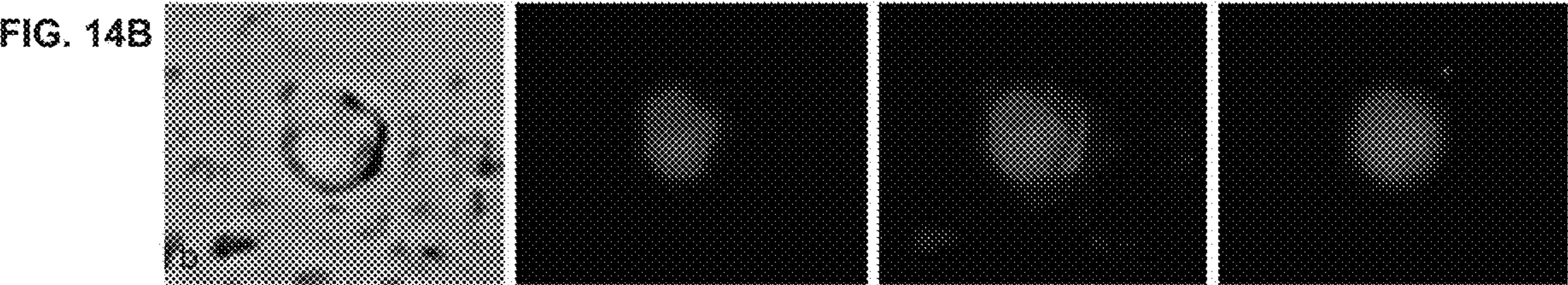
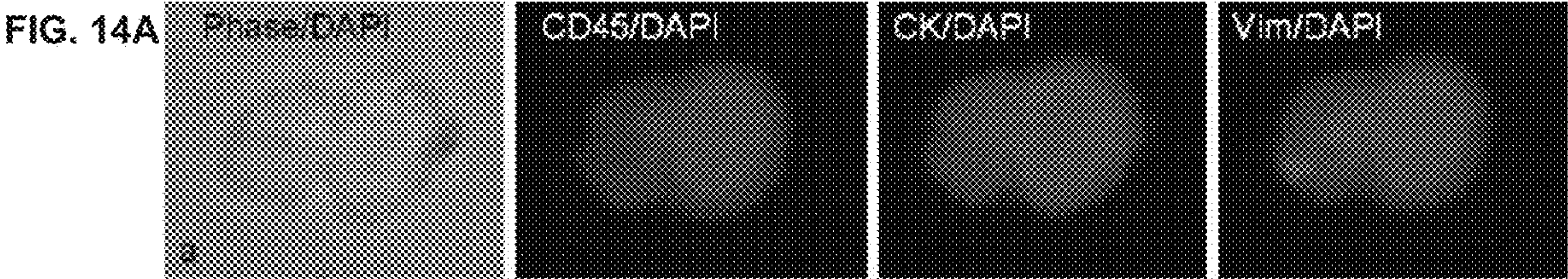
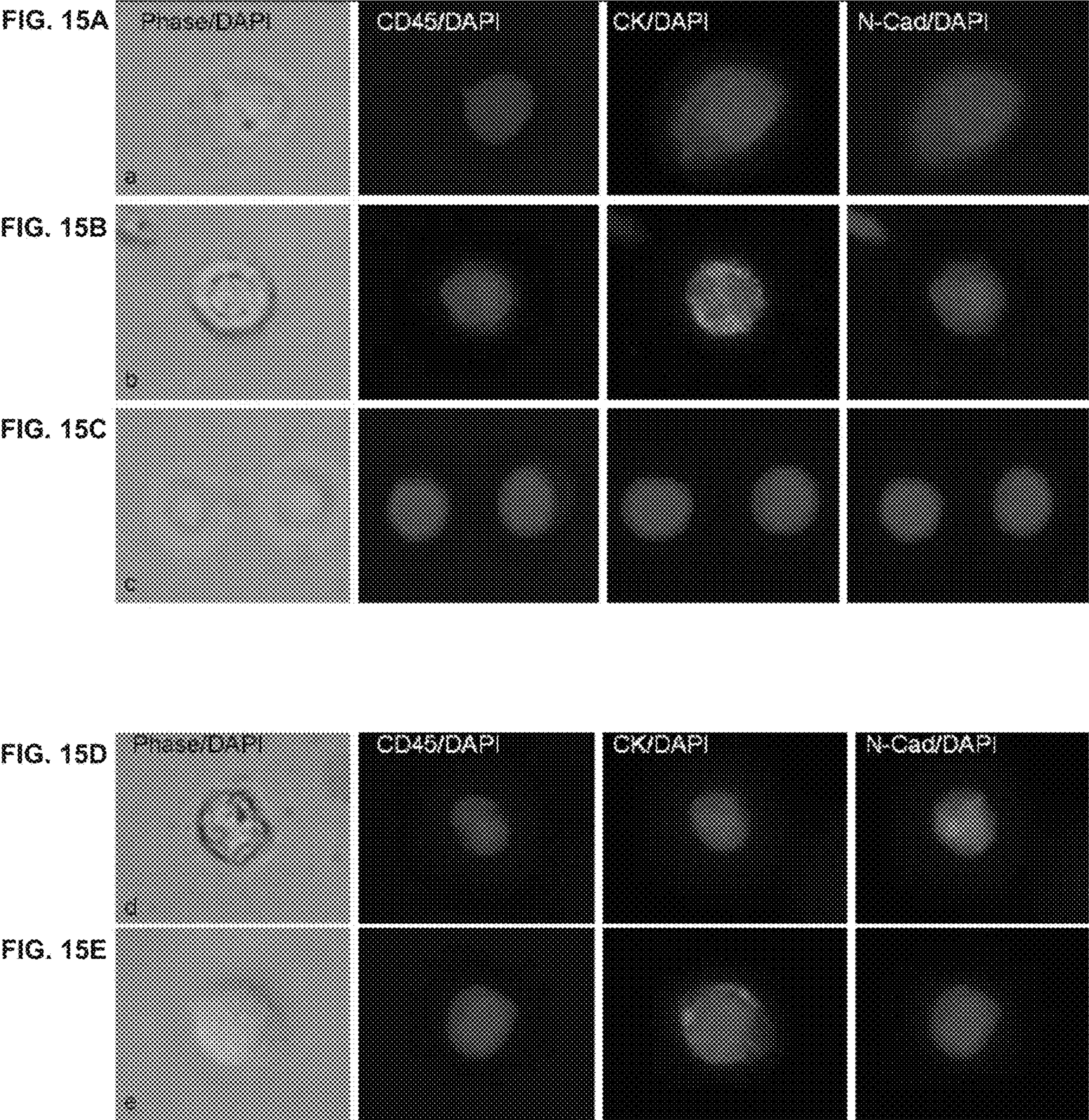


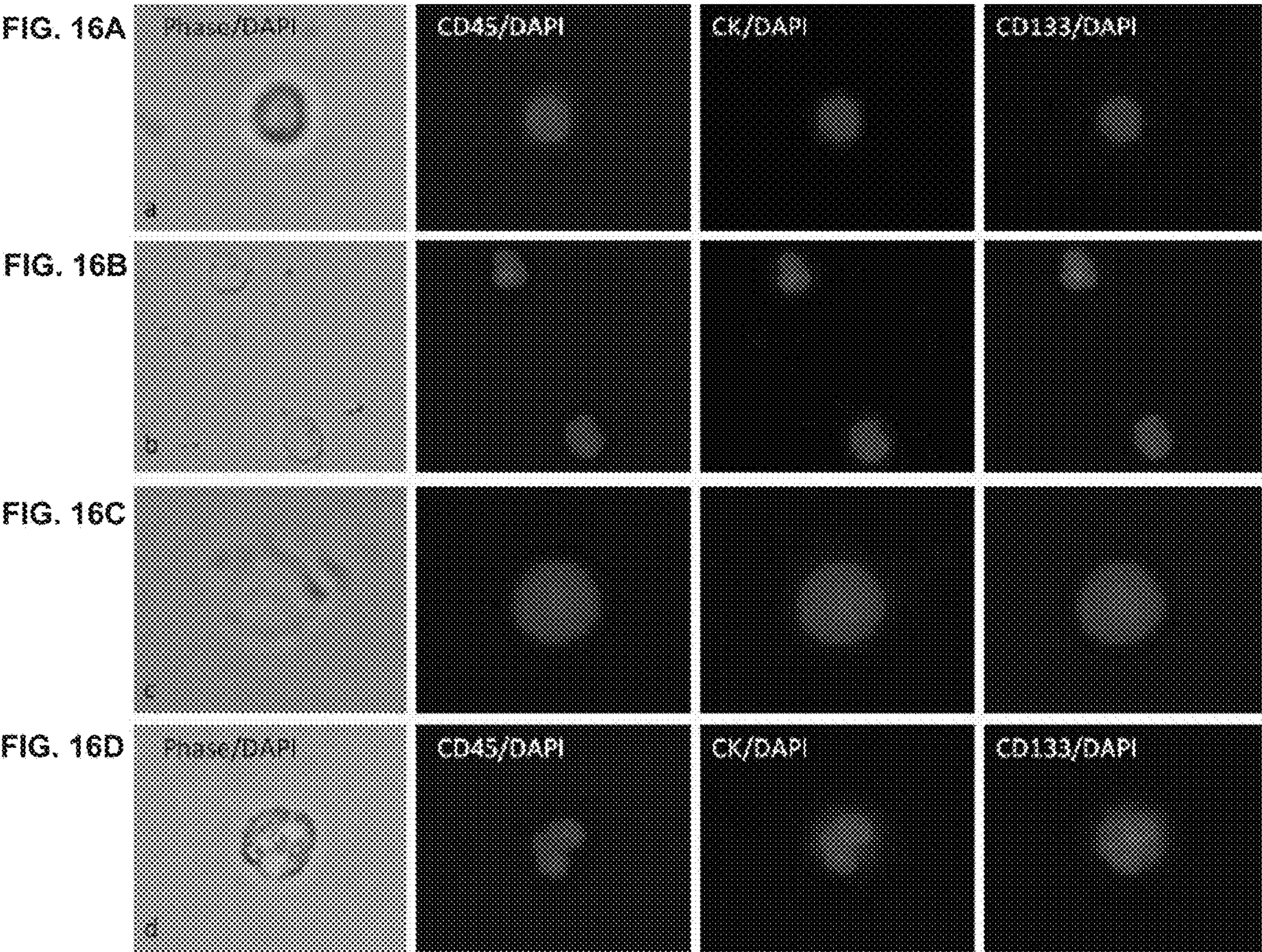
FIG. 11D











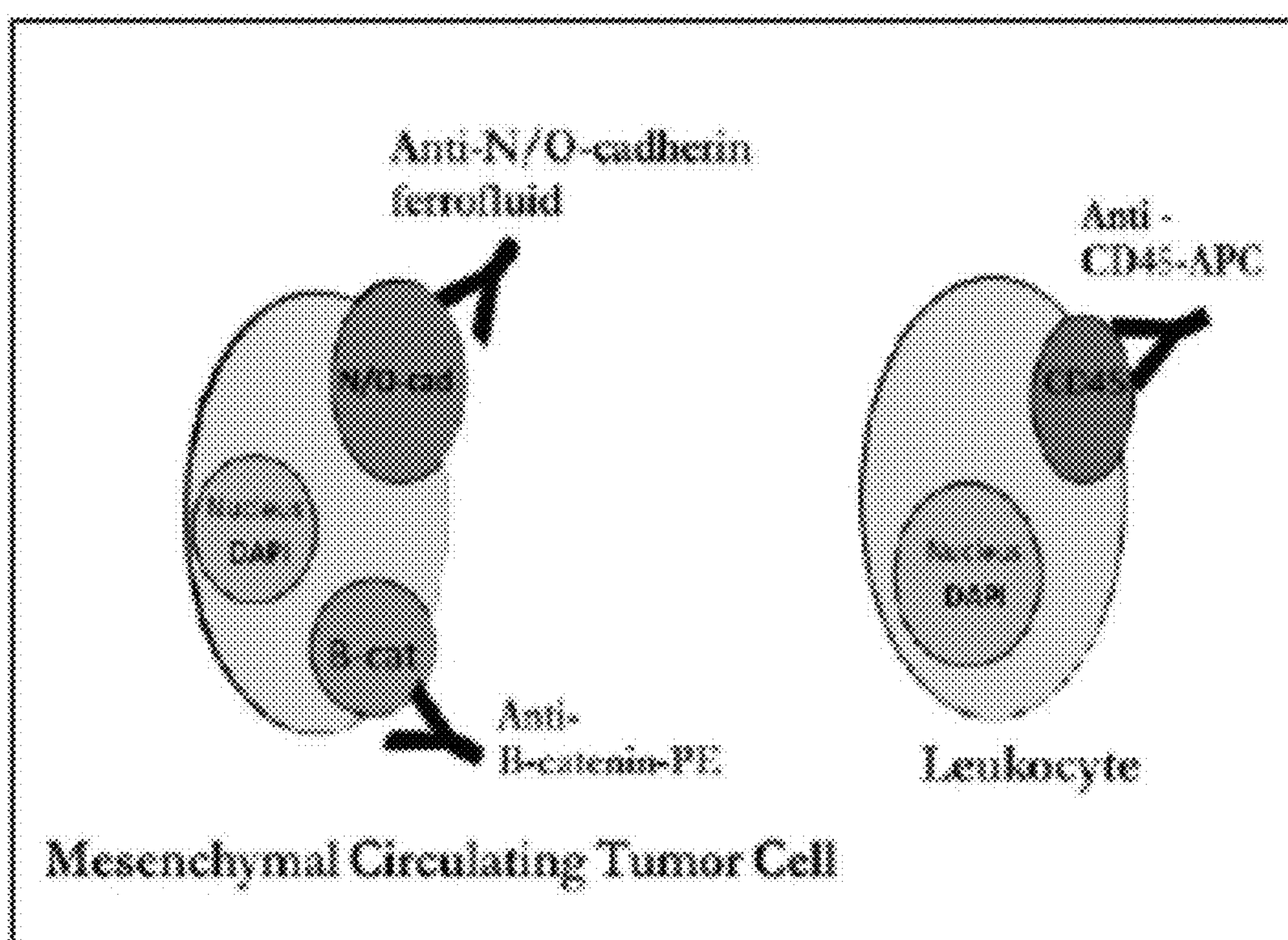
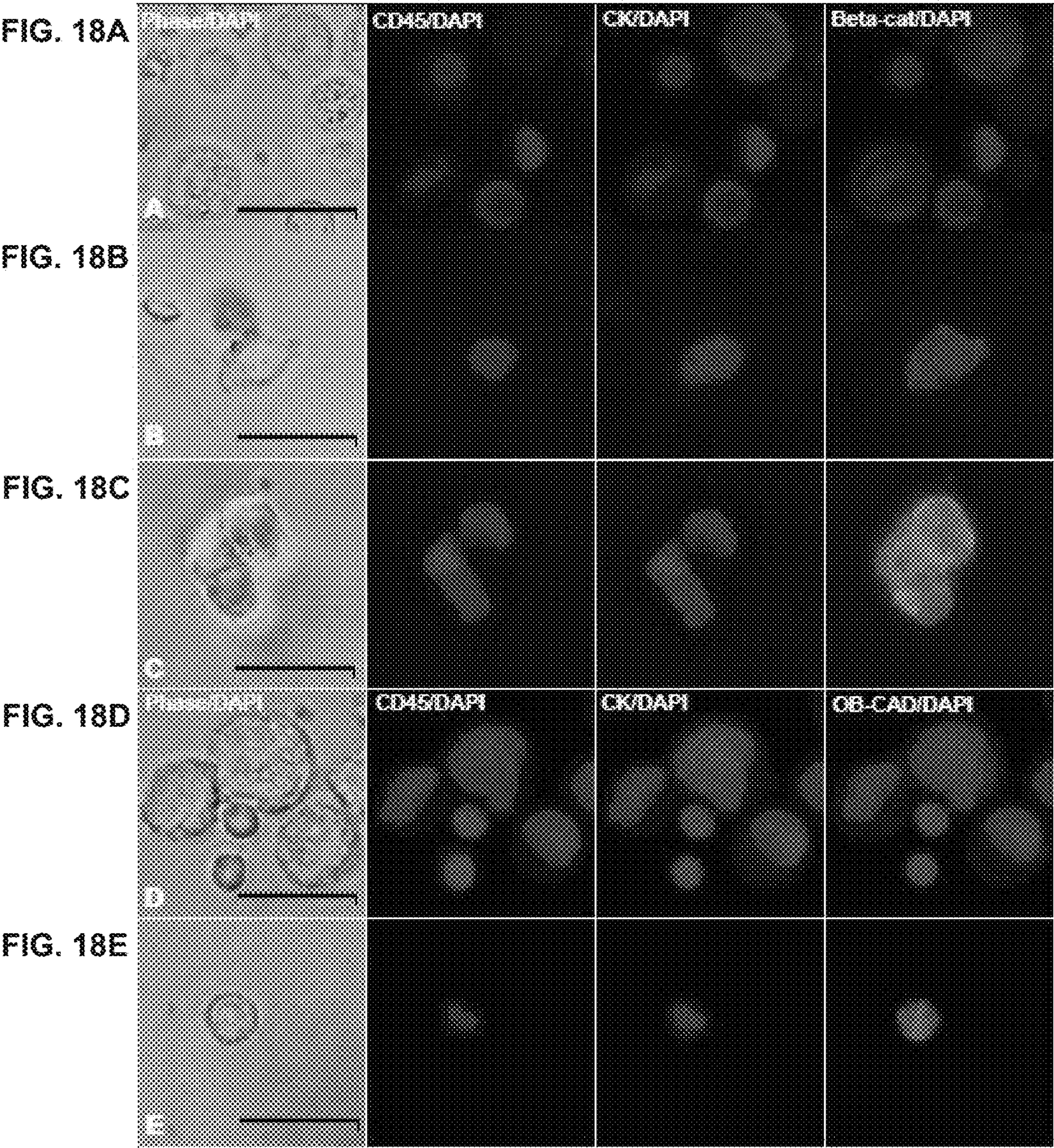


FIG. 17



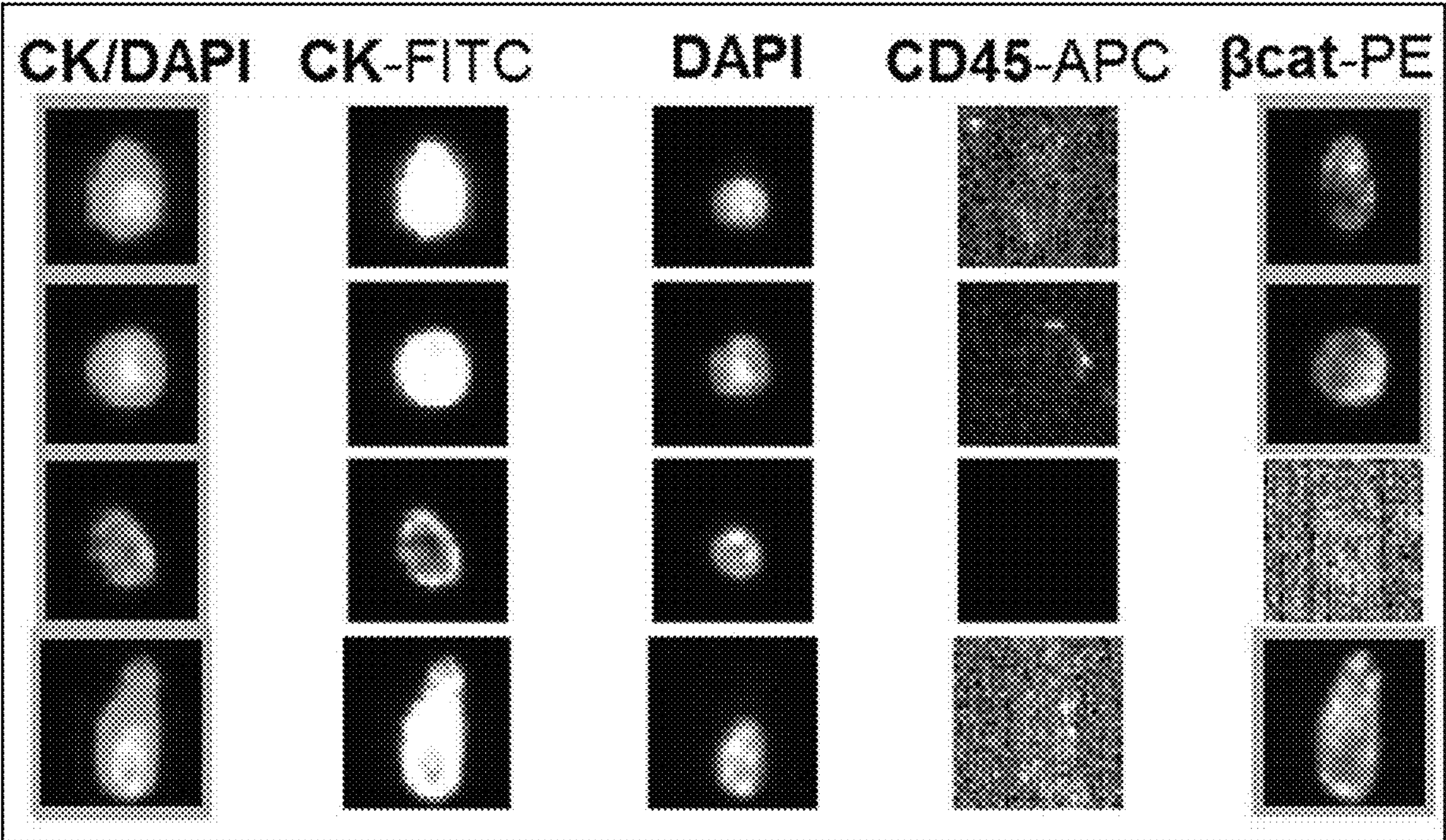
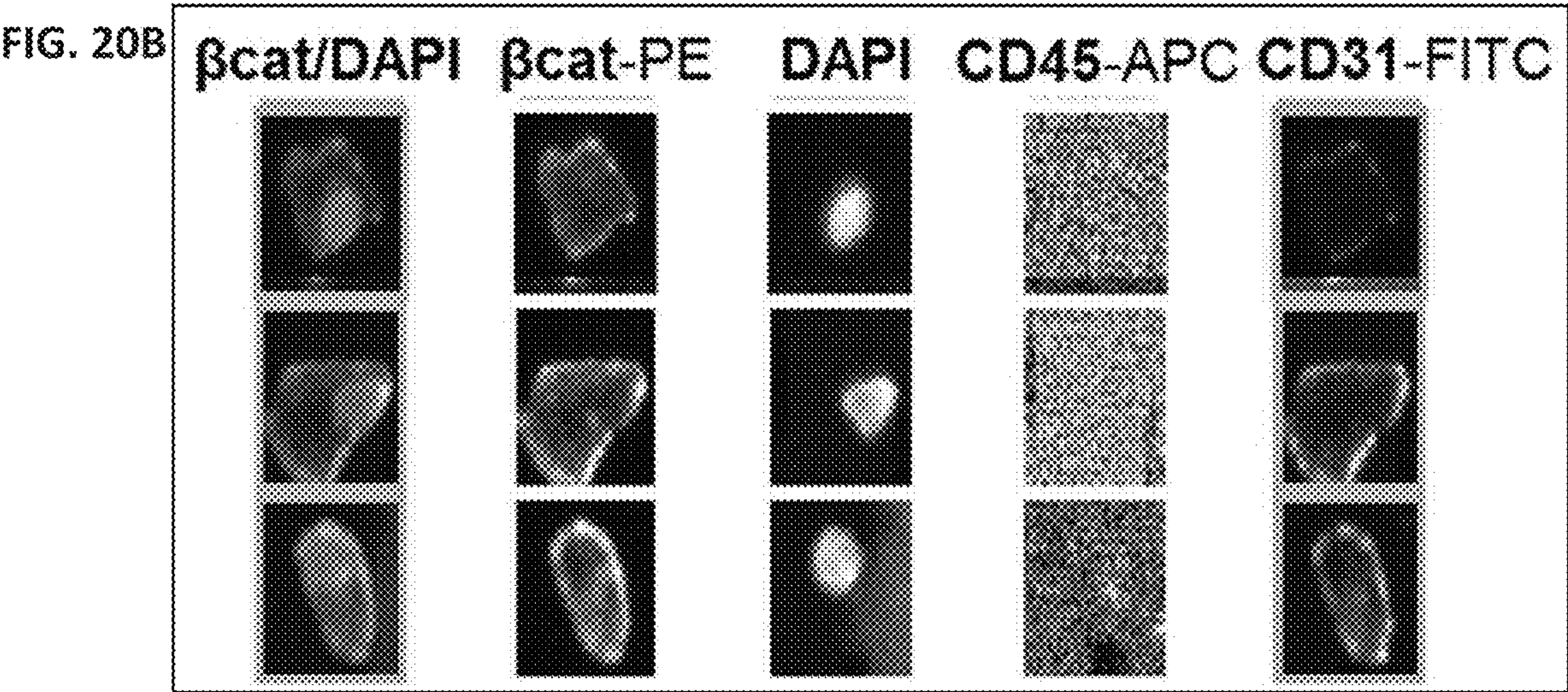
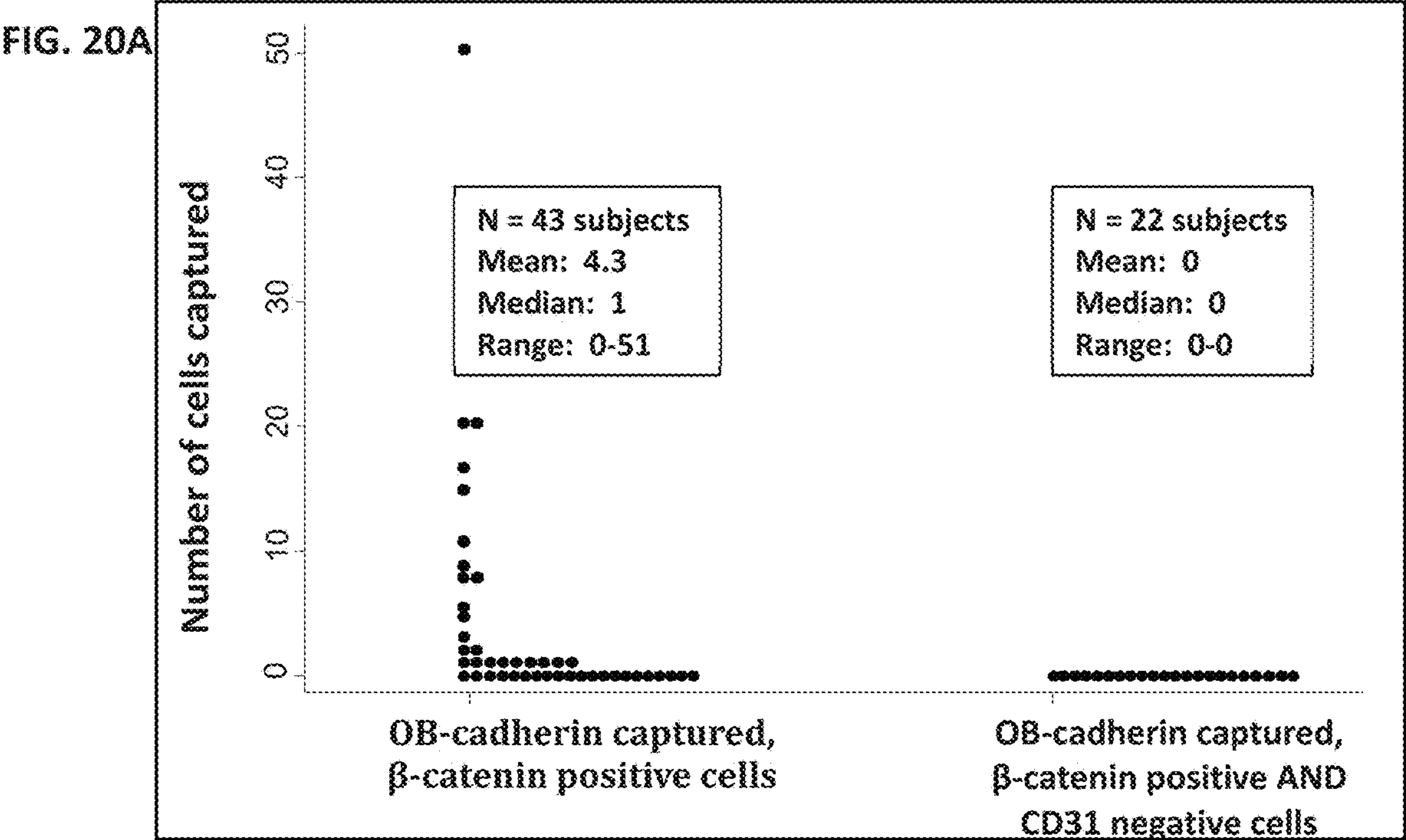


FIG. 19



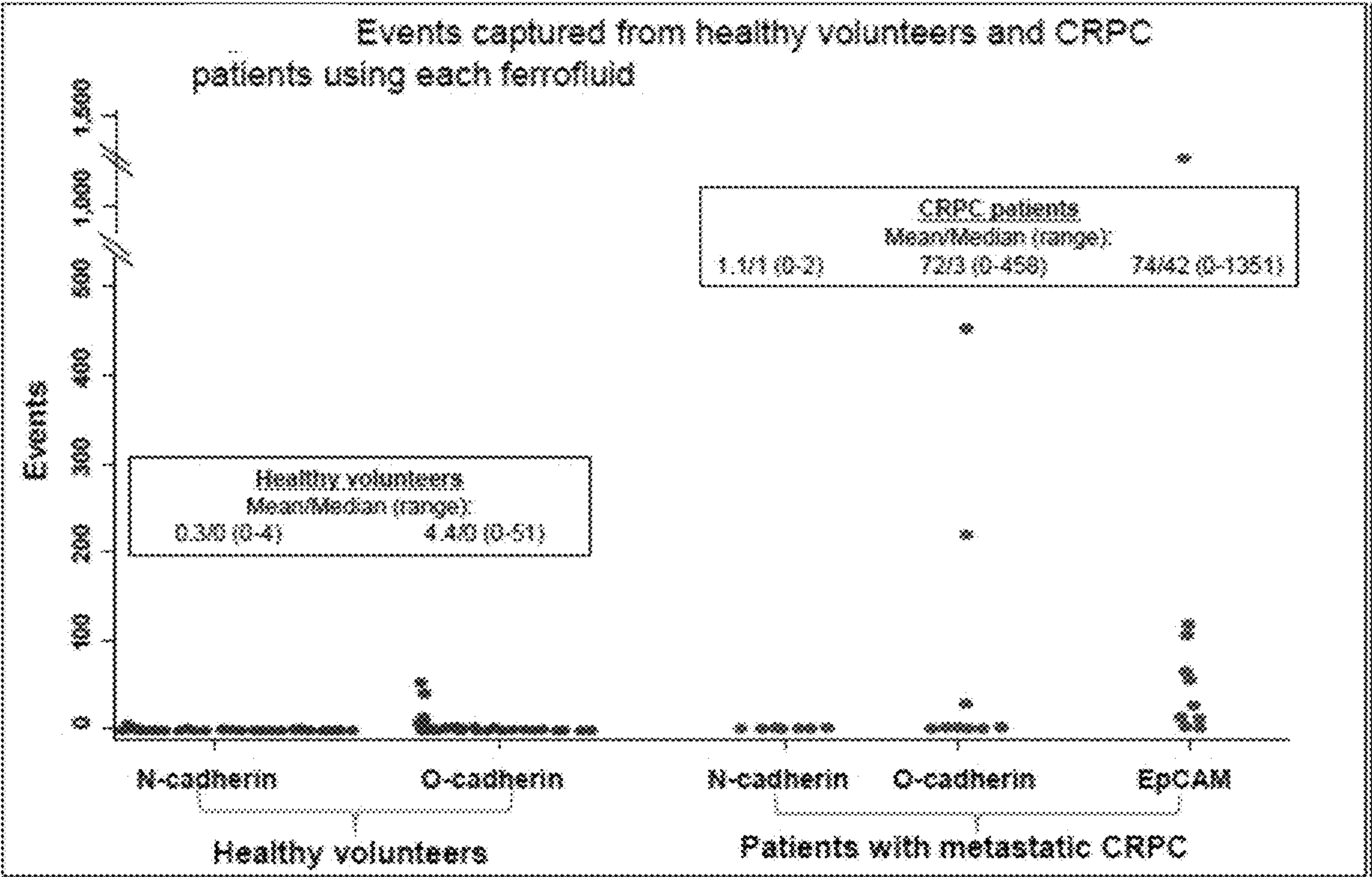


FIG. 21

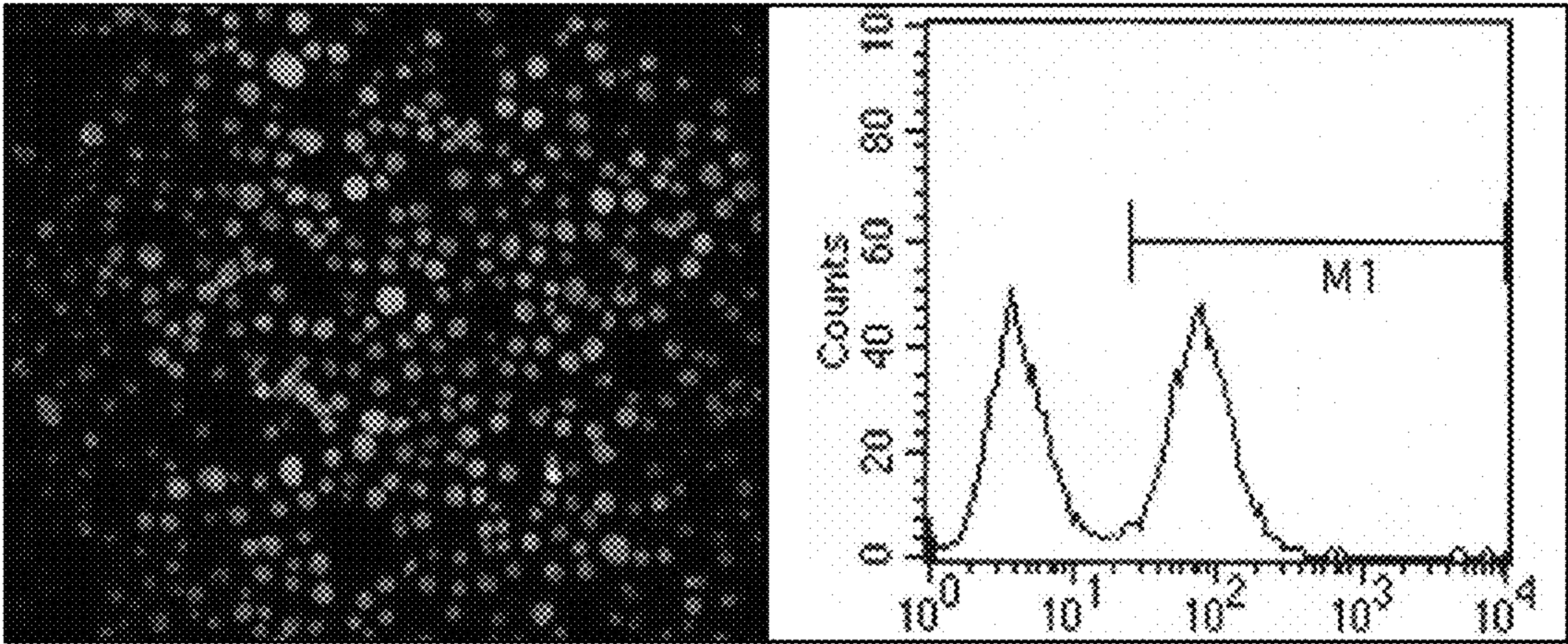


FIG. 23A

FIG. 23B

	FIG. 24A	FIG. 24B	FIG. 24C
	A	B	C
Capture Ferrofluid	EpCAM	OB-cadherin	EpCAM
Detection Antibody (PE)	Cytokeratin	β-catenin	Cytokeratin
CD45 staining	Negative	Negative	Positive
CD31 staining	Negative	Negative	Negative
CellSearch: DAPI/PE overlay			
AR FISH			
ERG break FISH			
PTEN FISH			

METHOD OF ISOLATING CIRCULATING TUMOR CELLS

CROSS REFERENCE TO RELATED APPLICATION

[0001] This patent application is a divisional of U.S. patent application Ser. No. 16/206,932, filed Nov. 30, 2018, which is a divisional of U.S. patent application Ser. No. 14/765,071, filed Jul. 31, 2015, which is a national stage filing under 35 U.S.C. § 371 of International Application No. PCT/US2013/043745, filed on May 31, 2013, which application claims priority to U.S. Provisional Application No. 61/760,042, filed Feb. 2, 2013, and U.S. Provisional Application No. 61/806,358, filed Mar. 28, 2013, each of which is incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under federal grant number W81XWH-10-1-0483 awarded by the Department of Defense and federal grant number 5R01-CA127727-03 awarded by NIH. The U.S. Government has certain rights to this invention.

SEQUENCE LISTING

[0003] The sequence listing is filed with the application in electronic format only and is incorporated by reference herein. The sequence listing text file “028193-9147-US04 Sequence Listing.xml” was created on Dec. 8, 2022, and is 66,845 bytes in size.

FIELD

[0004] The disclosure relates to methods for the detection and prognosis of cancer. Moreover, the disclosure provides methods for capturing and isolating circulating tumor cells (CTCs) that include the identification, detection, and optional enumeration of the CTCs that can be used in methods relating to a prognosis, diagnosis, or the treatment of cancer in a subject.

BACKGROUND

[0005] Most metazoan cells can be classified as either epithelial or mesenchymal based on morphology, behavior and molecular signatures. Epithelial tumor cells can become mesenchymal cells and vice versa via phenotypic transitions, a process known as epithelial plasticity. Epithelial cells are generally polar in the apico-basal direction, adherent to adjacent cells in the plane perpendicular to the polarity, and non-motile in the polar direction. Mesenchymal cells, in contrast, lack polarity, do not form tight interactions with neighboring cells, and are motile. In adult animals, epithelial and mesenchymal cells remain stably in one state or the other; that is, an epithelial cell does not change its properties and become mesenchymal. During development, however, epithelial cells of the early embryo give rise to all three embryonal layers (endoderm, mesoderm and ectoderm), which include mesenchymal cells (Hay, E. D., et al. *Am. J. Kidney Dis.* 1995, 26, 678-690). Therefore, these early embryonal cells have the ability to transition between epithelial and mesenchymal states. Embryos have been shown to undergo epithelial-mesenchymal transitions

(EMTs) as well as mesenchymal-epithelial transitions (METs) (Acloque, H., et al. *J. Clin. Invest.* 2009, 119, 1438-1449).

[0006] Epithelial plasticity (EP) refers to the reversible loss of the epithelial cellular phenotype, a process known to occur during cancer metastasis. This EP biology has been linked in multiple studies to the risk of cancer metastasis and the acquisition of mesenchymal and/or stemness properties through the EMT process. EMT has been linked to chemoresistance, invasion, intravasation, and dissemination in multiple preclinical models of cancer. The MET process, which results in the re-expression of the epithelial phenotype, is also likely of great importance in development and metastasis and has been linked to metastatic colonization and survival of tumor cells in the metastatic niche. For example, in prostate cancer, mesenchymal biomarkers may be upregulated during androgen deprivation in prostate cancer cell lines, animal models, and in patient tumor specimens. Moreover, these biomarkers are plastic, revert upon replacement of testosterone, and are linked to an increased metastatic propensity and chemoresistance. Mesenchymal-like tumor cells may better promote local tumor invasion and intravasation/extravasation, but epithelial tumor cells may be necessary for eventual survival and proliferation in the metastatic niche, illustrating the potential relevance of the dual nature of EP in mediating the full process of metastasis.

[0007] Circulating tumor cells (CTCs), which are cells that have detached from a primary tumor and circulate in the bloodstream, have potential prognostic, predictive and surrogate implications in oncology. CTCs may constitute seeds for subsequent growth of additional tumors (metastasis) in different tissues. Thus, detection of CTCs can provide a diagnosis and/or prognosis for overall survival and therapeutic implications in subjects with cancers such as metastatic prostate and breast cancer. The number of CTCs in any patient sample (e.g., a blood sample) can be very small, which can make detection difficult. Current methods for detecting CTCs are based on the detection of epithelial cell adhesion molecule (EpCAM) expression, which is a biomarker associated with epithelial cells. However, during the process of metastasis, circulating tumor cells (CTCs) may lose their epithelial phenotype and acquire a mesenchymal phenotype that is not sufficiently captured by existing epithelial-based CTC technologies. During metastasis, tumor cells may exist as a spectrum of epithelial to mesenchymal phenotypes. CTCs may lose their epithelial phenotype and acquire a mesenchymal phenotype, which may not be captured with existing epithelial-based CTC technology and thus lead to the under-detection of CTCs under circumstances where cells undergo a decrease or loss of EpCAM expression, such as during biologic processes including EMT. Because of the role CTCs can play in the diagnosis, monitoring, and prognosis of disease in patients having cancer, any shortcoming in the detection technology needs to be addressed by the art.

[0008] There is recent evidence to suggest that CTCs with a mesenchymal phenotype are missed by CELLSEARCH® and other epithelial-based technologies. Accordingly, there is a need for methods and systems for capturing CTCs that do not rely on existing capture technologies, and methods for correlating CTC detection to diagnosis, monitoring, and prognosis of disease in cancer patients.

SUMMARY

[0009] In an aspect, the disclosure provides a method for detecting a circulating tumor cell (CTC) in a biological sample, the method comprising detecting at least one epithelial mesenchymal transition (EMT) biomarker in the biological sample.

[0010] In an aspect, the disclosure provides a kit for detecting a circulating tumor cell (CTC) in a biological sample, the kit comprising an antibody to at least one EMT biomarker and instructions for use.

[0011] In an aspect, the disclosure provides a method of predicting responsiveness of a subject having cancer to a course of cancer treatment, the method comprising: determining the level or presence of expression of at least one EMT biomarker to obtain an EMT biomarker profile and/or optionally a gene expression pattern for a CTC; and predicting the responsiveness of the subject to the cancer drug based on the EMT biomarker profile and/or optional gene expression pattern. In some embodiments the method includes: determining the level or presence of expression of at least one EMT biomarker in a sample from the subject to obtain a biomarker profile and optionally a gene expression pattern in a CTC for the subject; identifying the type of cancer from the biomarker profile and/or optional gene expression pattern, and optionally characterizing the stage of the cancer; and predicting responsiveness of the subject to the cancer drug based on any one of the biomarker pattern, the optional gene expression pattern, the type of cancer, or the stage of the cancer. Embodiments of this aspect can include detecting a number of cells captured and enumerated from a blood sample using at least one EMT biomarker applied to a sample from the subject. These cells that express the EMT biomarker are thereby captured using the EMT biomarker and could then be used to obtain a gene expression pattern in CTCs for the subject; to predict responsiveness of the subject to the cancer drug based on the obtained gene expression pattern, and for the detection of other biomarkers in these CTCs to assist in guiding therapy of that subject. These cells could also be used to measure the level of the specified EMT biomarker or other EMT biomarkers.

[0012] In an aspect, the disclosure provides a method of assessing the number of CTCs using both the traditional EpCAM based capture methodology and an EMT-marker based capture methodology. This EMT-based capture may replace or complement existing CTC capture technologies. The further capture, enumeration, and characterization of these CTCs using EMT antigen capture may further targeting delivery of a cancer drug in a subject having cancer comprising administering to the subject a cancer drug linked to an antibody specific for at least one EMT biomarker or specific drugs based on a gene expression profile or presence of this EMT biomarker.

[0013] In an aspect, the disclosure provides a method of estimating the prognosis of a subject with cancer as well as permitting a further characterization of CTCs that may predict for therapeutic responsiveness, the method comprising: determining the level of or presence of expression of at least one EMT biomarker in a sample from the subject to determine the number of CTCs in the subject and to obtain a gene expression pattern for the subject; and providing a prognosis to the subject based on the gene expression or biomarker profile pattern obtained.

[0014] In an aspect, the disclosure provides a method for monitoring progression of cancer in a subject undergoing

therapeutic treatment, the method comprising detecting the level of expression or presence of expression of at least one EMT biomarker and the quantification of CTCs captured using this method in blood samples taken from the subject at a first and a second time; and comparing the first and second levels of expression; wherein a detected difference in the level of expression of the at least one EMT biomarker in the first and second samples over time indicates a change in the progression status of the cancer.

[0015] In an aspect, the disclosure provides a method for detecting cancer in a subject, the method comprising determining the presence of CTCs that express at least one EMT biomarker in a sample from the subject as compared to a normal or control sample, wherein an increased level of at least one EMT biomarker indicates presence of cancer progression or metastatic spread in the subject.

[0016] In an aspect, the disclosure provides a method of treating cancer in a subject comprising administering to the subject a cancer drug linked to an antibody that specifically binds at least one EMT biomarker.

[0017] The present disclosure is directed to a method for isolating, capturing, or enriching a circulating tumor cell from a patient, the method comprising: obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; and separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample, thereby isolating, capturing, or enriching the circulating tumor cell. The method may further comprise confirming the circulating tumor cell. Confirming the circulating tumor cell comprises at least one of DAPI staining, β -catenin detection, CD45 detection and CD31 detection. The circulating tumor cell may be confirmed if DAPI staining is positive, β -catenin expression is positive, CD45 expression is negative, and CD31 expression is negative. The circulating tumor cell may have a mesenchymal phenotype. The EMT biomarker may be at least one of OB-cadherin, N-cadherin, vimentin, E-cadherin, FGFR2 splice variant isoforms, or CD133. The patient may have cancer. The may comprise at least one of breast cancer, colon cancer, lung cancer, prostate cancer, testicular cancer, brain cancer, skin cancer, rectal cancer, gastric cancer, esophageal cancer, sarcomas, tracheal cancer, head and neck cancer, pancreatic cancer, liver cancer, ovarian cancer, lymphoid cancer, cervical cancer, vulvar cancer, melanoma, mesothelioma, renal cancer, bladder cancer, thyroid cancer, bone cancer, carcinoma, sarcoma, and soft tissue cancer. The method may further comprise determining the presence or absence of at least one prostate cancer-specific genomic event. The at least one prostate cancer-specific genomic event may be selected from the group consisting of androgen receptor amplification, phosphatase and tensin homolog (PTEN) loss, gene fusion of transmembrane protease, serine 2 (TMPRSS2) gene and ETS related (ERG) gene, and combinations thereof. The determining step may be performed by fluorescence in situ hybridization

(FISH). The biological sample may comprise a tissue sample or a fluid sample from an organism. The biological sample may comprise blood. The capture binding protein may be an antibody. The solid phase may be a microparticle. The microparticle may be magnetic or paramagnetic.

[0018] The present disclosure is directed to a method for detecting or identifying a circulating tumor cell in a patient, the method comprising: obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; and separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample, thereby detecting or identifying the circulating tumor cell. The method may further comprise confirming the circulating tumor cell. Confirming the circulating tumor cell comprises at least one of DAPI staining, β -catenin detection, CD45 detection and CD31 detection. The circulating tumor cell may be confirmed if DAPI staining is positive, β -catenin expression is positive, CD45 expression is negative, and CD31 expression is negative. The circulating tumor cell may have a mesenchymal phenotype. The EMT biomarker may be at least one of OB-cadherin, N-cadherin, vimentin, E-cadherin, FGFR2 splice variant isoforms, or CD133. The patient may have cancer. The may comprise at least one of breast cancer, colon cancer, lung cancer, prostate cancer, testicular cancer, brain cancer, skin cancer, rectal cancer, gastric cancer, esophageal cancer, sarcomas, tracheal cancer, head and neck cancer, pancreatic cancer, liver cancer, ovarian cancer, lymphoid cancer, cervical cancer, vulvar cancer, melanoma, mesothelioma, renal cancer, bladder cancer, thyroid cancer, bone cancer, carcinoma, sarcoma, and soft tissue cancer. The method may further comprise determining the presence or absence of at least one prostate cancer-specific genomic event. The at least one prostate cancer-specific genomic event may be selected from the group consisting of androgen receptor amplification, phosphatase and tensin homolog (PTEN) loss, gene fusion of transmembrane protease, serine 2 (TMPRSS2) gene and ETS related (ERG) gene, and combinations thereof. The determining step may performed by fluorescence in situ hybridization (FISH). The biological sample may comprise a tissue sample or a fluid sample from an organism. The biological sample may comprise blood. The capture binding protein may be an antibody. The solid phase may be a microparticle. The microparticle may be magnetic or paramagnetic. The present disclosure is directed to a method for isolating or capturing an intact cell from a patient, wherein the cell is β -catenin positive, DAPI positive, and CD45 negative, the method comprising: obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the intact cell to form a solid phase-capture

binding protein-intact cell complex; and separating the solid phase-capture binding protein-intact cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample, thereby isolating or capturing the intact cell. The intact cell may have a mesenchymal phenotype. The EMT biomarker may be at least one of OB-cadherin, N-cadherin, vimentin, E-cadherin, FGFR2 splice variant isoforms, or CD133. The patient may have cancer. The may comprise at least one of breast cancer, colon cancer, lung cancer, prostate cancer, testicular cancer, brain cancer, skin cancer, rectal cancer, gastric cancer, esophageal cancer, sarcomas, tracheal cancer, head and neck cancer, pancreatic cancer, liver cancer, ovarian cancer, lymphoid cancer, cervical cancer, vulvar cancer, melanoma, mesothelioma, renal cancer, bladder cancer, thyroid cancer, bone cancer, carcinoma, sarcoma, and soft tissue cancer. The biological sample may comprise a tissue sample or a fluid sample from an organism. The biological sample may comprise blood. The capture binding protein may be an antibody. The solid phase may be a microparticle. The microparticle may be magnetic or paramagnetic.

[0019] The present disclosure is directed to a method of detecting and treating cancer in a subject, the method comprising: obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample, determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex; comparing the level of circulating tumor cell to a reference level of circulating tumor cell, detecting cancer in the subject if the level of circulating tumor cell is higher than the reference level of circulating tumor cell, cancer is detected in the subject, and administering a therapy against cancer to the subject identified as having cancer. Determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex comprises confirming the circulating tumor cell. Confirming the circulating tumor cell comprises at least one of DAPI staining, β -catenin detection, CD45 detection and CD31 detection. The circulating tumor cell may be confirmed if DAPI staining is positive, β -catenin expression is positive, CD45 expression is negative, and CD31 expression is negative. The EMT biomarker may be at least one of OB-cadherin, N-cadherin, vimentin, E-cadherin, FGFR2 splice variant isoforms, or CD133. The patient may have cancer. The may comprise at least one of breast cancer, colon cancer, lung cancer, prostate cancer, testicular cancer, brain cancer, skin cancer, rectal cancer, gastric cancer, esophageal cancer, sarcomas, tracheal cancer, head and neck cancer, pancreatic cancer, liver cancer, ovarian cancer, lymphoid cancer, cervical cancer, vulvar cancer, melanoma, mesothelioma, renal cancer, bladder cancer, thyroid cancer, bone cancer, carcinoma, sarcoma, and soft tissue cancer. The

method may further comprise determining the presence or absence of at least one prostate cancer-specific genomic event. The at least one prostate cancer-specific genomic event may be selected from the group consisting of androgen receptor amplification, phosphatase and tensin homolog (PTEN) loss, gene fusion of transmembrane protease, serine 2 (TMPRSS2) gene and ETS related (ERG) gene, and combinations thereof. The determining step may performed by fluorescence in situ hybridization (FISH). The biological sample may comprise a tissue sample or a fluid sample from an organism. The biological sample may comprise blood. The capture binding protein may be an antibody. The solid phase may be a microparticle. The microparticle may be magnetic or paramagnetic.

[0020] The present disclosure is directed to a method of monitoring progression of cancer in a subject undergoing therapeutic treatment, the method comprising: obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample; determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex; correlating the level of circulating tumor cell with the progression of cancer in the subject, wherein if the level of the circulating tumor cell is higher as compared to the level of the circulating tumor cell in an earlier biological sample from the subject, the subject is identified as having progression of cancer; and administering a therapy against cancer to the subject identified as having progression of cancer. Determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex comprises confirming the circulating tumor cell. Confirming the circulating tumor cell comprises at least one of DAPI staining, β -catenin detection, CD45 detection and CD31 detection. The circulating tumor cell may be confirmed if DAPI staining is positive, β -catenin expression is positive, CD45 expression is negative, and CD31 expression is negative. The EMT biomarker may be at least one of OB-cadherin, N-cadherin, vimentin, E-cadherin, FGFR2 splice variant isoforms, or CD133. The patient may have cancer. The may comprise at least one of breast cancer, colon cancer, lung cancer, prostate cancer, testicular cancer, brain cancer, skin cancer, rectal cancer, gastric cancer, esophageal cancer, sarcomas, tracheal cancer, head and neck cancer, pancreatic cancer, liver cancer, ovarian cancer, lymphoid cancer, cervical cancer, vulvar cancer, melanoma, mesothelioma, renal cancer, bladder cancer, thyroid cancer, bone cancer, carcinoma, sarcoma, and soft tissue cancer. The method may further comprise determining the presence or absence of at least one prostate cancer-specific genomic event. The at least one prostate cancer-specific genomic event may be selected from the group consisting of androgen receptor amplification, phosphatase and tensin homolog (PTEN) loss, gene fusion of transmembrane protease, serine 2 (TMPRSS2)

gene and ETS related (ERG) gene, and combinations thereof. The determining step may performed by fluorescence in situ hybridization (FISH). The biological sample may comprise a tissue sample or a fluid sample from an organism. The biological sample may comprise blood. The capture binding protein may be an antibody. The solid phase may be a microparticle. The microparticle may be magnetic or paramagnetic.

[0021] The present disclosure is directed to a method of determining a cancer prognosis in a subject, the method comprising: obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; and separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample; determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex; comparing the level of circulating tumor cell to a reference level of circulating tumor cell, determining the cancer prognosis in the subject, wherein if the level of circulating tumor cell is higher than the reference level of circulating tumor cell, the subject is identified as having cancer, and administering a therapy against cancer to the subject identified as having cancer. Determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex comprises confirming the circulating tumor cell. Confirming the circulating tumor cell comprises at least one of DAPI staining, β -catenin detection, CD45 detection and CD31 detection. The circulating tumor cell may be confirmed if DAPI staining is positive, β -catenin expression is positive, CD45 expression is negative, and CD31 expression is negative. The EMT biomarker may be at least one of OB-cadherin, N-cadherin, vimentin, E-cadherin, FGFR2 splice variant isoforms, or CD133. The patient may have cancer. The may comprise at least one of breast cancer, colon cancer, lung cancer, prostate cancer, testicular cancer, brain cancer, skin cancer, rectal cancer, gastric cancer, esophageal cancer, sarcomas, tracheal cancer, head and neck cancer, pancreatic cancer, liver cancer, ovarian cancer, lymphoid cancer, cervical cancer, vulvar cancer, melanoma, mesothelioma, renal cancer, bladder cancer, thyroid cancer, bone cancer, carcinoma, sarcoma, and soft tissue cancer. The method may further comprise determining the presence or absence of at least one prostate cancer-specific genomic event. The at least one prostate cancer-specific genomic event may be selected from the group consisting of androgen receptor amplification, phosphatase and tensin homolog (PTEN) loss, gene fusion of transmembrane protease, serine 2 (TMPRSS2) gene and ETS related (ERG) gene, and combinations thereof. The determining step may performed by fluorescence in situ hybridization (FISH). The biological sample may comprise a tissue sample or a fluid sample from an organism. The biological sample may comprise blood. The

capture binding protein may be an antibody. The solid phase may be a microparticle. The microparticle may be magnetic or paramagnetic.

[0022] The present disclosure is directed to a method of predicting responsiveness of a subject having cancer to a course of treatment, the method comprising: obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; and separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample; determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex; and comparing the level of circulating tumor cell to a reference level of circulating tumor cell. Determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex comprises confirming the circulating tumor cell. Confirming the circulating tumor cell comprises at least one of DAPI staining, β -catenin detection, CD45 detection and CD31 detection. The circulating tumor cell may be confirmed if DAPI staining is positive, β -catenin expression is positive, CD45 expression is negative, and CD31 expression is negative. The EMT biomarker may be at least one of OB-cadherin, N-cadherin, vimentin, E-cadherin, FGFR2 splice variant isoforms, or CD133. The patient may have cancer. The may comprise at least one of breast cancer, colon cancer, lung cancer, prostate cancer, testicular cancer, brain cancer, skin cancer, rectal cancer, gastric cancer, esophageal cancer, sarcomas, tracheal cancer, head and neck cancer, pancreatic cancer, liver cancer, ovarian cancer, lymphoid cancer, cervical cancer, vulvar cancer, melanoma, mesothelioma, renal cancer, bladder cancer, thyroid cancer, bone cancer, carcinoma, sarcoma, and soft tissue cancer. The method may further comprise determining the presence or absence of at least one prostate cancer-specific genomic event. The at least one prostate cancer-specific genomic event may be selected from the group consisting of androgen receptor amplification, phosphatase and tensin homolog (PTEN) loss, gene fusion of transmembrane protease, serine 2 (TMPRSS2) gene and ETS related (ERG) gene, and combinations thereof. The determining step may be performed by fluorescence in situ hybridization (FISH). The biological sample may comprise a tissue sample or a fluid sample from an organism. The biological sample may comprise blood. The capture binding protein may be an antibody. The solid phase may be a microparticle. The microparticle may be magnetic or paramagnetic.

[0023] The present disclosure is directed to a kit for isolating or capturing a circulating tumor cell in a biological sample, the kit comprising an antibody linked to a magnetic particle, wherein the antibody binds specifically to at least one EMT biomarker and at least one staining reagent. The at least one EMT biomarker may include at least one of OB-cadherin, N-cadherin, vimentin, E-cadherin, FGFR2 splice variant isoforms, or CD133. The at least one staining

reagent may include at least one of phycoerytherin-labeled anti- β -catenin antibody and an allophycocyanin-labeled anti-CD45 antibody.

[0024] Other aspects and embodiments of the disclosure will become apparent by consideration of the detailed description and accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] FIG. 1A depicts a schematic representation of the IIb and IIc alternatively spliced isoforms of FGFR2. FIG. 1B is a schematic of the pRIIc1² minigene and the fluorescence read-out. FIG. 1C is an RT-PCR analysis of the reporter (upper panel) and endogenous FGFR2 (lower panel). FIG. 1D are epifluorescence and phase-contrast pictures of clones AT3-M and AT3-T.

[0026] FIG. 2A depicts examples of clusters of DsRED positive cells formed by AT3-M cells upon treatment with conditioned media from clone AT3-T. FIG. 2B depicts flow cytometry analysis of the same experiment.

[0027] FIG. 3A depicts growth curves for clones AT3-T and AT3-M. FIG. 3B is graph of growth of AT3-M, AT3-T, and DT cells in soft agar. FIG. 3C depicts a sacrifice curve for rats injected with AT3-M or AT3-T cells. FIG. 3D depicts a comparison of tumor volumes resulting from AT3-T and AT3-M injection.

[0028] FIG. 4A depicts a representative example of cells that express both RFP and GFP at the periphery of an AT3-M tumor stably transfected with Gint and pRIIc1² reporters. FIG. 4B is a representative example of a section from an AT3-T tumor stably transfected with GFP and pRIIc1² reporters.

[0029] FIG. 5 depicts a representative example of cells that express both RFP and GFP at the periphery of an AT3-M tumor stably transfected with Gint and pRIIc1² reporters.

[0030] FIG. 6A depicts representative pictures of cells for the scratch-wound assay. FIG. 6B is a quantification of migration. FIG. 6C depicts an invasion assay using Matrigel coated membranes. FIG. 6D is a quantification of invasion assay results.

[0031] FIGS. 7A-7B are metastatic foci in lungs from animals with tumors from either AT3-T or AT3-M clones (stably transfected with GFP and pRIIc1² reporters). FIG. 7A (upper panel) is an example of a section exhibiting the pattern for clone AT3-T (i.e. GFP+, DsRED+) in a metastatic focus and (lower panel) an example of a section exhibiting a plastic pattern for clone AT3-T (i.e. GFP+, DsRED-) in a metastatic focus. FIG. 7B (upper panel) is an example of a section exhibiting the pattern for clone AT3-M (i.e. GFP+, DsRED-) in a metastatic focus and (lower panel) an example of a section exhibiting a plastic pattern for clone AT3-M (i.e. GFP+, DsRED+) in a metastatic focus.

[0032] FIG. 8A depicts a membrane with serial two-fold dilutions of whole cell lysates cut in half and immunoblotted for CD133 (upper panel) or β -actin (lower panel). FIG. 8B depicts a membrane with serial twofold dilutions of whole cell lysates cut in half and immunoblotted for CD44 (upper panel) or β -actin (lower panel).

[0033] FIG. 9 depicts a model comparing stem cell-like character and epithelial mesenchymal phenotype.

[0034] FIGS. 10A-10D depict CTCs from patients with prostate adenocarcinoma. FIG. 10A illustrates an example of a leukocyte from a human peripheral blood mononuclear cell (PMBC) sample: CD45 (+), CK (-), and vimentin (+). FIG. 10B illustrates an example of a CD45 (-), CK (+), and

vimentin (–) cell from a patient with metastatic breast cancer. FIG. 10C illustrates an example of a CD45 (–), CK (+), vimentin (+) from a patient with metastatic breast cancer (mBC). FIG. 10D illustrates an example of a CD45 (–), CK (+), vimentin (+) from a patient with metastatic progressive castrate-resistant prostate cancer (mCRPC).

[0035] FIGS. 11A-11D depict immunofluorescent images of CTCs from patients with mCRPC and mBC.

[0036] FIGS. 12A-12D depict immunofluorescent images of CTCs from patients with mCRPC and mBC.

[0037] FIGS. 13A-13H depict immunofluorescent images of CTCs from patients with mCRPC and mBC.

[0038] FIGS. 14A-14E depict immunofluorescent images of CTCs from patients with mCRPC and mBC.

[0039] FIGS. 15A-15E depict immunofluorescent images of CTCs from patients with mCRPC and mBC.

[0040] FIGS. 16A-16D depict immunofluorescent images of CTCs from patients with mCRPC and mBC.

[0041] FIG. 17 depicts after enrichment using anti-N-cadherin or anti-OB-cadherin ferrofluid, mesenchymal CTCs were differentiated from leukocytes by the presence of β -catenin expression and the lack of CD45 expression.

[0042] FIGS. 18A-18E depict immunofluorescent images of control cells (PC-3 cells mixed with peripheral blood mononuclear cells) in rows FIG. 18A and FIG. 18D and patient-derived EpCAM-captured cells in rows FIG. 18B, FIG. 18C, and FIG. 18E. Cells are stained for CD45 and cytokeratin and further characterized by either β -catenin or OB-cadherin expression. Columns represent phase microscopy with 4',6-diamidino-2-phenylindole (DAPI), CD45 with DAPI, Cytokeratin (CK) with DAPI, and either β -catenin (beta-cat) or OB-cadherin (OB-CAD) with DAPI. Row FIG. 18A shows CD45-positive control cells lacking β -catenin and CK-positive control cells expressing β -catenin. Row FIG. 18B shows a CTC from a man with prostate cancer with both CK and β -catenin expression, while row FIG. 18C shows a CD45-negative, CK-negative patient cell with β -catenin expression. Row FIG. 18D illustrates CD45-positive control cells lacking OB-cadherin and CK-positive control cells expressing OB-cadherin, and row FIG. 18E shows a CTC with both CK and OB-cadherin expression.

[0043] FIG. 19 depicts examples of β -catenin expression in EpCAM-captured CTCs from a man with castration-resistant prostate cancer.

[0044] FIG. 20A depicts the distribution of OB-cadherin-captured, β -catenin-positive events from healthy volunteers based on CD31 status. All samples in which CD31 was assessed are CD31 positive; and FIG. 20B depicts examples of CD31+ cellular events detected in healthy volunteers.

[0045] FIG. 21 depicts events captured from healthy volunteers and CRPC patients using N-cadherin, OB-cadherin, or EpCAM ferrofluid.

[0046] FIG. 22 depicts examples of OB-cadherin captured, β -catenin-positive cellular events from men with metastatic castration-resistant prostate cancer. The top rows show single cells which are mostly CD31-positive and may represent endothelial cells, while the bottom rows show clumps of CD31-negative cells which may represent mesenchymal tumor cells.

[0047] FIG. 23A depicts immunofluorescent images of PC-3 cells stained for OB-cadherin (green) and DAPI/nucleus (blue), illustrating that PC-3 cells are somewhat heterogeneous for OB-cadherin expression; and FIG. 23B

depicts fluorescence-activated cell sorting of PC-3 cells based on OB-cadherin expression shows that approximately 50% of the cells express OB-cadherin.

[0048] FIGS. 24A-24C depict immunostaining and fluorescent in situ hybridization (FISH) results from a representative patient with metastatic castration-resistant prostate cancer. Columns FIG. 24A and FIG. 24B—Circulating cells captured with OB-cadherin and stained with β -catenin show same multigene FISH pattern as a CTC captured with EpCAM and stained with cytokeratin from the same patient. Androgen receptor (AR) FISH shows extra copies of the androgen receptor gene. For the ERG break FISH, yellow arrows denote missing 5' Erg signals which is indicative of TMPRSS2:ERG fusion. PTEN FISH shows homozygous deletion of PTEN gene. Column FIG. 24C—leukocyte from the same patient shows a cytokeratin-negative cell with a normal FISH pattern of 1 AR signal, no ERG rearrangement, and two copies of PTEN.

DETAILED DESCRIPTION

[0049] Before any embodiments are described in detail, it is to be understood that the claims are not limited to the details of construction and the arrangement of components set forth in the following description or illustrated in the included drawings.

[0050] 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. In case of conflict, the present document, including definitions, will control. Preferred methods and materials are described below, although methods and materials similar or equivalent to those described herein can be used in practice or testing of the present invention. All publications, patent applications, patents and other references mentioned herein are incorporated by reference in their entirety. The materials, methods, and examples disclosed herein are illustrative only and not intended to be limiting.

[0051] The terms “comprise(s),” “include(s),” “having,” “has,” “can,” “contain(s),” and variants thereof, as used herein, are intended to be open-ended transitional phrases, terms, or words that do not preclude the possibility of additional acts or structures. The singular forms “a,” “and” and “the” include plural references unless the context clearly dictates otherwise. The present disclosure also contemplates other embodiments “comprising,” “consisting of” and “consisting essentially of,” the embodiments or elements presented herein, whether explicitly set forth or not.

[0052] In a general sense, the disclosure provides biomarkers that have been identified to be associated with circulating tumor cells (CTCs). As described herein, one or more biomarkers of epithelial mesenchymal transition (EMT) are detectable on CTCs of patients afflicted with common epithelial malignancies. These transitional cells often display stem cell-like characteristics (stemness) and/or plasticity. Further, the disclosure provides description that metastatic propensity and epithelial phenotypic changes correlate with alternative splicing of the FGFR2 gene. The disclosure also provides that, as illustrated in the non-limiting Examples, transitional cells are found in cancer patients where many CTCs co-expressed biomarkers associated with epithelial and mesenchymal cells.

[0053] Thus, as described below EMT biomarker expression can be used to detect and quantify CTCs in a biological sample. Accordingly, methods comprising detection of EMT

biomarker expression, or detection of CTCs, or a combination thereof, can be used to assess cancer prognosis, tumor invasiveness, risk of metastasis, or to stage tumors. As one of skill in the art will appreciate, any suitable method for evaluating EMT biomarker expression can be used to evaluate EMT biomarker expression according to the methods described herein including, but not limited to, detection with antibodies, real time RT-PCR, Northern analysis, Western analysis, and flow cytometry.

[0054] The disclosure also describes the development of a CTC capture method that is based on the biology of epithelial plasticity and isolates cells based on mesenchymal markers, such as N-cadherin or OB-cadherin cell surface expression. In patients with advanced breast and prostate cancer, EP biomarkers including OB-cadherin, N-cadherin, and vimentin can be detected in CTCs that are isolated by EpCAM-based ferromagnetic capture and co-express cytokeratin, which is expressed in epithelial cells. Similarly, CTCs expressing the mesenchymal markers twist and vimentin have been identified rarely in patients with early stage breast cancer but in the majority of patients with metastatic breast cancer, suggesting that transition to a mesenchymal phenotype may be important for metastasis. Furthermore, recent serial monitoring of CTCs with a mesenchymal phenotype, as defined by RNA fluorescence in situ hybridization (FISH), suggests that there may be an association between mesenchymal CTCs and disease progression in women with breast cancer.

[0055] As described herein, in patients with metastatic castration-resistant prostate cancer (CRPC) and breast cancer (BC), CTCs isolated using epithelial cell adhesion molecule (EpCAM) ferromagnetic capture expressed mesenchymal markers, including N- and OB-cadherin, suggesting phenotypic plasticity and the presence of EMT. The CTC capture method described herein involves a mesenchymal-based assay. This assay detected OB-cadherin cellular events present in men with metastatic prostate cancer but were less common in healthy individuals. This method may complement existing epithelial-based methods and may potentially be useful in patients with bone metastases.

Definitions

[0056] “Antibody” and “antibodies” as used herein refers to monoclonal antibodies, multispecific antibodies, human antibodies, humanized antibodies (fully or partially humanized), animal antibodies such as, but not limited to, a bird (for example, a duck or a goose), a shark, a whale, and a mammal, including a non-primate (for example, a cow, a pig, a camel, a llama, a horse, a goat, a rabbit, a sheep, a hamster, a guinea pig, a cat, a dog, a rat, a mouse, etc.) or a non-human primate (for example, a monkey, a chimpanzee, etc.), recombinant antibodies, chimeric antibodies, single-chain Fvs (“scFv”), single chain antibodies, single domain antibodies, Fab fragments, F(ab') fragments, F(ab')₂ fragments, disulfide-linked Fvs (“sdFv”), and anti-idiotypic (“anti-Id”) antibodies, dual-domain antibodies, dual variable domain (DVD) or triple variable domain (TVD) antibodies (dual-variable domain immunoglobulins and methods for making them are described in Wu, C., et al., *Nature Biotechnology*, 25(11):1290-1297 (2007) and PCT International Application WO 2001/058956, the contents of each of which are herein incorporated by reference), and functionally active epitope-binding fragments of any of the above. In particular, antibodies include immunoglobulin molecules

and immunologically active fragments of immunoglobulin molecules, namely, molecules that contain an analyte-binding site. Immunoglobulin molecules can be of any type (for example, IgG, IgE, IgM, IgD, IgA, and IgY), class (for example, IgG1, IgG2, IgG3, IgG4, IgA1, and IgA2), or subclass. For simplicity sake, an antibody against an analyte is frequently referred to herein as being either an “anti-analyte antibody” or merely an “analyte antibody.”

[0057] “Antibody fragment” as used herein refers to a portion of an intact antibody comprising the antigen-binding site or variable region. The portion does not include the constant heavy chain domains (i.e. CH₂, CH₃, or CH₄, depending on the antibody isotype) of the Fc region of the intact antibody. Examples of antibody fragments include, but are not limited to, Fab fragments, Fab' fragments, Fab'-SH fragments, F(ab')₂ fragments, Fd fragments, Fv fragments, diabodies, single-chain Fv (scFv) molecules, single-chain polypeptides containing only one light chain variable domain, single-chain polypeptides containing the three CDRs of the light-chain variable domain, single-chain polypeptides containing only one heavy chain variable region, and single-chain polypeptides containing the three CDRs of the heavy chain variable region. The term “administration” or “administering,” as used herein, refers to providing, contacting, and/or delivery of a cancer treatment by any appropriate route to achieve the desired effect. The cancer treatment may be administered to a subject in numerous ways including, but not limited to, orally, ocularly, nasally, intravenously, topically, as aerosols, suppository, etc. and may be used in combination.

[0058] “Binding Protein” is used herein to refer to a monomeric or multimeric protein that binds to and forms a complex with a binding partner, such as, for example, a polypeptide, an antigen, a chemical compound or other molecule, or a substrate of any kind. A binding protein specifically binds a binding partner. Binding proteins include antibodies, as well as antigen-binding fragments thereof and other various forms and derivatives thereof as are known in the art and described herein below, and other molecules comprising one or more antigen-binding domains that bind to an antigen molecule or a particular site (epitope) on the antigen molecule. Accordingly, a binding protein includes, but is not limited to, an antibody a tetrameric immunoglobulin, an IgG molecule, an IgG1 molecule, a monoclonal antibody, a chimeric antibody, a CDR-grafted antibody, a humanized antibody, an affinity matured antibody, and fragments of any such antibodies that retain the ability to bind to an antigen.

[0059] The term “biomarker” as used herein refers to any quantifiable biological component that is unique to a particular physiological condition (e.g., cancer). A biomarker may be a gene, an mRNA transcribed from said gene, or a protein translated from said mRNA. A measureable increase or decrease, of a biomarker level, relative to a control, such as an individual, group of individuals or populations, or alternatively, relative to subjects with cancer, may provide a diagnosis of a particular physiological condition.

[0060] “Breast cancer” as used herein refers to a type of cancer that originates from and develops in the breast.

[0061] “Cancer” as used herein refers to the uncontrolled and unregulated growth of abnormal cells in the body. Cancerous cells are also called malignant cells. Cancer may invade nearby parts of the body and may also spread to more distant parts of the body through the lymphatic system or

bloodstream. Cancers include Adrenocortical Carcinoma, Anal Cancer, Bladder Cancer, Brain Tumor, Breast Cancer, Carcinoid Tumor, Gastrointestinal, Carcinoma of Unknown Primary, Cervical Cancer, Colon Cancer, Endometrial Cancer, Esophageal Cancer, Extrahepatic Bile Duct Cancer, Ewings Family of Tumors (PNET), Extracranial Germ Cell Tumor, Intraocular Melanoma Eye Cancer, Gallbladder Cancer, Gastric Cancer (Stomach), Extragonadal Germ Cell Tumor, Gestational Trophoblastic Tumor, Head and Neck Cancer, Hypopharyngeal Cancer, Islet Cell Carcinoma, Kidney Cancer (renal cell cancer), Laryngeal Cancer, Acute Lymphoblastic Leukemia, Leukemia, Acute Myeloid, Chronic Lymphocytic Leukemia, Chronic Myelogenous Leukemia, Hairy Cell Leukemia, Lip and Oral Cavity Cancer, Liver Cancer, Non-Small Cell Lung Cancer, Small Cell Lung Cancer, AIDS-Related Lymphoma, Central Nervous System (Primary) Lymphoma, Cutaneous T-Cell Lymphoma, Hodgkin's Disease Lymphoma, Non-Hodgkin's Disease Lymphoma, Malignant Mesothelioma, Melanoma, Merkel Cell Carcinoma, Metastatic Squamous Neck Cancer with Occult Primary, Multiple Myeloma and Other Plasma Cell Neoplasms, Mycosis Fungoides, Myelodysplastic Syndrome, Myeloproliferative Disorders, Nasopharyngeal Cancer, euroblastoma, Oral Cancer, Oropharyngeal Cancer, Osteosarcoma, Ovarian Epithelial Cancer, Ovarian Germ Cell Tumor, Pancreatic Cancer, Exocrine, Pancreatic Cancer, Islet Cell Carcinoma, Paranasal Sinus and Nasal Cavity Cancer, Parathyroid Cancer, Penile Cancer, Pituitary Cancer, Plasma Cell Neoplasm, Prostate Cancer, Rhabdomyosarcoma, Rectal Cancer, Renal Cell Cancer (cancer of the kidney), Transitional Cell Renal Pelvis and Ureter, Salivary Gland Cancer, Sezary Syndrome, Skin Cancer, Small Intestine Cancer, Soft Tissue Sarcoma, Testicular Cancer, Malignant Thymoma, Thyroid Cancer, Urethral Cancer, Uterine Cancer, Unusual Cancer of Childhood, Vaginal Cancer, Vulvar Cancer, and Wilms' Tumor.

[0062] "Circulating tumor cells", "CTC" and "CTCs" as used interchangeably herein refers to cells that have shed into the vasculature from a primary tumor and circulate in the bloodstream. CTCs are considered seeds for subsequent growth of additional tumors (metastasis) in vital distant organs, triggering a mechanism that is responsible for the vast majority of cancer-related deaths.

[0063] "Component," "components," or "at least one component," refer generally to a capture antibody, a detection or conjugate a calibrator, a control, a sensitivity panel, a container, a buffer, a diluent, a salt, an enzyme, a co-factor for an enzyme, a detection reagent, a pretreatment reagent/solution, a substrate (e.g., as a solution), a stop solution, and the like that can be included in a kit for assay of a test sample, such as a patient urine, serum or plasma sample, in accordance with the methods described herein and other methods known in the art. Some components can be in solution or lyophilized for reconstitution for use in an assay.

[0064] The term "effective dosage" as used herein means a dosage of a drug effective for periods of time necessary, to achieve the desired therapeutic result. An effective dosage may be determined by a person skilled in the art and may vary according to factors such as the disease state, age, sex, and weight of the individual, and the ability of the drug to elicit a desired response in the individual.

[0065] "Epithelial" and "epithelial phenotype" as used herein refer to membranous tissue composed of one or more layers of cells separated by very little intercellular substance

and forming the covering the most internal and external surfaces of the body and its organs.

[0066] "Label" and "detectable label" as used herein refer to a moiety attached to an antibody or an analyte to render the reaction between the antibody and the analyte detectable, and the antibody or analyte so labeled is referred to as "detectably labeled." A label can produce a signal that is detectable by visual or instrumental means. Various labels include signal-producing substances, such as chromagens, fluorescent compounds, chemiluminescent compounds, radioactive compounds, and the like. Representative examples of labels include moieties that produce light, e.g., acridinium compounds, and moieties that produce fluorescence, e.g., fluorescein. Other labels are described herein. In this regard, the moiety, itself, may not be detectable but may become detectable upon reaction with yet another moiety. Use of the term "detectably labeled" is intended to encompass such labeling.

[0067] Any suitable detectable label as is known in the art can be used. For example, the detectable label can be a radioactive label (such as ^3H , ^{14}C , ^{32}P , ^{33}P , ^{35}S , ^{90}Y , ^{99}Tc , ^{111}In , ^{125}I , ^{131}I , ^{177}Lu , ^{166}Ho , and ^{153}Sm), an enzymatic label (such as horseradish peroxidase, alkaline peroxidase, glucose 6-phosphate dehydrogenase, and the like), a chemiluminescent label (such as acridinium esters, thioesters, or sulfonamides; luminol, isoluminol, phenanthridinium esters, and the like), a fluorescent label (such as fluorescein (e.g., 5-fluorescein, 6-carboxyfluorescein, 3' 6-carboxyfluorescein, 5(6)-carboxyfluorescein, 6-hexachloro-fluorescein, 6-tetrachlorofluorescein, fluorescein isothiocyanate, and the like)), rhodamine, phycobiliproteins, R-phycoerythrin, quantum dots (e.g., zinc sulfide-capped cadmium selenide), a thermometric label, or an immuno-polymerase chain reaction label. An introduction to labels, labeling procedures and detection of labels is found in Polak and Van Noorden, *Introduction to Immunocytochemistry*, 2nd ed., Springer Verlag, N.Y. (1997), and in Haugland, *Handbook of Fluorescent Probes and Research Chemicals* (1996), which is a combined handbook and catalogue published by Molecular Probes, Inc., Eugene, Oreg. A fluorescent label can be used in FPIA (see, e.g., U.S. Pat. Nos. 5,593,896, 5,573,904, 5,496,925, 5,359,093, and 5,352,803, which are hereby incorporated by reference in their entirety). An acridinium compound can be used as a detectable label in a homogeneous chemiluminescent assay (see, e.g., Adamczyk et al., *Bioorg. Med. Chem. Lett.* 16: 1324-1328 (2006); Adamczyk et al., *Bioorg. Med. Chem. Lett.* 4: 2313-2317 (2004); Adamczyk et al., *Biorg. Med. Chem. Lett.* 14: 3917-3921 (2004); and Adamczyk et al., *Org. Lett.* 5: 3779-3782 (2003)).

[0068] "Mesenchymal" and "mesenchymal phenotype" as used interchangeably herein refer to a type of undifferentiated loose connective tissue that can develop into the tissues of the lymphatic and circulatory systems and connective tissues throughout the body, such as bone and cartilage. Mesenchymal phenotypes may be characterized morphologically by a prominent ground substance matrix containing a loose aggregate of reticular fibrils and unspecialized cells. Mesenchymal cells can migrate easily, in contrast to epithelial cells, which lack mobility and are organized into closely adherent sheets, are polygonal in shape, and are polarized in an apical-basal orientation.

[0069] “Mesenchymal phenotypic CTC” and “mesenchymal CTC” as used interchangeably herein refer to a CTC which has a mesenchymal phenotype.

[0070] The term “normal control” or “healthy control” as used herein means a sample or specimen taken from a subject, or an actual subject who does not have cancer, or is not at risk of developing cancer.

[0071] The term “normal subject” as used herein means a healthy subject, i.e. a subject having no clinical signs or symptoms of cancer. The normal subject is clinically evaluated for otherwise undetected signs or symptoms of cancer, which evaluation may include routine physical examination and/or laboratory testing.

[0072] The term “predetermined cutoff” and “predetermined level” as used herein means an assay cutoff value that is used to assess diagnostic, prognostic, or therapeutic efficacy results by comparing the assay results against the predetermined cutoff/level, where the predetermined cutoff/level already has been linked or associated with various clinical parameters (e.g., presence of disease, stage of disease, severity of disease, progression, non-progression, or improvement of disease, etc.). The disclosure provides exemplary predetermined levels. However, it is well-known that cutoff values may vary depending on the nature of the immunoassay (e.g., antibodies employed, reaction conditions, sample purity, etc.). It further is well within the ordinary skill of one in the art to adapt the disclosure herein for other immunoassays to obtain immunoassay-specific cutoff values for those other immunoassays based on the description provided by this disclosure. Whereas the precise value of the predetermined cutoff/level may vary between assays, the correlations as described herein should be generally applicable.

[0073] “Pretreatment reagent,” e.g., lysis, precipitation and/or solubilization reagent, as used in a diagnostic assay as described herein is one that lyses any cells and/or solubilizes any analyte that is/are present in a test sample. Pretreatment is not necessary for all samples, as described further herein. A pretreatment reagent may be homogeneous (not requiring a separation step) or heterogeneous (requiring a separation step). With use of a heterogeneous pretreatment reagent, there is removal of any precipitated analyte binding proteins from the test sample prior to proceeding to the next step of the assay. The pretreatment reagent optionally can comprise: (a) one or more solvents and salt, (b) one or more solvents, salt and detergent, (c) detergent, (d) detergent and salt, or (e) any reagent or combination of reagents appropriate for cell lysis and/or solubilization of analyte.

[0074] “Prostate cancer” as used herein refers to a type of cancer that develops in the prostate. Prostate cancer may be slow growing or aggressive, in which the cancer cells metastasize from the prostate to other parts of the body, particularly the bones and lymph nodes. “Metastatic prostate cancer” refers to prostate cancer that spreads outside the prostate gland to the lymph nodes, bones, or other areas. “Castration resistant prostate cancer” refers to prostate cancer disease progression despite androgen-deprivation therapy which may present as one or any combination of a continuous rise in serum levels of prostate-specific antigen, progression of pre-existing disease, or appearance of new metastases.

[0075] “Quality control reagents” in the context of immunoassays and kits described herein, include, but are not limited to, calibrators, controls, and sensitivity panels. A

“calibrator” or “standard” typically is used (e.g., one or more, such as a plurality) in order to establish calibration (standard) curves for interpolation of the concentration of an analyte, such as an antibody or an analyte. Alternatively, a single calibrator, which is near a predetermined positive/negative cutoff, can be used. Multiple calibrators (i.e., more than one calibrator or a varying amount of calibrator(s)) can be used in conjunction to comprise a “sensitivity panel.”

[0076] The term “reference activity level” or “reference” as used herein means an activity level of the biomarker in a sample group that serves as a reference against which to assess the activity level in an individual or sample group.

[0077] The term “risk assessment,” “risk classification,” “risk identification,” or “risk stratification” as used herein interchangeably, means an evaluation of factors including biomarkers, to predict the risk of occurrence of future events including disease onset or disease progression, so that treatment decisions regarding the subject may be made on a more informed basis.

[0078] The term “sample,” “test sample,” “specimen,” “biological sample,” “sample from a subject,” or “subject sample” as used herein interchangeably, means a sample or isolate of blood, tissue, urine, serum, plasma, amniotic fluid, cerebrospinal fluid, placental cells or tissue, endothelial cells, leukocytes, or monocytes, can be used directly as obtained from a subject or can be pre-treated, such as by filtration, distillation, extraction, concentration, centrifugation, inactivation of interfering components, addition of reagents, and the like, to modify the character of the sample in some manner as discussed herein or otherwise as is known in the art.

[0079] The term also means any biological material being tested for and/or suspected of containing an analyte of interest. The sample may be any tissue sample taken or derived from the subject. In some embodiments, the sample from the subject may comprise protein. Any cell type, tissue, or bodily fluid may be utilized to obtain a sample. Such cell types, tissues, and fluid may include sections of tissues such as biopsy and autopsy samples, frozen sections taken for histological purposes, blood (such as whole blood), plasma, serum, sputum, stool, tears, mucus, saliva, hair, skin, red blood cells, platelets, interstitial fluid, ocular lens fluid, cerebral spinal fluid, sweat, nasal fluid, synovial fluid, menses, amniotic fluid, semen, etc. Cell types and tissues may also include lymph fluid, ascetic fluid, gynecological fluid, urine, peritoneal fluid, cerebrospinal fluid, a fluid collected by vaginal rinsing, or a fluid collected by vaginal flushing. A tissue or cell type may be provided by removing a sample of cells from an animal, but can also be accomplished by using previously isolated cells (e.g., isolated by another person, at another time, and/or for another purpose). Archival tissues, such as those having treatment or outcome history, may also be used. Protein or nucleotide isolation and/or purification may not be necessary.

[0080] Methods well-known in the art for collecting, handling and processing urine, blood, serum and plasma, and other body fluids, are used in the practice of the present disclosure. The test sample can comprise further moieties in addition to the analyte of interest, such as antibodies, antigens, haptens, hormones, drugs, enzymes, receptors, proteins, peptides, polypeptides, oligonucleotides or polynucleotides. For example, the sample can be a whole blood sample obtained from a subject. It can be necessary or desired that a test sample, particularly whole blood, be

treated prior to immunoassay as described herein, e.g., with a pretreatment reagent. Even in cases where pretreatment is not necessary (e.g., most urine samples, a pre-processed archived sample, etc.), pretreatment of the sample is an option that can be performed for mere convenience (e.g., as part of a protocol on a commercial platform). The sample may be used directly as obtained from the subject or following pretreatment to modify a characteristic of the sample. Pretreatment may include extraction, concentration, inactivation of interfering components, and/or the addition of reagents.

[0081] “Solid phase” refers to any material that is insoluble, or can be made insoluble by a subsequent reaction. The solid phase can be chosen for its intrinsic ability to attract and immobilize a capture agent. Alternatively, the solid phase can have affixed thereto a linking agent that has the ability to attract and immobilize the capture agent. For example, the linking agent can include a charged substance that is oppositely charged with respect to the capture agent itself or to a charged substance conjugated to the capture agent. In general, the linking agent can be any binding partner (preferably specific) that is immobilized on (attached to) the solid phase and that has the ability to immobilize the capture agent through a binding reaction. The linking agent enables the indirect binding of the capture agent to a solid phase material before the performance of the assay or during the performance of the assay. For examples, the solid phase can be plastic, derivatized plastic, magnetic, paramagnetic, or non-magnetic metal, glass or silicon, including, for example, a test tube, microtiter well, sheet, bead, microparticle, chip, and other configurations known to those of ordinary skill in the art.

[0082] “Specific binding” or “specifically binding” as used herein may refer to the interaction of an antibody, a protein, or a peptide with a second chemical species, wherein the interaction is dependent upon the presence of a particular structure (e.g., an antigenic determinant or epitope) on the chemical species; for example, an antibody recognizes and binds to a specific protein structure rather than to proteins generally. If an antibody is specific for epitope “A”, the presence of a molecule containing epitope A (or free, unlabeled A), in a reaction containing labeled “A” and the antibody, will reduce the amount of labeled A bound to the antibody.

[0083] The term “subject”, “patient” or “subject in the method” as used herein interchangeably, means any vertebrate, including, but not limited to, a mammal (e.g., cow, pig, camel, llama, horse, goat, rabbit, sheep, hamsters, guinea pig, cat, dog, rat, and mouse, a non-human primate (for example, a monkey, such as a cynomolgous or rhesus monkey, chimpanzee, etc.) and a human. In some embodiments, the subject or subject may be a human or a non-human. In some embodiments, the subject may be a human subject at risk for developing or already having cancer.

[0084] “Treat”, “treating” or “treatment” are each used interchangeably herein to describe reversing, alleviating, or inhibiting the progress of a disease, such as cancer, or one or more symptoms of such disease, to which such term applies. Depending on the condition of the subject, the term also refers to preventing a disease, and includes preventing the onset of a disease, or preventing the symptoms associated with a disease. A treatment may be either performed in an acute or chronic way. The term also refers to reducing the severity of a disease or symptoms associated with such

disease prior to affliction with the disease. Such prevention or reduction of the severity of a disease prior to affliction refers to administration of an antibody or pharmaceutical composition of the present invention to a subject that is not at the time of administration afflicted with the disease. “Preventing” also refers to preventing the recurrence of a disease or of one or more symptoms associated with such disease. “Treatment” and “therapeutically,” refer to the act of treating, as “treating” is defined above.

[0085] As described herein the ability for a cell to transition easily between epithelial-like and mesenchymal-like states (phenotypic plasticity) is a relevant determinant of malignant fitness more so than the properties of the end states. While these epithelial transitions are phenotypic, the propensity to transition (plasticity) among carcinoma cells may be determined by genotype. The majority of plastic cells may inhabit transitional intermediate states with properties of both epithelium and mesenchyme, and that these transitional cells may be particularly malignant. Such cells may be detected in: (1) tumors where the cancer cells have mixed histology, which indeed have been observed and have been classified as highly aggressive (e.g., clonal sarcomatous carcinomas of epithelial origin, which exhibit an extremely aggressive behavior, such as sarcomatoid renal cell carcinoma and carcinosarcoma of the prostate); and (2) cancer cells co-expressing epithelial and mesenchymal markers, as described herein.

[0086] The disclosure, as illustrated by the non-limiting embodiments in the Examples, provides for identification of cells that possess an intermediate phenotype—expressing epithelial and mesenchymal isoforms of FGFR2, having epithelial-like morphology and gene expression patterns, while also displaying mesenchymal cell-like migration, tumor formation, and metastases. In embodiments, these cells are identified in patients with advanced cancer, metastatic adenocarcinoma, and metastatic breast and prostate carcinomas. In some embodiments, the cells comprise CTCs. In some embodiments the CTCs co-expresses biomarkers including, for example, EpCAM, cytokeratin, and vimentin, which identify cells as both epithelial- and mesenchymal-like. In some embodiments, these CTCs in intermediate phenotypic states are identified by detecting EMT biomarkers and provide a diagnosis and/or prognosis of the state and/or degree of malignancy of a cancer.

[0087] In an aspect the disclosure provides a method for detecting CTCs in a biological sample, the method comprising detecting at least one EMT biomarker in the biological sample. In some embodiments such as illustrated in the Examples, biomarkers of EMT are present on the CTCs of patients with common epithelial malignancies. In some embodiments methods that include detection and identification of alternative splice variants of the FGFR2 gene are used to correlate to metastatic propensity and epithelial phenotypic in a CTC.

[0088] Thus, EMT biomarker expression may be used to detect CTCs. EMT biomarker expression, or detection of CTCs, or a combination thereof, may be used to assess cancer prognosis, tumor invasiveness, risk of metastasis, or to stage tumors. As mentioned above, the methods described herein can include any suitable method for evaluating EMT biomarker expression including, but not limited to, detection with antibodies, real time RT-PCR, Northern analysis, magnetic particles (e.g., microparticles or nanoparticles), Western analysis, and any method or system involving flow

cytometry. In some embodiments, the methods and EMT biomarkers can be used in a commercially available system such as a system that has been approved by a regulatory agency (e.g., FDA) including, for example, CELLSEARCH® technology (Veridex LLC). Thus, the methods can incorporate standard protocols that are known in the art. For example, embodiments comprising CELLSEARCH® technology can include detecting the presence of an EMT biomarker, and correlated to quantifying the number of circulating tumor cells (CTCs) in a biological sample, (e.g., blood collected from women in need of a new treatment regimen for metastatic breast cancer, or men in need of treatment for mCRPC). Typical protocols can include drawing blood sample sizes of about 15 mL that can be collected at any particular time (suitably when the patient starts the new therapy, and then again at three to four week intervals). The number of CTCs can be correlated with disease response or progression as determined by standard radiology studies (e.g., CT scans) performed every nine to 12 weeks.

[0089] In an aspect, the disclosure relates to a method for detecting a circulating tumor cell (CTC) in a biological sample, wherein the method comprises detecting at least one EMT biomarker in the biological sample. As noted above, a biological sample can be from any tissue or fluid from an organism. In some embodiments the biological sample is from a bodily fluid or tissue that is part of, or associated with, the lymphatic system or the circulatory system of the organism. In some embodiments the biological sample is a blood sample.

[0090] The EMT and cellular plasticity biomarkers used in the methods described herein are associated with circulating tumor cells (CTCs). Accordingly, in various embodiments the methods include detecting the presence of one or more EMT biomarker and correlating that detection with the presence of a CTC, optionally quantifying the number of CTCs in the sample. As discussed herein, EMT biomarkers can include any detectable biomolecule that is associated with a transitional cell that exhibits characteristics (e.g., phenotype, or surface antigen or gene expression profiles, etc.) of plasticity, stem-like properties, invasiveness, and/or chemo-resistance of a cell. In some non-limiting embodiments, the EMT biomarker includes any of vimentin, N-cadherin, O-cadherin, E-cadherin, FGFR2 splice variant isoforms (such as, for example FGFR2 that includes or excludes either exon IIIc or exon IIIb), or CD133, or any combination of two or more thereof. In some embodiments, the EMT biomarker can include one or more of vimentin (polypeptide SEQ ID NO: 14 encoded by polynucleotide SEQ ID NO: 13), N-cadherin (polypeptide SEQ ID NO: 2 encoded by polynucleotide SEQ ID NO: 1; polypeptide SEQ ID NO: 16 encoded by polynucleotide SEQ ID NO: 15), O-cadherin (polypeptide SEQ ID NO: 4 encoded by polynucleotide SEQ ID NO: 3; polypeptide SEQ ID NO: 18 encoded by polynucleotide SEQ ID NO: 17), E-cadherin (polypeptide SEQ ID NO: 12 encoded by polynucleotide SEQ ID NO: 11; polypeptide SEQ ID NO: 24 encoded by polynucleotide SEQ ID NO: 23), FGFR2 (polypeptide SEQ ID NO: 8 encoded by polynucleotide SEQ ID NO: 7; polypeptide SEQ ID NO: 10 encoded by polynucleotide SEQ ID NO: 9; polypeptide SEQ ID NO: 22 encoded by polynucleotide SEQ ID NO: 21), and CD133 (polypeptide SEQ ID NO: 6 encoded by polynucleotide SEQ ID NO: 5; polypeptide SEQ ID NO: 20 encoded by polynucleotide SEQ ID NO: 19). In some embodiments, the EMT bio-

marker can include one or more of N-cadherin, for example human N-cadherin (for example SEQ ID NO: 16, CCDS ID No: CCDS11891.1); O-cadherin, for example human O-cadherin (for example SEQ ID NO: 18, CCDS ID No: CCDS10803.0); E-cadherin, for example human E-cadherin (for example SEQ ID NO: 24, CCDS ID No: CCDS10869.1); CD133, for example human CD133 (for example SEQ ID NO: 20, CCDS ID No: CCDS47029.1); FGFR2, for example human FGFR2 (for example SEQ ID NO: 22, CCDS ID No: CCDS31298.1); and vimentin, for example human vimentin (for example SEQ ID NO: 14, Accession No. BC000163). It will be understood by one of skill in the art that when reference is made to polynucleotides that encode polypeptides in the above embodiments as well as embodiments throughout, the polynucleotide can be disclosed as either an RNA (e.g., mRNA) or a DNA (e.g., cDNA).

[0091] The EMT biomarkers can be associated with any organism (ortholog) and in certain embodiments are EMT biomarkers associated with a human. Any portion or the entirety of an EMT biomarker can be used for detecting in the methods described herein such as, for example, an epitope of an EMT biomarker protein that binds to an antibody, or a nucleic acid sequence of an EMT biomarker an expressed or transcribed mRNA molecule that is complementary to a reporter nucleic acid probe or primer. In some embodiments, the methods provide for detecting expression of at least two EMT biomarkers. In certain embodiments, expression of vimentin and E-cadherin are detected. In certain embodiments, expression of N-cadherin and O-cadherin are detected. This measure may be used alone or in combination with another method to detect CTCs. In certain embodiments, the methods described herein may be used as a supplemental method in conjunction with CELLSEARCH® Circulating Tumor Cell Test (noted above). Thus, embodiments provide for a method as part of a dual or complementary detection system that can be used to detect and optionally quantify CTCs in a sample (e.g., comprising the detection of EpCAM and at least one EMT biomarker). The expression of at least one EMT biomarker may be used to isolate CTCs. The expression of at least one EMT biomarker may be used to count or provide a relative number or amount of CTCs, using any known method for correlating detection of a biomarker to a cell, such as a CTC. CTCs may be detected at the time of, prior to, or after metastasis.

[0092] Cancers may include, but are not limited to, breast cancer, colon cancer, lung cancer, prostate cancer, testicular cancer, brain cancer, skin cancer, rectal cancer, gastric cancer, esophageal cancer, sarcomas, tracheal cancer, head and neck cancer, pancreatic cancer, liver cancer, ovarian cancer, lymphoid cancer, cervical cancer, vulvar cancer, melanoma, mesothelioma, renal cancer, bladder cancer, thyroid cancer, bone cancers, carcinomas, sarcomas, and soft tissue cancers. Thus, the disclosure is generally applicable to any type of cancer in which expression of an EMT biomarker occurs. In certain embodiments, the cancer is a solid tumor malignancy. In certain embodiments, the cancer is breast, colon, or prostate cancer.

[0093] Expression of at least one EMT biomarker may be detected using any suitable method known in the art, including but not limited to, binding with antibodies or fragment thereof, antibodies tethered to or associated with an imaging agent, expression reporter plasmids, flow cytometry, and any

suitable array scanner technology. The antibody or fragment thereof may suitably recognize a particular intracellular protein, protein isoform, or protein configuration.

[0094] As used herein, an “imaging agent” or “reporter molecule” is any entity which enhances visualization or detection of the cell to which it is delivered. Any type of detectable reporter molecule/imaging agent can be used in the methods disclosed herein for the detection of one or more EMT biomarker. Such detectable molecules are known in the art and include, for example, magnetic beads, fluorophores, radionuclides, nuclear stains (e.g., DAPI). For example, an imaging agent can include a compound that comprises an unstable isotope (i.e., a radionuclide) or a fluorescent moiety, such as Cy-5, Alexa 647, Alexa 555, Alexa 488, fluorescein, rhodamine, and the like. Suitable radionuclides include both alpha- and beta-emitters. In some embodiments, the targeting vehicle is labeled. In other embodiments, suitable radioactive moieties include labeled polynucleotides and polypeptides which can be coupled to the targeting vehicle. In some embodiments, the imaging agent comprises a radionuclide such as, for example, a radionuclide that emits low-energy electrons (e.g., those that emit photons with energies as low as 20 keV). Such nuclides can irradiate the cell to which they are delivered without irradiating surrounding cells or tissues. Non-limiting examples of radionuclides that can be delivered to cells include ^{137}Cs , ^{103}Pd , ^{111}In , ^{125}I , ^{211}At , ^{212}Bi and ^{213}Bi , among others known in the art. Further imaging agents suitable for delivery to a cell in accordance with some embodiments include paramagnetic species for use in MRI imaging, echogenic entities for use in ultrasound imaging, fluorescent entities for use in fluorescence imaging (including quantum dots), and light-active entities for use in optical imaging. A suitable species for Mill imaging is a gadolinium complex of diethylenetriamine pentacetic acid (DTPA). For positron emission tomography (PET), ^{18}F or ^{11}C may be delivered. Other non-limiting examples of reporter molecules are discussed throughout the disclosure.

[0095] In an aspect, the disclosure provides a kit for detecting CTCs in a sample. In embodiments, the kit comprises an antibody to at least one EMT biomarker. The antibody in the kit can be connected to or associated with an imaging agent. In embodiments, the kit can comprise an antibody to at least one EMT biomarker, wherein the antibody is associated a magnetic bead. The magnetic bead may be used for ferromagnetic separation and enrichment of CTCs.

[0096] Aspects also relate to methods of predicting responsiveness of a subject to a cancer drug. The methods may comprise determining the level of expression of at least one EMT biomarker in a sample from the subject. The level of expression of at least one EMT biomarker may be used to obtain a gene expression pattern in CTCs for the subject. The methods may further comprise predicting responsiveness of the subject to the cancer drug based on the gene expression pattern obtained. Genome variation in CTCs from the subject may also be determined.

[0097] Also provided are methods of providing a cancer prognosis to a subject. The methods may comprise determining the level of expression of at least one EMT biomarker in a sample from the subject. The level of expression of at least one EMT biomarker may be used to determine the number of CTCs in the sample. The CTCs may be captured using at least one EMT biomarker. The level of expression

of at least one EMT biomarker may be used to determine a gene expression pattern in the CTCs for the subject. A prognosis may be provided to the subject based on the gene expression pattern obtained.

[0098] Also provided are methods for following the progress of cancer in a subject. The methods may comprise determining the level of expression of at least one EMT biomarker in samples from the subject at a first and a second time, and comparing the first and second levels of expression. The level of expression of at least one EMT biomarker in the sample may be determined over time, such as following initiation of a new cancer therapy. The level of expression of at least one EMT biomarker in the sample may be used to determine the number or amount of CTCs. An increase between the first and second levels may indicate progression of the cancer. A decrease between the first and second levels may indicate remission or response of the cancer to the therapy. No difference between the first and second levels may indicate arrest or stability in the progression of the cancer.

[0099] Also provided are methods of screening for cancer in a subject. The methods may comprise determining the level of expression of at least one EMT biomarker in a sample from the subject. The level of expression of at least one EMT biomarker may be used to determine the amount or number of CTCs in the subject. The level of expression of at least one EMT biomarker may be compared to a normal or control sample. An increased level of at least one EMT biomarker may indicate presence of cancer in the subject.

[0100] Also provided are methods of arresting cell growth or inducing cell death of a cancer cell expressing an EMT biomarker. The methods include contacting the cancer cell with a conjugate capable of mediating intracellular delivery of an agent, such as the antibodies to EMT markers described herein. The agent is capable of arresting or attenuating the growth of the cell or inducing cell death through any mechanism after agent internalization. The cancer cell may be contacted with the conjugate in vitro, in vivo, or ex vivo. These methods may be useful in treating cancer by directly targeting cancer cells expressing an EMT biomarker for delivery of agents capable of decreasing or arresting cell growth or inducing cell death.

[0101] The disclosure also provides for targeted therapeutic methods and molecules that comprise an anti-cancer agent linked to a binding agent that targets at least one EMT as described herein. In some embodiments the link between the anti-cancer agent and the binding agent is a covalent bond. In some embodiments the link is formed by strong electrostatic interactions (hydrogen bonds, hydrophilic/hydrophobic interaction, or oppositely charged moieties, and the like). Any anti-cancer agent can be used in such molecules and therapeutic methods, and can be selected by one of skill in the art based on the type of cancer to be treated, the progress/stage of the cancer, potential adverse drug interactions, dosage requirements, administration schedule, and the like.

[0102] Method of Isolating, Capturing, or Enriching Circulating Tumor Cells

[0103] With the importance of EP in metastasis, the disclosure identifies CTCs that have lost their epithelial phenotype in patients, such as patients with advanced prostate cancer, using a mesenchymal-based capture method. Such a

method may complement or replace existing epithelial-based approaches by capturing cells that have reduced or absent EpCAM expression.

[0104] The present disclosure is directed to methods of isolating, capturing or enriching CTCs using any EMT biomarker, as described above. The method includes obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; and separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample, thereby isolating, capturing, or enriching the circulating tumor cell. The EMT biomarker may include OB-cadherin, N-cadherin, vimentin, E-cadherin, FGFR2 splice variant isoforms, or CD133.

[0105] a) OB-Cadherin

[0106] OB-cadherin, also known as cadherin-11 and O-cadherin, is encoded by the CDH11 gene and was first identified in mouse osteoblasts. OB-cadherin is a homophilic cell adhesion molecule that mediates osteoblast adhesion during bone development. Aberrant OB-cadherin expression has been recognized in breast, gastric, and prostate cancers. OB-cadherin may be involved in prostate cancer metastasis. OB-cadherin expression was shown in CTCs (see FIGS. 1A-1D).

[0107] In human prostate cancer, examination by immunohistochemistry showed increased expression of OB-cadherin in bone metastases compared with the primary tumor. Although the exact mechanism of how OB-cadherin expression facilitates bone metastasis is unclear, OB-cadherin is known to mediate adhesion between PC cells and osteoblasts. Preclinically, in PC-3 cells, when OB-cadherin is silenced by shRNA and the silenced cells are injected into mice, fewer bone metastases develop, while metastases in other organs was unaffected. In addition, induction of OB-cadherin expression has been linked to EMT/EP biology in other model systems. Furthermore, androgen depletion leads to OB-cadherin upregulation, suggesting a role for OB-cadherin in castration-resistant disease progression. Finally, given that lethal metastatic prostate cancer invariably spreads to bone in the vast majority of men, OB-cadherin positive CTCs may be detectable in the blood of men with metastatic CRPC.

[0108] Antibodies directed against OB-cadherin were attached to iron particles to form a novel ferrofluid that may replace or complement EpCAM ferrofluid in the FDA-approved CELLSEARCH® technology. After OB-cadherin expressing cells are enriched from whole blood immunomagnetically, additional characterization steps follow to ensure that the captured cells are the cells of interest, such as measuring levels of an EMT-independent characterization protein, such as β -catenin. The antibody may bind to the extracellular domain of OB-cadherin.

[0109] b) N-Cadherin

[0110] N-cadherin, also known as neural cadherin (NCAD) and cadherin-2 (CDH2) is encoded by the CDH2 gene. N-Cadherin is commonly found in cancer cells and

provides a mechanism for transendothelial migration. When a cancer cell adheres to the endothelial cells of a blood vessel, it up-regulates the src kinase pathway, which phosphorylates beta-catenins attached to both N-cadherin and E-cadherins. This causes the intercellular connection between two adjacent endothelial cells to fail and allows the cancer cell to slip through. The antibody may bind to the extracellular domain of N-cadherin.

[0111] Methods of Confirming Circulating Tumor Cells

[0112] Given that CTCs are extraordinarily rare relative to other circulating cells, the isolation of CTCs involves the identification and exclusion of cells expressing the pan-leukocyte marker CD45. Circulating CD45 negative cells are not necessarily tumor-derived, however, but instead may represent normal blood vessel or stromal cells, circulating mesenchymal cells or stem cells, or other host cells that exist in rare quantities in the circulation. Circulating endothelial cells result from blood vessel wall turnover, and bone marrow-derived endothelial progenitor cells may circulate in the setting of neovascularization of ischemic tissue and tumor formation. These cells are all CD45 negative and CD31 positive. Also CD45 negative but CD31 negative, mesenchymal stromal cells (MSCs) are a more diverse group of cells that may be bone marrow-, peripheral blood-, or fat-derived. MSCs are multipotent cells that may differentiate into a variety of stromal cell types, circulate in inflammatory disorders, and are under active investigation for use in regenerative medicine and other conditions. The significance of circulating MSCs in cancer remains unclear. Thus, any CTC detection method involves distinguishing tumor cells from a range of other rare non-tumor cells in the circulation that may express non-epithelial biomarkers. Confirmation of CTCs may include detecting β -catenin, CD31, CD45, cytokeratin, and/or PSA, staining with DAPI, and/or detecting a prostate cancer-specific genomic event. For example, a CTC may be confirmed if DAPI staining is positive, β -catenin expression is positive, CD45 expression is negative, and CD31 expression is negative. The detection of β -catenin, CD31, CD45, cytokeratin, and/or PSA may be performed using antibodies against β -catenin, CD31, CD45, cytokeratin, and/or PSA, wherein the antibodies are labeled.

[0113] a) β -Catenin

[0114] β -catenin, encoded by the CTNNB1 gene, has multiple functions in cancer cells, including cadherin-mediated cell adhesion and involvement in the Wnt signaling pathway. When Wnt ligands are absent, β -catenin is phosphorylated and degraded. When Wnt ligands are present, β -catenin moves to the nucleus and activates target genes linked to EMT/invasion, proliferation, and survival in multiple cancers. In prostate cancer specifically, β -catenin may act as cofactor with the androgen receptor and increased expression and change in localization has been observed in advanced disease. β -catenin expression may be a constant finding regardless of the epithelial or mesenchymal phenotypic nature of a CTC. Using β -catenin expression to identify CTC does not have the epithelial bias associated with using cytokeratin. An antibody that binds to β -catenin may be used to detect β -catenin.

[0115] b) CD31

[0116] Cluster of differentiation 31 (CD31), also known as platelet endothelial cell adhesion molecule (PECAM-1), is encoded by the PECAM1 gene and plays a role in removing aged neutrophils from the body. CD31 is found on the surface of platelets, monocytes, neutrophils, and some types

of T-cells, and makes up a large portion of endothelial cell intercellular junctions. CD31 is normally found on endothelial cells and used in immunohistochemistry to demonstrate the presence of endothelial cells in histological tissue sections. An antibody that binds to CD31 may be used to detect CD31.

[0117] c) CD45

[0118] Cluster of differentiation 45 (CD45), also known as protein tyrosine phosphatase, receptor type, C and leukocyte common antigen, is encoded by the PTPRC gene. CD45 is used to identify leukocytes. An antibody that binds to CD45 may be used to detect CD45.

[0119] d) Cytokeratin

[0120] Cytokeratins are keratin-containing intermediate filaments found in the intracytoplasmic cytoskeleton of epithelial tissue. Cytokeratin-expressing cancer cells lose their cytokeratin expression after undergoing epithelial-mesenchymal transition, with up to 20% of cells having no detectable cytokeratin. A protein other than cytokeratin may identify a pure mesenchymal CTC.

[0121] e) PSA

[0122] Prostate-specific antigen (PSA), also known as gamma-seminoprotein or kallikrein-3 (KLK3), is a glycoprotein enzyme encoded in humans by the KLK3 gene. PSA is a member of the kallikrein-related peptidase family and is secreted by the epithelial cells of the prostate gland. PSA is present in small quantities in the serum of men with healthy prostates, but is often elevated in the presence of prostate cancer or other prostate disorders.

[0123] f) DAPI

[0124] DAPI, also known as 4',6-diamidino-2-phenylindole, is a fluorescent stain that binds strongly to A-T rich regions in DNA. It is used extensively in fluorescence microscopy. DAPI can pass through an intact cell membrane therefore it can be used to stain both live and fixed cells.

[0125] g) Prostate Cancer-Specific Genomic Event

[0126] A CTC may be confirmed by detecting the presence or absence of at least one prostate cancer-specific genomic event. The at least one prostate cancer-specific genomic event may be androgen receptor amplification, phosphatase and tensin homolog (PTEN) loss, gene fusion of transmembrane protease, serine 2 (TMPRSS2) gene and ETS related (ERG) gene, and combinations thereof. For example, the detection of the presence of androgen receptor amplification, phosphatase and tensin homolog (PTEN) loss, gene fusion of transmembrane protease, serine 2 (TMPRSS2) gene and ETS related (ERG) gene may indicate and/or confirm that the CTC is a prostate cancer cell. The detection of a prostate cancer-specific genomic event may be performed using fluorescent in situ hybridization.

[0127] Method of Detecting or Identifying Circulating Tumor Cells

[0128] The present disclosure is directed to methods of detecting or identifying CTCs using EMT biomarkers, as described above. The method includes obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; and separating the solid phase-capture

binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample, thereby detecting or identifying the circulating tumor cell.

[0129] Methods for Isolating or Capturing an Intact Cell from a Patient

[0130] The present disclosure is directed to methods of isolating or capturing an intact cell from a patient, wherein the intact cell is β -catenin positive, DAPI positive, and CD45 negative, using EMT biomarkers, as described above. The method includes obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; and separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample, thereby isolating or capturing the intact cell. The intact cell may be a mesenchymal phenotypic cell, such as mesenchymal CTC.

[0131] Method of Detecting Cancer in a Subject

[0132] The present disclosure is directed to methods of detecting cancer in a subject using EMT biomarkers, as described above. The method includes obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample, thereby isolating or capturing the intact cell; determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex; comparing the level of circulating tumor cell to a reference level of circulating tumor cell; and detecting cancer in the subject if the level of circulating tumor cell is higher than the reference level of circulating tumor cell, cancer is detected in the subject. The cancer may be at least one cancer, as described above. The method may further include administering a therapy against cancer to the subject identified as having cancer.

[0133] Method of Monitoring Progression of Cancer in a Subject Undergoing Therapeutic Treatment

[0134] The present disclosure is directed to methods of monitoring progression of cancer in a subject undergoing therapeutic treatment using EMT biomarkers, as described above. The method includes obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-

capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample, thereby isolating or capturing the intact cell; determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex; and correlating the level of circulating tumor cell with the progression of cancer in the subject, wherein if the level of the circulating tumor cell is higher as compared to the level of the circulating tumor cell in an earlier biological sample from the subject, the subject is identified as having progression of cancer. The method may further include administering a therapy against cancer to the subject identified as having cancer.

[0135] Method of Determining a Cancer Prognosis in a Subject

[0136] The present disclosure is directed to methods of determining a cancer prognosis in a subject using EMT biomarkers, as described above. The method includes obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample, thereby isolating or capturing the intact cell; determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex; comparing the level of circulating tumor cell to a reference level of circulating tumor cell; and determining the cancer prognosis in the subject, wherein if the level of circulating tumor cell is higher than the reference level of circulating tumor cell, the subject is identified as having cancer. The method may further include administering a therapy against cancer to the subject identified as having cancer.

[0137] Method of Predicting Responsiveness of a Subject Having Cancer to a Course of Treatment

[0138] The present disclosure is directed to methods of predicting responsiveness of a subject having cancer to a course of treatment using EMT biomarkers, as described above. The method includes obtaining a biological sample from a patient; obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex; contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one EMT biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein

complexes by application of an external magnetic field on the sample, thereby isolating or capturing the intact cell; determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex; and comparing the level of circulating tumor cell to a reference level of circulating tumor cell.

[0139] Treatment

[0140] The subject identified in the methods described above as having a level of circulating tumor cell higher than or equal to a reference level is identified as a patient having cancer. The subject is then treated for the cancer.

[0141] a) Prostate Cancer

[0142] The subject identified in the methods described above having levels of circulating tumor cell greater than or equal to a reference level is identified as a patient having prostate cancer. The subject is then treated for the prostate cancer. Treatments may include watchful waiting or active surveillance, surgery, cryosurgery, high-intensity focused ultrasound, radiation, hormone therapy, chemotherapy, and targeted therapy. Examples of surgery include pelvic lymphadenectomy, radical prostatectomy (retropubic prostatectomy and perineal prostatectomy), transurethral resection of the prostate, and laparoscopic prostatectomy. Examples of radiation therapy include high-energy x-rays (external and internal radiation therapy), proton beam radiation, and intensity-modulated radiation therapy. Examples of hormone therapy include luteinizing hormone-releasing hormone agonists, such as leuprolide (Lupron, Eligard), goserelin, triptorelin (Trelstar), histrelin (Vantas), and buserelin, anti-androgens, such as flutamide, bicalutamide (Casodex), and nilutamide (Nilandron), ketoconazole, orchiectomy, estrogen, and aminoglutethimide. Examples of chemotherapeutic drugs include abiraterone acetate (Zytiga), cabazitaxel, degarelix, docetaxel, enzalutamide (Xtandi), cabazitaxel (Jevtana), leuprolide acetate (Lupron, Lupron Depot, Lupron Depot-3 Month, Lupron Depot-4 Month, Lupron Depot-Ped, Viadur), prednisone, sipuleucel-T (Provenge), estramustine (Emcyt), mitoxantrone (Novantrone), vinorelbine (Navelbine), paclitaxel (Taxol), cyclophosphamide (Cytoxan), etoposide (VP-16), G-1 (a GPR30 agonist/stimulator), and docetaxel (Taxotere).

[0143] b) Breast Cancer

[0144] The subject identified in the methods described above having levels of circulating tumor cell greater than or equal to a reference level is identified as a patient having breast cancer. The subject is then treated for the breast cancer. Treatment may include surgery, radiation therapy, bone-directed therapy, chemotherapy, hormone therapy, and targeted therapy. Examples of surgery include lumpectomy, quadrantectomy, mastectomy, such as simple mastectomy, skin-sparing mastectomy, modified radical mastectomy, prophylactic mastectomy, and radical mastectomy, prophylactic ovary removal, cryotherapy, and lymph node surgery, such as axillary lymph node dissection and sentinel lymph node biopsy. Examples of radiation therapy include external beam radiation, such as accelerated breast irradiation and 3D-conformal radiotherapy, and brachytherapy (internal radiation), such as interstitial brachytherapy, intracavitary brachytherapy, and intraoperative radiation. Examples of bone-directed therapy include bisphosphonates and denosumab. Examples of chemotherapy include anthracyclines, such as doxorubicin (Adriamycin, Doxil), epirubicin (Elevance), and daunorubicin (Cerubidine, DaunoXome), capecitabine (Xeloda), carboplatin (Paraplatin), cisplatin,

cyclophosphamide (Cytoxan), eribulin (Halaven), fluorouracil (also called 5-fluorouracil or 5-FU; Adrucil), gemcitabine (Gemzar), ixabepilone (Ixempra), methotrexate (Ame-thopterin, Mexate, Folex), mitoxantrone (Novantrone), mutamycin (Mitomycin), taxanes, such as paclitaxel (Taxol, Abraxane), and docetaxel (Taxotere), thiotepa (Thioplex), vincristine (Oncovin, Vincasar PES, Vincrex), and vinorelbine (Navelbine). Examples of hormone therapy include aromatase inhibitors, such as anastrozole (Arimidex), exemestane (Aromasin), and letrozole (Femara), selective estrogen receptor modulators (SERMs), such as tamoxifen (Nolvadex), raloxifene (Evista), and toremifene (Fareston), and estrogen-receptor downregulators, such as fulvestrant (Faslodex). Examples of targeted therapy include trastuzumab (Herceptin), lapatinib (Tykerb), bevacizumab (Avastin), pertuzumab (Perjeta), and everolimus (Afinitor).

[0145] Kits

[0146] Provided herein is a kit, which may be used for isolating, capturing or enriching CTCs from a sample. The kit comprises at least one component for isolating, capturing or enriching CTCs from a sample and instructions for isolating, capturing or enriching CTCs from a sample. Instructions included in kits can be affixed to packaging material or can be included as a package insert. While the instructions are typically written or printed materials they are not limited to such. Any medium capable of storing such instructions and communicating them to an end user is contemplated by this disclosure. Such media include, but are not limited to, electronic storage media (e.g., magnetic discs, tapes, cartridges, chips), optical media (e.g., CD ROM), and the like. As used herein, the term "instructions" can include the address of an internet site that provides the instructions.

[0147] The at least one component may include at least one antibody that specifically binds to at least one EMT biomarker. The antibody may be an EMT biomarker capture antibody. The antibody may include an antibody against OB-cadherin or N-cadherin. The antibody may be associated with a solid support. The solid support may be a magnetic bead. The magnetic bead may be used for ferromagnetic separation and enrichment of CTCs.

[0148] The kit may also comprise at least one staining reagent. The at least one staining reagent may comprise an antibody which is detectably labeled. The antibody may include an antibody against a biomarker which is used to confirm the CTC. The biomarker may be β -catenin, CD45 or CD31. The at least one staining reagent may include of phycoerytherin-labeled anti- β -catenin antibody and an allophycocyanin-labeled anti-CD45 antibody.

[0149] The kit may also comprise a calibrator or control and/or at least one container for conducting the isolation, capturing, or enrichment of CTCs. The kit may comprise all components, i.e., reagents, standards, buffers, diluents, etc., which are necessary to perform the isolation, capturing, or enrichment of CTCs.

[0150] Unless otherwise defined herein, scientific and technical terms used in connection with the present disclosure shall have the meanings that are commonly understood by those of ordinary skill in the art. For example, any nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those that are well known and commonly used in the art. The meaning and scope of the terms should be clear; in the event however of

any latent ambiguity, definitions provided herein take precedent over any dictionary or extrinsic definition. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular.

[0151] For the recitation of numeric ranges herein, each intervening number there between with the same degree of precision is explicitly contemplated. For example, for the range of 6-9, the numbers 7 and 8 are contemplated in addition to 6 and 9, and for the range 6.0-7.0, the number 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 7.0 are explicitly contemplated.

EXAMPLES

Example 1. Materials and Methods

[0152] Plasmids and cell culture. The minigene used (pRIIICl²) was previously described (S. Oltean et al., *Proc Natl Acad Sci USA* 2006, 103, 14116, incorporated herein by reference in its entirety). All cell lines were cultured in low glucose DMEM (Invitrogen) with 10% FBS and 15 IJg/mL blasticidin. Single cell progenies were isolated from a population of AT3 cells stably transfected with pRIIICl² minigene by limiting dilution to produce a concentration of 1 cell/10 wells and plated on 96-well plates. Cells were counted using a hemocytometer to obtain an initial concentration of 1×10^5 cells/mL. Through a series of progressive dilutions a final concentration of 1 cell/mL was obtained and 100 IJI were pipetted in each well of three 96-well plates. All wells were monitored through bright field microscopy, those appearing to contain more than one cell were excluded, and those containing single cells were further cultured into 25 mL flasks. 16 of an expected 27 clones were obtained using this procedure in a first round.

[0153] To measure cell population growth rate in vitro, cells were plated at 50,000/well in 6-well dishes. Viable cells were counted using Trypan Blue staining at 24, 48, 72, and 96 h. 100.1551 Animals and tumor cell implantation. Cells were trypsinized, washed, and resuspended in PBS at a final concentration of 3×10^5 cells/mL, and kept on ice for less than 30 minutes before implantation. Cells (3×10^5) were injected subcutis in both flanks of Copenhagen 2331 rats (Harlan Labs, Indianapolis, Ind.; 75-90 g, 2 months of age). Animals were continuously monitored for tumor growth. All animal procedures were approved by the Duke University Institutional and Animal Care and Use committee and followed NIH guidelines. Sacrifice curves were compared using a Mantel-Haenszel logrank test. Tumor volume was compared using an unpaired t test. Prism 4.0c for the Macintosh (Graphpad, La Jolla, Calif.) was used for statistical analyses.

[0154] Histological sections and analysis. Excised tumors and lungs were washed in PBS at room temperature. Depending on the size of the lungs, they were frozen either together or separately. The tumor sections and the lungs were placed in cryomolds, embedded in optimal-cutting-temperature tissue sectioning medium (Sakura Finetek, Torrance, Calif.), snap-frozen in liquid nitrogen, and stored at 80° C. Slides for fluorescence imaging were prepared as follows: the tissue was incubated for 2-3 h at -20° C. to equilibrate the temperature and then sectioned with a microtome. The sections (15 μ m) were placed on glass slides, fixed in 4% (wt/vol) paraformaldehyde for 30 min at room temperature, and rinsed in PBS at room temperature. The

slides were mounted with gel/mount media (Biomed, Foster City, Calif.). The sections were analyzed by using an Olympus (Melville, N.Y.) IX 71 epifluorescence microscope, and images were acquired by using an Olympus DP70 digital camera. Image processing was done with DP Controller software (Olympus). For hematoxylin-eosin staining after fluorescence imaging, the slides were incubated in warm water for 15-20 minutes for the cover slip to come off, slides were dried, and staining was performed according to standard procedure.

[0155] RNA extraction from tumor sections. Sections were fixed in 4% (wt/vol) paraformaldehyde for 5 minutes, rinsed in PBS, and imaged. DsRED+ and DsRED- regions of the sections were marked on the slide. The slide was immersed in warm water for 5 minutes to remove the coverslip and the DsRED+ and DsRED- regions scraped off. RNA isolation was further performed as described before (N. Masuda, T. Ohnishi, S. Kawamoto, M. Monden, K. Okubo, *Nucleic Acids Res* 1999, 27, 4436, incorporated herein by reference in its entirety). Briefly, samples were treated with proteinase K in digestion buffer containing SDS, and further isolation of RNA was performed using the RNeasy kit (QIAGEN, Valencia, Calif.).

[0156] Immunoblots. Cells were collected from confluent 25 cm² tissues flasks by scraping, washed in PBS, and lysed in sample buffer. Whole cell lysates were serially diluted in sample buffer, fractionated via 7.5% SDS-PAGE, and transferred to PVDF. Membranes were cut in half. The bottom half was probed with anti- β -actin at 1:1000 or 1:5000 (Santa Cruz Biotechnology, CA, 47778) as an internal loading control, while the top half was probed with anti-CD 133 (Santa Cruz Biotechnology, CA, 30219) at 1:200 or anti-CD44 (Santa Cruz Biotechnology, CA, 7946) at 1:200.

[0157] Gene expression analysis. Triplicate cultures of AT3-M and AT3-T cells were grown to ~60% confluency. Total RNA was isolated using the RNeasy kit (Qiagen, Valencia, Calif.), and triplicate samples were submitted to the Duke Microarray Facility. Gene expression analysis was performed using the R027K rat spotted arrays 3.0 (Operon, Huntsville, Ala.). Bioinformatical analysis of expression differences between AT3-M and AT3-T cells was done using the GeneSpring GX software version 7.3.1 (Agilent Technologies, Durham, N.C.). The data files (representing signals for 26,986 gene probes in all six data points, three for AT3-M and three for AT3-T) were normalized using the feature: per Spot and per Chip—intensity dependent (lowess) normalization. The resulting gene list was used to determine the significantly differentially expressed genes

between AT3-M and AT3-T using the “Filtering on Volcano plot” feature with the following characteristics: (1) Test type: Parametric test, did not assume variances equal; (2) Multiple testing correction: None; (3) Fold Difference: Two-fold or greater and a P-value cutoff of 0.05.

[0158] Analysis of human circulating tumor cells. Patients eligible for the CTC biomarker protocols included (1) men with progressive CRPC, with metastatic progression by PSA (two consecutive rises over nadir separated by >1 week) or radiologic criteria (RECIST or new bone scan lesions), a PSA ≥ 5 , age ≥ 18 years; or (2) women with mBC with disease progression or with initiation of a new systemic therapy, who were >18 years of age, and who were at least 7 days from treatment with an anthracycline-containing regimen. Blood (15 mL) was collected from patients and processed within 48 hours at the Duke University CTC lab using the Cell Search System (Veridex, Raritan, N.J.). Veridex profile kits were used, which isolate EpCAM positive cells without additional staining. The isolated cells were either processed immediately or stored overnight in 4% paraformaldehyde and processed the next day. Immunostaining was done on teflon coated slides. Briefly, cells were pipetted into the wells of the slides and left to settle for ~30 minutes followed by standard immunostaining procedures with careful aspiration to minimize cell loss. An initial ferromagnetic wash using a benchtop magnet was performed to further isolate CTCs, with resuspension of the cell pellet after magnet release 100 uL PBS. Following 4% PFA fixation and permeabilization with PBT (PBS with 2% Triton) and blocking with 10% goat serum for 30 minutes, triple immunostaining was performed using CD45 antibody (AbCam #33533-50) labeled with Alexa 647, cytokeratin (AbD Serotec #MCA 1907HT) labeled with Alexa 555, and Vimentin (BD Biosciences, San Jose, Calif. #550513) labeled with Alexa 488. Nuclear staining with 4',6-diamidino-2-phenylindole (DAPI) was then performed. A mesenchymal CTC (an event) was defined as an intact cell by microscopic examination, containing an intact nucleus and expressing cytokeratin but lacking CD45 staining, using appropriate controls (see Table 1 for antibodies and controls). Human peripheral blood mononuclear cells (PBMCs), obtained by Ficoll purification of buffy coats from normal donors, were kindly provided by Micah Luftig (Duke University, Durham N.C.) and used as control cells for CD45 expression. Linear regression analysis was performed to compare CTC count (standard Cellsearch method) against the proportion of CTCs that co-express vimentin. Goodness of fit was tested by analysis of variance.

TABLE 1

EMT/Stemness Antigens to be assessed in CTCs.					
Antigen	Product	Positive Control	Negative Control	Leukocyte Expression	Dilution
Vimentin	BD Biosciences, mouse monoclonal IgG1	PBMCs, PC-3, DU145	T47D, LnCAP	Yes	2:225
N-cadherin	DAKO, mouse monoclonal IgG1, 6G11	Sarcoma, rat brain, PC-3	DU145, T47D, mock	No	4:225
Cytokeratin (pan)	AbD Serotec, mouse monoclonal IgG1, MCA1907HT, clone AE1/AE3	T47D, DU145	PC-3, PBMCs	No	2:45

TABLE 1-continued

EMT/Stemness Antigens to be assessed in CTCs.					
Antigen	Product	Positive Control	Negative Control	Leukocyte Expression	Dilution
CD45	Invitrogen, mouse IgG1, HI30, MHCD4500	PBMC	PC-3, DU145	Yes	1:45
CD133	Santa Cruz mouse monoclonal IgG, sc-130127	CaCo-2 colon cancer cells	Mock	Variable	4:225

[0159] The slides were mounted with gel/mount media (Biomedex, Foster City, Calif.). The slides were analyzed with an Olympus (Melville, N.Y.) IX 71 epifluorescence microscope, and images were acquired using an Olympus DP70 digital camera. Image processing was done with DP controller software (Olympus). All fields were analysed, with each cytokeratin positive nucleated cell that was CD45 negative being counted as a CTC. Positive control cells for each antibody included PC-3 cells for vimentin, peripheral blood mononuclear cells (PBMCs) for CD45, and T47D breast cancer cell lines for cytokeratin. A similar volume of reaction mix without antibody was used for negative controls.

[0160] Media exchange experiments. The cells of AT3-T or AT3-M clones were plated at a concentration of 150,000 cells/2 mL of media in 6-well plates and allowed to incubate for 24 h. The conditioned media was then filtered using a 0.22 μ m filter, and then immediately allowed to incubate with cells of the other clone, which was plated at the same concentration and had its media aspirated and cells washed with 2 mL of PBS. All cells with media replaced were incubated for 72 h, and phase and epifluorescent microscopy was used to monitor cell phenotypes 24, 48, and 72 h after treatment. Control plates, in which media was conditioned, cells washed with PBS and media added back to the same cells, were also used.

[0161] Scratch-wound assay. Cells were plated and left to grow to nearly 100% confluency in 6-well dishes. A wound was simulated by scratching the cells with a sterile 200 μ l pipette tip. The wells were washed twice with PBS and fresh media added. Pictures were taken in the same marked spot at 0, 24, and 48 h. Percent migration was calculated as (width at 0 h-width at 24 or 48 h) 1 width at 0 h $\times$ 100. Relative migration was compared using two-way analysis of variance via Prism 4.0c for the Macintosh (Graphpad, La Jolla, Calif.).

[0162] Matrigel assay. Matrigel assay was performed per manufacturer's indications (BD Biosciences). Briefly, after rehydration, 2×10^5 cells were plated either in the control or in the matrigel-coated inserts and incubated for 22 h. Following incubation, the non-invading cells from the upper-part of the inserts were removed using cotton-tipped swabs. The cells from the lower part of the membrane were stained with hematoxylin-eosin, membranes were removed, placed on a slide and observed under the microscope.

[0163] Immunohistochemical (IHC) analysis of metastases. Under the same informed consent protocol as the analysis of human circulating tumor cells described above, men undergoing CTC collection additionally consented to have a radiologic-guided metastatic biopsy for analysis of biomarker expression by IHC. Samples were obtained

through core needle biopsies during light sedation, and immediately formalin-fixed and paraffin embedded. For analysis, slides were deparaffinized, rehydrated, and endogenous peroxidase was inactivated for 30 min. in 0.3% H₂O₂ (hydrogen-peroxide) in methanol. Specific antigen retrieval steps were performed for individual antigens. Three markers were evaluated by IHC: vimentin (M7020, Dako, 1:150; antigen retrieval with pepsin treatment at 37° C. for 15 minutes), cytokeratin cocktail (18-0132, Invitrogen, 1:50 and 349205, BD Biosciences 1:50, antigen retrieval with pepsin treatment at 37° C. for 15 minutes), and CD45 (M0701, Dako, 1:200; antigen retrieval with sodium citrate 10 mM, pH 6.0 at 100° C. for 30 minutes). Primary antibody was incubated for 60 minutes at room temperature. Dako Envision horseradish peroxidase secondary antibody was used for 30 minutes at room temperature and the signal was detected with DAB reagent (Vector kit SK 4100). Slides were counter stained with hematoxylin and eosin and assessed by a trained pathologist for expression using appropriate positive (localized prostate tissue microarray sections) and negative controls (mock antibody) for each marker.

[0164] Statistical analyses. To determine the significantly differentially expressed genes between AT3-M and AT3-T the GeneSpring GX "Filtering on Volcano plot" feature was used with the following characteristics: (1) Test type: Parametric test, don't assume variances equal; (2) Multiple testing correction: None; (3) Fold Difference: Twofold or greater and a P-value cutoff of 0.05. To compare CTC count (standard CELLSEARCH® method) against the proportion of CTCs that co-express vimentin, N-cadherin, or CD133, linear regression analysis was performed. Goodness of fit was tested by analysis of variance.

Example 2. Isolation of Individual AT3 Clones that Inhabit an Intermediate Phenotypic State

[0165] The alternative splicing of FGFR2 transcripts, which produces either FGFR2-IIIb or -IIIc variants in epithelial and mesenchymal cells respectively, is exquisitely regulated (FIG. 1A). In FIG. 1A is a schematic representation of the IIIb and IIIc alternatively spliced isoforms of FGFR2. FGFR2 contains an extracellular domain (with three IgG-like domains), a transmembrane domain (TM), and two intracellular tyrosine kinase domains. The IIIb isoform is found in epithelial cells while the IIIc isoform in mesenchymal cells. Exons IIIb and IIIc are regulated coordinately to provide mutually exclusive expression of the two isoforms and transcripts including both exons are destabilized by nonsense-mediated decay. We have previously used FGFR2 alternative splicing reporters, in particular constructs that measure the epithelial-specific silencing of exon IIIc (e.g., pRIIIc² in FIG. 1B), to report on the phenotypic

state of cells in vitro and in vivo. In FIG. 1B is a schematic of the pRIIIc1² minigene and the fluorescence read-out. The minigene contains the DsRED open reading frame interrupted by exon IIIc and flanking introns of the FGFR2 gene. In epithelial cells exon IIIc is skipped, DsRED open reading frame is formed and results in fluorescence signal. In mesenchymal cells, exon IIIc is included and the DsRED open reading frame is disrupted, resulting in low or close-to-background fluorescence signal. The pRIIIc1² splicing reporter, which produces a variant red fluorescence protein (DsRED) when exon IIIc is silenced, revealed MET in primary tumors derived from AT3 cells implanted in the flanks of Copenhagen white rats. While most tumors contained MET foci, each tumor had very few foci and these were not randomly distributed but rather were associated with collagenous stroma. In contrast to the low frequency of MET in primary tumors, a high incidence of MET among lung metastases in these animals was observed, suggesting an unexpected association between the more epithelial phenotype and aggressive behavior. These studies could not ascertain whether the epithelial-like AT3 cells found in the lungs had undergone MET in the primary tumors or during the process of metastasis.

[0166] In an attempt to find post-MET cells in vitro, limiting dilution was used to obtain clones from AT3 cells stably transfected with the pRIIIc1² reporter. A total of 16 clones of a maximum calculated recovery of 27 were obtained, which is ~60% cloning efficiency. Eleven of these sixteen clones expressed RIIIc1² transcripts (italicized in Table 2), and of these, eight expressed DsRED (Table 2). Some of the clones had an epithelial-like morphology (cells with cobblestone appearance and adherent to each other), while others had a mesenchymal-like morphology (spindle-shaped), as well as clones that displayed a mixed phenotype. It is important to note that given the high cloning efficiency and the high frequency of DsRED+ clones, it is highly unlikely for these epithelial-like clones to come from a very small population within the parental AT3 cells. Rather, the process of subcloning induced a phenotypic transition in a significant number of the AT3 cells.

TABLE 2

Properties of AT3 clones.				
AT3 Clones	Cellular morphology ³	DsRED expression ²	Detection of exon IIIc skipping among RIIIc13 transcripts ¹	FGFR2 transcripts detected ³
1	Epithelial	High	+	IIIc
2	Epithelial	High	+	IIIc > IIIb
3	Epithelial	Low	ND	IIIc > IIIb
4	Epithelial	Low	ND	IIIc
5	Epithelial	High	+	IIIc > IIIb
6	Mesenchymal	Low	ND	IIIc
7	Mixed	Low	ND	IIIc
8	Mixed	High	+	IIIc
9	Mixed	Low	ND	IIIc
10	Mixed	High	+	IIIc
11	Mesenchymal	Low	-	IIIc
12	Mesenchymal	Low	-	IIIc
13	Epithelial	High	-1	IIIc > IIIb

TABLE 2-continued

Properties of AT3 clones.				
AT3 Clones	Cellular morphology ³	DsRED expression ²	Detection of exon IIIc skipping among RIIIc13 transcripts ¹	FGFR2 transcripts detected ³
14	Epithelial	Low	-	IIIc
15	Epithelial	High	+	IIIc
16	Mixed	High	-2	IIIc

¹See FIG. 1C. A “+” indicates detection of RIIIc1² transcripts missing exon IIIc, a “-” all RIIIc1² transcripts include exon IIIc, ND means that no RIIIc1² transcripts were detected.
²Determined by epifluorescence microscopy (high is defined as fluorescence above background of naive AT3 cells and low undistinguishable from the same cells).
³Discussed further herein and illustrated in FIG. 1C.

[0167] All of the clones obtained by limiting dilution were analyzed to determine the splicing status of RIIIc1² and endogenous FGFR2 transcripts. We could not detect exon IIIc skipping among pRIIIc1² transcripts or any evidence of exon IIIb inclusion among endogenous FGFR2 transcripts in clones with a mesenchymal-like morphology (FIG. 1C and Table 2). FIG. 1C shows RT-PCR analysis of the reporter (upper panel) and endogenous FGFR2 (lower panel). Primers used for the reporter are designed in the DsRED regions flanking exon IIIc. RT-PCR shows a higher percentage of the skipped product in clone AT3-T compared to clone AT3M. Reactions that did not include RT (-RT) reveal a contaminating product that is out-competed by the presence of a bona fide cDNA template (AT3-M lanes). Since exons IIIb and IIIc differ in size by only 3 nucleotides, analysis of the presence of IIIb or IIIc exons in FGFR2 gene was done by using primers in the flanking exons and specific restriction digestion of the resulting RT-PCR products. Exon IIIb is digested by Aval (A) and IIIc by HincII (H). There is a higher percentage of exon IIIb in clone AT3-T. The RT-PCR are replicates from three different cultures of the two clones. These clones did not express detectable levels of DsRED (FIG. 1D and Table 2). FIG. 1D shows epifluorescence and phase-contrast pictures of clones AT3-M and AT3-T shows the difference in fluorescence intensity and morphology between the two clones. Epifluorescence pictures were taken at the same exposure. All pictures were acquired at 200x magnification. While the skipping of exon IIIc among pRIIIc1² transcripts from epithelial-like clones could be expected, the observation that all of these clones both skipped and included exon IIIc was unexpected (FIG. 1C, Table 2 and data not shown). Analysis of endogenous FGFR2 transcripts revealed that four of the clones with epithelial morphology and DsRED expression had clear evidence of coexpression of both IIIb and IIIc isoforms (Table 2, and FIG. 1C and FIG. 1D). As shown in FIGS. 1A-1D, AT3-T cells expressed epithelial and mesenchymal isoforms of FGFR2. The expression of DsRED in all the cells suggested that each cell in the culture was expressing both isoforms (FIG. 1C).

[0168] We followed two clones with epithelial morphology, high DsRED levels and co-expression of FGFR2-IIIb and -IIIc transcripts (clone 2 and clone 5 (clone 5 herein AT3-T)) and noted that the phenotypic characteristics described above were stable for over six months. Equally, we followed clone 11 (clone 11 herein AT3-M) and clone 12 for six months, and noted that the mesenchymal morphology, undetectable DsRED expression and exclusive production of FGFR2-IIIc were also stable. We concluded from

these observations that AT3 cells were plastic and were coaxed by sub-cloning to populate intermediate phenotypic states, with properties of epithelial and mesenchymal cells. [0169] A media exchange experiment was used to investigate whether or not the splicing of pRIIICl^2 transcripts in the DsRED expressing clones was regulated by soluble factors. Media conditioned by DsRED expressing clones (clone 5 in Table 2) was filtered and added to DsRED negative clones (clone 11 in Table 2). DsRED+ cells were observed among DsRED- cells incubated with DsRED+ conditioned media (FIGS. 2A-2B). FIG. 2A shows examples of clusters of DsRED positive cells formed by AT3-M cells upon treatment with conditioned media from clone AT3-T. Media was conditioned for 24 h, filtered and added on AT3-M cells. Pictures (acquired at 200 $\times$) are taken 48 h following media exchange. FIG. 2B shows results from flow cytometry analysis of the same experiment. Left upper panel represents clone AT3-M conditioned with media from the same clone, as a negative control. Right upper panel represents clone AT3-T, which is DsRED positive. The lower panel represents clone AT3-M 48 h after conditioned media from clone AT3-T was added. Different lots of fetal bovine serum caused variation in this effect. This effect was quantified by flow cytometry and these data suggested that about half of the DsRED- cells were induced to express DsRED at levels equivalent to those seen in DsRED+ cells (FIGS. 2A-2B). The changes observed were not due to prolonged culture of the cells in the same wells because conditioned media from a separate DsRED- culture did not induce DsRED expression. As shown in FIGS. 2A-2B, AT3-T conditioned media induced AT3-M cells to express DsRED. These observations suggest that soluble factors secreted by the DsRED+ clones or dilution of factors extant in the DsRED- conditioned media may contribute to plasticity.

Example 3. AT3-M and AT3-T Cells are Tumorigenic

[0170] The initial characterization of the AT3-T revealed that these transitional cells grew slower and reached a lower confluent density than the AT3-M (FIG. 3A). FIG. 3A shows growth curves for clones AT3-T and AT3-M. Cells were plated at 0 h time-point, trypsinized, and counted at the indicated times. Data are the mean $\pm$ S.D. (n=3). To investigate their growth in vivo AT3-M and AT3-T cells were co-transfected with pGint a plasmid that expresses EGFP (herein GFP) in both mesenchymal and epithelial cells, and sorted stable populations of each cell line using flow cytometry for uniform GFP intensity. The GFP expressing cells maintained the morphological characteristics, the differential DsRED expression, and the differences in the splicing of pRIIICl^2 and FGFR2 transcripts first observed after sub-cloning.

[0171] We injected 3×10^5 GFP-expressing AT3-T or AT3-M cells subcutis in both flanks of Copenhagen white 2331 male rats. All of the animals developed bilateral tumors, indicating that both AT3-M and AT3-T cells were highly tumorigenic in these syngeneic rats. As a humane endpoint, rats were sacrificed when tumor length estimated by palpation reached 1 cm. The in vivo growth curves for the AT3-M and AT3-T tumors were significantly different, as determined by a logrank test ($p=0.0020$; FIG. 3B). FIG. 3B is a sacrifice curve for rats injected with AT3-M or AT3-T cells. FIG. 3C shows comparison of tumor volumes resulting from AT3-T and AT3-M injection. The Y-axis represents

tumor volumes at the time of sacrifice of the animals and the X-axis days from the time of implantation to the time of sacrifice. Average tumor volumes and average days until sacrifice are represented with S.D. bars. Some points represent more than one tumor with the same volume on the same day. Tumor volume was measured (FIG. 3C) and although most AT3-T animals were sacrificed later, there was no significant difference in tumor size ($p=0.76$). As shown in FIGS. 3A-3D, AT3-T cells grew more slowly than the mesenchymal-like AT3-M cells in vitro and in vivo, but both were equally tumorigenic. We concluded that whereas AT3-T cells grew more slowly in vitro and in vivo relative to their more mesenchymal siblings, these transitional cells were capable of forming tumors.

Example 4. Both AT3-M and AT3-T are Plastic

[0172] Since the implanted AT3-M and AT3-T cells could be tracked by GFP expression, and epithelial character could be interrogated by DsRED expression, the plasticity of the tumors were able to be investigated. The overwhelming majority of cells in AT3-M tumors expressed GFP but not DsRED (FIG. 4A). As shown in FIGS. 4A-4B, tumors from both AT3-T and AT3-M clones have evidence of plasticity. FIG. 4A shows representative example of cells that express both RFP and GFP at the periphery of an AT3-M tumor stably transfected with Gint and pRIIICl^2 reporters. Pictures were taken at 200 $\times$ magnification. To compensate for a low RFP signal, the color curve of the entire picture was adjusted. Nonetheless, groups of cells were observed expressing both GFP and DsRED in many AT3-M tumor sections, especially near the tumor capsule, (FIG. 4A; see also FIG. 5). FIG. 5 shows a representative example of cells that express both RFP and GFP at the periphery of an AT3-M tumor stably transfected with Gint and pRIIICl^2 reporters. Pictures were taken at 200 $\times$ magnification. In this version, overall RFP signal was not adjusted via color curve after the image was captured. RFP positive cells were clearly above background level.

[0173] Many sections from AT3-T tumors co-expressed GFP and DsRED; however, large areas were observed that expressed GFP but not DsRED in all 64 sections surveyed (FIG. 4B). FIG. 4B shows representative example of a section from an AT3-T tumor stably transfected with GFP and pRIIICl^2 reporters. Pictures were taken at 200 $\times$ magnification. RNA extracted from these regions of AT3-T tumors confirmed the presence of the pRIIICl^2 transcripts. Both AT3-T and AT3-M cells were plastic and produced tumors with cells that displayed a range of epithelial-mesenchymal properties.

Example 5. AT3-T Cells are Motile In Vitro and Metastatic In Vivo

[0174] Comparison of AT3-T and AT3-M mobility and invasive potential was performed in culture. Motility was measured in culture by a "wound closure" assay, and no significant motility difference ($p=0.59$) was found between cell lines 24 and 48 hours after a scratch-wound had been made in the cultures (FIGS. 6A-6D). FIG. 6A shows representative pictures for the scratch-wound assay (experiment done in triplicate for each clone). Pictures were taken at 40 $\times$ magnification. FIG. 6B shows quantification of migration as explained in Methods. Mean and SO values were derived from triplicate experiments. FIG. 6C shows invasion assay

using Matrigel coated membranes. Representative pictures of each clone and for both control membranes and Matrigel-coated membranes (n=5). Cells were stained with hematoxylin-eosin. Pictures were taken at 40× magnification. FIG. 6D shows quantification of invasion assay results. Mean and SD values were derived from five individual experiments. To gauge invasive properties of the cells we measured the number of cells traversing through Matrigel membranes in a 22-hour period. The same number of AT3-T and AT3M cells was observed on the Matrigel membranes suggesting that the two cell lines were equally capable of invading this membrane (FIGS. 6A-6D). While a higher number of cells from clone AT3-T were observed on the control membrane compared to clone AT3-M, these studies nevertheless indicated that the more epithelial AT3-T cells had similar motility and invasive potential as the AT3-M cells. As shown in FIGS. 6A-6D, AT3-M and AT3-T cells exhibited similar migration in vitro.

[0175] In order to assess invasiveness in vivo lungs from the twenty animals harboring AT3-M and AT3-T tumors were examined for presence of metastatic foci. No macroscopic metastatic nodules were observed in any of the lungs, which was likely due to the sacrificing protocol used on the animals when the tumors reached a specified size instead of using survival as the end-point. The GFP expression from the Gint reporter was examined to evaluate the presence of micrometastases by epifluorescence microscopy. To assure a comprehensive evaluation, 7-8 equally spaced sections from each lung were surveyed (total of 150 sections for each clone). The presence of metastatic foci was determined by GFP fluorescence, followed by counter-staining of the sections with hematoxylineosin (FIGS. 7A-7B). FIG. 7A shows (upper panel) an example of a section exhibiting the expected pattern for clone AT3-T (i.e. GFP+, DsRED+) in a metastatic focus, and (lower panel) an example of a section exhibiting a plastic pattern for clone AT3-T (i.e. GFP+, DsRED-) in a metastatic focus. FIG. 7B shows (upper panel) an example of a section exhibiting the expected pattern for clone AT3-M (i.e. GFP+, DsRED-) in a metastatic focus, and (lower panel) an example of a section exhibiting a plastic pattern for clone AT3-M (i.e. GFP+, DsRED+) in a metastatic focus. As shown in FIGS. 7A-7B, metastatic foci in lungs from animals with tumors from either AT3-T or AT3-M clones (stably transfected with GFP and pRIIcI² reporters) had evidence of plasticity. Metastatic foci were found in 7 out of 10 lungs for clone AT3-M and 6 out of 10 lungs for clone AT3-T.

[0176] Evaluation of the plasticity of the metastatic foci using the combined output of the GFP and DsRED reporters revealed plastic foci (DsRED+ for AT3-M and DsRED- for AT3-T) in the case of both clones: 3 out of 12 for clone AT3-T and 13 out of 16 for clone AT3-M (FIGS. 7A-7B). These studies indicated phenotypic plasticity for the AT3-M cells and suggested it for the AT3-T cells. Importantly, both cell lines were metastatic despite differences in the original epithelial vs. mesenchymal phenotype.

[0177] Plasticity and metastatic behavior of cancer cells. Both the mesenchymal AT3-M and the more epithelial AT3-T cells metastasized efficiently. The drivers of metastasis, however, may be different in these two cells. The gene expression comparison between the AT3-M and AT3-T clones revealed at least one intriguing possibility: microarray analysis showed a 12-fold increase in the expression of junctional adhesion molecule C (JAM-C) in AT3-T com-

pared to AT3-M, and this was confirmed by RT-PCR and immunoblot analysis. JAMs were present in leukocytes and at the tight junctions of epithelial and endothelial cells and have been shown to be involved in transendothelial migration of monocytes. JAM-C is expressed in several cell lines with high metastatic potential and knock-down of this molecule in the HT1080 human fibrosarcoma line significantly decreases its metastatic properties in vivo. Moreover, JAM-C is also present in the gene sets associated with stemness that had significant overlaps with genes that define clone AT3-T. Therefore clone AT3-T, by over-expression of different adhesion molecules may acquire metastatic capabilities. In addition, the overexpression of the downstream Hedgehog pathway effector GLI3 may be significantly upregulated in the more epithelial and stem cell-like AT3-T cells as compared to the more mesenchymal AT3-M cells. Hedgehog signaling has been linked to EMT, stemness, and metastasis/aggressiveness in several tumor types, and thus differential expression or regulation of developmental programs may underly these phenotypical differences across these cell lines. Increased expression of Patched, a Hedgehog pathway component, has been linked to prostate tumors during progression to androgen independence and in circulating tumor cells of men with metastatic castration-resistant prostate cancer.

Example 6. AT3-T Cells Display a Stem Cell-Like Gene Expression Signature

[0178] AT3-T cells sometimes formed tight clusters resembling protospheres. While sphere formation is not an exclusive property of stem cells, it has been associated with stemness in many different systems. Given these observations and the high tumorigenicity of AT3-T and AT3-M cells, they were tested for the expression of markers associated with cancer stem-like cells. Also included were the parental AT3 cells and another Dunning tumor cell line, DT cells, which display epithelial markers and are only weakly tumorigenic in Copenhagen white rats. The DT cells expressed very low levels of CD44 and CD133, which are associated with highly malignant cancer stem-like cells (FIGS. 8A-8B). CD133 was detectable in DT lysates only when four fold more lysate was loaded. The mesenchymal-like AT3 cells expressed much higher levels of both CD44 and CD133 than the DT cells (note that the lanes for the DT samples are overloaded in FIG. 8A), which is consistent with recent reports that EMT induces stemness in mammary epithelial carcinoma cells. FIG. 8A shows a membrane with serial twofold dilutions of whole cell lysates was cut in half and immunoblotted for CD133 (upper panel) or β -actin (lower panel). Size markers are in kDa. A faster migrating CD133 band repeatably detected only in DT lysates is marked (*), suggesting possible post-translational regulation. FIG. 8B shows a membrane with serial twofold dilutions of whole cell lysates was cut in half and immunoblotted for CD44 (upper panel) or β -actin (lower panel). Representative blots from two independent sets of lysates are shown. AT3-T expressed CD44 and CD133. Interestingly, the AT3-T cells expressed overall higher levels of CD44 and CD133 than the more mesenchymal AT3-M. Moreover, AT3-T cells expressed a higher ratio of CD44H to CD44E when compared to AT3-M cells. The CD44H isoform has been associated with malignancy while CD44E is not. This suggests a more complex relationship between epithelial transitions and acquisition of stem cell-like prop-

erties. Consistent with expression of stem-like markers, both AT3-M and AT3-T cells formed colonies in soft agar and tumors when injected into Copenhagen white rats, and these tumors led to extensive metastases similar to parental AT3 cells (FIG. 3B).

[0179] To further explore these connections between transitions and stemness, global gene expression in AT3-M and AT3-T cells was compared. This analysis showed that 422 genes were differentially expressed (≥ 2 -fold; p-value < 0.05) in these two cells (Table 3). Many of the genes that were upregulated in AT3-T relative to AT3-M were preferentially expressed in epithelial cells and vice versa for those preferentially expressed in mesenchymal cells (Table 4). There were exceptions to this, however. Expression of the gene disintegrin-like and metalloprotease was consistent with a mesenchymal phenotype, but this mRNA level was 4-fold higher in AT3-T compared to AT3-M. Integrin β -4, normally associated with epithelial-like cells, was expressed 3-fold lower in AT3-T compared to AT3-M. These observations were consistent with the characterization of AT3-T cells as displaying more epithelial features than AT3-M cells and as populating an intermediate phenotypic state.

TABLE 3		
x Fold change (AT3-T/AT3-M)	Gene Symbol (Human)	Gene Symbol (Rat)
0.00771	P2RX5	P2rx5
0.011	CCNB1IP1	#N/A
0.0296	STRA6	Stra6
0.0327	G0S2	G0s2
0.0835	SERPINF1	Serpinf1
0.101	GSTA1	#N/A
0.107	RSNL2	Clip4
0.115	ADAMTS7	#N/A
0.134	GZMB	#N/A
0.137	SPON2	#N/A
0.156	MMP3	#N/A
0.191	ATP8A1	#N/A
0.197	EVPL	Evpl
0.21	LGALS3BP	Lgals3bp
0.216	SERPINB2	Serpinb2
0.219	NETO2	Neto2
0.223	PTX3	#N/A
0.23	SERPINB7	Serpinb7
0.233	RASIP1	#N/A
0.235	OMD	#N/A
0.239	HLA-G	#N/A
0.239	HLA-A	#N/A
0.247	CD97	Cd97
0.251	GJA4	Gja4
0.254	DSU	#N/A
0.257	MGLL	Mgll
0.261	SPHK1	#N/A
0.268	HRBL	Zcwpw1
0.268	ZCWPW1	Zcwpw1
0.27	ENPP3	Enpp3
0.275	PTGS1	Ptgs1
0.278	RAMP1	Ramp1
0.281	DHRS3	Dhrs3
0.282	FAM117A	Fam117a
0.284	TUBB2A//TUBB2B	Tubb2b
0.284	TUBB2B	Tubb2b
0.285	C10orf10	LOC500300
0.289	SYTL2	#N/A
0.291	SLC39A4	Slc39a4
0.292	CHRD	Chrd
0.292	GIP	Gip
0.293	CKLF	Cklf
0.294	PLAU	Plau
0.295	GUF1	#N/A
0.307	CGI-38	Tppp3

TABLE 3-continued		
x Fold change (AT3-T/AT3-M)	Gene Symbol (Human)	Gene Symbol (Rat)
0.311	LECT2	Lect2
0.318	NQO2	#N/A
0.32	C11orf75	RGD1309410
0.324	DOCK2	#N/A
0.325	LGALS2	#N/A
0.326	CASP4	Casp1
0.326	LTBP4	Ltbp4
0.334	HSPB1	Hspb1
0.335	ITGB4	Itgb4
0.34	BPHL	Bphl
0.341	FOXF2	#N/A
0.345	MYH1	#N/A
0.345	SMAD6	Smad6
0.348	TGFB1	Tgfb1
0.351	MMP10	#N/A
0.363	MMP9	Mmp9
0.363	COL18A1	Col18a1
0.366	HES1	#N/A
0.369	SLC35D2	#N/A
0.377	ADORA2B	Adora2b
0.377	COL3A1	Col3a1
0.379	DPEP2	Dpep2
0.382	GPR153	Gpr153_predicted
0.383	LOC55908	#N/A
0.389	SELPLG	#N/A
0.394	P2RX1	Atp2a3
0.394	ATP2A3	Atp2a3
0.394	ADD3	Add3
0.395	TSPAN9	Tspan9
0.399	LOC54103	#N/A
0.4	BFSP2	#N/A
0.4	FLJ14213	RGD1309969
0.4	PGGT1B	Pggt1b
0.401	HCN2	Hcn2
0.403	C2orf33	RGD1310230
0.404	TMEPAI	#N/A
0.405	INHA	Inha
0.406	HPSE	#N/A
0.409	CRY1	Cry1
0.413	IL3RA	Il3ra
0.413	CDC42EP1	#N/A
0.416	ARG1	Arg1
0.417	MAPK14	Mapk14
0.419	FLJ22028	#N/A
0.421	GALR2	Galr2
0.422	TSPAN8	Tspan8
0.422	FAM77C	RGD1561205
0.422	USP2	Usp2
0.422	LAMA3	#N/A
0.424	CCNE1	Ccne1
0.424	NSF	Nsf
0.428	ST3GAL5	St3gal5
0.429	SYNJ2	Synj2
0.43	ADA	Ada
0.43	PCBP3	Pcbp3
0.433	ZNF43	#N/A
0.433	C14orf130	Ubr7
0.436	SOS2	#N/A
0.436	RASSF3	#N/A
0.436	GLMN	Glmn
0.438	OSR2	Osr2
0.44	AGTPBP1	Agtpbp1
0.444	DBNDD2	RGD1311642
0.445	SGCB	#N/A
0.446	HBLD2	Isca1
0.448	SCARB1	Scarb1
0.448	EVI2A	Evi2a
0.448	AP4M1	#N/A
0.451	IGF2BP3	#N/A
0.452	FLJ10404	Ddx41
0.454	TGFB2	Tgfb2
0.459	PASK	Pask
0.461	C19orf37	Zfp428
0.462	BMP1	Bmp1

TABLE 3-continued			TABLE 3-continued		
x Fold change (AT3-T/AT3-M)	Gene Symbol (Human)	Gene Symbol (Rat)	x Fold change (AT3-T/AT3-M)	Gene Symbol (Human)	Gene Symbol (Rat)
0.464	PTPN13	Ptpn13	2.264	OXCT1	#N/A
0.47	PTPRG	#N/A	2.266	TEK	#N/A
0.47	EFNB1	Efnb1	2.268	HIST2H4///H4/o///	#N/A
0.472	PER2	Per2		LOC648164	
0.472	IRS3L///	Irs3	2.271	TMF1	Tmfl
	LOC442715		2.273	ZBTB7B	Zbtb7b
0.472	HRBL	Irs3	2.274	CAMSAP1L1	RGD1310950
0.472	MAP3K3	Kcnh6	2.279	CYP3A5	Cyp3ai
0.472	WDR68	Kcnh6	2.279	CYP3A7	Cyp3a9
0.472	KCNH6	Kcnh6	2.279	CYP3A4	Cyp3a9
0.472	CCDC44	Kcnh6	2.282	PENK	Penk1
0.473	CIB2	Cib2	2.283	KIAA2010	Smek1
0.475	MPZL1	Mpzl1	2.284	CHRNA1	#N/A
0.475	FADS2	#N/A	2.299	BAT3	Bat3
0.48	ZNF185	#N/A	2.302	ROM1	Rom1
0.482	SLC29A1	Slc29a1	2.306	HOXB8	#N/A
0.487	RUNX3	Runx3	2.309	KLK14	#N/A
0.488	NINJ1	Ninj1	2.31	SUV39H1	#N/A
0.489	RASL11B	Rasl11b	2.315	LOC440354///BOLA2///	RGD1564579
0.49	ECE2	Ece2		LOC595101	
0.49	TNNC2	Tnnc2	2.315	UBN1	Ubn1
0.491	WASPIP	Wipf1	2.323	C1orf103	#N/A
0.492	FN1	Fn1	2.333	EYA2	Eya2
0.494	NDE1	Nde1	2.347	MT2A	#N/A
0.494	CAMK2G	Camk2g	2.353	KIAA1815	Ermp1
0.495	CUTL1	Cux1	2.355	SETD1B	#N/A
0.495	ABHD6	Abhd6	2.369	MPHOSPH1	Kif20b
0.495	PTPN14	Ptpn14	2.38	EFNA1	Efna1
0.497	FLJ13946	#N/A	2.392	ABCF2	Abcf2
0.498	BAIAP2	Baiap2	2.397	LIMA1	Lima1
0.499	MSL3L1	Msl3l1	2.418	EXTL3	Extl3
0.499	DYNLT1	Dynlt1	2.418	ARL6IP2	Arl6ip2
0.499	GSTM3	Gstm5	2.442	GRAMD3	Gramd3
2	CHES1	Foxn3	2.456	JARID1A	Jarid1a
2.004	AQR	Agr///Znf770	2.476	ARHGEF9	Arhgef9
2.006	EPN1	Epn1	2.485	CAD	Cad
2.011	PPBP	Ppbp	2.493	RAI17	#N/A
2.019	SLC35D1	#N/A	2.526	KIAA0284	#N/A
2.022	PTPRC	#N/A	2.529	SGPP1	Sgpp1
2.031	USP47	Usp47	2.531	ABCB1	#N/A
2.041	DHX29	#N/A	2.531	ABCB1///ABCB4	#N/A
2.047	HMOX1	#N/A	2.542	KIF1C	#N/A
2.05	CAV1	Cav1	2.553	KIAA0020	LOC499339
2.053	BUB1B	Bub1b	2.563	ADAM15	Adam15
2.069	KCNIP4	#N/A	2.577	UBE1	Uba1
2.072	—	#N/A	2.577	INE1	Uba1
2.072	ADAM10	#N/A	2.58	GRIP2	Grip2
2.073	KIAA115	#N/A	2.59	PPEF1	#N/A
2.074	PSTPIP2	#N/A	2.619	SC65	Sc65
2.083	MAML1	#N/A	2.62	FER1L3	#N/A
2.084	RAB32	#N/A	2.62	NOC3L	#N/A
2.089	FAM111A	#N/A	2.62	RBP4	#N/A
2.095	ATRNL1	#N/A	2.645	SPINK4	Spink4
2.101	PPIC	Ppic	2.653	ATXN2L	#N/A
2.101	CHD4	Chd4	2.711	AHCYL1	Ahcyl1
2.109	IDE	Ide	2.723	TUBB3	Tubb3
2.117	PITPNM3	#N/A	2.723	MC1R	Tubb3
2.121	NFE2L1	Nfe2l1	2.729	AGPAT7	Lpcat4
2.121	MFSD1	#N/A	2.749	HOXC11	#N/A
2.133	KITLG	Kitlg	2.766	APH1A	Aph1a
2.161	ING3	Ing3	2.785	CNOT1	RGD1308009
2.167	CD24	#N/A	2.785	CSNK2A2	RGD1308009
2.169	IDS	#N/A	2.794	STAC	#N/A
2.177	MGC3196	LOC686289///	2.904	STAG1	#N/A
		LOC690285	2.942	MBNL1	#N/A
2.185	FBXL11	Fbxl11	2.982	MNT	Mnt
2.185	—	Fbxl11	3.007	RANBP5	Ipo5
2.191	ZC3H12A	#N/A	3.014	HERC1	Herc1
2.195	RKHD2	#N/A	3.065	ALDOC	Aldoc
2.201	LAMC2	Lamc2	3.122	KIAA0460	—
2.217	KIF11	Kif11	3.174	FLT3	#N/A
2.242	SNAPC5	Snape5	3.278	CXCL6	Cxcl6
2.252	THRAP3	#N/A	3.366	GLI3	#N/A
2.261	HS6ST1	#N/A	3.489	SSR3	#N/A

TABLE 3-continued

x Fold change (AT3-T/AT3-M)	Gene Symbol (Human)	Gene Symbol (Rat)
3.585	BCAN	Bcan
3.824	FKBP10	Fkbp10
3.903	GSTK1	Gstk1
3.931	PSCDBP	#N/A
3.974	ALCAM	Alcam
4.056	ADAMTS13	
4.203	SPRR2B	#N/A
4.276	GPR126	#N/A
5.169	SULF1	Sulf1
5.529	TFF1	Tff1
6.52	PTN	Ptn
8.591	MLF1	Mlf1
9.012	THBS2	Thbs2
10.79	HEPH	Heph
12.53	JAM3	Jam3

TABLE 4

Examples of epithelial or mesenchymal genes in the expression data analysis of clones AT3-T and AT3-M.	
Gene name	x Fold change in AT3-T vs. AT3-M
Junctional adhesion molecule C	12.53
Disintegrin-like and metalloprotease	4.05
Activated leukocyte cell adhesion molecule	3.97
Tubulin	2.73
Epithelial protein lost in neoplasm	2.39
Laminin	2.20
TGFβ2	0.45
MMP9	0.36
Collagen, type XVIII	0.36
MMP10	0.35
Integrin β4	0.33
TGFβ1	0.31

TABLE 4-continued

Examples of epithelial or mesenchymal genes in the expression data analysis of clones AT3-T and AT3-M.	
Gene name	x Fold change in AT3-T vs. AT3-M
Urokinase plasminogen activator	0.29
MMP3	0.15

[0180] Two gene sets were assembled: one composed of gene products upregulated in AT3-T (relative to AT3-M) and the second of those downregulated in AT3-T (relative to AT3-M). The two gene sets were compared for overlap with 5,452 gene sets from the Molecular Signature Database collections (Gene Set Enrichment Analysis (GSEA) <http://www.broad.mit.edu/gsea/>). Analysis of genes over-expressed in AT3-T relative to AT3-M for overlap with 5,452 gene sets from the Molecular Signature Database collections via Gene Set Enrichment Analysis (GSEA) did not show any significant enrichment of sets associated with EMT or MET. In this regard, both AT3-M and AT3-T resembled the mesenchymal-like, parental AT3 line. Among the 15 most significant overlaps for the genes overexpressed in AT3-T there were three sets of genes activated in hematopoietic stem cells ($p=3.24\times10^{-8}$), neural stem cells ($p=3.07\times10^{-7}$) and embryonal murine stem cells ($p=5.14\times10^{-6}$), (Table 5) while among the 20 most significant overlaps for the genes that are relatively downregulated in AT3-T cells were two gene sets associated with development of mature cell types. Expression of the downstream hedgehog pathway effector GLI3 was found to be 3.4-fold overexpressed in AT3-T cells compared to AT3-M cells, indicating that regulation of this developmental/stemness pathway in prostate cancer may be tied to the underlying phenotypic state during EMT/MET, similar to what has been reported in other tumors. These data indicated that AT3-T cells have gene expression profiles similar to stem cells, and, in concordance with the analysis of CD44 and CD133 protein expression, suggested that AT3-T cells exist in a more stem cell-like state than the more mesenchymal AT3-M cells.

TABLE 5

GSEA Collections: C1, C3, C2, C5, C4 # overlaps shown: 20 # gene sets in collections: 5452 # genes in comparison (N) 127 # genes in collections (N) 39655					
gene set name	# genes in gene set (k)	Description	# genes in overlap (k)	k/K	p value
TATAAA_V\$TATA_O1	1333	Genes with promoter regions [-2kb, 2kb] around transcription start site containing the motif TATAAA which matches annotation for TAF TATA	20	0.015	8.07E-09
STEMCELL_HEMATOPOIET IC_UP	1452	Enriched in mouse hematopoietic stem cells, compared to differentiated brain and bone marrow cells	20	0.0138	3.24E-08

TABLE 5-continued

GSEA Collections: C1, C3, C2, C5, C4					
# overlaps shown: 20					
# gene sets in collections: 5452					
# genes in comparison (N) 127					
# genes in collections (N) 39655					
gene set name	# genes in gene set (k)	Description	# genes in overlap (k)	k/K	p value
GNF2__RAP1B	37	Neighborhood of RAP1B	5	0.1351	1.23E-07
STEMCELL__NEURAL__UP	1838	Enriched in mouse neural stem cells, compared to differentiated brain and bone marrow cells	21	0.0114	3.07E-07
module__2	383	Genes in Module__2	10	0.0261	4.34E-07
CTTTGA__V\$LEF1__Q2	1270	Genes with promoter regions [−2kb, 2kb] around transcription start site containing the motif CTTTGA which matches annotation for LEF1: lymphoid enhancer-binding factor 1	17	0.0134	5.48E-07
SIGNAL__TRANSDUCTION	1637	Genes annotated by the GO term GO:0007165. The cascade of processes by which a signal interacts with a receptor, causing a change in the level or activity of a second messenger or other downstream target, and ultimately effecting a change in the functioning of the cell.	19	0.0116	9.33E-07
module__385	28	Genes in module 385	4	0.1429	1.91E06
V\$MYCMAX__O1	261	Genes with promoter regions [−2kb, 2kb] around transcription start site containing the motif NNACCACGTGG TNN which matches annotation for MYC: v-myc myelocytomatosis viral oncogene homolog (avian) MAX: MYC associated factor X	8	0.0307	1.98E06
GGGCGGR__V\$SP1__Q6	3053	Genes with promoter regions [−2kb, 2kb] around transcription start site containing the motif GGGCGGR which matches annotation for SP1: Sp1 transcription factor	26	0.0085	2.59E-06

TABLE 5-continued

GSEA Collections: C1, C3, C2, C5, C4					
# overlaps shown: 20					
# gene sets in collections: 5452					
# genes in comparison (N) 127					
# genes in collections (N) 39655					
gene set name	# genes in gene set (k)	Description	# genes in overlap (k)	k/K	p value
AACTTT_UNKNOWN	1963	Genes with promoter regions [−2kb, 2kb] around transcription start site containing motif AACTTT. Motif does not match any known transcription factor	20	0.0102	3.29E−06
V\$API_C	281	Genes with promoter regions [−2kb, 2kb] around transcription start site containing the motif NTGASTCAG which matches annotation for JUN: jun oncogene	8	0.0285	3.38E−06
MEMBRANE_PART	1673	Genes annotated by the GO term GO:004425 Any constituent part of a membrane, a double layer of lipid molecules that encloses all cells, and, in eukaryotes, many organelles; may be a single or double lipid bilayer; also includes associated proteins.	18	0.0108	5.09E−06
STEMCELL_EMBRYONIC_UP	1344	Enriched in mouse embryonic stem cells, compared to differentiated brain and bone marrow cells	16	0.0119	5.14E−06
INTRINSIC_TO_MEMBRANE	1350	Genes annotated by the GO term GO:0031224. Located in a membrane such that some covalently attached portion of the gene product, for example part of a peptide sequence or some other covalently attached moiety such as a GPI anchor, spans or is embedded in one or both leaflets of the membrane.	16	0.0119	5.43E−06
CELL_SURFACE	79	Genes annotated by the GO term GO:0009986. The external part of the cell wall and/or plasma membrane.	5	0.0633	5.58E−06

TABLE 5-continued

GSEA Collections: C1, C3, C2, C5, C4 # overlaps shown: 20 # gene sets in collections: 5452 # genes in comparison (N) 127 # genes in collections (N) 39655					
gene set name	# genes in gene set (k)	Description	# genes in overlap (k)	k/K	p value
UVC_XPCS_8HR_DN	408	Down-regulated at 8 hours following treatment of XPB/CS fibroblasts with 3 J/m ² UVC	9	0.0221	6.35E-06
NOTCH_SIGNALING_PATHWAY	12	Genes annotated by the GO term GO:0007219. The series of molecular signals initiated by binding of an extracellular ligand to a Notch receptor on the surface of the target cell.	3	0.25	6.86E-06
LEI_MYB_REGULATED_GENES	325	Myb-regulated genes	8	0.0246	9.62E-06
MORF_DDBI	246	Neighborhood of DDBI	7	0.0285	1.40E-05

[0181] Epithelial plasticity and stem cell-like behavior. It is well appreciated that cells induced to undergo EMT activate stem cell pathways. Work presented here shows that AT3 cells that transitioned towards a more epithelial state, i.e. were involved in MET, also activated expression of stem cell-like markers. This finding suggested a broader relationship between plasticity and stem cell-like character or stemness, which was modeled using a Gibbs free energy diagram (FIG. 9). FIG. 9 shows a model comparing stem cell-like character and epithelial-mesenchymal phenotype. The x-axis represents the spectrum of epithelial to mesenchymal phenotypes and the y-axis represents the stem cell-like character of the cells. The left arrow represents an EMT and the right arrow represents an MET. The model posits that as cells transition back and forth along the epithelial and mesenchymal x-axis they course through states of varying stemness, and this property peaks at intermediate states between epithelial and mesenchymal phenotypes. The number of different states and the exact height of the barriers between states are speculative and are not meant to be taken as proportional. Two phenotypic transitions are shown, the first is a partial EMT (left arrow) and the second is a partial MET (right arrow). Both of these transitions result in states with higher stem cell-like character. It should be noted that the model also predicts that some EMTs, and equally some METs, will result in a decrease in stemness and indeed this has been observed when the highly aggressive human DKAT basal-type breast cancer cell line is induced to undergo EMT (N. D'Amato and V. Seewaldt, personal communication). The model also suggests a link between stemness, plasticity, and metastatic propensity, perhaps explained by activation of certain oncogenic pathways (e.g., PI3 kinase/Akt) and developmental pathways.

[0182] The model also predicts that cells with maximal stem-cell character, which by definition will be highly malignant, should display both epithelial and mesenchymal

traits, because they inhabit intermediate states in the epithelial-mesenchymal axis. The highly malignant rat adenocarcinoma AT3-T cells are in this type of state. Importantly, in humans with metastatic breast and prostate carcinomas many CTCs also exist in these intermediate states. These cells correlate with disease progression and are believed to be highly aggressive. A population of cells enriched in CTCs expressed RNAs encoding mesenchymal markers; however, the data did not indicate whether or not epithelial and mesenchymal markers were co-expressed in the same cell. Another clinical example of cells in intermediate states is found in sarcomatoid renal cell carcinomas, which have been shown to co-express epithelial markers, such as epithelial membrane antigen, and mesenchymal ones, like vimentin. These tumors, though rare (1-8% of renal tumors) are highly aggressive and difficult to treat. A similar situation may be found in carcinosarcomas of both the prostate and breast, highly aggressive, rare tumors with mixed epithelial and mesenchymal components but of clonal origin. It is not completely clear whether or not single cells in these tumor co-express epithelial and mesenchymal markers and are thus truly in intermediate states.

[0183] Finally, the model suggests that as sarcomas undergo MET they will activate stem cell-like pathways and become more aggressive. Indeed, there are many descriptions of sarcomas with mixed epithelial and mesenchymal components in close proximity as seen in some synovial- and osteo-sarcomas. New genetically-defined mouse models of soft tissue sarcoma should shed light on the existence and importance of cells intermediate cell states in progression of these tumors.

Example 7. Phenotypic Plasticity Among Human Circulating Tumor Cells

[0184] The experiments described above indicated that Dunning rat prostate adenocarcinoma cells that inhabit an

intermediate phenotypic state are tumorigenic, metastatic, and possess stem cell-like antigens and cellular programs. To investigate whether or not similar transitional cells could play a role in human cancer, cancer cells isolated from blood of men with metastatic castrate resistant progressive prostate cancer (CRPC) or women with progressive metastatic breast cancer (mBC) were examined. Circulating tumor cells (CTCs) represent an ideal source of tissue to investigate evidence of this plasticity in vivo, given that these cells are likely to be in circulation prior to and during metastatic colonization. CTCs have both independent prognostic and predictive significance in multiple epithelial malignancies, including breast and prostate cancer. These cells can be collected, isolated, and analyzed for a variety of biomarkers relevant to cancer biology.

[0185] It was tested whether there was a high likelihood of finding transitional cells within a population of CTCs captured by FDA-approved EpCAM (Epithelial Cell Adhesion Molecule)-targeted ferromagnetic antibodies. These cells were interrogated for expression of CD45 (expressed in many leukocytes; FIG. 10A), cytokeratin (CK; an epithelial marker), and vimentin (a mesenchymal marker) by immunofluorescence. CTCs were defined as CD45-negative and CK-positive nucleated intact cells (FIG. 10B) and transitional CTCs were so defined if they additionally co-expressed vimentin (FIGS. 10C-10D). FIGS. 10A-10D show that CTCs from patients with prostate adenocarcinoma stained positive for epithelial and mesenchymal markers. Triple staining was performed using anti-CD45 antibody labeled with Alexa 647, anti-cytokeratin (CK) antibody labeled with Alexa 555, and anti-vimentin antibody labeled with Alexa 488. Nuclei were labeled with DAPI. FIG. 10A shows an example of a leukocyte from a human peripheral blood mononuclear cell sample: CD45 (+), CK (-), and vimentin (+). Additionally, CD45 (+), CK (-), and vimentin (-) cells were observed. FIG. 10B shows an example of a CD45 (-), CK (+), and vimentin (-) cell from a patient with metastatic breast cancer. Such cells were counted as vimentin (-) CTCs in Table 6. FIG. 10C shows an example of a CD45 (-), CK (+), vimentin (+) from a patient with metastatic breast cancer. Such cells were counted as vimentin (+) CTCs in Table 6. FIG. 10D shows an example of a CD45 (-), CK (+), vimentin (+) from a patient with metastatic progressive castrate-resistant prostate cancer. Such cells were counted as vimentin (+) CTCs in Table 6.

[0186] Transitional CTCs co-expressed vimentin and CK in many of the patients with elevated CTC counts (≥ 5 CTCs/7.5 mL by standard testing) (Table 6, FIGS. 10A-10D). In fact, among nine patients with progressive metastatic CRPC and eight patients with progressive mBC, it was found that approximately 75% (range 0-100%, 85.5% in CRPC, 54% in mBC) of the CTCs stained for both CK and vimentin (FIGS. 10C-10D), indicating a transitional phenotype. These data indicated that circulating tumor cells in patients with metastatic breast and prostate cancer co-express epithelial (EpCAM and cytokeratin) and mesenchymal (vimentin) markers, and thus exist in a transitional phenotypic state, similar to that observed in our preclinical models.

TABLE 6

Circulating tumor cell (CTC) counts and vimentin expression in patients with metastatic castration resistant prostate or metastatic breast cancer.		
Subject Number	CTC Count (Cellsearch)*	Ratio: vimentin (+) CTCs/ Total CTC Count
Castrate-Resistance Metastatic Prostates Cancer		
1	5	4/6
2	41	11/11
3	45	6/10
4	626	5/8
5	110	17/21
6	182	5/6
7	17	13/16
8	19	33/34
9	34	12/12
Total		106/124 (85.5%)
Metastatic Breast Cancer		
1	21	0/6
2	7	2/2
3	8	4/4
4	21	1/2
5	12	2/2
6	188	21/22
7	138	8/20
8	377	6/23
Total		44/81 (54.3%)
Overall Total	—	150/205 (73.1%)

*Column 2 represents the CTC count as determined by the standard Cellsearch EpCAM based method for each subject, while column 3 represents the number and proportion of CTCs counted manually that were found to express cytokeratin and co-express vimentin, expressed as a ratio and percentage.

[0187] Plasticity and CTCs. The identification of plasticity among CTCs in a significant subset of patient samples offers several important clinical opportunities. Expression of plasticity may have prognostic or predictive value in patients with metastatic cancers, especially mBC where a significant range of values were shown for plasticity. Thus, the subset of patients with very high plasticity may have a more aggressive natural history and exhibit greater resistance to systemic treatments. In terms of diagnosis and utility as predictive biomarkers the data suggested that in addition to cells expressing both epithelial and mesenchymal markers there may be an unknown number of CTCs that have moved further towards the mesenchymal pole and are EpCAM negative. These cells will be missed by the FDA approved CELLSEARCH® System and also by the Adna Test (Adna-Gen AG) system and current microfluidic technologies, which enrich for CTCs by immunoabsorption of cells expressing MUC1 or EpCAM. Indeed, recent studies in breast cancer have suggested that “normal” type breast cancer cell lines that overexpress both EMT and stem cell antigens (CD44+, CD24-) may lack EpCAM and are thus not detectable by currently approved CTC detection systems. Therefore it is possible that the number of CTCs in patients with metastatic cancer is much higher than currently appreciated. Identification of this additional subset of CTC can provide greater prognostic value than CTC counts as currently determined, as well as earlier detection of CTCs and the metastatic potential in patients with earlier stage disease.

[0188] Furthermore, CTCs in intermediate states, which comprise the 50-75% of cells isolated herein from patients with metastatic breast and prostate cancer as well as those cells that may go undetected because they have undergone

a more complete EMT, represent a therapeutic problem. It has been well documented that EMT alters drug sensitivity of lung cancer cells and it has been challenging to direct therapy to cancer cells with stem cell-like properties, perhaps because of their recalcitrance to undergo apoptosis.

[0189] While recent studies suggest both a screening method and actual compounds (e.g., salinomycin) that can selectively target cancer stem cells, these aggressive cells still represent a formidable therapeutic challenge. Thus, molecules comprising a binding agent that has binding specificity to an EMT biomarker described herein and linked to an anti-cancer agent provide additional therapeutic options.

Example 8. CTCs from Patients with Metastatic Breast and Prostate Cancer Express Vimentin and N-Cadherin

[0190] Eligible men had progressive metastatic CRPC (progression despite testosterone <50 ng/dL) and were about to begin a new systemic therapy. Eligible women had progressive metastatic breast cancer (mBC) and were about to begin a new systemic therapy. Baseline characteristics of patients (n=29) are presented in Table 7.

TABLE 7

Baseline characteristics of patients (n = 29)		
	Metastatic Prostate (n = 17)	Metastatic Breast (n=12)
DEMOGRAPHICS		
Age, median	69 (59-82)	61.5 (48-81)
Race, Ethnicity		
White, non-hispanic	76%	58%
Other, non-hispanic	23%	42%
BASELINE DISEASE HISTORY		
Gleason Score, median	7 (7-9)	—
ER/PR, %	—	75%/67%
Baseline median PSA, Range	396.4 (14-13, 419.5)	—
Baseline Pain Score (0-10), median	1 (0-7)	0 (0-6)
Karnofsky Performance Status, median	90 (70-100)	90 (70-100) (n = 6)
# of Prior Hormonal Therapies	2 (0-5)	2 (0-4)
Prior Chemotherapy	47%	83%
Baseline CTC Count, median	40 (4-828)	13 (0-1062)
METASTATIC SITES		
Lymph Node	65%	50%
Liver	24%	50%
Lung	47%	42%
Bone	94%	75%

[0191] CTCs were drawn into standard FDA-approved Cellsave tubes and processed within 48 hours using the CELLSEARCH® methodology using EpCAM-based ferro-magnetic capture. A CTC was defined as an intact nucleated (DAPI+) cell that expressed pan-CK and lacked expression of the leukocyte antigen CD45, and was enumerated using standard methods. A second CELLSEARCH® tube was collected and processed using EpCAM capture, and isolated cells were stained for CK (IgG1, AbD Serotec) labeled with Alexa 555, CD45 (IgG1, AbCam) labeled with Alexa 647, and either vimentin (IgG1, BD Biosciences) or N-Cadherin (IgG1, DAKO) using immunofluorescent labeling with

Alexa 488. The proportion of CTCs staining positive for an EMT antigen was calculated from the total number of CTCs manually scored from the second tube. Positive controls using American Red Cross-derived PBMCs (CD45), PC3 prostate cancer cells (vimentin, N-cadherin), and T47D breast cancer cells (CK) were used for each marker. Negative controls using mock antibody were used to optimize the staining/scoring of each antigen.

[0192] Prevalence of vimentin and CK co-expression in CTCs, and prevalence of N-cadherin and CK co-expression in CTCs are presented in Tables 8 and 9, respectively. Vimentin co-expression was detected in 17/20 (85%) patients with mCRPC or mBC and 78% of all CTCs. N-Cadherin co-expression was detected in 8/9 (89%) patients and 81% of CTCs. Immunofluorescent images of CTCs from patients with mCRPC and mBC are shown in FIGS. 11A-11D (FIG. 11A, a leukocyte; FIG. 11B, vimentin negative CTC (CRPC); FIG. 11C, vimentin positive CTC (BC); and FIG. 11D) vimentin positive CTC (CRPC)). Immunofluorescent images of CTCs from patients with mCRPC and mBC are shown in FIGS. 12A-12D (FIG. 12A, leukocyte; FIG. 12B, Ncad positive CTC (BC); FIG. 12C, Ncad negative CTC (BC); and FIG. 12D, two NCad positive

CTCs (arrows) and 1 Ncad negative CTC (CRPC)). Immunofluorescent images of CTCs from patients with mCRPC and mBC are shown in FIGS. 13A-13H (FIG. 13A, Phase/DAPI; FIG. 13B, CD45/DAPI; FIG. 13C, CK/DAPI; FIG. 13D, Vimentin/DAPI positivity in a man with mCRPC; FIG. 13E, Phase/DAPI; FIG. 13F, CD45/DAPI; FIG. 13G, CK/DAPI; and FIG. 13H, Vimentin/DAPI negativity in a second man with mCRPC).

[0193] The data showed the co-expression of cytokeratin with the EMT antigens vimentin and N-cadherin in CTCs from men with metastatic CRPC and women with metastatic

breast cancer. A majority of CTCs examined co-expressed CK and EMT proteins by immunofluorescent labeling. The majority of patients in this study had CTCs that co-expressed vimentin or N-cadherin suggesting potential epithelial plasticity during metastasis. The data suggests that CTCs can lack epithelial markers and provide methods for assessing patients with breast and prostate cancer as well as for the optimal detection of circulating tumor cells in other common malignancies.

TABLE 8

Subject Number	CTC Count (Cellsearch)	Ratio of: Vimentin (+) CTCs/ Total Manual CTC Count
castrate-resistant metastatic prostate cancer	1	5
	2	4
	3	54
	4	45
	5	626
	6	110
	7	182
	8	17*
	9	19
	10	34
Total	1127	108/126 (86%)
metastatic breast cancer	1	13
	2	85
	3	8
	4	21
	5	12
	6	188
	7	324**
	8	377
	9	0
	10	3
Total	884	65/97 (67%)
Overall Total	—	173/223 (78%)

TABLE 9

Subject Number	CTC Count (Cellsearch)	Ratio of: N-Cadherin (+) CTCs/ Total Manual CTC Count
castrate-resistant metastatic prostate cancer	1	45
	2	12
	3	10
	4	5
	5	12
	6	221
	7	828
Total	1132	130/156 (83%)
metastatic breast cancer	1	1062
	2	2
Total	1064	9/16 (56%)
Overall Total	—	139/172 (81%)

*Count from 3 months prior to baseline (no intervening therapy)
**Count from time point #2

[0194] In a second trial to test for the existence of transitional CTCs, blood was collected from 31 men with mCRPC and 16 women with mBC (see baseline characteristics for the patients in Table 10 and Table 11). CTCs were processed using the CELLSEARCH® EpCAM-based immunocapture method and profiled for expression of CD45 (PTPRC) (a leukocyte marker), cytokeratins (CK) (epithelial markers), vimentin (VIM) and N-cadherin (CDH2) (mesenchymal markers), and CD133 (a stem cell marker) by immunofluo-

rescence (IF) (Table 2). Leukocytes were defined as nucleated (DAPI positive), CD45-positive and CK-negative cells, whereas CTCs were defined as nucleated (DAPI positive), CD45-negative and CK-positive cells. Among CTCs we identified transitional cells as those that additionally expressed vimentin or N-cadherin.

TABLE 10

Baseline demographic and clinical characteristics of the men with metastatic CPRC.	
n = 31	
DEMOGRAPHICS	
Age, years (range)	71 (59-89)
Race, Ethnicity	
White, non-Hispanic	71%
Black, non-Hispanic	29%
BASELINE DISEASE HISTORY	
Median Gleason Score (range)	8 (5-10)
Median Baseline PSA ¹ (ng/dl, range)	267.5 (14.0-13, 419.5)
Median Baseline Pain (range) ²	1 (0-7)
Median Karnofsky Performance Status (range)	90 (60-100)
Median Number of Prior Hormonal Therapies (range)	3 (0-5)
Prior Chemotherapy	65%
Prior Bisphosphonates	71%
SITES OF METASTATIC DISEASE	
Visceral (lung + liver)	35%
Lymph Node Only	0%
Bone metastatic:	
Bone Metastatic With Lymph Nodes (no visceral metastases)	39%
Bone Metastatic Without Lymph Nodes (no visceral metastases)	26%

¹PSA: prostate specific antigen.
²Pain is scored as a linear analog scale (0-10 range).

TABLE 11

Baseline characteristics of mBC patients.	
n = 16	
DEMOGRAPHICS	
Median age (range)	61 (48-81)
Race, Ethnicity	
White, non-Hispanic	44%
Black, non-Hispanic	50%
Asian, non-hispanic	6%
BASELINE DISEASE HISTORY	
ER and/or PR positive disease	56%
HER2 positive disease (HER2 3+)	0%
Median Karnofsky Performance Status (range)	90 (70-90)
Median Number of Prior Endocrine Therapies (range)	1 (0-4)
Median Number of Prior Chemotherapies	2 (0-7)
SITES OF METASTATIC DISEASE	
Visceral (lung or liver)	75%
Lymph Node Only	0%
Lymph Node, soft tissue, or contralateral breast only	13%

TABLE 11-continued	
Baseline characteristics of mBC patients.	
	n = 16
Bone metastases only:	
Bone Metastatic With Lymph Nodes (no visceral metastases)	0%
Bone Metastatic Without Lymph Nodes (no visceral metastases)	13%

[0195] Among ten men with mCRPC, CTCs co-expressed vimentin and CK in 10/10 (100%) patients, and by this criterion 108/126 (86%) of enumerated CTCs were transitional (Table 12, FIGS. 14A-14E). Biopsies of bony metastases performed within one week of CTC collection in two of these patients revealed no vimentin expression in the CK positive tumor foci, but strong vimentin expression in the surrounding bone stroma, which lacks CK expression. These same patients had CTCs taken at the same time as the CT-guided tumor biopsy that commonly expressed co-expressed CK and vimentin. These findings are consistent with invasion and metastasis by transitional CTCs that subsequently undergo MET; alternatively, vimentin expression may be heterogeneously expressed in metastases, similar to CTC expression.

TABLE 12		
Circulating tumor cell (CTC) and transitional CTCs in patients with metastatic CRPC.		
Subject Number	CTC Count (Cellsearch) ⁱ	Ratio: Vimentin (+) CTCs/ Total Manual CTC Count ⁱⁱ
1	5	4/6
2	4	2/2
3	54	11/11
4	45	6/10
5	626	5/8
6	110	17/21
7	182	5/6
8	17	13/16
9	19	33/34
10	34	12/12
Total	1127	108/126 (86%)
Subject Number	CTC Count (Cellsearch)	Ratio: N-Cadherin (+) CTCs/ Total Manual CTC Count
11	45	13/19
12	12	5/7
13	10	8/8
14	5	7/8
15	12	3/4
16	220	11/13
17	828	81/96
18	26	6/11
19	12	18/22
20	42	15/18
Total	1224	167/206 (81%)
Subject Number	CTC Count (Cellsearch)	Ratio: CD133 (+) CTCs/ Total Manual CTC Count
21	485	38/38
22	16	6/11
23	91	15/21
24	6	0/0
25	36	29/29

TABLE 12-continued		
Circulating tumor cell (CTC) and transitional CTCs in patients with metastatic CRPC.		
26	27	9/9
27	43	10/15
28	2	0/0
29	23	12/14
30	38	23/26
31	30	12/17
Total	797	154/180 (86%)

ⁱThe middle column represents the CTC Count from the FDA-approved CELLSEARCH® enumeration of CTCs for each subject.
ⁱⁱRight column represents the ratio of vimentin (co-expression of vimentin ranged from 60-100% of cells in a given individual and did not correlate with CTC count ($R^2 = 0.11$)), N-cadherin (Co-expression of N-cadherin ranged from 55-100% of cells in a given individual, and did not correlate with CTC count ($R^2 = -0.09$)), or CD133 (CD133 co-expression ranged from 55-100% of evaluable cells in a given individual and did not correlate with CTC number ($R^2 = 0.04$)) expressing CTCs among the total number of CTCs that were manually enumerated. A CTC was defined as an intact DAPI positive (nucleated) cell that lacked CD45 expression and expressed cytokeratin.

TABLE 13		
CTCs and transitional CTCs in patients with mBC.		
Subject Number	CTC Count (Cellsearch) ⁱ	Ratio: Vimentin (+) CTCs/ Total Manual CTC Count ⁱⁱ
1	21	0/6
2	7	2/2
3	8	4/4
4	21	1/2
5	12	2/2
6	188	21/22
7	324	29/33
8	377	6/23
9	0	0/0
10	3	0/3
Total	961	65/97 (67%)
Subject Number	CTC Count (Cellsearch)	Ratio: N-Cadherin (+) CTCs/ Total Manual CTC Count
11	1062	9/13
12	2	0/3
13	147	52/59
14	6	2/5
15	33	15/15
16	2	0/0
Total	1252	78/95 (82%)

[0196] Among the next cohort of 10 men with mCRPC, CTCs co-expressed N-cadherin and CK in 10/10 (100%) patients, and by this criterion 167/206 (81%) of CTCs were identified as transitional (Table 12, FIGS. 15A-15E). Among 10 women with mBC, nine had detectable CTCs and of these, we found evidence of vimentin co-expression in seven (78%) patients, and 55/88 CTCs overall (63%) co-expressed vimentin (Table 13, FIGS. 14A-14E). Among another six women with detectable CTCs and mBC, four had evidence of CK and N-cadherin co-expression, and overall 78/95 CTCs (82%) had N-cadherin expression, with significant heterogeneity in expression in a given individual (Table 13, FIGS. 15A-15E). These data indicate that many CTCs in patients with mBC and mCRPC co-express epithelial (Ep-CAM and cytokeratin) and mesenchymal (vimentin, N-cadherin) markers, and thus exist in a transitional phenotypic state, similar to that observed in our preclinical models.
[0197] Given the expression of the stem cell associated antigen CD133 in transitional AT3-T cells, CD133 expres-

sion in CTCs from men with mCRPC was evaluated. CD133 was expressed in 11/11 (100%) men with CTCs, and in 154/180 (86%) of CTCs from these men (Table 12, FIGS. 16A-16D). These data suggest that CTCs from patients with common epithelial malignancies inhabit transitional states characterized by co-expression of epithelial and mesenchymal markers as well as CD133, biomarkers that have been associated with stem-like properties, invasiveness, and chemoresistance.

Example 9. CTC Isolation

[0198] Antibody selection. For cell capture, several antibodies against the extracellular domain of OB-cadherin or N-cadherin were tested using positive control (PC3) and negative control (LNCaP) cell lines and analyzed by flow cytometry. The antibodies with the highest signal in the PC3 cells and minimum background with LNCaP cells were selected for conjugation to magnetic particles. The anti-OB-cadherin antibody (R&D Systems, clone 283416) and anti-N-cadherin antibody (Santa Cruz Biotechnology, D-4) were conjugated to magnetic particles using procedures similar to previously described methods (Allard et al. (2004) Clin Cancer Res 10:6897-6904; see also U.S. Pat. No. 6,645,731).

[0199] Mesenchymal capture assay. A mesenchymal capture assay was used with the CELLSEARCH® platform (Veridex LLC), including the CELLTRACKS® AUTOPREP® for sample preparation and the CELLTRACKS ANALYZER II® for analysis of the captured cells (Allard et al. (2004) Clin Cancer Res 10:6897-6904). The mesenchymal capture assay included ferrofluid coated with anti-OB-cadherin antibodies or anti-N-cadherin antibodies to produce a capture reagent, which would immunomagnetically enrich mesenchymal cells, and staining reagents, such as a phycoerythrin (PE)-conjugated antibody that binds to β -catenin (clone L54E2 from Cell Signaling Technology, Inc.); an antibody to CD45 conjugated to allophycocyanin (APC); and nuclear dye 4',6-diamidino-2-phenylindole (DAPI) to fluorescently label the cells. The assay included buffers to wash, permeabilize, and resuspend the cells such as the following components from the CELLSEARCH® kit: capture enhancement reagent, permeabilization reagent, cell fixative, and dilution buffer.

[0200] Sample Preparation. 7.5 mL of blood was transferred to 15 mL CELLTRACKS® AUTOPREP® sample tubes and mixed with 6.5 mL of dilution buffer, centrifuged at 800 g for 10 min, and then placed on the CELLTRACKS® AUTOPREP® (Veridex LLC) for automated sample preparation using the mesenchymal capture assay. After aspiration of the plasma and buffer layer by the instrument, ferrofluid was added. After the incubation period and subsequent magnetic separation, unbound cells and remaining plasma were aspirated. The target cells were enriched, fluorescently labeled, and resuspended using the CELLTRACKS® AUTOPREP® in the MAGNEST® Cell Presentation Device (Veridex LLC). The magnetic field generated by the MAGNEST® device caused the magnetically labeled cells to distribute uniformly over the analysis surface of the cartridge for analysis using the CELLTRACKS ANALYZER II®.

[0201] Sample Analysis. After ferromagnetic antibody capture, the staining reagents were added, along with a permeabilization buffer, to fluorescently label the immunomagnetically enriched cells. The cells were stained with

labeled anti- β -catenin antibodies, labeled anti-CD45 antibodies, and DAPI for visualization. The anti- β -catenin antibodies was used to identify mesenchymal CTCs. After incubation on the system, the magnetic separation was repeated, and excess staining reagents were aspirated. In the final processing step, the cells were resuspended in the MAGNEST® device, which included a chamber and two magnets that oriented the immunomagnetically labeled cells for analysis using the CELLTRACKS ANALYZER II®.

[0202] The MAGNEST® was placed on the CELLTRACKS ANALYZER II®, a four-color semi-automated fluorescence microscope. Image frames covering the entire surface of the cartridge for each of the four fluorescent filter cubes were captured. The captured images that contained PE as well as DAPI positive events in the same frame were presented in a web-enabled browser gallery for classification of the events based on cell fluorescence and morphology. The final selection of the cells was made by the operator. The criteria to classify the object as a mesenchymal cell (designated as “events”) included round to oval morphology, an intact cell greater than 4 μ m with a visible nucleus (DAPI positive), positive staining for anti- β -catenin-PE, and negative staining for anti-CD45-APC, as depicted in FIG. 17. Results of cell enumeration were expressed as the number of cells per 7.5 mL of blood.

[0203] An isolated “CTC” using this assay was defined as a β -catenin positive, CD45 negative, nucleated intact cell, based on our preliminary data that beta-catenin was visualized in tumor cells but not in leukocytes, as illustrated in FIGS. 18A-18E. In comparison, 50-75% of EpCAM-captured CTCs stained for β -catenin, as shown in FIG. 19 (Bitting et al. (2012) J Clin Oncol 30S:abstr 10533). A mesenchymal CTC phenotype was identified using the OB-cadherin antibody or N-cadherin antibody ferrofluid to capture CTCs with positive β -catenin expression, lack of CD45 expression, and positive nuclear DAPI staining to characterize the cells.

[0204] To determine if circulating mesenchymal-like tumor cells meeting the above criteria were present in healthy individuals, blood was drawn from healthy adults age over 18 years of age into 10 mL EDTA tubes. Subjects were not eligible if they had any chronic medical condition requiring medication or a history of cancer. Samples were processed as described above within 8 hours of blood collection. All subjects were enrolled using an institutional review board-approved protocol and provided informed consent. The OB-cadherin-captured, β -catenin-positive events (“rare events”) were detected in healthy volunteers. The rare events detected in healthy volunteers were stained with the endothelial marker CD31 (BD Biosciences, clone WM59) for further characterization as previously described (Pusztaszeri et al. (2006) J Histochem Cytochem 54:385-395). All detected events in healthy volunteers were CD31 positive, which indicates that these cells may represent endothelial cells. Examples of the OB-cadherin captured, β -catenin-positive and CD31-positive events in healthy individuals are shown in FIGS. 20A-20B.

[0205] After establishing a threshold for detection in healthy volunteers (zero OB-cadherin, β -catenin-positive cells if CD31 was included as an additional characterization marker), the prevalence of these cellular events was determined in men with progressive, metastatic CRPC that were enrolled in a correlative clinical blood-drawing study prior to initiating a new systemic therapy. See Table 14. This

population largely was composed of men with bone metastases (>90%) and would thus theoretically be enriched for OB-cadherin positive cells, if present.

TABLE 14	
Baseline characteristics of CRPC patients	Results (n = 10)
Median age, years (range)	68 (57-74)
Race	
Caucasian, n (%)	7 (70)
Black, n (%)	3 (30)
Karnofsky performance status, median (range)	90 (80-100)
Gleason score, median (range)	8 (7-10)
Pain score >4, n (%)	4 (40)
Initial local therapy	
Prostatectomy, n (%)	3 (30)
External beam radiation, n (%)	3 (30)
None, n (%)	4 (40)
Laboratory values	
PSA ng/mL, median (range)	408 (7-4377)
LDH U/L, median (range)	220 (206-291)
Hemoglobin g/dL, median (range)	908 (8.8-12.1)
Alkaline phosphate U/L, median (range)	197 (57-463)
CTC count, median (range)	34 (1-1000)
Sites of metastasis	
Bone, n (%)	10 (100)
Liver, n (%)	2 (20)
Lung, n (%)	4 (40)
Lymph nodes only	0
Prior therapies	
Number of hormonal therapies, median (range)	4 (1-5)
Abiraterone, MDV3100, or TAK700, n (%)	7 (70)
Siputeucel-T, n (%)	3 (30)
Docetaxel, n (%)	8 (80)
Cabazitaxel, n (%)	2 (20)
>1 chemotherapy, n (%)	2 (20)
Bone targeted therapy, n (%)	9 (90)
Palliative radiation, n (%)	3 (30)

TABLE 14-continued	
Baseline characteristics of CRPC patients	Results (n = 10)
Type of Progression prior to study enrollment	
Imaging	8 (80)
Clinical (symptoms, PSA increase)	2 (20)

[0206] One CELLSAVE® and two EDTA 10 mL tubes of blood were collected at baseline, at treatment cycle 3, and at progression. Blood obtained in EDTA tubes was used for OB-cadherin capture in duplicate and was processed as soon as possible, with no more than 8 hrs elapsing from the time of collection. Blood obtained in the CELLSAVE® tubes was used for the standard EpCAM capture only and was processed within 72 hours, per the established protocol (Allard et al. (2004) Clin Cancer Res 10:6897-6904). Using the CELLSEARCH® system, circulating mesenchymal cells were captured with the anti-OB-cadherin ferrofluid, then permeabilized and stained for further characterization, as described above. Cells were enumerated per 7.5 mL of blood, and the mesenchymal CTC enumeration was compared with the standard EpCAM-based capture method. Any discrepancy in the scoring of events was resolved with discussion between two independent reviewers.

[0207] The EpCAM-based capture assay detected more tumor cells in the majority of patients, however there were exceptions in which more mesenchymal events were seen. Also, OB-cadherin capture resulted in more sheets or clumps of cells and multiple cells per field than EpCAM capture, as illustrated in FIG. 22.

[0208] As shown in Table 15, rare events were detected in healthy volunteers using the mesenchymal capture methods. In CRPC patients, OB-cadherin capture (“O-cad capture”) detected more events in 3 of 5 subjects than EpCAM-based capture, and the majority of captured cells were cytokeratin negative.

TABLE 15			
	O-cad capture, beta-catenin+	N-cad capture, beta-catenin+	EpCAM capture, cytokeratin+
Healthy Volunteers	0-51 events (mean 5.95) n = 21	0-4 events (mean 0.28) n = 25	NA
CRPC patients	0-465 events (mean 138.4) n =5	0-2 events (mean 1.2) n = 5	1-123 CTCs (mean 50.8) n = 5

[0209] Table 16 shows events captured of cells in the blood from CRPC patients using the N-cadherin (“N-cad”) and OB-cadherin (“O-cad”) ferrofluid. The number of cells captured from the CRPC patients appeared to be higher in some patients as compared with the EpCAM-based technology, and the rate of detection appeared higher than that in healthy volunteers for at least OB-cadherin capture.

TABLE 16							
Subject	N-cad Capture		O-cad capture		EpCAM Capture		CELLSEARCH ®
	#events	CD31+	#events	CD31+	#events	CD31+	standard-of-care
1	2	na	4	na	102	na	46
2	1	na	458	na	71	na	45
3	2	na	3	na	2	na	7
4	0	—	0	—	31	na	50

TABLE 16-continued

Subject	N-cad Capture		O-cad capture		EpCAM Capture		CELLSEARCH ® standard-of-care
	#events	CD31+	#events	CD31+	#events	CD31+	
5	1	na	220	na	17	na	23
6	1	1	3	0	1351	na	>1000
7	na	na	1	0	0	—	1
8	2	0	0	—	53	1	22
9	0	—	29	29	9	0	12
10	na	na	2	1	111	1	69

Example 10

[0210] To verify that the mesenchymal capture assay using anti-OB-cadherin antibodies could isolate and detect the cells of interest, spiking studies of positive and negative control cells were performed. Preliminary data investigating OB-cadherin expression on the prostate cancer PC-3 cell line revealed OB-cadherin expression on approximately 40-50% of PC3 cells by both immunofluorescence and flow cytometry, as shown in FIGS. 23A-23B.

[0211] The prostate cancer cell line PC-3 was cultured in flasks containing DMEM high-glucose supplemented with 10% fetal calf serum and subsequently harvested using cell dissociation buffer (Gibco Cat. No. 13150-016) per package insert. Cells were counted on a hemocytometer and either 500 or 1,000 cells were spiked into 7.5 mL of blood obtained from healthy volunteers as described above. A median of 31.4% (range 16.1-103.4, n=13) of spiked cells were recovered using OB-cadherin capture and characterization as beta-catenin positive, CD45 negative cells as above.

[0212] The collection of the mesenchymal phenotypic CTC left open the possibility that cellular events were not cancer cells, but rather host cells. For example, in healthy volunteers, the OB-cadherin positive cells were most likely endothelial cells derived from phlebotomy as they were CD31 positive. In men with metastatic CRPC (mCRPC), these OB-cadherin positive cells could be endothelial or circulating osteoblasts, bone marrow derived mesenchymal cells, or other circulating mesenchymal-like cells expressing OB-cadherin. To determine the significance of OB-cadherin captured events from patients and whether these cellular events represented host cells or prostate cancer cells, DNA fluorescence in situ hybridization (FISH) was performed for prostate cancer-specific genomic events. Cellular events were identified using the CELLTRACKS ANALYZER II® as described above and then fixed and dried on the cartridge for DNA FISH. Using a 4-color FISH assay for androgen receptor (AR) amplification, PTEN loss, and gene fusion involving the TMPRSS2-ERG locus (ERG break-apart assay), as previously described (Attard et al. (2009) Cancer Res 69:2912-2918), the captured mesenchymal cells were evaluated for these prostate-cancer specific changes. As shown in FIGS. 24A-24C, AR amplification and the TMPRSS2-ERG fusion were present in both EpCAM and OB-cadherin captured cells, indicating that these cellular events were tumor-derived.

[0213] As these genomic amplification or deletion events were unlikely to be found in normal tissues, and were found in the EpCAM positive cells, these findings suggest that at least some of the OB-cadherin, beta-catenin positive cells were prostate cancer derived and not derived from the tumor microenvironment or normal host cells. Further analysis of the prevalence of these mesenchymal phenotypic CTCs in a broader population of men with mCRPC and other tumor types indicate the usefulness of this assay to complement existing CTC assays. These analyses may include comparison with EpCAM based approaches, particularly in men with low CTC counts despite progressive metastatic disease, to define the clinical utility of OB-cadherin positive events. In addition, correlations of OB-cadherin positive cellular events with clinical and pathologic characteristics and patient outcomes with systemic therapies further define the independent role of this assay in the context of other prognostic biomarkers. These methods and data suggested that these cells were detectable in men with mCRPC, were absent in healthy volunteers, and that the disclosed methods detected OB-cadherin positive human prostate cancer cells in blood.

[0214] The common expression of OB-cadherin in CTCs suggested osteomimicry, and provided some insight into the mechanism of prostate cancer homing to bone and the development of osteoblastic bone metastases. OB-cadherin events in healthy volunteers uniformly expressed the endothelial marker CD31, whereas the CD31 status in patients with cancer was more variable. Additional markers may be used to confirm that these mesenchymal-like cells were tumor cells (e.g., FISH, cytokeratin, PSA).

[0215] It is understood that the foregoing detailed description and accompanying examples are merely illustrative and are not to be taken as limitations upon the scope of the invention, which is defined solely by the appended claims and their equivalents.

[0216] Various changes and modifications to the disclosed embodiments will be apparent to those skilled in the art. Such changes and modifications, including without limitation those relating to the chemical structures, substituents, derivatives, intermediates, syntheses, compositions, formulations, or methods of use of the invention, may be made without departing from the spirit and scope thereof.

SEQUENCE LISTING

Sequence total quantity: 24
SEQ ID NO: 1 moltype = DNA length = 4321

-continued

FEATURE		Location/Qualifiers				
source		1..4321				
		mol_type = genomic DNA				
		organism = Mus musculus				
SEQUENCE: 1						
cacacacaca	cgcacacaca	cacacacaca	cactttctcgg	cgcgcacgac	gcccgccectt	60
ctccccgccc	cctccccagc	tccttgatct	cccgtctgtt	ttattactcc	tggtgcgagt	120
ccggcggact	ccgaggcccc	ctatttgtta	ccaactcgct	ctcattggcg	gggaggagag	180
cagcggagaa	gggggtgggg	aggggagggg	aagggaaggg	gtggccactg	ccggagccga	240
ctccgcgctg	ctgttggtgc	cgtctgcgct	tctgtgcctg	ctgctgccgc	cgccgcgcgc	300
tccggctcct	cgctcggccc	ctctccgcct	ccatgtgccg	gatagcggga	gcgcgcggga	360
ccctgctgcc	gctttctggcg	gccttgcttc	aggcgtctgt	ggaggcttct	ggtgaaattg	420
cattatgcaa	gactggattt	cctgaagatg	tttacagcgc	agtcttaccg	aaggatgtgc	480
acgaaggaca	gcccccttct	aatgtgaaat	tcagcaactg	caatagaaaa	aggaaagttc	540
agtatgaaag	cagcgagcca	gcagatttca	aggtggacga	ggacggcacg	gtgtatgctg	600
tgagaagctt	ccctctcact	gcagagcagg	caaagttcct	gatatatgcc	caagacaaaag	660
aaacccagga	aaagtggcag	gtagctgtaa	acctgagccg	ggagccaacc	ctgactgagg	720
agcctatgaa	ggaaccacat	gaaattgaag	aaatagtatt	ccctagacaa	cttgccaagc	780
acagtggagc	tctacaaagg	cagaagagag	actgggtcat	ccgcaccaatc	aacttgccag	840
aaaactccag	aggacccttt	cctcaagagc	ttgtcagaat	caggtctgat	agagataaaa	900
acctttccct	gagatacagc	gtcactgggc	caggagctga	ccagcctcca	acgggcatct	960
tcattatcaa	ccccatctca	ggacagctgt	cagtcacaaa	gcctctggat	cgagagctga	1020
tagccccggt	tacttgaga	gcacatgcag	tggacatcaa	tggcaatcaa	gtggagaacc	1080
ccattgacat	tgtcatcaat	gttattgaca	tgaatgataa	cagacctgag	tttctgcacc	1140
agggtttgaa	tgggtctgtt	ccagagggat	caaagcctgg	gacgtatgtg	atgacgggtca	1200
ctgccattga	tgcggatgat	ccaaatgccc	tgaatggaat	gctgcggtac	aggatcctgt	1260
cccaggcgcc	cagcacacct	tcacccaaca	tgtttacaat	caacaatgag	actggggaca	1320
tcatcactgt	ggcagctggg	ctggatcgag	agaaagtgca	acagtatacg	ttaataattc	1380
aagccacaga	catggaaggc	aatcccactt	atggcctttc	aaacacagcc	acagccgtca	1440
tcacggtgac	agatgtcaat	gacaatcctc	cagagtttac	tgccatgact	ttctacggag	1500
aagtccctga	gaacaggggtg	gacgtcattg	tagccaacct	aactgtcacg	gacaaagatc	1560
agccccacac	gccggcctgg	aatgcggcat	acagaatcag	tggtggagac	cctacaggaa	1620
gggtttgccat	cctgacagac	cccaacagca	atgatgggct	agtcacagtg	gtaaaaccaa	1680
ttgactttga	aacgaatagg	atgtttgtcc	ttactgttgc	tgcagaaaac	caagtgccat	1740
tagctaaagg	cattcagcac	ccacctcagt	cgacagccac	tgtgtctgtg	acagttattg	1800
atgtcaatga	aaatccttat	tttgcccaaa	atcctaaaat	cattcgccaa	gaggaaggcc	1860
tccacgcagg	tacctgctg	accacgctca	ctgctcagga	ccccgatcga	tatatgcaac	1920
agaatatcag	atacacaaaa	ttgtctgatc	ctgccactg	gctgaaaata	gaccccgtag	1980
atgggcagat	cactactatt	gccgttttgg	acagagaatc	gccaaatgta	aaaaacaaca	2040
tctataatgc	taccttccct	gcttctgaca	atggaatccc	gcctatgagt	gggacaggaa	2100
cactgcaaat	ctatttactt	gatatacaat	acaacgcccc	tcagggtgta	cctcaagagg	2160
cggagacctg	tgaaactcca	gaacccaact	caattaacat	cacagcactt	gattatgaca	2220
tagacccaaa	cgccggggcg	ttcgcgtttg	atcttccctt	atctccagtg	actattaaaa	2280
gaaactggac	catcaaccgg	cttaatgggtg	attttgctca	gctcaattta	aagataaaaat	2340
ttttggaagc	tggtatctat	gaagttccca	tcattatcac	agattcaggg	aatcccccca	2400
agtccaacat	ttccatcctg	cgtgtgaaag	tttgtcagtg	tgactccaat	ggagactgca	2460
cggacgtgga	caggatcgtg	gggtgcagggc	ttggcacggg	cgccatcatc	gctatccttc	2520
tgtgtatcat	catcctgctg	atccttgttc	tcattgtttg	ggtatggatg	aaacggcggg	2580
ataaagagcg	ccaagccaag	cagcttttaa	ttgacccaga	agatgatgta	agagataata	2640
tattgaaata	tgatgaagaa	ggtggaggag	aagaagacca	ggactatgac	ttgagccagc	2700
tccagcaacc	agatactgtg	gagcctgatg	ccatcaagcc	cgtgggaatc	agacggctag	2760
acgagaggcc	tatccatgct	gagccacagt	accagtcgg	atccgcagcc	ccacaccctg	2820
gggatattgg	ggacttcatt	aatgagggcc	ttaaagctgc	tgacaacgac	cccacggcgc	2880
caccgtatga	ctccctctta	gtctttgact	acgagggcag	cggctccacg	gctggctcct	2940
tgagctccct	caactcctcc	agtagcgggtg	gggaccagga	ctatgactac	ctgaatgact	3000
ggggaccccc	cttcaagaaa	ctggcggaca	tgtacggcgg	tggtgacgac	tgaacggcag	3060
gacggacttg	gcttttgtag	aaagtatgaac	agtttcacct	gatattccca	aaaaaaagca	3120
tacagaagct	aggctttaac	tctgtagtcc	actagcaccg	tgcttgctgg	aggetttggc	3180
gtaggctgcg	aaccagtttg	ggctcccagg	gaatatcagt	gatccaatac	tgtctggaaa	3240
acaccgagct	cagctacact	tgaattttac	agtaaagaag	cactgggatt	tatgtgcctt	3300
tttgtacctt	tttcagattg	gaattagttt	tctgttttaa	gctttaatgg	tactgatttc	3360
tgaaatgata	aggaaaagac	aaaatatatt	gtggcgggag	cagaaaagta	aatgtgatac	3420
gcttcaaccc	acttttggtt	caatgcattt	gcttttggtt	agatacagaa	cgaaacaacc	3480
agattaaaaa	aaattaactc	atggagtgat	tttgttacct	ttgggggtggg	ggggatgaga	3540
ccacaagata	ggaaaatgta	cattacttct	agtttttagac	tttagatttt	tttttttcac	3600
taaaatctta	aaacttacgc	agctgggttg	agataaaggg	agttttcata	tcaccaattt	3660
gtagcaaaat	gaattttttc	ataaactaga	atgttagaca	cattttgggtc	ttaatccatg	3720
tacacttttt	tattttctgt	attttttcca	cctcgctgta	aaaatggtgt	gtgtacataa	3780
tgtttatcag	catagactat	ggaggagtgc	agagaactcg	gaacatgtgt	atgtattatt	3840
attggactttg	tttcaggttt	tttgcattgt	aatatccttc	gttatgggta	aagtatttac	3900
aaaaacaaagt	gacatttgat	tcaactgttg	agctgtagtt	agaatactca	atttttaatt	3960
ttttaatttt	ttttaaat	ttttattttc	tttttggttg	tttcgttttg	gggaggggta	4020
aaagtcttta	gcacaatgtt	ttacataatt	tgtacaaaaa	aaattacaca	caaaaaaaaaa	4080
aaaaaagaaaa	gaaaagaaaa	gtgaaagggg	tggcctgttt	cttgacgac	tagcaagtgt	4140
gtgtttttta	aaaacaaaaa	aaacaaaaca	aaaaataaat	aaaaagagga	aaaagaaaaa	4200
aaaaaaaagct	tttaactcgg	agagacttct	gaaacagctt	tgcgtctgtg	ttgtgtacca	4260

-continued

gaatacaaac	aatacacctc	tgaccccagc	gttctgaata	aaaagcta	tttggatctg	4320
g						4321
SEQ ID NO: 2						
FEATURE						
source						
moltype = AA length = 906						
Location/Qualifiers						
1..906						
mol_type = protein						
organism = Mus musculus						
SEQUENCE: 2						
MCRAGAPRT	LLPLLAALLQ	ASVEASGEIA	LCKTGFPEDV	YSAVLPKDVH	EGQPLLNVKF	60
SNCNRKRKVQ	YESSEPADFK	VDEDGTVYAV	RSFPLTAEQA	KFLIYAQDKE	TQEKWQVAVN	120
LSREPTLTEE	PMKEPHEIEE	IVFPRQLAKH	SGALQRQKRD	WVIPPINLPE	NSRGFPFQEL	180
VRIRSDRDKN	LSLRYSVTGP	GADQPPTGIF	IINPISGQLS	VTKPLDRELI	ARFHLRAHAV	240
DINGNQVENP	IDIVINVIDM	NDNRPEFLHQ	VWNGSVPEGS	KPGTYVMTVT	AIDADDPNAL	300
NGMLRYRILS	QAPSTPSPNM	FTINNETGDI	ITVAAGLDRE	KVQQYTLIIQ	ATDMEGNPTY	360
GLSNTATAVI	TVTVDVNDNP	EFTAMTFYGE	VPENRVDVIV	ANLTVTDKDQ	PHTPAWNAAY	420
RISGGDPTGR	FAILTDPNSN	DGLVTVVKPI	DFETNRMFVL	TVAAENQVPL	AKGIQHPPQS	480
TATVSVTVID	VNENPYFAPN	PKIIRQEEGL	HAGTMLTTLT	AQDPDRYMQQ	NIRYTKLSDP	540
ANWLKIDPVN	GQITTIIVLD	RESPNVKNNI	YNATFLASDN	GIPPMSTGT	LQIYLLDIND	600
NAPQVLPQEA	ETCETPEPNS	INITALDYDI	DPNAGPFAFD	LPLSPVTIKR	NWTINRLNGD	660
FAQNLNLIKIF	LEAGIYEVPI	IITDSGNPPK	SNISILRVKV	CQCDSNGDCT	DVDRIVGAGL	720
GTGAIIAILL	CIIILLILVL	MFVVMKRRD	KERQAKQLLI	DPEDDVRDNI	LKYDEEGGGE	780
EDQDYDLSQL	QQPDTVEPDA	IKPVGIRRLD	ERPIHAEPQY	PVRSAAHPHG	DIGDFINEGL	840
KAADNDPTAP	PYDSLLVFDY	EGSGSTAGSL	SSLNSSSSGG	DQDYDYLNWD	GPRFKKLADM	900
YGGGDD						906
SEQ ID NO: 3						
FEATURE						
source						
moltype = DNA length = 3237						
Location/Qualifiers						
1..3237						
mol_type = genomic DNA						
organism = Xenopus laevis						
SEQUENCE: 3						
tcggcacgag	ctggagtgtg	caggactttt	aagatgctgc	tgggtgtctg	cactgtgtcc	60
atgtgaatgt	ggcattttta	ttttgaattc	cctccggaga	caagatttca	tcaagagttt	120
cctttggata	ttaaagtcaa	gtgcaagcaa	tggagattct	ctataagaag	gcaataatct	180
gggggattta	ctaaaattaa	acaaacagat	tgacattcgc	tggattttat	aagcaatttt	240
gcattttaca	cactaccaaa	aatgaagaaa	gacttttgct	tacacggttt	actttttatg	300
ttgggaattg	cgtattgtag	tcatgccaca	tctttaagaa	aaaacaataa	actaaggcaa	360
tattcccatg	gtcaccatga	aaaaggcaaa	gaagggcaag	ttttacatag	gtcaaagaga	420
ggatgggttt	ggaatcaatt	ttttgttaata	gaagaataca	ccggaccaga	tcctgtactc	480
gttggaaggc	ttcactcaga	tgttgactct	ggagattgga	agataaaaata	catactctca	540
ggagaggggt	ctgggaccat	ttttgtcatt	gatgacaaat	cagggaatat	ccatgcaacc	600
aagaccctgg	atcgagaaga	aagggtcag	tataccttaa	tggctcaggc	agttgacaga	660
gaaacaaata	aaccactgga	accaccatca	gagtttatcg	ttaaagtcca	agacataaat	720
gataatcccc	cggagtcttt	gcatgaaaac	taccacgcaa	atgtgcctga	gatgtccaat	780
gtgggtacat	cagtaattca	agtaacagcc	tctgatgcag	atgatccaac	atatggaaac	840
agcgctaagc	ttgtgtatag	tattctcgaa	gggcagccat	atttttcagt	cgaagcacia	900
tagggaatca	ttaggactgc	ccttccaaac	atggacagag	aagccaagga	agaataccat	960
gttgttattc	aagcaaagga	tatgggagga	catatgggag	gactctcagg	gacaactaaa	1020
gtgacaataa	cgctgacaga	tgtcaatgac	aatccacca	agtttccaca	aagtgcgtac	1080
cccatgtctg	tgtcagaagc	tgtgtgtcca	ggggaagagg	ttggcagaat	aaaagctaaa	1140
gatccagaca	ttggagaaaa	tggcttaata	agtaaccgta	ttcttgaaag	agatggggca	1200
gagatggttg	aaatcacagc	tgattatgta	actcaggaag	gcgttgtaaa	gctaaaaaag	1260
gtgggtggatt	atgaaaccac	gaagttctac	agtatgaagg	ttgaagctgt	caacgttcat	1320
attgatccca	gattccctag	ccggggacca	ttcaaagaca	ctgctactgt	taagatctca	1380
gtagaggatt	ttgatgaacc	gcctattttc	ttagaaagaa	gttacatttt	ggaagtatat	1440
gaaaatgctc	catcggtac	tgtggtcgga	agagtgcacg	ctaaagaccc	agatgctgct	1500
aacagcccaa	ttaggtattc	aatcgatcgc	cacactgacc	ttgacagatt	cttcagcatc	1560
aaccagagg	atggtgtcat	caaaaccaca	aagggtttgg	atagagagga	aagcccttgg	1620
cacaacatct	cagtcattgc	aactgaagtc	cacaatcgaa	ttcatgaaac	tagagttcca	1680
gtagctatta	aagtcttgg	taagaatgac	aatgctccgg	aatttgcaaa	gccctatgaa	1740
gcttttgtct	gtgaaaatgc	tccaatcaat	caggagtttt	tgaccatcac	tgacagtagat	1800
aaagatgata	cagccaatgg	acttcgtttt	ctcttttagt	ttccccccaga	aattgtacat	1860
ccaaatccaa	atttcaccat	aatagacaaa	cgagataaca	cagcaagcat	ccgtgttggc	1920
cgtggagt	tcagccgaca	gaaacaagac	ttgtatttgg	ttcctattgt	tataagtgat	1980
gggggaagcc	caccgatgag	cagcaccaat	accctttctg	tccgaatctg	cagttgcaat	2040
agtgatggat	cccaactatc	ttgtaatgct	gaaccccaat	cccttaacgc	tggaactcagt	2100
actggagcac	tgattgcaat	ccttgcttgc	attgtaattt	tattagtgat	tgtgggtttg	2160
tttgtgactc	tgaggagaga	gaagaaggaa	cctctaattg	tctttgaaga	ggaagatatc	2220
cgggaaaata	taattacata	tgatgatgaa	ggtggtggag	aggaagacac	cgaagcattt	2280
gacattgcaa	cactgcagaa	tcctgatggg	attaatggat	ttatgccacg	gaaagatatc	2340
aaaccogaat	ttcaatataa	ccccagagat	attggaataa	gaccagcacc	aaacagtgtt	2400
gacgttgatg	acttcattaa	cacaaggata	catgaggccg	ataatgaccc	tgacagctccg	2460
ccttatgact	ccattcagat	ctatggatac	gaaggagag	gttctgtggc	tggtctctct	2520
agttcattag	agtcagcctc	tacagattca	gatttggaat	atgattatct	acaaaactgg	2580
ggacctcgat	ttaagaaact	agcaaattta	tatgggtcca	aagacacttg	tgaagatgat	2640

-continued

tcttaacaaa	taagttctga	atttggcctt	atgaactgca	taatgtactg	aaatatccag	2700
agtaaacatt	aacaggtatt	tttttaaagg	aaaacatgaa	aaaggcttct	ttaaccttcc	2760
aaggtttaca	aacaggattc	cttccaaaac	aagaactggt	aaatggtggt	ggatactgtg	2820
aaaaccctat	ggcctgtgta	gaagttgtgt	attcattttt	ttttttgttt	tttgtttttt	2880
ttccaagaaa	ccacttgtaa	aatgcagcct	atttaaggga	atggaaatgc	aggaaaaacg	2940
caacaaaaaa	ggggaatctt	tacagtatta	aacataacca	tcaaactctc	tcaaacaaag	3000
cttccacaca	aaaaaaaaaa	aagataacag	ttttgagctg	taatttcgcc	ttaaactatg	3060
gacactttat	atgtagtgca	tttttaaact	tgaaaaaaat	atatatataa	tatccagcca	3120
gcttcaatcc	atataatgta	tgtacagtaa	aatgtacaat	tattctgtct	cttgagcatc	3180
agacttgtta	ctgctgattc	ttgtaaatct	tttttgctta	taatcccctc	gtgccga	3237
SEQ ID NO: 4	moltype = AA length = 794					
FEATURE	Location/Qualifiers					
source	1..794					
	mol_type = protein					
	organism = Xenopus laevis					
SEQUENCE: 4						
MKKDFCLHGL	LLCLGIAYCS	HATSLRKNK	LRQSFHGHHE	KGKEGQVLHR	SKRGWVWNQF	60
FVIEEYTGPD	PVLVGR LHSD	VDSGDWKIKY	ILSGEGAGTI	FVIDDKSGNI	HATKTL DREE	120
RAQYTLMAQA	VDRETNKPLE	PPSEFIVKVQ	DINDNPPEFL	HENYHANVPE	MSNVGTSVIQ	180
VTASDADDPT	YGNSAKLVYS	ILEGQPYFSV	EAQSGIIRTA	LPNMDREAKE	EYHVVIQAKD	240
MGGHMGGLSG	TTKVTITLTD	VNDNPPKFPQ	SAYPMSVSEA	AVPGEEVGRI	KAKDPDIGEN	300
GLIKYRILEG	DGAEMFEITA	DYVTQEGVVK	LKKVVDYETK	KFYSMKVEAV	NVHIDPRFLS	360
RGPFKDTATV	KISVEDFDEP	PIFLERSYIL	EVYENAPSDT	VVGRVHAKDP	DAANSPIRYS	420
IDRHTDLDRF	FSINPEDGVI	KTTKGLDREE	SPWHNISVIA	TEVHNRIHET	RVPVAIKVLD	480
KNDNAPEFAK	PYEAFVCENA	PINQEFLTIT	AVDKDDTANG	LRFLFSFPPE	IVHPNPNFTI	540
IDKRDNTASI	RVGRGVFSRQ	KQDLYLVPIV	ISDGGSPMS	STNTLSVRIC	SCNSDGSQLS	600
CNAEPQSLNA	GLSTGALIAI	LACIVILLVI	VVLVFTLRRE	KKEPLIVFEE	EDIRENIITY	660
DDEGGGEEDT	EAFDIATLQN	PDGINGFMPR	KDIKPEFQYN	PRDIGIRPAP	NSVDVDDFIN	720
TRIHEADNDP	AAPPYDSIQI	YGYEGRGSVA	GSLSSLESAS	TDSLDLDYDL	QNWGPRFKKL	780
ANLYGSKDTC	EDDS					794
SEQ ID NO: 5	moltype = DNA length = 3701					
FEATURE	Location/Qualifiers					
source	1..3701					
	mol_type = genomic DNA					
	organism = Mus musculus					
SEQUENCE: 5						
gtccaatcag	tgcgctcaga	ctcagagccc	taggctcctg	ctctttaaat	taccgagcct	60
tgtggagacc	cgggcacctg	gccttaagct	cagccctgag	gatggactct	tgagtgaatg	120
accaccttgg	agaccgttct	tctgtttccc	ttgttaccag	ccaggaggca	gaagagtcca	180
ccggtccagg	aaagacccat	ttcccttgag	tttccagaaa	gtacctcatg	cttgagagat	240
caggccaaca	actatggctc	tcgtcttcag	tgccctgctg	ttactggggc	tgtgtggaaa	300
gatctcttca	gaaggtcagc	ctgcattcca	taacactcct	ggggctatga	attatgaatt	360
gcctaccacc	aaatatgaga	cccaagatac	cttcaatgct	gggattgttg	gccctctcta	420
caaaatggtg	cacatcttcc	tcaacgtggg	ccagccgaat	gacttccctc	tagatttgat	480
caaaaaactc	atacagaaca	agaactttga	catctcagtt	gattccaagg	agccagaaat	540
catagtcttg	gctctgaaga	ttgccctcta	tgagatcgga	gtccttatct	gcgccatcct	600
gggactgctg	ttcattatcc	tcattgcctc	ggtgggctgc	ttcttttgta	tgtgccgttg	660
ctgcaacaaa	tgcggcggag	agatgcacca	gcggcagaag	cagaatgcgc	catgcaggag	720
gaagtgcttg	ggcctctccc	tcctgggtgat	ttgtctgctc	atgagccttg	gcattatata	780
tggctttgtg	gctaaccagc	agaccaggac	tcggatcaaa	gggaccaga	aactggcaaa	840
gagcaatttc	agagactttc	aaacactcct	gactgaaaca	ccaaagcaaa	ttgactatgt	900
agtggagcag	tacaccaaca	ccaagaacaa	ggcattctca	gacctggatg	gcacggctc	960
cgtgctggga	ggcagaataa	aggaccaact	aaaacccaaa	gtaactcctg	tcctcgaaga	1020
gattaaggcc	atggcgacag	ccatcaaaac	gaccaaggat	gccctgcaga	acatgagcag	1080
cagcctgaaa	agtctccaag	atgcagccac	ccagctcaat	accaacctga	gctctgtgag	1140
aaacagcatc	gagaattcgc	tcagcagcag	tgactgtacc	tcagatccag	ccagcaagat	1200
ctgcgatagc	atcagaccaa	gcctaagcag	tctggggagc	agcctcaatt	caagtcaagc	1260
cccacagtg	gatagagaac	tcaacactgt	tactgaagtc	gacaaaactg	atctggagag	1320
cctcgtcaaa	aggggggtata	cgacaattga	tgaaataccc	aatacaatac	aaaaccaaac	1380
tgtggatgtc	atcaaagacg	tcaaaaatac	cttggactcc	attagctcca	acattaagga	1440
catgagccaa	agtattccta	ttgaggatat	gctgttacag	gtctcccatt	accttaataa	1500
cagcaacaga	tacttaaacc	aggagctgcc	caagctggaa	gaatatgact	cgtactgggtg	1560
gctgggtggc	ttgattgtct	gctttctgct	gactctcatt	gtgaccttct	ttttcctggg	1620
cttgctgtgt	ggtgtgtttg	gctatgacaa	gcatgccacc	ccaactagaa	gaggctgtgt	1680
gtccaacact	ggaggcatct	tcctcatggc	tggggttgga	ttcggcttcc	ttttttgctg	1740
gatattgatg	atccttgttg	ttcttaactt	tgttgtttgt	gcaaagtgtg	aaaagttgct	1800
ctgcgaacct	tatgaaaaga	agaaattatt	acaggttttg	gacactccct	atctgctcaa	1860
ggaacaatgg	caattttatc	tttctggcat	gctattcaat	aaccagaca	ttaacatgac	1920
ctttgagcaa	gtctacaggg	attgcaaaa	aggtcgagggt	atatatgctg	cttttcagct	1980
tgagaatgtc	gtcaacgtca	gtgatcattt	caacattgac	cagatttctg	aaaacataaa	2040
tacggagtgtg	gaaaacctga	atgtgaacat	tgatagcatt	gaactgttgg	ataacacagg	2100
aaggaagagc	ctcgaggact	ttgcacattc	tgggatagat	acaatcgatt	attccacata	2160
cttgaaggag	actgagaaat	cccctactga	agtgaatctg	ctgacatttg	cctctaccct	2220
ggaagcaaaa	gcaaaccagt	tgctgaagg	aaagctgaaa	caggccttct	tactggatgt	2280

-continued

acagaatata	agagccatcc	accagcatct	cctccctcct	gtgcagcaat	caactgaaatt	2340
tgtgaggggtg	aggaatacgt	taagacaaaag	tgtctggacc	ctccagcaaa	caagcaacaa	2400
gttgccggag	aaagtgaaga	agatccttgc	ctctttggac	tctgttcagc	atttcctcac	2460
caataacggt	tccctcatcg	ttatcgggga	aacgaagaag	tttgggaaaa	caatactagg	2520
ctactttgaa	cattatctgc	actgggtcct	ttatgccatc	acagagaaga	tgacatcctg	2580
caaaccatg	gccaccgcga	tggactctgc	tgttaatggc	attctgtgtg	gctatgttgc	2640
ggaccctctg	aatttgttct	ggttcggcat	agggaaagcc	acggtgctct	tacttcgggc	2700
tgtaatcatt	gctatcaagc	tggccaagta	ctatcgcagg	atggattcag	aggatgtata	2760
cgacgaccgc	tctcgatact	gacaactgga	gttgaagctg	cttgaacaac	aagatagtca	2820
acatggaaaag	catcacagat	tttggatagt	ttctgagtct	tctagaacgt	tccaagtgca	2880
gaagaaacct	ggtggagact	caggcgggca	ctaggaacat	ggcatcagt	gtcttagggg	2940
agcactttgt	caggaatgaa	cagtcatcat	ggtataatc	cacatatcca	ttgcaactca	3000
tgaatgattc	tctcctgttt	tgtttttaac	tttctttttt	acactgattt	tctatttaga	3060
caactaaaaca	tataggggtg	cttattcccc	ctggatacat	ttacctgtga	accagctatt	3120
ccggtgtcat	agctgggtac	ctaacttact	tccatatgtg	aagtgtgcta	aacacaaaacc	3180
agtttacaga	agagatgtat	tttgtgtata	gtaaactgta	tataaccct	tttaccacag	3240
tcagttttttt	aaacaaatga	atactctaga	tttttcttct	aaatgaggtt	actggtgggg	3300
tggttgtgac	ctagtgtatg	tgtagaaagg	agtctgcatt	caactaaaagt	gtgtcaacct	3360
agagcaggca	atgcccttcc	ttgtggattt	ctgtctgctc	gttttggagc	tacctgcggg	3420
ttagaaatag	aattcaagaa	caatcacgga	gtttccact	tgatgccact	gccaaagtca	3480
gaacaagggga	tcttgagaga	aggaactgtc	gctcagctgg	gagcggaatc	attatcgcaa	3540
tcacagggtcc	tggttcacag	tttagtggca	ctctctgggt	tgtaagaatg	ggcattacgt	3600
tcagtgtcat	ctggtcatct	gtgatgtgtg	tcatcagcct	gtcctgatgt	tgagatttaa	3660
aataaagcat	gaatgaacag	aaaaaaaaaa	aaaaaaaaaa	a		3701
SEQ ID NO: 6						
FEATURE						
source						
moltype = AA length = 842						
Location/Qualifiers						
1..842						
mol_type = protein						
organism = Mus musculus						
SEQUENCE: 6						
MALVFSALLL	LGLCGKISSE	GQPAFHNTPG	AMNYELPTTK	YETQDTFNAG	IVGPLYKMHV	60
IFLNVVQPN	FPLDLIKKLI	QNKNFDISVD	SKEPEIIVLA	LKIALYEIGV	LICAILGLLF	120
IILMPLVGC	FCMCRCCKNC	GGEMHQRQKQ	NAPCRRKCLG	LSLLVICLLM	SLGIIYGFVA	180
NQOTRTRIK	GQKLAKSNFR	DFQTLTETP	KQIDYVVEQY	TNTKNKAFSD	LDGIGSVLGG	240
RIKDQLKPK	TPVLEEIKAM	ATAIKQTKDA	LQNMSSSLKS	LQDAATQLNT	NLSSVRNSIE	300
NSLSSSDCT	DPASKICDSI	RPSLSSLGSS	LNSSQLPSVD	RELNTVTEVD	KTDLESVLR	360
GYTTIDEIP	TIQNQTVDVI	KDVKNLTDSI	SSNIKDMSQS	IPIEDMLLQV	SHYLNNSNRY	420
LNQELPKLE	YDSYWWLGL	IVCFLLTLIV	TFFFLGLLCG	VFGYDKHATP	TRRGCVSNTG	480
GIFLMAGVG	GFLFCWILMI	LVVLTFFVGA	NVEKLLCEPY	ENKLLQVLD	TPYLLKEQWQ	540
FYLSGMLFN	PDINMTFEQV	YRDCKRGRGI	YAAFQLENV	NVSDHFNIDQ	ISENINTELE	600
NLNVNIDSIE	LLDNTGRKSL	EDFAHSGIDT	IDYSTYLKET	EKSPTEVNLL	TFASTLEAKA	660
NQLPEGKLG	AFLLDVQNIR	AIHQHLPPV	QKSLKQVVR	NTLRQSVWTL	QQTSNKLPEK	720
VKKILASLDS	VQHFLTNNVS	LIVIGETKKF	GKTLIGYFEH	YLHWVFYAIT	EKMTSCPKMA	780
TAMDSAVNGI	LCGYVADPLN	LFWFGIGKAT	VLLLPAVIAA	IKLAKYYRRM	DSEDVYDDPS	840
RY						842
SEQ ID NO: 7						
FEATURE						
source						
moltype = DNA length = 3306						
Location/Qualifiers						
1..3306						
mol_type = genomic DNA						
organism = Mus musculus						
SEQUENCE: 7						
gaattcccg	gcggccgcca	gagctccggc	ccgggggctg	cctgtgtgtt	cctggcccgg	60
cgtagcgact	gctctccggg	ctggcggggg	ccggcggtga	gcccgggcct	cagcgcttct	120
gagcgctgcg	agtgttcact	actcgccagc	aaagtttgga	gtaggcaacg	caagctccag	180
tccctttctt	tgctgctgcc	cagatccgag	agcagctccg	gtgtatgtct	agctgttctg	240
cgatcccggc	gcgcgtgaag	cctcggaacc	ttggcgccgg	ctgctacca	aggaatcggt	300
ctcttttttg	agttttcctc	cgagatcatc	gcctgtctca	tcccgatcca	ctctgggctc	360
cggcgcagca	ccgagcgag	aggagcgctg	ccattcaagt	ggcagccaca	gcagcagcag	420
cagcagcagt	gggagcagga	acagcagtaa	caacagcaac	agcagcacag	ccgcctcaga	480
gctttgctcc	tgagccctcg	tgggctgaag	gcattgcagg	tagcccatgg	tctcagaaga	540
agtgtgcaga	tgggattacc	gtccacgtgg	agatatggaa	gaggaccagg	gattggcact	600
gtgaccatgg	tcagctgggg	gcgcttcctc	tgctgtgtct	tggtcaccat	ggcaaccttg	660
tccctggccc	ggcctcctt	cagtttagtt	gaggatacca	ctttagaacc	agaagagcca	720
ccaaccaaat	accaaactc	ccaaccagaa	gcgtacgtgg	ttgcccccg	ggaatcgcta	780
gagttgcagt	gcatgttgaa	agatgccgcc	gtgatcagtt	ggactaagga	tggggtgcac	840
ttggggccca	acaataggac	agtgtctatt	ggggagtatc	tccagataaa	aggtgccaca	900
cctagagact	ccggcctcta	tgcttgtagt	gcagctagga	cggtagacag	tgaaacttgg	960
atcttcatgg	tgaatgtcac	agatgccatc	tcactctggg	atgatgagga	cgacacagat	1020
agctccgaag	acgttgtcag	tgagaacagg	agcaaccaga	gagcaccgta	ctggaccaac	1080
accgagaaga	tggagaagcg	gctccacgct	tgtcctgccg	ccaacactgt	gaagttccgc	1140
tgtccggctg	gggggaatcc	aacgtccaca	atgaggtggg	taaaaaacgg	gaaggagttt	1200
aagcaggagc	atcgcatagg	aggctataag	gtacgaaacc	agcactggag	ccttattatg	1260
gaaagtgtgg	tcccgtcaga	caaaggcaac	tacacctgcc	tgggtggagaa	tgaatacggg	1320
tccatcaacc	acacctacca	cctggatgtc	gttgaacggt	caccacaccg	tcccatcctc	1380

-continued

caagctggac	tgcttgcaaa	tgctccacg	gtggtcggag	gggatgtgga	gtttgtctgc	1440
aaggtttaca	gcatgcccc	gccccacatc	cagtggatca	agcacgtgga	aaagaacggc	1500
agtaaaaacg	ggcctgatgg	gctgccctac	ctcaagggttc	tgaaagctgc	cggtgttaac	1560
accacggaca	aagagattga	ggttctctat	attcggaatg	taacttttga	ggatgctggg	1620
gaatatacgt	gcttggcggg	taattctatc	gggatatacct	ttcactctgc	atggttgaca	1680
gttctgccag	cgctgtgag	agagaaggag	atcacggctt	ccccagatta	tctggagata	1740
gctatttact	gcataggggt	cttcttaatc	gcctgcatgg	tggtgacagt	catcttttgc	1800
cgaatgaaga	ccacgaccac	gaagccagac	ttcagcagcc	agccagctgt	gcacaagctg	1860
accaagcgca	tccccctgcg	gagacaggta	acagtttcgg	ccgagtccag	ctcctccatg	1920
aactccaaca	ccccgctggt	gaggataaca	acgcgtctgt	cctcaacagc	ggacaccccc	1980
atgctagcag	gggtctccga	gtatgagttg	ccagaggatc	caaagtggga	attccccaga	2040
gataagctga	cgctgggcaa	acccctgggg	gaaggttgct	tcgggcaagt	agtcattggc	2100
gaagcagtg	gaatcgataa	agacaaaacc	aaggaggcgg	tcaccgtggc	agtgaagatg	2160
ttgaaagatg	atgccacaga	gaaggacctg	tctgatctgg	tatcagagat	ggagatgatg	2220
aagatgattg	ggaaacataa	gaacattatc	aaacctcctg	gggcctgcac	gcaggatgga	2280
cctctctacg	tcatagttga	atatgcacg	aaaggcaacc	tccgggaata	cctccgagcc	2340
cggaggccac	ctggcatgga	gtactcctat	gacattaacc	gtgtccccga	ggagcagatg	2400
accttcaagg	acttggtgtc	ctgcacctac	cagctggcta	gaggcatgga	gtacttggt	2460
tccccaaaa	gtatccatcg	agatttggct	gccagaaacg	tgttggtaac	agaaaaaca	2520
gtgatgaaga	tagcagactt	tggcctggcc	agggatatca	acaacataga	ctactataaa	2580
aagaccacaa	atgggcgact	tccagtcaag	tggatggctc	ctgaagccct	ttttgataga	2640
gtttacactc	atcagagcga	tgtctggctc	ttcgggggtg	taatgtggga	gatctttact	2700
ttagggggct	caccctaccc	agggattccc	gtggagggaac	tttttaagct	gctcaaagag	2760
ggacacagga	tggacaagcc	caccaactgc	accaatgaac	tgtacatgat	gatgagggat	2820
tgttggcatg	ctgtaccctc	acagagaccc	acattcaagc	agtttggtcg	agacttggat	2880
cgaattctga	ctctcacaac	caatgaggaa	tacttggatc	tcaccagcc	tctcgaacag	2940
tattctccta	gttaccgccg	cacaagtagc	tcttgttctt	caggggacga	ttctgtgttt	3000
tctccagacc	ccatgcctta	tgaaccctgt	ctgcctcagt	atccacacat	aaacggcagt	3060
gttaaaacat	gagtgaatgt	gtcttcctgt	ccccaaacag	gacagcacca	ggaacctact	3120
tacactgagc	agagaggctg	tgtccagag	cctgtgacac	gcctccactt	gtatatatgg	3180
atcagaggag	taaatagtgg	gaagcatatt	tgtcacgtgt	gtaaagattt	atacagttgg	3240
aacatgtact	acaggaagga	gactgttctg	atagtgcag	ccgccaccat	gccacctttg	3300
accaca						3306
SEQ ID NO: 8						
FEATURE						
source						
moltype = AA length = 821						
Location/Qualifiers						
1..821						
mol_type = protein						
organism = Mus musculus						
SEQUENCE: 8						
MVSWGRFICL	VLVTMATLSL	ARPSFSLVED	TITLEPEEPPT	KYQISQPEAY	VVAPGESLEL	60
QCMLKDAAVI	SWTKDGVHLG	PNNRTVLIGE	YLQIKGATPR	DSGLYACTAA	RTVDSETWIF	120
MVNVTDIAIS	GDDDDTDSS	EDVVSERNRN	QRAPYWTNTE	KMEKRLHACP	AANTVKFRCP	180
AGGNPTSTMR	WLKNGKEFKQ	EHRIGGYKVR	NQHWSLIMES	VVPSDKGNYT	CLVENEYGSI	240
NHTYHLDVVE	RSPHRPILQA	GLPANASTVV	GGDVEFVCKV	YSDAQPHIQW	IKHVEKNGSK	300
NGPDGLPYLK	VLKAAGVNTT	DKEIEVLYIR	NVTFEDAGEY	TCLAGNSIGI	SFHSAWLTVL	360
PAPVREKEIT	ASPDYLEIAI	YCIGVFLIAC	MVVTVIFCRM	KTTTKKPDFS	SQPAVHKLTK	420
RIPLRQVTV	SAESSSMNS	NTPLVRITTR	LSSTADTPML	AGVSEYELPE	DPKWEFPRDK	480
LTGKPLGEG	CFGQVMAEA	VGIDKDKPKE	AVTVAVKMLK	DDATEKDLS	LVSEMEMMKM	540
IGKHKNIIIN	LGAQTQDGPL	YVIVEYASKG	NLREYLRRAR	PPGMEYSYDI	NRVPEEQMTF	600
KDLVSCITYQ	ARGMEYLASQ	KCIHRDLAAR	NVLVTENNVM	KIADFGGLAR	INNIDYYKKT	660
TNGRLPVKWM	APEALFDRVY	THQSDVWSFG	VLMWEIFTLG	GSPYPGIPVE	ELFKLLKEGH	720
RMDKPTNCTN	ELYMMMRDCW	HAVPSQRPTF	KQLVEDLDRI	LTLTTNEEYL	DLTQPLEQYS	780
PSYPDTSSSC	SSGDDSVFSP	DPMPYEPCLP	QYPHINGSVK	T		821
SEQ ID NO: 9						
FEATURE						
source						
moltype = DNA length = 3037						
Location/Qualifiers						
1..3037						
mol_type = genomic DNA						
organism = Mus musculus						
SEQUENCE: 9						
ggcgagggga	gagagccggg	agagggcagc	ggcgccggcg	caggcgcgga	acggggcgac	60
ggacgatcga	acgcgcggcc	gccagagctc	cgcgcggggg	gctgcctgtg	tgttctctgg	120
ccggcggtgg	gactgctctc	cgggctggcg	ggggccgggc	gtgagcccg	gcctcagcgt	180
tcctgagcgc	tgcgagtgtt	cactactcgc	cagcaaaagt	tggagtaggc	aacgccaaag	240
tccagtcctt	tcttctgctg	ctgcccagat	ccgagagcag	ctccggtgtc	atgtcctagc	300
tgttctgcga	tccccggcgc	gcgtgaagcc	tcggaacctt	cgcgccggct	gctacccaag	360
gaatcgttct	ctttttggag	ttttcctccg	agatcatcgc	ctgctccatc	ccgatccact	420
ctgggctccg	gcgcagaccg	agcgagaggg	agcgctgcca	ttcaagtggc	agccacagca	480
gcagcagcag	cagcagtggg	agcaggaaca	gcagtaacaa	cagcaacagc	agcacagccg	540
cctcagagct	ttggctcctg	agccccctgt	gggtcgaagg	cattgcaggt	agcccatggt	600
ctcagaagaa	gtgtgcagat	gggattaccg	tccacgtgga	gatatggaag	aggaccaggg	660
attggcactg	tgaccatggt	cagctggggg	cgcttcatct	gcctggtctt	ggtcaccatg	720
gcaaccttgt	ccctggcccc	gcccctcttc	agtttagttg	aggataccac	tttagaacca	780
gaaggagcac	cgtactggac	caacaccgag	aagatggaga	agcggtcca	cgtgtccct	840
cccgccaa	ctgtgaagtt	ccgctgtccg	gctgggggga	atccaacgcc	cacaatgagg	900

-continued

tggttaaaaa	acgggaagga	gtttaagcag	gagcatcgca	ttggaggcta	taaggtagca	960
aaccagcact	ggagccttat	tatggaaagt	gtggtcccgt	cagacaaagg	caactacacc	1020
tgcttggtgg	agaatgaata	cgggtccatc	aaccacacct	accacctcga	tgctcgttgaa	1080
cggtcaccac	accggcccat	cctccaagct	ggactgcctg	caaatgcctc	cacggtggtc	1140
ggaggggatg	tggagtttgt	ctgcaagggt	tacagcgatg	cccagcccca	catccagtgg	1200
atcaagcacg	tggaaaagaa	cggcagtaaa	tacgggcctg	atgggctgcc	ctacctcaag	1260
gtcctgaagc	actcggggat	aaatagctcc	aatgcagaag	tgctggctct	gttcaatgtg	1320
acggagatgg	atgctgggga	atatatatgt	aaggtctcca	attatatagg	gcaggccaac	1380
cagtctgcct	ggctcactgt	cctgccccaa	cagcaagcgc	ctgtgagaga	gaaggagatc	1440
acggcttccc	cagattatct	ggagatagct	atttactgca	taggggtctt	cttaatcgcc	1500
tgcatggtgg	tgacagtcac	cttttgccga	atgaagacca	cgaccaagaa	gccagacttc	1560
agcagccagc	cagctgtgca	caagctgacc	aagcgcaccc	ccctgcggag	acaggtaaca	1620
gtttcggccg	agtccagctc	ctccatgaac	tccaacaccc	cgctggtgag	gataacaacg	1680
cgtctgtcct	caacagcgga	caccccgatg	ctagcagggg	tctccgagta	tgagttgccca	1740
gaggatccaa	agtgggaatt	ccccagagat	aagctgacgc	tgggcaaacc	cctgggggaa	1800
ggttgcttcg	ggcaagtagt	catggctgaa	gcagtgggaa	tcgataaaga	caaacccaag	1860
gaggcgggtca	ccgtggcagt	gaagatgttg	aaagatgatg	ccacagagaa	ggacctgtct	1920
gatctgggat	cagagatgga	gatgatgaag	atgattggga	aacataagaa	cattatcaac	1980
ctcctggggg	cctgcacgca	ggatggacct	ctctacgtca	tagttgaata	tgcatcgaaa	2040
ggcaacctcc	gggaataact	cagagcccgg	agggccacctg	gcatggagta	ctcctatgac	2100
attaaccgtg	tccccgagga	gcagatgacc	ttcaaggact	tggtgtcctg	cacctaccag	2160
ctggctagag	gcatggagta	cttggcttcc	caaaaatgta	tccatcgaga	tttggctgcc	2220
agaaacgtgt	tggtaacaga	aaacaatgtg	atgaagatag	cagactttgg	cctggccagg	2280
gatatcaaca	acatagacta	ctataaaaag	accacaaatg	ggcgacttcc	agtcaagtgg	2340
atggctcctg	aagccctttt	tgatagagtt	tacactcatc	agagcgatgt	ctggtccttc	2400
ggggtgttaa	tgtgggagat	ctttacttta	gggggctcac	cctaccagag	gattcccgtg	2460
gaggaacttt	ttaagctgct	caaagaggga	cacaggatgg	acaagcccac	caactgcacc	2520
aatgaactgt	acatgatgat	gagggattgc	tggcatgctg	tacctcaca	gagaccaca	2580
ttcaagcagt	tggtcgaaga	cttggatcga	attctgactc	tcacaaccaa	tgaggaatac	2640
ttggatctca	cccagcctct	cgaacagtat	tctcctagtt	accccgacac	aaggagctct	2700
tgttcttcag	gggacgattc	tgtgttttct	ccagacccca	tgctttatga	acctgtctct	2760
cctcagtatc	cacacataaa	cggcagtggt	aaaacatgag	tgaatgtgtc	ttcctgtccc	2820
caaacaggac	agcaccagga	acctacttac	actgagcaga	gaggctgtct	cagagcctgt	2880
gacacgcctc	cacttgtata	tatggatcac	aggagtaaat	agtgggaagc	atattgtcac	2940
gtgtgtaaag	atttatacac	ttcggaaaca	tgttacctaa	ccaggaaaag	aagactgttt	3000
tctgataag	tggacagccg	caagccacca	tgccacc			3037
SEQ ID NO: 10 moltype = AA length = 707						
FEATURE Location/Qualifiers						
source 1..707						
mol_type = protein						
organism = Mus musculus						
SEQUENCE: 10						
MVSWGRFICL	VLVTMATLSL	ARPSFSLVED	TITLEPEGAPY	WTNTEKMEKR	LHAVPAANTV	60
KFRCPAGGNP	TPTMRWLKNG	KEFKQEHRI	GYKVRNQHWS	LIMESVVP	SGD KGN	120
EYGSINHTYH	LDVVERSPHR	PILQAGLPAN	ASTVVG	GDVE FVCKV	YSDAQ	180
KNGSKYGP	PDG LPYLKVLKHS	GINSSNAEVL	ALFNV	TEMDA GEYICKV	SNY IGQANQ	240
TVLPKQQAPV	REKEITASPD	YLEIAIY	CIG VFLIAC	MVVT VIFCRM	KTTT KKP	300
VHKLTKRIPL	RRQVT	VS AES SSSMNS	NTPL VRIT	TRLSST ADT	PMLAGVS	360
EFPRDKLTLG	KPLGEGCFGQ	VVMAEAVGID	KDKPK	EAVTV AVKML	KDDAT EK	420
MEMMKMIGKH	KNIINLLGAC	TQDGPLYVIV	EYASKGNLRE	YLRARR	PPGM EYSYD	480
EEQMTFKDLV	SCTYQLARGM	EYLASQKCIH	RDLAARNVLV	TENNVMKIAD	FGLARDINNI	540
DYYKKT	TNGR LPVKWMAPEA	LFDRVYTHQS	DVWSFGVLMW	EIFTLGGSPY	PGIPVEELFK	600
LLKEGHRMDK	PTNCTNELYM	MMRDCWHAVP	SQRPTFKQLV	EDLDRILTLT	TNEEYLDLTQ	660
PLEQYSPSYP	DTRSSCSSGD	DSVFS	PDMP YEPCLPQYPH	INGSVKT		707
SEQ ID NO: 11 moltype = DNA length = 3344						
FEATURE Location/Qualifiers						
source 1..3344						
mol_type = genomic DNA						
organism = Xenopus tropicalis						
SEQUENCE: 11						
agagcagggga	agtacagcgc	tgcgctacaa	gaactgagca	aacgagcaga	aaagtagaca	60
ttcctgatcc	ttcgggtcttt	ccaaaagtcc	ccaatgggggt	cacacaggcc	atggttactt	120
ggtgctgtgg	tgctgctggc	actccttcag	gtacagggag	gactggcaga	atggacacag	180
tgtcaaatgg	gatttttcaa	ggaaaggtac	agcttttcgg	tacctaaaga	cttggagaca	240
gacaaagcac	tgggtagagt	gatctttaac	agctgtgagg	gaccagtgag	aattcagttt	300
gcctctaag	atcctaattt	tgaaattcac	aaagatggca	cagtttatgt	taagaatcct	360
accaagatga	aagacaacag	aaaaacattc	cgtgtcctgg	cttgggagaa	tcaaggatcat	420
gtatactcta	ccagtgtaac	cttgaaaggg	gaagggcatc	accataagca	ggacatttct	480
tctgtgaaac	attcccacca	cccaaatct	gagactgggt	taaaaagaca	aaaaagagac	540
tggtgtgatc	caccaatcgt	aacatctgag	aatgaaaagg	gcccatttcc	caaacggctt	600
gtgcagatca	agtccagtaa	tgcaaaggaa	atcaaggttt	tttacagtat	cacaggccag	660
ggtgccgata	cccctccaga	aggagtgttc	actattggac	gggaggatgg	atggctaaat	720
gtgacacgac	ctttggacag	agaagccatt	gatagttaca	ctcttttttc	tcatgctgtg	780
tcagtaaatg	ggcaaaatgt	ggaagatccc	atggaaatcc	aaattaaagt	acaagatcag	840

-continued

aatgataatg	acccagtttt	cacacaggag	gtctttgaag	gctatgtgcc	tgaagggctc	900
aagccaggta	cgcccgatcat	gactgtatct	gcaacagatg	ccgatgatgc	tatagacatg	960
tacaatgggtg	tgattactta	ctccattctc	aaccaagacc	ctaaagagcc	caacaatcaa	1020
atgttccacta	ttgattccca	gtctgggttg	atcagcgtag	ttacaactgg	attagacaga	1080
gagaaaatac	cagtgtacac	actgactatt	caagctgcag	atggagaatt	tgggaaagat	1140
cgcacaacaa	ctgcaaaagc	tgtgatcatt	gtgacagaca	ccaatgataa	ccctcctgtg	1200
tttaacccaa	cgcaatacat	tgcagagggt	cctgaaaatg	aagttggata	tgagggttgca	1260
cgtcttacgg	taacagatgc	agatattgaa	gggtcagatg	cctggaatgc	tgtgtacaag	1320
atcattaaag	gaaatgaggc	tggctttttc	agcatccaaa	cagatattga	caacattggg	1380
ctactgaaaa	cagtgaaggg	tctggactat	gagctgaaga	agcagtatat	tctgtcagtc	1440
attgtgacaa	acaaagctaa	cttttctgtt	ccactacaaa	cttcaactgc	aacggtcact	1500
gtaactgtca	cagatgtgaa	tgaggcccca	gtatttgtac	cagtgttgaa	agacgtgtct	1560
gtgccagagg	atctgccag	tggccaagtt	gttgctacct	ataccgcaca	ggatccagac	1620
aaggaacaga	accagaaaat	aagttacttc	attggaaaatg	acccagcagg	gtgggtgtct	1680
gtgaacagag	ataatgggat	tgtcactgga	aatggaaaact	tggatcgga	atcaaagttt	1740
gtgctaaaca	acacctacaa	agtcataatc	ttggccgctg	acagtggcac	tccttctgcc	1800
actgggactg	gaacccttgt	gcttaatctc	attgatgtta	atgataatgg	cccatTTTTg	1860
gatccccaac	aaaatagttt	gtgccagaag	gatccaggct	ttcgtgtatt	taatatcatt	1920
gacaaagatc	tttaccctaa	cacataccca	tatacagtag	acctgactgg	tgaatccaat	1980
gaaaactgga	ctgctacagt	gacagaacag	agtttacttg	agctgagacc	taaaaaggaa	2040
ctggatattg	gacgatacga	agttttgatc	tcattgagag	acaatcaggg	actgacagat	2100
gtgacaaagc	tacagattac	aatctgtcaa	tgtaatgggtg	accaaagtca	atgtgaggaa	2160
aaggctgctc	aagcaggagg	tttggggata	tcagccatag	ttggaatcct	tggagggatc	2220
ctagcgcttc	ttttattggt	gttgctgctc	ttactgtttg	tacgacgaaa	gaaagtggta	2280
aaagaacctt	tattaccacc	agaagatgag	actcgggaca	atgtattttt	ctatgatgaa	2340
gaaggcggtg	gtgaggaaga	ccaggatttt	gatctaagcc	agcttcaccg	tgggtctagat	2400
gctcgccag	atataatccg	taatgatgtc	gttccagttt	tagctgctcc	ccagtatcga	2460
ccccgtcctg	ccaatccaga	tgaaattgga	aatttctattg	atgagaactt	gcatgcagct	2520
gacaatgacc	ccactgctcc	tccatacgac	tcgctccttg	tgttcgatta	cgaaggcagt	2580
ggctctgagg	ccgcatcact	cagctctctt	aactcttcca	actctgattt	agatcaggat	2640
tacagtgcct	tgaataactg	gggacctcgt	ttcaccaaac	tggcagaaat	gtatggagga	2700
gatgaggatt	agaatgtgca	ctgcaatacc	attttgatcc	taaacagtaa	actaaaaacc	2760
ataatttgtg	atgcagctct	tgggaattcac	tttgttttct	cctgctctta	aaacagagat	2820
aaggactgct	caaaagttac	tcctcctgct	tttgtaaaaat	cgttcaaaaa	tattttatgt	2880
atatgtatat	atgaaaaaat	cgtatttttt	gtactatttg	tgttcttata	tccttgcaat	2940
ttgtaataca	agaggatctt	tatctgctta	attataaata	taaaatgccc	gatatgattc	3000
actatgattt	taatgtgttg	agaaatcttt	ttttaaaaag	gtttccagac	acctgacgct	3060
tggaagggaa	ttccataaaa	atataattga	attgggggga	gattgtgttt	tgccatggtc	3120
tgatatacat	tttcatatat	atacatatga	tcattcacag	agtacagtca	acatttggaa	3180
tttgatgagc	ttgctggtcg	aactgaaaaa	aaaatgtatt	atagctgggg	taaaaattaa	3240
tgtatgagct	aaatggggca	caattttgat	atctctgcat	ttgtatttta	cttggcatgt	3300
atacttttgt	aataaaaataa	agatatacat	taatatacaa	cata		3344
SEQ ID NO: 12						
FEATURE						
source						
moltype = AA length = 872						
Location/Qualifiers						
1..872						
mol_type = protein						
organism = Xenopus tropicalis						
SEQUENCE: 12						
MGSHRPWLLG	AVVLLALLQV	QGGLAEWTQC	QMGFSKERYS	FSVPKNLETD	KALGRVIFNS	60
CEGPNVRIQFA	SKDPNFEIHK	DGTVYVKNPT	KMKDNRKTFR	VLAWENQGHV	YSTSVTLKGE	120
GHHHKQDISS	VKHSHHPKSE	TGLKRQKRDW	VIPPIVTSEN	EKGFPFKRLV	QIKSSNAKEI	180
KVFYSITGQG	ADTPPEGVFT	IGREDGWLNV	TRPLDREAI	SYTLFSAVS	VNGQNVEDPM	240
EIQIKVQDQN	DNDPVFTQEV	FEQYVPEGSK	PGTPVMTVSA	TDADDAIDMY	NGVITYSILN	300
QDPKEPNQM	FTIDSQSLI	SVVTTGLDRE	KIPVYTLTIQ	AADGEFGKDR	TTAKAVIIV	360
TDNDNPPVF	NPTQYIAEVP	ENEVGYEVAR	LTVTDADIEG	SDAWNAYKI	IKGNEAGFFS	420
IQTIDIDNIGL	LKTVKGLDYE	LKKQYILSVI	VTKANFSVP	LQTSTATVTV	TVTDVNEAPV	480
FVPVLKDVSV	PEDLPSGQV	ATYTAQDPDK	EQNQKISYFI	GNDPAGWVS	NRDNGIVTGN	540
GNLDRESKFV	LNNTYKVIL	AADSGTPSAT	GTGTLVLNLI	DVNDNGPFLD	PQONSFCQKD	600
PGFRVFNID	KDLYPNTYPY	TVDLTGESNE	NWTATVTEQS	LLELRPKKEL	DIGRYEVLIS	660
LRDNQGLTDV	TKLQITICQC	NGDQMQCEEK	AAQAGGLGIS	AIVGILGGIL	ALLLLLLLLL	720
LFVRRKKVVK	EPLLPPEDET	RDNVFFYDEE	GGGEEDQDFD	LSQLHRGLDA	RPDIIRNDVV	780
PVLAAPQYRP	RPANPDEIGN	FIDENLHAAD	NDPTAPPYDS	LLVFDYEGSG	SEAASLSSLN	840
SSNSDLDDQDY	SALNNWGPRF	TKLAEMYGGD	ED			872
SEQ ID NO: 13						
FEATURE						
source						
moltype = DNA length = 1862						
Location/Qualifiers						
1..1862						
mol_type = genomic DNA						
organism = Homo sapiens						
SEQUENCE: 13						
gtccccgcgc	cagagacgca	gccgcgctcc	caccacccac	acccaccgcg	ccctcgttcg	60
cctcttctcc	gggagccagt	ccgcgccacc	gccgcgcgcc	aggccatcgc	caccctccgc	120
agccatgtcc	accaggtccg	tgtcctcgtc	ctcctaccgc	aggatgttcg	gcggccccgg	180
caccgcgagc	cggccgagct	ccagccggag	ctacgtgact	acgtccaccc	gcacctacag	240
cctgggcagc	gcgctgcgcc	ccagcaccag	ccgcagcctc	tacgcctcgt	ccccgggcgg	300

-continued

cgtgtatgcc	acgcgctcct	ctgcegtgcg	cctgcggagc	agcgtgcccg	gggtgceggct	360
cctgcaggac	tcggtggact	tctcgctggc	cgacgccatc	aacaccgagt	tcaagaacac	420
ccgcaccaac	gagaaggtgg	agctgcagga	gctgaatgac	cgcttcgcca	actacatcga	480
caaggtgcgc	ttcctggagc	agcagaataa	gatcctgctg	gccgagctcg	agcagctcaa	540
gggccaaaggc	aagtcgcgcc	tgggggacct	ctacgaggag	gagatgcggg	agctgcgccg	600
gcaggtggac	cagctaacca	acgacaaagc	ccgcgtcgag	gtggagcgcg	acaacctggc	660
cgaggacatc	atgcgcctcc	gggagaaatt	gcaggaggag	atgcttcaga	gagaggaagc	720
cgaaaacacc	ctgcaatctt	tcagacagga	tgttgacaat	gcgtctcttg	cacgtcttga	780
ccttgaacgc	aaagtggaat	ctttgcaaga	agagattgcc	tttttgaaga	aactccacga	840
agaggaaatc	caggagctgc	aggetcagat	tcaggaaacag	catgtccaaa	tcgatgtgga	900
tgtttccaag	cctgacctca	cggctgccct	gcgtgacgta	cgtcagcaat	atgaaagtgt	960
ggctgccaaag	aacctgcagg	aggcagaaga	atggtacaaa	tccaagtttg	ctgacctctc	1020
tgaggctgcc	aaccggaaca	atgacgccct	gcgccaggca	aagcaggagt	ccactgagta	1080
ccggagacag	gtgcagtccc	tcacctgtga	agtggatgcc	cttaaaggaa	ccaatgagtc	1140
cctggaacgc	cagatgcgtg	aaatggaaga	gaactttgcc	gttgaagctg	ctaactacca	1200
agacactatt	ggccgcctgc	aggatgagat	tcagaatatg	aaggaggaaa	tggctcgtca	1260
ccttcgtgaa	taccaagacc	tgctcaatgt	taagatggcc	cttgacattg	agattgccac	1320
ctacaggaag	ctgctggaag	gcgaggagag	caggatttct	ctgcctcttc	caaacttttc	1380
ctccctgaac	ctgagggaaa	ctaactctgga	ttcactccct	ctggttgata	cccactcaaa	1440
aaggacactt	ctgattaaga	cggttgaaac	tagagatgga	caggttatca	acgaaacttc	1500
tcagcatcac	gatgaccttg	aataaaaaatt	gcacacactc	agtgcagcaa	tatattacca	1560
gcaagaataa	aaaagaaatc	catatcttaa	agaaacagct	ttcaagtgcc	tttctgcagt	1620
ttttcaggag	cgcaagatag	atttggaaata	ggaataagct	ctagttctta	acaaccgaca	1680
ctcctacaag	atttagaaaa	aagttttacaa	cataatctag	tttacagaaa	aatcttgtgc	1740
tagaatactt	tttaaaaggt	attttgaata	ccattaaaac	tgcttttttt	tttccagcaa	1800
gtatccaacc	aacttggttc	tgcttcaata	aatctttgga	aaaactcaaa	aaaaaaaaaa	1860
aa						1862
SEQ ID NO: 14						
FEATURE						
source						
moltype = AA length = 466						
Location/Qualifiers						
1..466						
mol_type = protein						
organism = Homo sapiens						
SEQUENCE: 14						
MSTRSVSSSS	YRRMFGPGT	ASRPSSRSY	VTSTRITYSL	GSALRPSTSR	SLYASSPGGV	60
YATRSSAVRL	RSSVPGVRL	QDSVDFSLAD	AINTEFKNTR	TNEKVELQEL	NDRFANYIDK	120
VRFLEQQNKI	LLAELEQLKG	QGKSRLGLDY	EEEMRELRRQ	VDQLTNDKAR	VEVERDNLAE	180
DIMRLREKLQ	EEMLQREEAE	NTLQSFRQDV	DNASLARLDL	ERKVESLQEE	IAFLKKLHEE	240
ETQELQAQIQ	EQHVQIDVDV	SKPDLTAALR	DVRQYQYESVA	AKNLQEAEEW	YKSKFADLSE	300
AANRNNDALR	QAKQESTEYR	RQVQSLTCEV	DALKGTNESL	ERQMREMEEN	FAVEAANYQD	360
TIGRLQDEIQ	NMKEEMARHL	REYQDLLNVK	MALDIEIATY	RKLLEGEESR	ISLPLPNFSS	420
LNLRETNLDS	LPLVDTHSKR	TLLIKTVETR	DGQVINETSQ	HHDDLE		466
SEQ ID NO: 15						
FEATURE						
source						
moltype = DNA length = 2721						
Location/Qualifiers						
1..2721						
mol_type = genomic DNA						
organism = Homo sapiens						
SEQUENCE: 15						
atgtgccgga	tagcgggagc	gctgcggacc	ctgctgccgc	tgctggcggc	cctgcttcag	60
gcgtctgtag	aggcttctgg	tgaaatcgca	ttatgcaaga	ctggatttcc	tgaagatgtt	120
tacagtgcag	tcttatcgaa	ggatgtgcat	gaaggacagc	ctcttctcaa	tgtgaagttt	180
agcaactgca	atggaaaaag	aaaagtacaa	tatgagagca	gtgagcctgc	agattttaag	240
gtggatgaag	atggcatggt	gtatgccgtg	agaagctttc	cactctcttc	tgagcatgcc	300
aagtctctga	tatatgccc	agacaaagag	accagggaaa	agtggcaagt	ggcagtaaaa	360
ttgagcctga	agccaacctt	aactgaggag	tcagtgaagg	agtcagcaga	agttgaagaa	420
atagtgttcc	caagacaatt	cagtaagcac	agtggccacc	tacaaaggca	gaagagagac	480
tgggtcatcc	ctccaatcaa	cttgccagaa	aactccaggg	gaccttttcc	tcaagagctt	540
gtcaggatca	ggtctgatag	agataaaaa	ctttcactgc	ggtacagtgt	aactggggcca	600
ggagctgacc	agcctccaac	tggtatcttc	attatcaacc	ccatctcggg	tcagctgtcg	660
gtgacaaaagc	ccttgatcgc	cgagcagata	gcccggtttc	atttgagggc	acatgcagta	720
gatattaatg	gaaatcaagt	ggagaacccc	attgacattg	tcataaatgt	tattgacatg	780
aatgacaaca	gacctgagtt	cttacaccag	gtttggaatg	ggacagtctc	tgagggatca	840
aagcctggaa	catatgtgat	gaccgtaaca	gcaattgatg	ctgacgatcc	caatgccctc	900
aatgggatgt	tgaggtacag	aatcgtgtct	caggctccaa	gcacctcttc	acccaacatg	960
tttacaatca	acaatgagac	tggtgacatc	atcacagtgg	cagctggact	tgatcgagaa	1020
aaagtgcaac	agtatacgtt	aataattcaa	gctacagaca	tggaaggcaa	ttccacatat	1080
ggcctttcaa	acacagccac	ggccgtcatc	acagtgacag	atgtcaatga	caatcctcca	1140
gagtttactg	ccatgacgtt	ttatggtgaa	gttcctgaga	acagggtaga	catcatagta	1200
gctaactctaa	ctgtgaccga	taaggatcaa	ccccatacac	cagcctggaa	cgcagtgtac	1260
agaatcagtg	geggagatcc	tactggacgg	ttcgccatcc	agaccgaccc	aaacagcaac	1320
gacgggttag	tcaccgtggt	caaaccaatc	gactttgaaa	caaataggat	gtttgtcctt	1380
actgttgctg	cagaaaatca	agtgccatta	gccaaaggga	ttcagcacc	ccctcagtca	1440
actgcaaccg	tgtctgttac	agttattgac	gtaaatgaaa	acccttattt	tgcccccaat	1500
cctaagatca	ttcgccaaga	agaagggtct	catgccggtg	ccatgttgac	aacattcact	1560
gctcaggacc	cagatcgata	tatgcagcaa	aatattagat	acactaaatt	atctgatcct	1620

-continued

gccaattggc	taaaaataga	tctgtgaat	ggacaaataa	ctacaattgc	tgttttggac	1680
cgagaatcac	caaagtgtgaa	aaacaatata	tataatgcta	ctttccttgc	ttctgacaat	1740
ggaattcctc	ctatgagtgg	aacaggaacg	ctgcagatct	atttacttga	tattaatgac	1800
aatgccccctc	aagtgttacc	tcaagaggca	gagacttgcg	aaactccaga	cccccaattca	1860
attaatatta	cagcacttga	ttatgacatt	gatccaaaatg	ctggaccatt	tgtttttgat	1920
cttcctttat	ctccagtgc	tattaagaga	aattggacca	tcactcggct	taatggtgat	1980
tttgctcagc	ttaatttaaa	gataaaatth	cttgaagctg	gtatctatga	agttcccatc	2040
ataatcacag	attcgggtaa	tcttcccaaa	tcaaatatth	ccatcctgcg	tgtgaagggt	2100
tgccagtgtg	actccaacgg	ggactgcaca	gatgtggaca	ggattgtggg	tgcggggctt	2160
ggcaccgggtg	ccatcattgc	catcctgctc	tgcatcatca	tcttgcttat	ccttggtgctg	2220
atgtttgtgg	tatggatgaa	acgccgggat	aaagaacgcc	aggccaaaca	acttttaatt	2280
gatccagaag	atgatgtaag	agataatatt	ttaaaatatg	atgaagaagg	tggaggagaa	2340
gaagaccagg	actatgactt	gagccagctg	cagcagcctg	acactgtgga	gcctgatgcc	2400
atcaagcctg	tgggaatccg	acgaatggat	gaaagaccca	tccacgccga	gccccagtat	2460
ccggtccgat	ctgcagcccc	acaccctgga	gacattgggg	acttcattaa	tgagggcctt	2520
aaagcggctg	acaatgacct	cacagctcca	ccatagact	ccctgttagt	gtttgactat	2580
gaaggcagtg	gctccactgc	tgggtccttg	agctccctta	attcctcaag	tagtgggtgt	2640
gagcaggact	atgattacct	gaacgactgg	gggccacggg	tcaagaaact	tgtgacatg	2700
tatggtggag	gtgatgactg	a				2721
SEQ ID NO: 16						
FEATURE						
source						
moltype = AA length = 906						
Location/Qualifiers						
1..906						
mol_type = protein						
organism = Homo sapiens						
SEQUENCE: 16						
MCRAGALRT	LLPLLAALLQ	ASVEASGEIA	LCKTGFPEDV	YSAVLSKDVH	EGQPLLNVKF	60
SNCNGKRKVQ	YESSEPADFK	VDEdGMVYAV	RSFPLSSEHA	KFLIYAQDKE	TQEKWQVAVK	120
LSLKPTLTEE	SVKESAEVEE	IVFPRQFSKH	SGHLQRQKRD	WVIPPINLPE	NSRGPFPQEL	180
VRIRSDRDKN	LSLRYSVTGP	GADQPPTGIF	IINPISGQLS	VTKPLDREQI	ARFHLRAHAV	240
DINGNQVENP	IDIVINVIDM	NDNRPEFLHQ	VWNGTVPEGS	KPGTYVMTVT	AIDADDPNAL	300
NGMLRYRIVS	QAPSTPSPNM	FTINNETGDI	ITVAAGLDRE	KVQQYTLIIQ	ATDMEGNPTY	360
GLSNTATAVI	TVTVDVNDNP	EFTAMTFYGE	VPENRVDIIV	ANLTVTDKDQ	PHTPAWNAVY	420
RISGGDPTGR	FAIQTDPNNS	DGLVTVVKPI	DFETNRMFVL	TVAENQVPL	AKGIQHPPQS	480
TATVSVTVID	VNENPYFAPN	PKIIRQEEGL	HAGTMLTTFT	AQDPDRYMQQ	NIRYTKLSDP	540
ANWLKIDPVN	GQITTIIVLD	RESPNVKNNI	YNATFLASDN	GIPPMSTGT	LQIYLLDIND	600
NAPQVLPQEA	ETCETPDNS	INITALDYDI	DPNAGPFAFD	LPLSPVTIKR	NWTITRLNGD	660
FAQLNLKIKF	LEAGIYEVPI	IITDSGNPPK	SNISILRVKV	CQCDSDNGDCT	DVDRIVGAGL	720
GTGAIIAILL	CIIILLILVL	MFVVMKRRD	KERQAKQLLI	DPEDDVRDNI	LKYDEEGGGE	780
EDQDYDLSQL	QQPDTVEPDA	IKPVGIRMD	ERPIHAEPQY	PVRSAPHPG	DIGDFINEGL	840
KAADNDPTAP	PYDSLIVFDY	EGSGSTAGSL	SSLNSSSSGG	EQDYDYLNWD	GPRFKKLADM	900
YGGGDD						906
SEQ ID NO: 17						
FEATURE						
source						
moltype = DNA length = 2391						
Location/Qualifiers						
1..2391						
mol_type = genomic DNA						
organism = Homo sapiens						
SEQUENCE: 17						
atgaaggaga	actactgttt	acaagccgcc	ctggtgtgcc	tgggcatgct	gtgccacagc	60
catgcctttg	ccccagagcg	gcggggggcac	ctgcggccct	ccttccatgg	gcaccatgag	120
aaggggcaagg	aggggcaggt	gctacagcgc	tccaagcgtg	gctgggtctg	gaaccagttc	180
ttcgtgatag	aggagtacac	cgggcctgac	cctgtgcttg	tgggcaggct	tcattcagat	240
attgactctg	gtgatgggaa	cattaaatac	attctctcag	gggaaggagc	tggaaccatt	300
tttgtgattg	atgacaaatc	agggaacatt	catgccacca	agacgttggg	tcgagaagag	360
agagcccagt	acacgttgat	ggctcaggcg	gtggacaggg	acaccaatcg	gccactggag	420
ccaccgtcgg	aattcattgt	caaggtccag	gacattaatg	acaaccctcc	ggagttcctg	480
cacgagacct	atcatgccaa	cgtgcctgag	aggtccaatg	tgggaacgtc	agtaatccag	540
gtgacagctt	cagatgcaga	tgacccact	tatggaaata	gcgccaaagt	agtgtacagt	600
atcctcgaag	gacaacccta	tttttcgggtg	gaagcacaga	caggtatcat	cagaacagcc	660
ctacccaaca	tggacaggga	ggccaaggag	gagtaccacg	tggatgatcca	ggccaaggac	720
atgggtggac	atatgggcgg	actctcaggg	acaaccaaaag	tgacgatcac	actgaccgat	780
gtcaatgaca	accacccaaa	gtttccgcag	agcgtatacc	agatgtctgt	gtcagaagca	840
gccgtccctg	gggagggaagt	agggaagagt	aaagctaaag	atccagacat	tggagaaaaat	900
ggcttagtca	catacaatat	tgttgatgga	gatggatgga	aatcgtttga	aatcacaacg	960
gactatgaaa	cacaggaggg	ggtgataaag	ctgaaaaagc	ctgtagattt	tgaaacccaaa	1020
agagcctata	gcttgaaggt	agaggcagcc	aacgtgcaca	tcgaccggaa	gtttatcagc	1080
aatggccctt	tcaaggacac	tgtgaccgtc	aagatctcag	tagaagatgc	tgatgagccc	1140
cctatgttct	tggccccaag	ttacatccac	gaagtccaag	aaaatgcagc	tgttggcacc	1200
gtggttggga	gagtgcagtc	caaagaccct	gatgctgcca	acagcccgat	aaggatttcc	1260
atcgatcgtc	acactgacct	cgacagattt	ttcactatta	atccagagga	tggttttatt	1320
aaaactacaa	aacctctgga	tagagaggaa	acagcctggc	tcaacatcac	tgtctttgca	1380
gcagaaatcc	acaatcggca	tcagggaagcc	aaagtcccag	tggccattag	ggctccttgat	1440
gtcaacgata	atgctcccaa	gtttgctgcc	ccttatgaag	gtttcatctg	tgagagtgat	1500
cagaccaagc	cactttccaa	ccagccaatt	gttacaatta	gtgcagatga	caaggatgac	1560
acggccaatg	gaccaagatt	tatcttcagc	ctacccccctg	aatcatttca	caatccaaat	1620

-continued

ttcacagtca	gagacaaccg	agataacaca	gcaggcgtgt	acgcccggcg	tggaggggttc	1680
agtcggcgaga	agcaggactt	gtaccttctg	cccatagtga	tcagcgatgg	cggcatcccc	1740
cccatgagta	gcaccaaac	cctcaccatc	aaagtctgcg	ggtgcgacgt	gaacgggggca	1800
ctgctctcct	gcaacgcaga	ggcctacatt	ctgaacgccg	gcctgagcac	aggcgccctg	1860
atcgccatcc	tcgcctgcat	cgtcattctc	ctgggtcattg	tagtattgtt	tgtgaccctg	1920
agaaggcaaa	agaaagaacc	actcattgtc	tttgaggaag	aagatgtccg	tgagaacatc	1980
attactttatg	atgatgaagg	gggtggggaa	gaagacacag	aagcctttga	tattgccacc	2040
ctccagaatc	ctgatggtat	caatggattt	atcccccgca	aagacatcaa	acctgagtat	2100
cagtacatgc	ctagacctgg	gctccggcca	gcgcccacaa	gcgtggatgt	cgatgacttc	2160
aatcaacacga	gaatacagga	ggcagacaat	gacccccagg	ctcctcctta	tgactccatt	2220
caaactctacg	gttatgaagg	caggggctca	gtggccgggt	ccctgagctc	cctagagtcg	2280
gccaccacag	attcagactt	ggactatgat	tatctacaga	actggggacc	tcgttttaag	2340
aaaactagcag	atgtgtatgg	ttccaaagac	acttttgatg	acgattctta	a	2391

SEQ ID NO: 18

FEATURE

source

moltype = AA length = 796

Location/Qualifiers

1..796

mol_type = protein

organism = Homo sapiens

SEQUENCE: 18

MKENYCLQAA	LVCLGMLCHS	HAFAPERRGH	LRPSFHGHHE	KGKEGQVLQR	SKRGWVWNQF	60
FVIEEYTGPD	PVLVGR LHSD	IDSGDGNIKY	ILSGEGAGTI	FVIDDKSGNI	HATKTL DREE	120
RAQYTLMAQA	VDRDTNRPLE	PPSEFIVKVQ	DINDNPPEFL	HETYHANVPE	RSNVGTSVIQ	180
VTASDADDPT	YGNSAKLVYS	ILEGQPYFSV	EAQTGIIRTA	LPNMDREAKE	EYHVVIQAKD	240
MGGHMGGLSG	TTKVITITLD	VNDNPPKFPQ	SVYQMSVSEA	AVPGEEVGRV	KAKDPDIGEN	300
GLVTYNIVDG	DGMESFEITT	DYETQEGVIK	LKKPVDFTK	RAYSLKVEAA	NVHIDPKFIS	360
NGPFKDTVTV	KISVEDADEP	PMFLAPSYIH	EVQENAAAGT	VVGRVHAKDP	DAANSPIRYS	420
IDRHTDLDRF	FTINPEDGFI	KTTKPLDREE	TAWL NITVFA	AEIHN RHQEA	KVPVAIRVLD	480
VNDNAPKFAA	PYEGFICESD	QTKPLSNQPI	VTISADDKDD	TANGPRFIFS	LPPEI IHNPN	540
FTVRDNRDNT	AGVYARRGGF	SRQKQDLYLL	PIVISDGGIP	PMSSTNTLTI	KVCGCDVNGA	600
LLSCNAEAYI	LNAGLSTGAL	IAILACIVIL	LVIVVLFVTL	RRQKKEPLIV	FEEEDVRENI	660
ITYDDEGGGE	EDTEAFDIAT	LQNP DGINGF	IPRKDIKPEY	QYMPRPGLRP	APNSVDVDDF	720
INTRIQEADN	DPTAPPYDSI	QIYGYEGRGS	VAGSLSSLES	ATTDSDL DYD	YLQNWGPRFK	780
KLADLYGSKD	TFDDDS					796

SEQ ID NO: 19

FEATURE

source

moltype = DNA length = 2598

Location/Qualifiers

1..2598

mol_type = genomic DNA

organism = Homo sapiens

SEQUENCE: 19

atggccctcg	tactcggtc	cctgttgctg	ctggggctgt	gcgggaactc	cttttcagga	60
gggcagcctt	catccacaga	tgctcctaag	gcttggaatt	atgaattgcc	tgcaacaaat	120
tatgagaccc	aagactcca	taaagctgga	cccatggca	ttctctttga	actagtgc at	180
atctttctct	atgtgttaca	gccgcgtgat	ttcccagaag	atactttgag	aaaattctta	240
cagaaggcat	atgaatccaa	aattgattat	gacaagccag	aaactgtaat	cttaggtcta	300
aagattgtct	actatgaagc	agggattatt	ctatgctgtg	tcctggggct	gctgtttatt	360
attctgatgc	ctctggtggg	gtatttcttt	tgtatgtgtc	gttgctgtaa	caaatgtgg t	420
ggagaaatgc	accagcgaca	gaaggaaaat	gggcccttcc	tgaggaaatg	ctttgcaatc	480
tcctgtttgg	tgatttgtat	aataataagc	attggcatct	tctatggttt	tgtggcaaat	540
caccaggtaa	gaaccgggat	caaaaggagt	cggaaactgg	cagatagcaa	tttcaaggac	600
tttgcgaactc	tcttgatga	aactccagag	caaatcaaat	atatattggc	ccagtacaac	660
actaccaagg	acaaggcgtt	cacagatctg	aacagtatca	attcagtgtc	aggaggcgga	720
attcttgacc	gactgagacc	caacatcatc	cctgttcttg	atgagattaa	gtccatggca	780
acagcgatca	aggagaccaa	agaggcgttg	gagaacatga	acagcacctt	gaagagcttg	840
caccaacaaa	gtacacagct	tagcagcagt	ctgaccagcg	tgaaaactag	cctgcggtca	900
tctctcaatg	accctctgtg	cttggtgcat	ccatcaagtg	aaacctgcaa	cagcatcaga	960
ttgtctctaa	gccagctgaa	tagcaaccct	gaactgaggc	agcttccacc	cgtggatgca	1020
gaacttgaca	acgttaataa	cgttcttagg	acagatttgg	atggcctgg t	ccaacagggc	1080
tatcaatccc	ttaatgatat	acctgacaga	gtacaacgcc	aaaccacgac	tgtcgtagca	1140
ggtatcaaaa	gggtcttgaa	ttccattggt	tcagatatcg	acaatgtaac	tcagcgctct	1200
cctattcagg	ataactctc	agcattctct	gtttatgtta	ataacactga	aagttacatc	1260
cacagaaatt	tacctacatt	ggaagagtat	gattcatact	ggtggctggg	tggcctggtc	1320
atctgctctc	tgctgaccct	catcgtgatt	ttttactacc	tgggcttact	gtgtggcgtg	1380
tgcggtatg	acaggcatgc	caccccgacc	acccgagggt	gtgtctccaa	caccggaggc	1440
gtcttctcca	tggttgaggt	tggattaagt	ttcctctttt	gctggatatt	gatgatcatt	1500
gtggttctta	cctttgtctt	tggtgcaaat	gtggaaaaac	tgatctgtga	accttacacg	1560
agcaaggaat	tattccgggt	tttgataca	ccctacttac	taaatgaaga	ctgggaatac	1620
tatctctctg	ggaagctatt	taataaatca	aaaatgaagc	tcacttttga	acaagtttac	1680
agtgactgca	aaaaaaatag	aggcacttac	ggcactcttc	acctgcagaa	cagcttcaat	1740
atcagtgaac	atctcaacat	taatgagcat	actggaagca	taagcagtga	attggaaagt	1800
ctgaaggtaa	atcttaatat	ctttctgttg	ggtgcagcag	gaagaaaaaa	ccttcaggat	1860
tttctgctt	gtggaataga	cagaatgaat	tatgacagct	acttggtcca	gactggtaaa	1920
tcccccgag	gagtgaatct	tttatcattt	gcatatgatc	tagaagcaaa	agcaaacagt	1980
ttgccccag	gaaatttgag	gaactccctg	aaaagagatg	cacaaactat	taaaacaatt	2040
caccagcaac	gagtccttcc	tatagaacaa	tcactgagca	ctctatacca	aagcgtcaag	2100

-continued

atacttcaac	gcacagggaa	tggattgttg	gagagagtaa	ctaggattct	agcttctctg	2160
gatttttgc	agaacttcat	cacaaacaat	acttcctctg	ttattattga	ggaaactaag	2220
aagtatggga	gaacaataat	aggatatttt	gaacattatc	tgcagtggat	cgagttctct	2280
atcagtgaga	aagtggcatc	gtgcaaacct	gtggccaccg	ctctagatac	tgctgttgat	2340
gtcttttct	gtagctacat	tatcgacccc	ttgaatttgt	tttggtttgg	cataggaaaa	2400
gctactgtat	ttttacttcc	ggctctaatt	tttgcggtaa	aactggctaa	gtactatcgt	2460
cgaatggatt	cggaggacgt	gtacgatgat	gttgaaacta	tacccatgaa	aaatatggaa	2520
aatggtaata	atgggttatca	taaagatcat	gtatatggta	ttcacaatcc	tgttatgaca	2580
agcccatcac	aacattga					2598
SEQ ID NO: 20	moltype = AA length = 865					
FEATURE	Location/Qualifiers					
source	1..865					
	mol_type = protein					
	organism = Homo sapiens					
SEQUENCE: 20						
MALVLGSLLL	LGLCGNSFSG	GQPSSTDAPK	AWNYELPATN	YETQDSHKAG	PIGILFELVH	60
IFLYVVQPRD	FPEDTLRKFL	QKAYESKIDY	DKPETVILGL	KIVYYEAGII	LCCVLGLLFI	120
ILMPLVGYFF	CMCRCCNKC	GEMHQRQKEN	GPFLRKCFAI	SLLVICIIIS	IGIFYGFVAN	180
HQVTRIKRS	RKLADSNFKD	LRTLNETPE	QIKYILAQYN	TTKDKAFTDL	NSINSVLGGG	240
ILDRLRPNII	PVLDEIKSMA	TAIKETKEAL	ENMNSTLKSL	HQQSTQLSSS	LTSVKTSLRS	300
SLNDPLCLVH	PSSETCNSIR	LSLSQLNSNP	ELRQLPPVDA	ELDNVNNVLR	TDLDGLVQQG	360
YQSLNDIPDR	VQRQTTTVVA	GIKRVLNSIG	SDIDNVTQRL	PIQDILSAFS	VYVNNTESYI	420
HRNLPTLEEY	DSYWWLGGLV	ICSLTLTIVI	FYYLGLLCGV	CGYDRHATPT	TRGCVSNTGG	480
VFLMVGVLGS	FLFCWILMII	VVLTFVFGAN	VEKLICEPYT	SKELFRVLDT	PYLLNEDWEY	540
YLSGKLFNKS	KMKLTFEQVY	SDCKKNRGTY	GTLHLQNSFN	ISEHLNINEH	TGSISSELES	600
LKVNLNIFLL	GAAGRKNLQD	FAACGIDRMN	YDSYLAQTGK	SPAGVNLLSF	AYDLEAKANS	660
LPPGNLRNSL	KRDAQTIKTI	HQQRVLPSEQ	SLSTLYQSVK	ILQRTGNGLL	ERVTRILASL	720
DFAQNFITNN	TSSVIIETK	KYGRTIIGYF	EHYLQWIEFS	ISEKVASCKP	VATALDTAVD	780
VFLCSYIIDP	LNLFWFGIGK	ATVFLLPALI	FAVKLAKYYR	RMDSEDVYDD	VETIPMKNME	840
NGNNGYHKDH	VYGIHNPVMT	SPSQH				865
SEQ ID NO: 21	moltype = DNA length = 2466					
FEATURE	Location/Qualifiers					
source	1..2466					
	mol_type = genomic DNA					
	organism = Homo sapiens					
SEQUENCE: 21						
atggtcagct	ggggtcgttt	catctgcctg	gtcgtggtea	ccatggcaac	cttgctccctg	60
gcccggccct	ccttcagttt	agttgaggat	accacattag	agccagaaga	gccaccaacc	120
aaataccaaa	tctctcaacc	agaagtgtac	gtggctgcgc	cagggggagtc	gctagagggtg	180
cgctgcctgt	tgaaagatgc	cgccgtgatc	agttggacta	aggatggggg	gcacttggggg	240
cccaacaata	ggacagtgtc	tattggggag	tacttgccaga	taaaggggcg	cacgcctaga	300
gactccggcc	tctatgcttg	tactgccagt	aggactgtag	acagtgaaac	ttggtagcttc	360
atggtgaatg	tcacagatgc	catctcatcc	ggagatgatg	aggatgacac	cgatgggtgcg	420
gaagattttg	tcagtgaaga	cagtaacaac	aagagagcac	catactggac	caacacagaa	480
aagatggaaa	agcggctcca	tgctgtgcct	gcggccaaca	ctgtcaagtt	tcgctgcccc	540
gccggggggg	acccaatgcc	aaccatgcgg	tggctgaaaa	acgggaagga	gtttaagcag	600
gagcatcgca	ttggaggcta	caaggtagca	aaccagcact	ggagcctcat	tatggaaagt	660
gtgggtcccat	ctgacaaggg	aaattatacc	tgtgtagtgg	agaatgaata	cgggtccatc	720
aatcacacgt	accacctgga	tgttgtggag	cgatcgccct	accggcccat	cctccaagcc	780
ggactgccgg	caaatgcctc	cacagtggtc	ggaggagacg	tagagtttgt	ctgcaagggt	840
tacagtgatg	cccagcccca	catccagtgg	atcaagcacg	tggaaaagaa	cggcagtaaa	900
tacggggccg	acgggctgcc	ctacctcaag	gttctcaagg	ccgcccgtgt	taacaccacg	960
gacaaagaga	ttgaggttct	ctatatctcg	aatgtaactt	ttgaggacgc	tggggaatat	1020
acgtgcttgg	cgggtaattc	tattgggata	tcctttcact	ctgcatgggt	gacagttctg	1080
ccagcgccct	gaagagaaaa	ggagattaca	gcttccccag	actacctgga	gatagccatt	1140
tactgcatag	gggtcttctt	aatcgccctg	atggtggtaa	cagtcatcct	gtgccgaatg	1200
aaagaacacga	ccaagaagcc	agacttcagc	agccagccgg	ctgtgcacaa	gctgacccaa	1260
cgtatcccc	tgcggagaca	ggtaaacagt	tcggctgagt	ccagctcctc	catgaactcc	1320
aaacaccccgc	tggtgaggat	aacaacacgc	ctctcttcaa	cggcagacac	ccccatgctg	1380
gcaggggtct	ccgagtatga	acttccagag	gacccaaaat	gggagtttcc	aagagataag	1440
ctgacactgg	gcaagcccct	gggagaaggt	tgctttgggc	aagtgggtcat	ggcggaagca	1500
gtgggaattg	acaaagacaa	gcccaggag	gcggtcaccg	tggccgtgaa	gatgttgaaa	1560
gatgatgcca	cagagaaaga	cctttctgat	ctgggtgtcag	agatggagat	gatgaagatg	1620
attgggaaac	acaagaatat	cataaatctt	cttggagcct	gcacacagga	tgggcctctc	1680
tatgtcatag	ttgagtatgc	ctctaaaggc	aacctccgag	aatacctccg	agcccggagg	1740
ccacccggga	tggagtactc	ctatgacatt	aaccgtgttc	ctgaggagca	gatgaccttc	1800
aaggacttgg	tgtcatgcac	ctaccagctg	gccagaggca	tggagtactt	ggcttcccaa	1860
aaatgtattc	atcgagattt	agcagccaga	aatgttttgg	taacagaaaa	caatgtgatg	1920
aaaatagcag	actttggact	cgccagagat	atcaacaata	tagactatta	caaaaagacc	1980
accaatgggc	ggcttccagt	caagtggatg	gctccagaag	ccctgtttga	tagagtatac	2040
actcatcaga	gtgatgtctg	gtccttcggg	gtgttaatgt	gggagatctt	cactttaggg	2100
ggctcgccct	acccagggat	tcccgtggag	gaacttttta	agctgctgaa	ggaaggacac	2160
agaatggata	agccagccaa	ctgcaccaac	gaactgtaca	tgatgatgag	ggactgttgg	2220
catgcagtgc	cctcccagag	accaacgttc	aagcagttgg	tagaagactt	ggatcgaatt	2280

-continued

ctcactctca	caaccaatga	ggaatacttg	gacctcagcc	aacctctcga	acagtattca	2340
cctagttacc	ctgacacaag	aagttcttgt	tcttcaggag	atgattctgt	tttttctcca	2400
gaccccatgc	cttacgaacc	atgccttcct	cagtatccac	acataaacgg	cagtgttaaa	2460
acatga						2466
SEQ ID NO: 22	moltype = AA length = 821					
FEATURE	Location/Qualifiers					
source	1..821					
	mol_type = protein					
	organism = Homo sapiens					
SEQUENCE: 22						
MVSWGRFICL	VVVTMATLSL	ARPSFSLVED	TITLEPEEPPT	KYQISQPEVY	VAAPGESLEV	60
RCLLKDAAVI	SWTKDGVHLG	PNNRTVLIGE	YLQIKGATPR	DSGLYACTAS	RTVDSETWYF	120
MVNVTDAISS	GDDEDDTDGA	EDFVSENSNN	KRAPYWTNTE	KMEKRLHAVP	AANTVKFRCP	180
AGGNPMPMTMR	WLKNGKEFKQ	EHRIGGYKVR	NQHWSLIMES	VVPSDKGNYT	CVVENEYGS	240
NHTYHLDVVE	RSPHRPILQA	GLPANASTVV	GGDVEFVCKV	YSDAQPHIQW	IKHVEKNGSK	300
YGPDGLPYLK	VLKAAGVNTT	DKEIEVLVIR	NVTFEDAGEY	TCLAGNSIGI	SFHSAWLTVL	360
PAPGREKEIT	ASPDYLEIAI	YCIGVFLIAC	MVVTVILCRM	KNTTKKPDFS	SQPAVHKLTK	420
RIPLRRQVTV	SAESSSSMNS	NTPLVRITTR	LSSTADTPML	AGVSEYELPE	DPKWEFPRDK	480
LTGKPLGEG	CFGQVMAEA	VGIDKDKPKE	AVTVAVKMLK	DDATEKDLS	LVSEMEMMKM	540
IGKHKNIINL	LGACTQDGPL	YVIVEYASKG	NLREYLRARR	PPGMEYSYDI	NRVPEEQMTF	600
KDLVSCITYQL	ARGMEYLASQ	KCIHRDLAAR	NVLVTENNVM	KIADFGGLARD	INNIDYYKKT	660
TNGRLPVKWM	APEALFDRVY	THQSDVWSFG	VLMWEIFTLG	GSPYPGIPVE	ELFKLLKEGH	720
RMDKPANCTN	ELYMMMRDCW	HAVPSQRPTF	KQLVEDLDRI	LTLTTNNEEYL	DLSQPLEQYS	780
PSYPDTRSSC	SSGDDSVFSP	DPMPYEPCLP	QYPHINGSVK	T		821
SEQ ID NO: 23	moltype = DNA length = 2649					
FEATURE	Location/Qualifiers					
source	1..2649					
	mol_type = genomic DNA					
	organism = Homo sapiens					
SEQUENCE: 23						
atggggccctt	ggagccgcag	cctctcggcg	ctgctgctgc	tgctgcaggt	ctcctcttgg	60
ctctgccagg	agccggagcc	ctgccaccct	ggctttgacg	ccgagagcta	cacgttcacg	120
gtgccccggc	gccacctgga	gagaggccgc	gtcctgggca	gagtgaattt	tgaagattgc	180
accggtcgac	aaaggacagc	ctatttttcc	ctcgacaccc	gattcaaagt	gggcacagat	240
ggtgtgatta	cagtcaaaa	gcctctacgg	tttcataacc	cacagatcca	tttcttggtc	300
tacgcctggg	actccaccta	cagaaagttt	tccaccaaa	tcacgctgaa	tacagtgggg	360
caccaccacc	gccccccgcc	ccatcaggcc	tccgtttctg	gaatccaagc	agaattgctc	420
acatttccca	actcctctcc	tggcctcaga	agacagaaga	gagactgggt	tattcctccc	480
atcagctgcc	cagaaaatga	aaaaggccca	tttcctaaaa	acctggttca	gatcaaatcc	540
aacaaagaca	aagaaggcaa	ggttttctac	agcatcactg	gccaaggagc	tgacacaccc	600
cctgttggtg	tctttattat	tgaaagagaa	acaggatggc	tgaaggtgac	agagcctctg	660
gatagagaac	gcattgccac	atacactctc	ttctctcacg	ctgtgtcatc	caacgggaat	720
gcagttgagg	atccaatgga	gattttgatc	acggtaaccg	atcagaatga	caacaagccc	780
gaattcacc	aggaggtctt	taaggggtct	gtcatggaag	gtgctcttcc	aggaacctct	840
gtgatggagg	tcacagccac	agacgcggac	gatgatgtga	acacctacaa	tgccgccatc	900
gcttacacca	tcctcagcca	agatcctgag	ctccctgaca	aaaatatgtt	caccattaac	960
aggaacacag	gagtcatcag	tgtggtcacc	actgggctgg	accgagagag	tttccctacg	1020
tataccctgg	tggttcaagc	tgtgacctt	caaggtgagg	ggttaagcac	aacagcaaca	1080
gctgtgatca	cagtcactga	caccaacgat	aatcctccga	tcttcaatcc	caccacgtac	1140
aagggtcagg	tgccctgagaa	cgaggctaac	gtcgtaatca	ccacactgaa	agtgactgat	1200
gctgatgccc	ccaatacccc	agcgtgggag	gctgtataca	ccatattgaa	tgatgatggg	1260
ggacaatttg	tcgtcaccac	aaatccagtg	aacaacgatg	gcattttgaa	aacagcaaag	1320
ggcttggaat	ttgaggccaa	gcagcagtac	attctacacg	tagcagtgac	gaatgtggta	1380
ccttttgagg	tctctctcac	cacctccaca	gccaccgtca	ccgtggatgt	gctggatgtg	1440
aatgaagccc	ccatctttgt	gcctcctgaa	aagagagtgg	aagtgtccga	ggactttggc	1500
gtggggccagg	aaatcacatc	ctacactgcc	caggagccag	acacatttat	ggaacagaaa	1560
ataacatatc	ggatttgga	agacactgcc	aactggctgg	agattaatcc	ggacactggt	1620
gccatttcca	ctcgggctga	gctggacagg	gaggattttg	agcacgtgaa	gaacagcacg	1680
tacacagccc	taatcatagc	tacagacaat	ggttctccag	ttgctactgg	aacagggaca	1740
cttctgctga	tcctgtctga	tgtgaatgac	aacgccccca	taccagaacc	tcgaactata	1800
ttcttctgtg	agaggaatcc	aaagcctcag	gtcataaaca	tcattgatgc	agaccttcct	1860
cccaatacat	ctcccttcac	agcagaacta	acacacgggg	cgagtgccaa	ctggaccatt	1920
cagtacaacg	acccaaccca	agaatctatc	attttgaagc	caaagatggc	cttagagggtg	1980
ggtgactaca	aaatcaatct	caagctcatg	gataaccaga	ataaagacca	agtgaccacc	2040
ttagagggtca	gcgtgtgtga	ctgtgaaggg	gccgctggcg	tctgtaggaa	ggcacagcct	2100
gtcgaagcag	gattgcaa	tcctgccatt	ctggggattc	ttggagggaat	tcttgctttg	2160
ctaattctga	ttctgtgtct	cttgctgttt	cttcggaggga	gagcgggtgt	caaagagccc	2220
ttactgcccc	cagaggatga	caccggggac	aacgtttatt	actatgatga	agaaggaggc	2280
ggagaagagg	accaggactt	tgacttgagc	cagctgcaca	ggggcctgga	cgctcggcct	2340
gaagtgaact	gtaacgacgt	tgcaccaacc	ctcatgagtg	ttccccggta	tcttccccgc	2400
cctgccaatc	ccgatgaaat	tggaaatttt	attgatgaaa	atctgaaagc	ggctgatact	2460
gacccacag	ccccgcctta	tgattctctg	ctcgtgtttg	actatgaagg	aagcgggttc	2520
gaagctgcta	gtctgagctc	cctgaactcc	tcagagtcag	acaaagacca	ggactatgac	2580
tacttgaacg	aatggggcaa	tcgcttcaag	aagctggctg	acatgtacgg	aggcggcgag	2640

-continued

gacgactag		2649				
SEQ ID NO: 24	moltype = AA length = 882					
FEATURE	Location/Qualifiers					
source	1..882					
	mol_type = protein					
	organism = Homo sapiens					
SEQUENCE: 24						
MGPWSRSLSA	LLLLLQVSSW	LCQEPEPCHP	GFDAESYTFT	VPRRHLEGR	VLGRVNFEDC	60
TGRQRTAYFS	LDTRFKVGTD	GVITVKRPLR	FHNPQIHFLV	YAWDSTYRKF	STKVTLNTVG	120
HHHRPPPHQA	SVSGIQAELL	TFPNSSPGLR	RQKRDWVIP	ISCPENEKGP	FPKNLVQIKS	180
NKDKEGKVFY	SITGQGADTP	PVGVFIIERE	TGWLKVTEPL	DRERIATYTL	FSHAVSSNGN	240
AVEDPMEILI	TVTDQNDNKP	EFTQEVFKGS	VMEGALPGTS	VMEVTATDAD	DDVNTYNAAI	300
AYTILSQDPE	LPDKNMFTIN	RNTGVISVVT	TGLDRESFPT	YTLVVQAADL	QGEGLSTTAT	360
AVITVTD TND	NPPIFNPTTY	KGQVPENEAN	VVITTLKVT	ADAPNTPAWE	AVYTILNDDG	420
GQFVVTTNPV	NNDGILKTAK	GLDFEAKQQY	ILHVAVTNVV	PFEVSLTTST	ATVTVDVLDV	480
NEAPIFVPPE	KRVEVSEDFG	VGQEITSYTA	QEPDTFMEQK	ITYRIWRDTA	NWLEINPDTG	540
AISTRAELDR	EDFEHVKNST	YTALIIATDN	GSPVATGTGT	LLLILSDVND	NAPIPEPTI	600
FFCERNPKPQ	VINIIDADLP	PNTSPFTAEL	THGASANWTI	QYNDPTQESI	ILKPKMALEV	660
GDYKINLKLM	DNQNKDQVTT	LEVSVCDC	EG AAGVCRKAQ	P VEAGLQIPAI	LGILGGILAL	720
LILILLLLLLF	LRRRAVVKEP	LLPPEDDTRD	NVYYYDEEGG	GEEDQDFDLS	QLHRGLDARP	780
EVTRNDVAPT	LMSVPRYLPR	PANPDEIGNF	IDENLKAADT	DPTAPPYDSL	LVFDYEGSGS	840
EAASLSSLNS	SESDKDQDYD	YLNEWGNR	FK KLADMYGGGE	DD		882

- What is claimed is:
1. A method of detecting the level of circulating tumor cell in a subject, the method comprising:
 - a) obtaining a biological sample from the subject;
 - b) obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex;
 - c) contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to allow the solid phase-capture binding protein complex to bind at least one epithelial-mesenchymal transition (EMT) biomarker on a circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex;
 - d) separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample, thereby isolating or capturing the intact cell;
 - e) determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex.
 2. The method of claim 1, wherein determining the level of circulating tumor cell in the solid phase-capture binding protein-circulating tumor cell complex comprises at least one of DAPI staining, β -catenin detection, CD45 detection, and CD31 detection.
 3. The method of claim 1, wherein the EMT biomarker is at least one of OB-cadherin, N-cadherin, vimentin, E-cadherin, FGFR2 splice variant isoforms, or CD133.
 4. The method of claim 1, further comprising determining the presence or absence of at least one prostate cancer-specific genomic event.
 5. The method of claim 4, wherein the at least one prostate cancer-specific genomic event is selected from the group consisting of androgen receptor amplification, phosphatase and tensin homolog (PTEN) loss, gene fusion of transmembrane protease, serine 2 (TMPRSS2), and ETS related gene (ERG), and a combination thereof.
 6. The method of claim 1, wherein the biological sample comprises a tissue sample or a fluid sample from the subject.
 7. A kit for isolating or capturing a circulating tumor cell in a biological sample, the kit comprising:
 - (1) a capture agent comprising an antibody specifically binding at least one epithelial-mesenchymal transition (EMT) biomarker; and
 - (2) a staining reagent for confirming the capture of the circulating tumor cell.
 8. The kit of claim 7, wherein the at least one EMT biomarker is at least one of OB-cadherin, N-cadherin, vimentin, E-cadherin, FGFR2 splice variant isoforms, or CD133.
 9. The kit of claim 7, wherein the at least one staining reagent comprises DAPI or a detectably labeled antibody selected from a detectably labeled anti- β -catenin antibody, a detectably labeled anti-CD45 antibody, or a detectably labeled anti-CD31 antibody.
 10. The kit of claim 7, wherein the antibody specifically binding at least one EMT biomarker is linked to a solid phase or an imaging agent.
 11. The kit of claim 10, wherein the solid phase is a microparticle, magnetic particle, or paramagnetic particle.
 12. The method of claim 1, wherein the circulating tumor cell is from breast cancer, colon cancer, lung cancer, prostate cancer, testicular cancer, brain cancer, skin cancer, rectal cancer, gastric cancer, esophageal cancer, sarcomas, tracheal cancer, head and neck cancer, pancreatic cancer, liver cancer, ovarian cancer, lymphoid cancer, cervical cancer, vulvar cancer, melanoma, mesothelioma, renal cancer, bladder cancer, thyroid cancer, bone cancer, carcinoma, sarcoma, and/or soft tissue.
 13. A method for detecting or identifying a circulating tumor cell or a level of a circulating tumor cell in a patient, the method comprising:
 - (a) obtaining a biological sample from the patient;
 - (b) obtaining at least one capture binding protein, wherein the capture binding protein is linked to a solid phase to form a solid phase-capture binding protein complex;
 - (c) contacting the biological sample with the solid phase-capture binding protein complex for a time sufficient to

allow the solid phase-capture binding protein complex to bind at least one epithelial-mesenchymal transition (EMT) biomarker on the circulating tumor cell to form a solid phase-capture binding protein-circulating tumor cell complex; and

- (d) separating the solid phase-capture binding protein-circulating tumor cell complex from the sample and unbound magnetic particle-capture binding protein complexes by application of an external magnetic field on the sample, thereby detecting or identifying the circulating tumor cell or the level of the circulating tumor cell in the patient.

14. The method of claim **13**, further comprising at least one of DAPI staining, β -catenin detection, CD45 detection, and CD31 detection.

15. The method of claim **14**, wherein the circulating tumor cell is further detected or identified as a circulating tumor cell if DAPI staining is positive, β -catenin expression is positive, CD45 expression is negative, and CD31 expression is negative.

* * * * *