

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
28 April 2011 (28.04.2011)

PCT

(10) International Publication Number  
**WO 2011/050103 A1**

(51) International Patent Classification:  
G01N 33/574 (2006.01)

(21) International Application Number:  
PCT/US2010/053431

(22) International Filing Date:  
20 October 2010 (20.10.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
61/253,787 21 October 2009 (21.10.2009) US

(71) Applicant (for all designated States except US): THE SCRIPPS RESEARCH INSTITUTE [US/US]; 10550 North Torrey Pines Road, La Jolla, CA 92037 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): KUHN, Peter [US/US]; 708 Barbara Avenue, Solano Beach, CA 92075 (US). KOLATKAR, Anand [US/US]; 5360 Toscana Way, G-219, San Diego, CA 92121 (US). KUNKEN, Joshua [US/US]; 9525 Genesee Avenue, #224, San Diego, CA 92121 (US). MARRINUCCI, Dena [US/US]; 2731 Caminito Verdugo, Del Mar, CA 92014 (US). YANG, Xing [US/US]; 10728 Heather Ridge Drive, San Diego, CA 92130 (US). STUELPNAGEL, John [US/US]; 38 Briggs Avenue, Encinitas, CA 92024 (US).

(74) Agent: HAILE, Lisa, A.; Dla Piper LLP (US), 4365 Executive Drive, Suite 1100, San Diego, CA 92121-2133 (US).

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: METHOD OF USING NON-RARE CELLS TO DETECT RARE CELLS

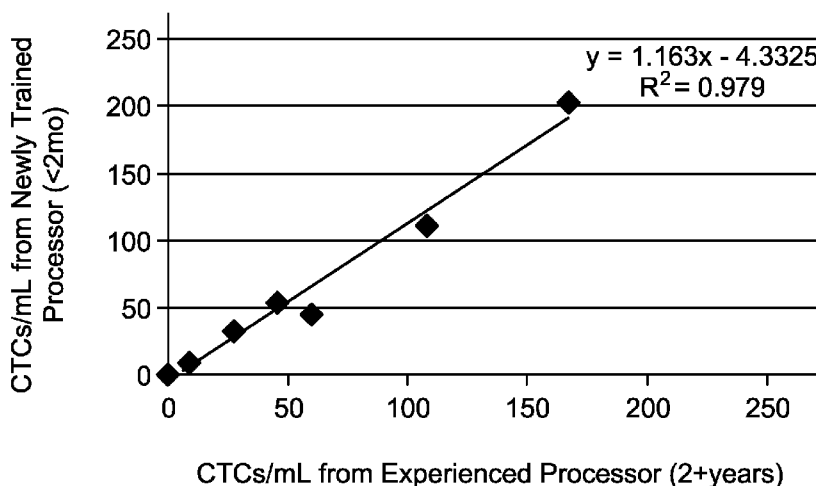


FIG. 3

(57) Abstract: The invention provides seminal computational approaches utilizing data from non-rare cells to detect rare cells, such as circulating tumor cells (CTCs). The invention is applicable at two distinct stages of CTC detection; the first being to make decisions about data collection parameters and the second being to make decisions during data reduction and analysis. Additionally, the invention utilizes both one and multi-dimensional parameterized data in a decision making process.

WO 2011/050103 A1

## **METHOD OF USING NON-RARE CELLS TO DETECT RARE CELLS**

### **BACKGROUND OF THE INVENTION**

#### **FIELD OF THE INVENTION**

**[0001]** The invention relates generally to medical diagnostics and more specifically to detection and categorization of rare cells, such as circulating tumor cells (CTCs).

#### **BACKGROUND INFORMATION**

**[0002]** Significant unmet medical need exists for the longitudinal disease monitoring in patients with epithelial cancers at the cellular level. Predicting and monitoring therapy response and disease progression are particularly important in epithelial cancer patients due to the natural history of the disease and the selective selection process in response to the therapeutic pressure. While progress has been made in understanding the primary and metastatic tumors in their respective microenvironments, a substantial barrier exists in understanding carcinoma behavior during the fluid phase, as it spreads within and occupies the bloodstream. The circulating component of cancer contains within it the cells giving rise to future metastases, and as such, represents a compelling target for investigation.

**[0003]** Research to fully characterize the clinical significance of this fluid phase of solid tumors has been hindered by the lack of easily accessible and reliable experimental tools for the identification of CTCs. The unknown character and low and unknown frequency of CTCs in the blood, combined with the difficulty of distinguishing between cancerous versus normal epithelial cells, has significantly impeded research into how the fluid phase might be clinically important. The ideal fluid phase biopsy should find significant numbers of a specific CTC population in most epithelial cancer patients and preserve and present CTCs to a pathologist and/or researcher in a format that enables not only enumeration but further molecular, morphologic and/or phenotypic analysis. In addition, it should preserve the remaining rare populations for further analysis.

**[0004]** CTCs are generally, although not exclusively, epithelial cells that originate from a solid tumor in very low concentration and enter into the blood stream of patients with various types of cancer. CTCs are also thought to be capable of originating in the blood, forming small colonies throughout the body. The shedding of CTCs by an existing tumor or

metastasis often results in formation of secondary tumors. Secondary tumors typically go undetected and lead to 90% of all cancer deaths. Circulating tumor cells provide the link between the primary and metastatic tumors. This leads to the promise of using the identification and characterization of circulating tumor cells for the early detection and treatment management of metastatic epithelial malignancies. Detection of CTCs in cancer patients offers an effective tool in early diagnosis of primary or secondary cancer growth and determining the prognosis of cancer patients undergoing cancer treatment because number and characterization of CTCs present in the blood of such patients has been correlated with overall prognosis and response to therapy. Accordingly, CTCs serve as an early indicator of tumor expansion or metastasis before the appearance of clinical symptoms.

**[0005]** While the detection of CTCs has important prognostic and potential therapeutic implications in the management and treatment of cancer, because of their occult nature in the bloodstream, these rare cells are not easily detected. CTCs were first described in the 1800s, however only recent technological advances have allowed their reliable detection. The challenge in the detection of circulating tumor cells is that they are present in relatively low frequency compared to other nucleated cells, commonly less than 1:100,000. To compensate for this challenge, most conventional approaches for detecting circulating tumor cells rely on experimental enrichment methods, whereby the CTCs are preferentially separated from the other cellular components (*e.g.*, non-CTCs), most importantly other nucleated cells that are the most similar to CTCs.

**[0006]** Currently, the most utilized methods of positive enrichment for enumeration/characterization of CTCs are immunomagnetic enrichment methods targeting the surface protein EpCAM and the “CTC chip”. The most widely used methodology to detect CTCs, J&J’s Veridex technology, utilizes immunomagnetic enrichment. The technology relies upon immunomagnetic enrichment of tumor cell populations using magnetic ferrofluids linked to an antibody which binds epithelial cell adhesion molecule (EpCAM), expressed only on epithelial derived cells. This methodology requires 7.5mL of blood for analysis and finds greater than 2 CTCs in only some metastatic cancer patients.

**[0007]** Microfluidic or “CTC-Chip” technology, is another positive enrichment method for enumeration/characterization of CTCs. The methods utilizes 1-3mL of blood in which whole blood flows past 78,000 EpCAM-coated microposts. EpCAM+ cells stick to the posts

and are subsequently stained with cytokeratin, CD45, and DAPI. With this methodology, CTCs are found in virtually all metastatic cancer patients at a relatively high purity and not in healthy controls. Additionally, CTC-chip technology identifies CTCs in all patients and in higher numbers than other technologies by a factor of approximately 10 to 100 fold as reported in two recent publications.

**[0008]** The only routinely used technology for CTC detection is based on immunomagnetic enrichment. This current “gold standard” and FDA approved test is called CellSearch® and employs an immunomagnetic enrichment step to isolate cells that express the epithelial cells adhesion molecule (EpCAM). Additionally, to be identified as a CTC, the cell must contain a nucleus, express cytoplasmic cytokeratin, and have a diameter larger than five microns. This system has uncovered the prognostic utility of enumerating and monitoring CTC counts in patients with metastatic breast, prostate, and colorectal cancers; however, the sensitivity of this system is low, finding no or few CTCs in most patients. Most follow-on CTC technologies have reported higher sensitivity and are pursuing variations of the enrichment strategy, however this directly biases the detectable events towards those that have sufficient expression of the protein selected for the initial enrichment step.

**[0009]** A standardized microscope based approach has also been previously utilized to identify and morphologically characterize and credential CTCs in case studies of breast, colorectal, and lung cancer patients.

**[0010]** Although many CTC detection approaches are currently in use, significant limitations have been identified with the current approaches. For example, one significant limitation of positive selection methods to enumerate/characterize CTCs is that positive physical selection invariably leads to loss of CTCs and is less than 100% efficient. Thus the number of CTCs detected per sample using current methods is often too low to provide robust interpretation or clinically meaningful content of a particular sample. Additional limitations of current methods include low CTC detection due to CTC heterogeneity. For example, differences in individual CTC features within the CTC population of interest further hinder the number of CTCs detected using current methodologies. Such differences may include size variations between individual CTCs, and variable or down regulated expression between individual CTCs of the cell surface markers used to detect CTCs. A further limitation of existing methodologies includes limitations in purity levels and variable purity.

Any enrichment will have a certain number of false positives, for instance other nucleated blood cells that stick to the enrichment. For example, the Veridex magnet has typically 5,000 to 10,000 false positives on top of the 5 to 10 positives.

### SUMMARY OF THE INVENTION

[0011] The present invention is based in part on the discovery of innovative methods for analyzing samples to detect, enumerate and characterize rare cells, such as CTCs.

Accordingly, the present invention provides methods for improved detection and characterization allowing for clinically meaningful analysis of samples for use in clinical, research and development settings.

[0012] Accordingly, the present invention provides methods for the improved detection and characterization of rare cells in a sample by utilizing data from non-rare cells (cells present at a concentration of 10, 50, 100, 200, 300, 400, 500, 1,000, 5,000, 10,000 times or greater as compared to the rare cell) in the sample. Thus the method of the invention utilizes similarity measures to assess non-similarity of cells, requiring both the biggest distance exclusion, *e.g.*, events that are clearly non-rare cell related and the fine distinction of a cutoff based on similarities of surrounding non-rare cells.

[0013] The method includes providing a sample suspected of having at least one rare cell and at least one cell that is present at a concentration that is at least 10 times that of the rare cell; contacting the sample with at least one detectable agent, such as an agent that binds a cell marker; performing cell imaging on the sample to generate an image; and detecting the at least one rare cell as compared with other cells in the sample by analyzing the cell from the image, thereby detecting the rare cell in the sample. In various aspects of the invention, the method further includes plating of the suspected rare cell and at least one cell on a solid support, such as a slide, to facilitate contacting the cells with the detectable agent and cell imaging. In various aspects of the invention, the detectable agent is any agent used to stain the cells, such as an agent that binds a cell marker, including, but not limited to, a positive marker, negative marker, nuclear marker, content marker, or any combination thereof.

[0014] In various aspects of the invention, the methods described herein are performed on an apparatus for efficiently imaging a slide containing a detectable signal, such as a fluorescent signal. The apparatus may typically include a computer having at least one

system processor with image processing capability, a computer monitor, an input device, a power supply and a microscope subsystem. Thus the apparatus includes a computer having executable code for performing the various analysis required to practice the invention. The microscope subsystem includes an optical sensing array for acquiring images. A two-dimensional motion stage for sample movement and for focus adjustment, and input and output mechanisms for multiple sample analysis and storage. The apparatus may also include a transmitted light source as well as an illuminating/fluorescent excitation light source for fluorescing samples.

**[0015]** In one embodiment of the invention, the method includes establishing optimal exposure limits for performing the cell imaging that facilitate detection of rare cells present. In one aspect, the exposure limit for the detectable agent is determined using a signal from at least one cell. In various aspects, the detectable marker may be a positive marker, negative marker, nuclear marker or content marker. In a related aspect, the exposure limits may be set using data relating to the cells and/or suspected rare cells gathered from a first image, to re-image the slide.

**[0016]** In another embodiment, the method includes minimizing exposure settings to minimize data collection time and maximize throughput to facilitate detection of rare cells.

**[0017]** In another embodiment, the method includes utilizing data associated with non-rare cells to generate a quality control parameter that facilitates detection of rare cells. In various aspects, the quality control parameter is distribution of at least one non-rare cell on the slide, alignment of multiple cell images via alignment of non-rare cell markers, quality of cell staining, distribution of a positive marker throughout the non-rare cells, or cell loss from repeated processing.

**[0018]** In another embodiment, the method includes determining intensity cut-off limits to minimize false negatives, as well as false positives and to facilitate rare cell detection. In one aspect, the detectable agent is a positive marker and the intensity limits are determined using mean, standard deviation, coefficient of variation, other statistical parameters or any combination thereof, for a background signal of the positive marker. In another aspect, the detectable agent is a positive marker and the intensity limits are determined within a single image, or portions of that image, by identifying the highest signal event from a positive

marker and comparing the highest signal to the mean and standard deviation calculated from signals of all, or a subset of events. In yet another aspect, the detectable agent is a negative marker and the intensity limit for the negative marker is determined using mean and standard deviation of signals from the negative markers from non-rare cells (either all non-rare cells or a specific subset).

**[0019]** In another embodiment, cytological features of non-rare cells, such as cellular and nuclear size (absolute and relative; overall and apparent) and distribution, are utilized to facilitate detection of non-rare cells.

**[0020]** In another embodiment, the method includes utilizing data associated with non-rare cells to enumerate rare cells, thus facilitating their detection. In various embodiments, data may include, but is not limited to, total intensity, mean intensity, segmented intensity, fixed circle, variable circle, or any combination thereof.

**[0021]** In another embodiment, the method includes determination of the expression level of a content marker in rare cells and non-rare cells to facilitate detection of rare cells.

**[0022]** In various aspects of the invention, a rare cell is a CTC or subpopulation thereof.

**[0023]** As such, in another embodiment, the invention provides a method for diagnosing or prognosing cancer in a subject. The method includes performing the method of improved detection and characterization of CTCs as described herein and analyzing detected CTCs and provide a diagnosis or prognosis based on analysis of the CTCs, thereby diagnosing or prognosing cancer in a subject.

**[0024]** In another embodiment, the invention provides a method for determining responsiveness of a subject to a therapeutic regime. The method includes performing the method of improved detection and characterization of CTCs as described herein and analyzing the CTCs, thereby determining the responsiveness of the subject to a therapeutic regime.

**[0025]** In another embodiment, the invention provides a method for determining a candidate subject for a clinical trial. The method includes performing the method of improved detection and characterization of CTCs as described herein and analyzing the CTCs, thereby determining a candidate subject for a clinical trial.

## BRIEF DESCRIPTION OF THE DRAWINGS

[0026] Figure 1 is a graphical representation of mean observed SKBR3s plotted against expected SKBR3s. Four aliquots of normal control blood were spiked with varying numbers of SKBR2 cells to produce 4 slides with approximately 10, 30, 100, and 300 cancer cells per slide. The mean of each quadruplicate is displayed as well as error bars noting standard deviation.

[0027] Figure 2 is a pictorial representation of a gallery of a representative subpopulation of CTCs found in cancer patients. Each CTC of the subpopulation is cytokeratin positive, CD45 negative, contains a DAPI nucleus, and is morphologically distinct from surrounding white blood cells which are circular in shape.

[0028] Figure 3 is a graphical representation comparing CTC counts between two separate processors on 9 different cancer patient samples. CTC/mL counts ranged from 0 to 203.

[0029] Figure 4 is a graphical representation including four graphs plotting CTC and PSA levels of serial blood draws from 4 different prostate cancer patients over a three month time period. Two patients had increasing CTC and PSA levels and two patients had decreasing/stable CTC and PSA levels. PSA levels increased in patients that had increasing CTC counts and decreased in patients that had decreasing/stable CTC counts.

[0030] Figure 5 is a graphical representation showing the incidence rate of a putative rare cell population across patients relative to a CTC subpopulation (HD-CTC).

## DETAILED DESCRIPTION OF THE INVENTION

[0031] The present invention provides a method which omits physical methods for positively enriching for rare cells, such as CTCs, from a mixed population, thereby minimizing the loss of rare cells. This methodology further allows for the capture/identification of subsets of cell populations, such as subpopulations of CTCs or other rare populations by detection of the same or different markers using different parameters, such as cutoff values, that allow for distinguishing between events and non-events. For example, as discussed in detail herein, different cutoffs may be utilized to characterize different cell subpopulations.



[0032] While the disclosure highlights CTCs and subpopulations thereof, the same methodologies may be used to find any other rare cell type in a background of non-rare cells. As used herein, a “rare cell” is intended to include a cell that is either 1) of a cell type that is less than about 5%, 4%, 3%, 2%, 1%, 0.1%, 0.01% or 0.001% of the total nucleated cell population in a fluid sample, or 2) of a cell type that is present at less than one million cells per milliliter of fluid sample. Exemplary rare cells include, but are not limited to CTCs, circulating endothelial cells (CECs), white blood cells in emboli, cancer stem cells, activated or infected cells, such as activated or infected blood cells, and fetal cells.

[0033] Accordingly, it will be understood by one in the art that references to CTCs throughout the specification include reference to rare cells and vice versa.

[0034] The present method allows for identification of rare cells, such as CTCs or subpopulations of CTCs from the background of other blood cells using microscopy, cytometry, automation, and computation. The present invention utilizes these components, individually and collectively, to identify rare cells. The benefits include the ability to find more rare cells, to present them in a way that enables subsequent analyses for content markers, and to do so in a time and resource efficient manner.

[0035] Further, the present disclosure is based in part on a next generation assay capable of identifying subpopulations of CTCs in cancer patients. One particular subpopulation identified was from a small cohort of cancer patients. In addition to using specific parameters defining subpopulations of CTCs, such as one referred to herein as the High-Definition-CTC (HD-CTC) subpopulation, the assay affords greater sensitivity with a smaller volume of blood than previous efforts. The key innovative aspects of this assay are driven by the need for simplicity and minimal processing of the blood specimen as well as conforming to the need to enable professional interpretation with diagnostic quality imagery.

[0036] The approach used to identify a rare cell population, such as CTCs, or subpopulation thereof, is distinct in that it does not rely on any single protein enrichment strategies. All nucleated blood cells are imaged in multiple colors to locate and morphologically evaluate rare events. This enrichment-free strategy results in an assay capable of ‘tunable specificity/sensitivity’ allowing high sensitivity and high specificity while still enabling the study of a rare cell population known to be heterogeneous. A key advantage

and difference to physical enrichment is that one may 'tune' the outcome, while physical enrichment is 'yes' or 'no'. Another key advantage of this approach is that one or multiple analysis parameters can be pursued to identify and characterize specific populations of interest.

**[0037]** Before the present compositions and methods are described, it is to be understood that this invention is not limited to particular compositions, methods, and experimental conditions described, as such compositions, methods, and conditions may vary. It is also to be understood that the terminology used herein is for purposes of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only in the appended claims.

**[0038]** As used in this specification and the appended claims, the singular forms "a", "an", and "the" include plural references unless the context clearly dictates otherwise. Thus, for example, references to "the method" includes one or more methods, and/or steps of the type described herein which will become apparent to those persons skilled in the art upon reading this disclosure and so forth.

**[0039]** Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods and materials are now described.

**[0040]** In general, reference to "a circulating tumor cell" is intended to refer to a single cell, while reference to "circulating tumor cells" or "cluster of circulating tumor cells" is intended to refer to more than one cell. However, one of skill in the art would understand that reference to "circulating tumor cells" is intended to include a population of circulating tumor cells including one or more circulating tumor cells.

**[0041]** The term "circulating tumor cell" (CTC) or CTC "cluster" is intended to mean any cancer cell or cluster of cancer cells that is found in a subject's sample. Typically CTCs have been exfoliated from a solid tumor. As such, CTCs are often epithelial cells shed from solid tumors found in very low concentrations in the circulation of patients with advanced cancers. CTCs may also be mesothelial from sarcomas or melanocytes from melanomas. CTCs may

also be cells originating from a primary, secondary, or tertiary tumor. CTCs may also be circulating cancer stem cells. While the term "circulating tumor cell" (CTC) or CTC "cluster" includes cancer cells, it also is intended to include non-tumor cells that are not commonly found in circulation, for example, circulating epithelial or endothelial cells. Accordingly tumor cells and non-tumor epithelial cells are encompassed within the definition of CTCs.

**[0042]** The term "cancer" as used herein, includes a variety of cancer types which are well known in the art, including but not limited to, dysplasias, hyperplasias, solid tumors and hematopoietic cancers. Many types of cancers are known to metastasize and shed circulating tumor cells or be metastatic, for example, a secondary cancer resulting from a primary cancer that has metastasized. Additional cancers may include, but are not limited to, the following organs or systems: brain, cardiac, lung, gastrointestinal, genitourinary tract, liver, bone, nervous system, gynecological, hematologic, skin, breast, and adrenal glands. Additional types of cancer cells include gliomas (Schwannoma, glioblastoma, astrocytoma), neuroblastoma, pheochromocytoma, paraganglioma, meningioma, adrenalcortical carcinoma, medulloblastoma, rhabdomyosarcoma, kidney cancer, vascular cancer of various types, osteoblastic osteocarcinoma, prostate cancer, ovarian cancer, uterine leiomyomas, salivary gland cancer, choroid plexus carcinoma, mammary cancer, pancreatic cancer, colon cancer, and megakaryoblastic leukemia; and skin cancers including malignant melanoma, basal cell carcinoma, squamous cell carcinoma, Karposi's sarcoma, moles dysplastic nevi, lipoma, angioma, dermatofibroma, keloids, sarcomas such as fibrosarcoma or hemangiosarcoma, and melanoma.

**[0043]** Using the methods described herein, rare cells, such as CTCs may be detected and characterized from any suitable sample type. As used herein, the term "sample" refers to any sample suitable for the methods provided by the present invention. The sample may be any sample that includes rare cells suitable for detection. Sources of samples include whole blood, bone marrow, pleural fluid, peritoneal fluid, central spinal fluid, urine, saliva and bronchial washes. In one aspect, the sample is a blood sample, including, for example, whole blood or any fraction or component thereof. A blood sample, suitable for use with the present invention may be extracted from any source known that includes blood cells or components thereof, such as venous, arterial, peripheral, tissue, cord, and the like. For example, a sample

may be obtained and processed using well known and routine clinical methods (e.g., procedures for drawing and processing whole blood). In one aspect, an exemplary sample may be peripheral blood drawn from a subject with cancer.

**[0044]** The term "blood component" is intended to include any component of whole blood, including red blood cells, white blood cells, platelets, endothelial cells, mesothelial cells or epithelial cells. Blood components also include the components of plasma, such as proteins, lipids, nucleic acids, and carbohydrates, and any other cells that may be present in blood, due to pregnancy, organ transplant, infection, injury, or disease.

**[0045]** As used herein, a "white blood cell" is a leukocyte, or a cell of the hematopoietic lineage that is not a reticulocyte or platelet. Leukocytes can include nature killer cells ("AK cells") and lymphocytes, such as B lymphocytes ("B cells") or T lymphocytes ("T cells"). Leukocytes can also include phagocytic cells, such as monocytes, macrophages, and granulocytes, including basophils, eosinophils and neutrophils. Leukocytes can also comprise mast cells.

**[0046]** As used herein, a "red blood cell" or "RBC" is an erythrocyte. Unless designated a "nucleated red blood cell" ("nRBC") or "fetal nucleated red blood cell", as used herein, "red blood cell" is used to mean a non-nucleated red blood cell.

**[0047]** The present invention provides a method whereby a biological sample may be assayed or examined in many different ways to detect and characterize rare cells. A sample may be stained or labeled with one or more detectable markers and examined by fluorescent microscopy and/or light microscopy. Unlike conventional enrichment schemes whose goal it is to eliminate the non-rare or non-CTCs from evaluation, the present invention relies on the non-rare cells or non-CTCs present in the sample to aid in the identification and characterization of the rare cells or CTCs. In the presently described non-enrichment method, the sample (e.g., blood or other body fluid, including urine, peritoneal, pleural, saliva, cerebral spinal, and the like) is minimally processed, and the rare cells, such as CTCs are not separated from other nucleated cells (e.g., non-rare or non-CTCs).

**[0048]** As used herein, the terms "non-rare cell" and "non-rare cells", generally refer to any cell that is not a rare cell as defined herein. Similarly, as used herein, the terms "non-CTC" and "non-CTCs", generally refer to any cell that is not a CTC as defined herein. Non-

rare and non-CTCs may include nucleated or enucleated cells, such as, in the case of blood, white blood cells (also called leukocytes) including neutrophils, eosinophils, basophils, lymphocytes, and monocytes; red blood cells (also known as erythrocytes); and platelets.

**[0049]** In the case of blood, while the CTCs may not be separated from other nucleated cells, red blood cells, which are typically only found nucleated in the blood of newborns, are removed from the sample before plating. This is commonly performed by lysing the red blood cells, although several alternative approaches are well known in the literature and may be utilized with the present methods, for example, removing the cells by filtration or density gradient centrifugation. After removing the red blood cells, the remaining cells may be processed by spinning, re-suspending, and plating the cells onto a solid support that may be used in cell imaging.

**[0050]** A variety of solid supports are well known in the art and include slides that may be treated to promote cellular attachment to the slide surface. The slide may be constructed from a variety of materials sufficient to provide a support for performing a biological assay. In an exemplary aspect, the support is composed of a material that may be coated with a compound that promotes electrostatic interaction of biological material to the support. A variety of substrate materials are well known in the art and suitable for use with the present invention. Such materials may include one or more of glass; organoplastics such as polycarbonate and polymethylmethacrylate, polyolefins; polyamides; polyesters; silicones; polyurethanes; epoxies; acrylics; polyacrylates; polyesters; polysulfones; polymethacrylates; polycarbonate; PEEK; polyimide; polystyrene; and fluoropolymers. In an exemplary aspect, the slide is manufactured from glass or plastic and includes one or more biologically interactive coatings.

**[0051]** Slides may include one or more active areas defined on the surface thereof. An active field, as used herein, is intended to include areas in which the slide has been chemically or electrically treated, such as with a biologically interactive coating, for example to promote the adhesion of cells to the slide. For example, the slide may be treated such that the surface is positively charged which allows for cells to be anchored to the surface through the electrostatic adhesion of a negatively charged cell. The slide may include from 1 to any number of active areas depending on the size of the slide and the intended application. In various aspects, the slide includes a single active area.

**[0052]** The total number of rare cells or CTCs that are adhered to a given slide is dependent, in part, on the initial sample volume. In various aspects, a wide range of initial sample volumes may be used to practice the present method and provide clinically significant results. As such, the initial sample volume may be less than about 1  $\mu$ l, 2  $\mu$ l, 2.5  $\mu$ l, 3  $\mu$ l, 4  $\mu$ l, 5  $\mu$ l, 6  $\mu$ l, 7  $\mu$ l, 7.5  $\mu$ l, 8  $\mu$ l, 9  $\mu$ l, 10  $\mu$ l, 12.5  $\mu$ l, 15  $\mu$ l, 17.5  $\mu$ l, 20  $\mu$ l, 25  $\mu$ l, 50  $\mu$ l, 75  $\mu$ l, 100  $\mu$ l, 125  $\mu$ l, 150  $\mu$ l, 175  $\mu$ l, 200  $\mu$ l, 225  $\mu$ l, 250  $\mu$ l, 300  $\mu$ l, 400  $\mu$ l, 500  $\mu$ l, 750  $\mu$ l, 1 ml, 2 ml, 3 ml, 4 ml, 5 ml, 6 ml, 7 ml, 8 ml, 9 ml or greater than about 10 ml. In an exemplary aspect, the initial sample volume is between about 200 and 500  $\mu$ l, 200 and 1000  $\mu$ l, 1000 to 2000  $\mu$ l, 1000 to 3000  $\mu$ l or 1000 to 5000  $\mu$ l. In another exemplary aspect, a sample processed as described herein includes greater than about 1, 2, 5, 7, 10, 15, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, or even 1000 rare cells or CTCs.

**[0053]** After adhering the minimally processed cells to a solid support, for example by plating the cells on a slide, the cells are contacted with one or more detectable markers to facilitate cell imaging via examination of the cells by fluorescent microscopy and/or light microscopy. In general, detectable markers include a variety of agents useful in detecting and characterizing cellular phenomenon. For example, detectable markers may include agents such as polynucleotides, polypeptides, small molecules, and/or antibodies that specifically bind to a marker present in a sample and which are labeled such that the agent is detectable when bound or hybridized to its target marker or ligand. For example, detectable markers may include enzymatic, fluorescent, or radionuclide labels. Additional reporter means and labels are well known in the art.

**[0054]** A marker can be any cell component present in a sample that is identifiable by known microscopic, histologic, or molecular biology techniques. Markers can be used, for example, to detect and characterize rare cells, including CTCs, and distinguish rare cells from non-rare cells and non-CTCs. In general a marker can be, for example, a molecule present on a cell surface, an overexpressed target protein, a nucleic acid mutation or a morphological characteristic of a cell present in a sample. Thus markers may include any cellular component that may be detected within or on the surface of a cell, or a macromolecule bound or aggregated to the surface of the cell. As such, markers are not limited to markers physically on the surface of a cell. For example, markers may include, but are not limited to surface antigens, transmembrane receptors or coreceptors, macromolecules bound to the

surface, such as bound or aggregated proteins or carbohydrates, internal cellular components, such as cytoplasmic or nuclear components, and the like. A marker may also include a blood component that binds preferentially to specific cell types, such as platelets or fibrin.

**[0055]** In one aspect, a detectable marker may be a detectably labeled antibody. Antibodies useful in the methods of the invention include intact polyclonal or monoclonal antibodies, as well as any fragments thereof, such as Fab and F(ab')<sub>2</sub>, as well as combinations of such antibodies or fragments. Methods for generating fluorescently labeled antibodies are well known in the art, for example, fluorescent molecules may be bound to an immunoglobulin either directly or indirectly by using an intermediate functional group. In related aspects, a detectable marker may be a nucleic acid molecule (*e.g.*, an oligonucleotide or polynucleotide). For example, *in situ* nucleic acid hybridization techniques are well known in the art and can be used to identify an RNA or DNA marker present in a sample or subsample (*e.g.*, individual cell).

**[0056]** In various aspects of the invention, the detectable markers used to stain the cells include one or more detectable markers that are tissue specific and thus used as a positive marker for a specific type of cell and/or tissue. As used herein, a “positive marker” is a detectable marker that specifically binds to a rare cell such as a CTC, but not a non-rare cell or non-CTC. For instance the positive marker may be epithelial and/or tissue specific, for example, cytokeratin and/or EpCAM marker may be used which bind preferentially to epithelial cells. Similarly, markers that are tissue specific may be employed. There are numerous examples of tissue-specific markers known in the art and suitable for use in practicing the invention, such as PSA and PSMA for prostate tissue, CDX2 for colon tissue and TTF1 for lung tissue (of the subpopulation of lung cancer patients that are TTF1 positive). As used herein a “positive marker” may also be a detectable marker that specifically binds to subpopulations of rare cells or CTCs, but not all rare cells or CTCs of a population. For example, a “positive marker” may specifically bind to HD-CTCs, but not all CTCs.

**[0057]** In various aspects of the invention, the detectable markers used to stain the cells include one or more detectable markers that specifically bind to non-rare cells or non-CTCs and may be used as a negative selector. As used herein a “negative marker” is a detectable marker that specifically binds to non-rare cells or non-CTCs and is a negative selector. The

most commonly used negative marker for non-CTCs is CD45, which binds preferentially to WBCs. There are other detectable markers or combinations of detectable markers that bind to the various subpopulations of WBCs. These may be used in various combinations, including in combination with or as an alternative to CD45. As used herein a “negative marker” may also be a detectable marker that specifically binds to subpopulations of non-rare cells or non-CTCs and is a negative selector.

**[0058]** In addition to positive and negative detectable markers to identify CTCs, additional detectable markers may be used to stain cells that specifically bind to the nucleus of the cell allowing differentiation of cells from non-cellular material. As used herein, a “nuclear marker” is a detectable marker that binds to a nuclear component of a cell and allows differentiation of cells from non-cellular material. The most common nuclear marker for use in the present invention is DAPI.

**[0059]** In various aspects of the invention, the detectable markers used to stain the cells include one or more detectable markers referred to herein as “content markers”. Content markers typically may include, detectably labeled oligonucleotide probes, such as FISH probes or immunohistochemistry probes. In one embodiment, content markers are applied to the slide at the same time as the positive and negative markers, or are applied to the slide after the positive and negative markers and after the identification of the rare cells by imaging. Content markers include detectable markers directed to EGFR, HER2, ERCC1, CXCR4, EpCAM, E-Cadherin, Mucin-1, Cytokeratin, PSA, PSMA, RRM1, Androgen Receptor, Estrogen Receptor, Progesterone Receptor, IGF1, cMET, EML4, or leukocyte associated receptor (LAR). In some cases, a content marker may also be a positive marker.

**[0060]** The intensity of signal from a positive marker, or any marker, is detectable on a scale of intensities, which based on the methodology of the disclosure, is highly quantifiable. The scale of intensity allows for vastly improved quantification and ranking of detectable events enabling further categorization. For example, a CTC that emits a low intensity signal for cytokeratin may be a cancer stem cell; or the change in the number of high/low cytokeratin cells might be either predictive of response or a readout of response (or resistance course). The same is true for positive, negative and content markers.



**[0061]** The present invention utilizes detectable markers to facilitate cell imaging via examination of the cells by fluorescent microscopy and/or light microscopy. In an exemplary aspect, the minimally processed cells are stained with several fluorescent markers, and then imaged using a fast, automated microscope. Typically, a prepared slide may be loaded onto the automated system or may be placed in a slide carrier that holds any number of additional slides. The slide carriers are loaded into an input hopper of the automated system. An operator may then enter data identifying the size, shape, and location of a scan area on each slide or the system can automatically locate a scan area for each slide during slide processing. The processing parameters of the slide may be identified by a bar code present on the slide or slide carrier. At system activation, a slide carrier is positioned on an X-Y stage, the entire slide, or portion thereof, is rapidly scanned. This may be done at low or high magnification and may be repeated at various levels of magnification and/or for various regions of the slide. Images may be stored on an appropriate storage medium and analyzed using executable code as is well known in the art for performing the various analysis discussed herein. As discussed herein, various parameters may be adjusted throughout the imaging process to facilitate detection of rare cells, for example, CTCs using data regarding non-rare or non-CTCs, such as exposure limits and intensity settings.

**[0062]** As used herein, the terms “image” and “sample image” generally refer to an image, digital or otherwise, of a minimally processed sample including various cells, such as rare cells and CTCs. Typically, a sample image is an image of all or a portion of a sample slide having cells adhered to its surface and optionally stained with one or more detectable markers.

**[0063]** One advantage of the present invention, which allows for tunable specificity/sensitivity and focuses on data reduction and analysis rather than enrichment, is that minimal processing is expected to minimize bias. In alternative techniques that require enrichment, rare cells are invariably lost in the process. Specifically, in the use of immunocapture or size filtration to distinguish between WBCs and CTCs, variation in the expression of the targeted antigen in the case of immunocapture or variation in the size differential between the WBC and CTC causes some CTCs to be lost during the enrichment phase. This can lead to (i) inaccurate counts of CTCs; (ii) too few CTCs for downstream

characterization or content analysis; and (iii) the creation of a selection bias as some types of CTCs are preferentially lost based upon their type of variation.

**[0064]** The challenge with the minimal processing approach is that it is difficult to find the low frequency rare cells or CTCs in the background of the non-rare cells or non-CTCs. The low frequency may be 1 rare cell or CTC : 1,000 non-rare cells or non-CTCs, 1:10,000, 1:100,000, 1:1,000,000, and even 1:10,000,000, or anywhere between those ratios. Complicating the ability to find and characterize the rare cells is that the positive and negative markers, while very selective, are not perfect resulting in either false positives or false negatives. In other words, it is common to have some background staining of the negative markers on the rare cells and/or some background staining of the positive markers on the non-rare cells. While assay optimization is used to minimize this background staining, it is challenging to completely eliminate the phenomenon with assay optimization.

**[0065]** As mentioned previously, most other approaches for finding rare cells attempt to remove the non-rare cells. The present invention uses the non-rare cells or non-CTCs to aid in finding and characterizing the rare cells or CTCs. The numerous ways in which non-rare cells and non-CTCs may be analyzed are discussed throughout the disclosure. Throughout this disclosure, non-rare cells or non-CTCs are typically referred to as a single group and may be analyzed using the methods described herein as such. However, the invention also recognizes that non-rare cells may contain various discrete subgroups. For example, in the case of CTCs, the various discrete subgroups may include neutrophils, macrophages, lymphocytes, eosinophils and basophils, and cells in varying states such as various states of apoptosis or cell division, that may be distinguished using the methods described herein by size, shape, nuclear characteristics, and staining pattern. In some embodiments of the invention, it may be useful to use one of these subgroups to aid in finding rare cells or CTCs, rather than to use the entire group. The use of non-rare or non-CTCs in the present invention is not meant to limit the invention to using only the entire group when it may be appropriate in some of the embodiments to use just one or more of the subgroups.

**[0066]** An enabling aspect of this invention is that the low frequency of rare cells or CTCs to non rare cells or non-CTCs allows one to treat the majority of cells as non-rare cells or non-CTCs even if they have not been definitively identified as such. The low frequency of rare cells and CTCs allows one to ignore such cells and assume the cells are non-rare cells or

non-CTCs to derive quality control, cut-off, normalization, and calibration metrics. Since the rare cells are in low abundance, if these metrics are to be refined taking into consideration the population of rare cells, outlier removal techniques may be utilized. The outlier removal techniques mathematically ensure that the population of rare cells does not factor into the metrics.

**[0067]** As discussed herein, the disclosed methodology allows detection, enumeration and characterization of populations of rare cells or subpopulations of rare cells. The methodology utilizes data from non-rare cells in the sample to identify and characterize rare cells by applying defined parameters pertaining to exposure limits, exposure settings, quality control, intensity cut-off limits, cell size and shape calibration, cell enumeration and content evaluation, each of which is further discussed in turn. In various aspects, the assay allows for simultaneous cytomorphic review of fluorescent images with individual channel images, augmented with cell-by-cell annotation with ancillary semi-quantitative data regarding size and fluorescent intensity of objects both absolute and relative to the non-rare cells or non-rare cell candidates, *e.g.*, non-CTCs or non-CTC candidates, from either the full experiment or the local environment.

**[0068] Establishing Exposure Limits.**

**[0069]** While variation should be minimized through assay optimization and instrument standardization, variation in the staining of the markers is common, slide-to-slide, batch-to-batch, operator-to-operator, and day-to-day. Thus selecting the right exposure for a particular slide is non-trivial, as setting it too low or too high will cause one to miss information. While standard approaches work for those markers that are common on the majority of events on the slide, it is challenging for those that are specific to rare cells or CTCs. Within the dynamic range of the imaging system, the signal in rare cells or CTCs and background in non-rare cells or non-CTCs are proportional to the exposure time. But noise which is random variation in both signal and background caused by electronics in the imaging system decreases when exposure increases. Ideally, exposure should be set to maximize the signal without saturating the imaging system. But this is impractical due to the impact on data collection time. Because a rare cell or CTC is present in very low frequency, it is unlikely that a rare cell or CTC would be found in a small number of Sample Images, preventing one from using the Sample Images to set the exposure for the positive marker. Complicating this

further, there is a natural variation in the expression of and staining of both positive and negative markers to their target cells. A small number of Sample Images to set exposure may not capture this natural variation on the target rare cells or CTCs.

[0070] In one embodiment of this invention, the signal from the non-rare cells or non-CTCs is utilized to set the exposure limit for the positive marker. This is somewhat counter-intuitive as the non-rare cell or non-CTC is not the target of choice for the positive marker. However in this embodiment, the exposure is adjusted so that a visible but low signal is observed from the non-rare cells or non-CTCs in the Sample Images originating from fluorescent sources such as nonspecific staining, autofluorescence and optical system properties. The brightfield imagery, nuclear marker and the negative marker may be used to identify the non-rare cells or non-CTCs in the Sample Images. The low signal is a distinguishable cellular signal in the non-rare cells or non-CTCs when compared to the non-cellular areas in the Sample Image. This process provides a method to set the exposure for the positive marker when the target of those markers are in low frequency and also helps to maximize the Signal/Background of the positive marker, both of which are aids to finding rare cells or CTCs while still minimizing the total time required to collect data. This phenomenon is especially true when the signal is low or dim. Once cellular background is statistically significant above non-cellular background, the exposure time for this particular marker is optimized for speed of data collection. All subsequent optimization can be performed *in silico*. Once the exposure is set, the entire slide is ready to be imaged at that setting.

[0071] While the above embodiment facilitates the setting of exposure for the positive marker, non-rare cells or non-CTCs may also be used to set the exposure for the negative markers, nuclear markers and content markers in a way that is relevant for the clinical interpretation of rare cells or CTCs. In one embodiment, the nuclear marker on non-rare cells or non-CTCs in the Sample Images is set to a level that allows the evaluation of the nuclear content of a cell, and in particular whether the cell is classified as live or dead, facilitating the calculations of live:dead ratios for cells by cell type. These exposure levels for the nuclear marker for the non-rare cells or non-CTCs will be satisfactory for the same evaluation of the nuclear marker for the rare cells or CTCs, in particular to distinguish nuclear shape on normal vs. malignant cells.

[0072] In another embodiment, the exposure for the negative marker is set from the Sample Images by looking at the distribution of the signal from that marker on the non-rare or non-CTCs where the exposure is chosen to maximize the signal/background ratio, especially at the critical low end of the dynamic range where a faint signal to a negative marker in a rare cell or CTC may occur.

[0073] In another embodiment, the exposure is set from the non-rare cells or non-CTCs for the content marker. The setting of the signal for the content marker using the non-rare cells or non-CTCs in Sample Images will depend on the specific content marker. For instance, some content markers may have relatively high expression in the non-rare cells or non-CTCs when compared to rare cells or CTCs, in which case one would use the information from the non-rare cells or non-CTCs in the Sample Images to set the upper boundary for the content marker. Conversely, the content marker may have relatively low expression in the non-rare cells or non-CTCs when compared to the rare cells or CTCs, in which case one would use the information from the non-rare cells or non-CTCs in the Sample Images to set the lower boundary for the content marker.

[0074] While the above embodiments use non-rare cells or non-CTCs to set the exposure limits from Sample Images for various markers prior to imaging the slide to find rare cells or CTCs, in another embodiment information from the non-rare cells or non-CTCs and/or rare cells or CTCs from the images taken during the first imaging event of the entire slide is used to set the exposure limits when selected areas of the slide are re-imaged. Selected areas are re-imaged for a variety of reasons, including collecting images that are in optimal focus or that are in a higher magnification. In this embodiment, the distribution of signals for the various markers in the non-rare cells or non-CTCs and the rare cells or CTCs across the entire slide may be used to calculate a better exposure that maximizes the desired signal or the desired dynamic range.

**[0075] Minimizing Exposure Setting to Minimize Data Collection Time and Maximize Throughput.**

[0076] As described above, exposure settings can be adjusted to optimize the signal or signal:background parameters. However, in another embodiment, exposure settings are adjusted with a goal of minimizing data collection time and maximizing throughput. For

example, one might determine that it takes 5 seconds of exposure time to fully utilize the dynamic range of the CCD camera but only 500 milliseconds to get the cellular background above the non-cellular background, hence saving 10x data collection time. Thus exposure times can be optimized either for maximum signal (or signal:background) or for minimum time.

**[0077] Quality Control.**

**[0078]** In another embodiment, the use of non-rare cells or non-CTCs to aid in identifying rare cells or CTCs also includes their use as quality control parameters. Since the non-rare cells or non-CTCs are represented in much higher frequency and distributed throughout the slide, they provide an available resource to evaluate the quality of the processing and imaging of the slide, both relative to a particular slide as well as across slides and across data sets.

**[0079]** In one embodiment, the invention provides observing the distribution of the non-rare cells or non-CTCs using the nuclear markers to identify the non-rare cells or non-CTCs. In this embodiment, the goal is to find cells, not necessarily to distinguish between non-rare cells or non-CTCs and rare cells or CTCs, and thus the positive or negative markers may not be utilized to distinguish between these categories; however, since the vast majority of the cells are non-rare cells or non-CTCs, most of the cells that utilized to determine distribution of cells on the slide are non-rare cells or non-CTCs. The distribution of the cells is important from a quality control standpoint as the desired distribution is an even distribution of cells with minimal overlap between the cells. If there is a substantial deviation from that ideal distribution, one may elect to reject the slide from further processing. While a nuclear marker is used in this example, any method for identifying the cells would suffice, including brightfield imaging and conventional stains such as Wright Giemsa.

**[0080]** In another embodiment, the co-location of the nuclear marker and the negative marker is used as a quality control method to evaluate whether the alignment of different images is satisfactory. In this embodiment, the nuclear marker and the negative marker should have significant overlap.

**[0081]** In another embodiment, the ratio between negative marker events and the nuclear marker events is a measure for the effectiveness of the negative marker staining, where the higher the ratio without exceeding 12 is desirable. A desirable negative marker may have a

0.8, 0.9, 1.0 or 1.1 ratio. In another embodiment, the distribution, including mean, standard deviation and coefficient of variation (CV) of the negative marker over the population of the non-rare cells or non-CTCs is used as a quality control parameter, where the distribution of the negative marker is consistent with expected distribution patterns of past experiments and/or consistent with the distribution of WBC's normal expression patterns.

**[0082]** In another embodiment, the distribution, including mean, standard deviation and CV of the positive marker over the population of the non-rare cells or non-CTCs is used as a quality control parameter, where the distribution of the positive marker is consistent with expected distribution patterns from past experiments.

**[0083]** While the quality control methods described above describe methods to evaluate overall slide quality, the same methods may be used to evaluate an image or a group of images. In some instances, the parameters derived from an image or a group of images in a region may be compared to the same parameters calculated over the entire slide. In another instance, the quality control parameters described above may be compared across different slides.

**[0084]** In another embodiment, cell loss may be calculated from the slide during processing by comparing the ratio of the nuclear marker events or negative marker events to the known number of non-rare cells or non-CTCs placed on the slide, where the known number of non-rare cells or non-CTCs is derived from the WBC count and the volume used in the experiment.

**[0085] Setting Intensity Cut-off Limits for CTC Detection (minimizing false negatives).**

**[0086]** As mentioned above, the challenges in this approach to rare cell and CTC detection are 1) that the relative frequency of rare cells, such as CTCs, to non-rare cells or non-CTCs is low; and 2) the imperfect staining of the positive and negative markers to CTCs and non-CTCs respectively. However, the present method takes those challenges and turns them into strengths. In one embodiment, the background signal from the positive markers on the highly abundant non-rare cells or non-CTCs is used to calculate mean, standard deviation and CV. Those metrics are subsequently used to determine detection cut-offs to separate rare cells, such as CTCs from non-rare cells or non-CTCs. In one aspect, the factor of 10 multiplied by

the standard deviation and added to the mean for the non-rare cell or non-CTC positive marker signal, is used as a cut-off to distinguish rare cells or CTCs, where putative rare cells or CTCs are determined to have a positive marker signal greater than that cut-off. In other aspects, the metric uses a factor of 5, 7.5, 12.5, 15, 17.5, 20 or more, or any number between those numbers. The calculation of the metric may be set on a global slide basis. Alternatively, it may be set on an image basis or a regional basis.

**[0087]** In another embodiment, the cut-off may be determined dynamically within each image by locating a signal of the highest positive marker events, then comparing that signal to the standard deviation between additional signal events. In an exemplary aspect, the cut-off may be determined dynamically within each image by locating signals of the five highest positive marker events, then comparing that signal to the standard deviation between the next fifty positive signal events. Positive marker events can also include multiple positive markers or inclusion of positive markers and exclusion of negative marker events. The number 'five' could be varied from 1 to 10 per field of view assuming a 10x magnification (*i.e.*, assuming no more than 5 rare cells or CTCs per field of view or a relative concentration of no more than 1 in 500). This approach has the advantage of being entirely numerical and not being based on shape analysis. It is expected to robustly and substantially reduce the number of possible events with minimal risk of missing events.

**[0088]** In another embodiment, the cut-off for the negative marker signal is set using the mean and standard deviation of the non-rare cells or non-CTCs. In this case, the cut-off is derived from a factor of 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 multiplied by the standard deviation of the negative marker signal for the non-rare cells or non-CTCs and subtracted from the mean negative marker signal. Putative CTCs will have a negative marker signal below that cut-off. As described above, this cut-off may be generated globally by using the signal from all non-rare cells or non-CTCs on the slide or at the image or regional level by using the signal only from the non-rare cells or non-CTCs in that image or that region.

**[0089] Size and Shape Calibration.**

**[0090]** In another embodiment of the invention, the cellular size and distribution of cell sizes for the non-rare cells or non-CTCs as measured on the slide is compared to the published sizes and distributions for the WBCs in the literature. In addition, the distribution



may be corrected for individual patient differences by using differential count of the subgroups of the WBCs obtained from an automated cell counter. The ratio between the published size and the calculated size of the non-rare cells or non-CTCs may then be used as a correction factor to determine an accurate size for the rare cells or CTCs by multiplying that ratio with by the size of the rare cell or CTC as measured on the slide. Calibration allows for standardization and comparisons across slides and across blood tubes and across patients and across indications.

**[0091]** In another embodiment, we compare one or more of the nuclear size, distribution of nuclear size, and contour patterns for non-rare cells or non-CTCs as measured on an image or slide to nuclear size and contour pattern of a putative rare cell or CTC and apply cut-off values to qualify that putative rare cell or CTC as a valid rare cell or CTC.

**[0092] Enumeration.**

**[0093]** In an embodiment of the invention, the non-rare cells or non-CTC information is used to accurately calculate the concentration of rare cells or CTCs in a bodily fluid. In one example of this embodiment, one determines the ratio of CTCs to total CTCs + non-CTCs (all nuclear marker events) and then divides that by the volume of the original body fluid used for that experiment, and finally multiply that by the original concentration of the cells in the body fluid. The later measurement may be obtained through the use of a standard automated cell counter (or cytometer).

**[0094] Content Evaluation.**

**[0095]** As mentioned above, the rare cell or CTC expression level of a content marker may be evaluated. However, because of the slide variability caused by processing and imaging variation, it is difficult to provide universal parameters to determine the expression level. To mitigate this challenge, the expression level of the content marker and its distribution in non-rare cells or non-CTCs in a patient population is first determined. Then for the individual cancer patient, the expression level of the content marker in rare cells or CTCs and in non-rare cells or non-CTCs is determined. The ratio of the expression level between rare cells or CTCs and non-rare cells or non-CTCs becomes a relative measure that normalizes slide-to-slide variation. In addition, multiplying that ratio by the mean non-rare

cell or non-CTC expression level from the control population provides for an absolute value for the rare cell or CTC expression corrected for slide-to-slide variation.

**[0096]** Nuclear shape and size is a potential rich source of information. It is expected to give detailed information about the type of cell in the case of a blood cell and about the state of the cell in the case of a rare cell or CTC. For example, nuclear shape could give insight to the malignant nature of the cell, it could give insight into the state of cell viability and/or cell cycle. For example, a patient undergoing a successful chemotherapy might see a spike in CTCs but nuclear interpretation might show that these CTCs are non-viable/apoptotic.

**[0097]** Another potential opportunity for using nuclear size and shape would be similar to FACS analyses using forward scatter and side scatter as ways to characterize cells. It may be possible to gather enough detailed information about the nucleus and other cellular components to recapitulate the forward and side scatter information from FACS.

**[0098]** In various embodiments of the invention, a single or combination of parameters may be utilized in performing the assay depending, in part on the data to be determined and the rare cell population being investigated. Additionally, subpopulations of specific rare cell populations may be identified using the disclosed methodology. For example, as shown in Example 1, subpopulations of CTCs may be identified and differentiated by further defining specific assay parameters. The Example discloses identification and classification of a CDC subpopulation referred to as HD-CTC.

**[0099]** In one embodiment of the invention, by further defining parameters of the disclosed assay, a strict classification the HD-CTC was established. Without being bound to any particular theory, it is believed that the HD-CTC subpopulation as classified herein, includes CTCs exhibiting the highest potential of becoming an intact tumor cell. All other CTCs partially fulfill the defined parameters but lack one or more of the strict inclusion criteria. Non-HD-CTCs are CTCs which may be less reliable in evaluation performed in further downstream methodologies.

**[0100]** In one embodiment, an HD-CTC is a cell that comprises a) a positive marker; b) has an intact nucleus; and c) is morphologically distinct from normal WBCs, wherein the cell is not positive for a negative marker. As discussed herein, the positive marker may be a marker that preferentially binds to epithelial cells, such as cytokeratin and/or EpCAM.

Further, as discussed herein, the negative marker may be any non-cancer specific marker, such as CD45 which preferentially binds to WBCs. Determination of an intact nucleus is typically determined by DAPI imaging, but other suitable nuclear markers are well known in the art. In an exemplary embodiment, an HD-CTC is a cell that is a) cytokeratin positive; b) CD45 negative; c) has an intact nucleus; and d) is morphologically distinct from normal WBCs.

**[0101]** In various embodiments, the positive marker may have an intensity that is greater than 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 times that of a nucleated white blood cells' cytokeratin intensity. In various embodiments, the intensity of the negative marker is in the lowest 10%, 5%, 4%, 3%, 2% or less of all cellular events.

**[0102]** In various embodiments, the nucleus of a HD-CTC is intact and non-apoptotic. Mild apoptotic changes in the cytoplasm are accepted, as long as the nucleus does not appear apoptotic.

**[0103]** In various embodiments HD-CTCs are morphologically distinct from normal WBCs. For example, HD-CTCs may have a morphology that is consistent with a malignant epithelial cell by criteria used in standard diagnostic cytopathology, predominantly embodied as enlarged size, but that may also include cytomorphologic features, such as, architectural organization of nucleus and cytoplasm, cytoplasmic shape, and nuclear shape. A gallery of representative HD-CTCs is displayed in Figure 2.

**[0104]** While the methods described in this invention are useful in detecting rare cells using data derived from analysis of non-rare cells, as discussed throughout, the invention also is useful in characterization of rare cells. In particular, use of the various combinations of detectable markers and computational methods for performing cell imaging and analysis allow for meaningful characterization useful in assessing cancer prognosis and in monitoring therapeutic efficacy for early detection of treatment failure that may lead to disease relapse. In addition, CTC analysis according to the invention enables the detection of early relapse in presymptomatic patients who have completed a course of therapy. This is possible because the presence of CTCs has been associated and/or correlated with tumor progression and spread, poor response to therapy, relapse of disease, and/or decreased survival over a period of time. Thus, enumeration and characterization of revealed CTCs provides methods to

stratify patients for baseline characteristics that predict initial risk and subsequent risk based upon response to therapy.

**[0105]** The term “subject” as used herein refers to any individual or patient to which the subject methods are performed. Generally the subject is human, although as will be appreciated by those in the art, the subject may be an animal. Thus other animals, including mammals such as rodents (including mice, rats, hamsters and guinea pigs), cats, dogs, rabbits, farm animals including cows, horses, goats, sheep, pigs, etc., and primates (including monkeys, chimpanzees, orangutans and gorillas) are included within the definition of subject.

**[0106]** Accordingly, in another embodiment, the invention provides a method for diagnosing or prognosing cancer in a subject. The method includes detecting CTCs as described herein. CTCs may then be analyzed to diagnose or prognose cancer in the subject. As such, the methods of the present invention may be used, for example, to evaluate cancer patients and those at risk for cancer. In any of the methods of diagnosis or prognosis described herein, either the presence or the absence of one or more indicators of cancer, such as, a cancer cell, or of any other disorder, may be used to generate a diagnosis or prognosis.

**[0107]** In one aspect, a blood sample is drawn from the patient and processed to detect CTCs as described herein. Using the method of the invention, the number of CTCs in the blood sample is determined and the CTCs are characterized by analysis of the detectable markers and other data gathered from imaging the cells. For example, analysis may be performed to determine the number and characterization of CTCs in the sample, and from this measurement, the number of CTCs present in the initial blood sample may be determined.

**[0108]** In various aspects, analysis of a subject’s CTC number and characterization may be made over a particular time course in various intervals to assess a subject’s progression and pathology. For example, analysis may be performed at regular intervals such as one day, two days, three days, one week, two weeks, one month, two months, three months, six months, or one year, in order to track level and characterization of circulating epithelial cells as a function of time. In the case of existing cancer patients, this provides a useful indication of the progression of the disease and assists medical practitioners in making appropriate therapeutic choices based on the increase, decrease, or lack of change in circulating epithelial

cells, such as the presence of CTCs in the patient's bloodstream. Any increase, be it 2-fold, 5-fold, 10-fold or higher, in the number of CTCs over time decreases the patient's prognosis and is an early indicator that the patient should change therapy. Similarly, any increase, be it 2-fold, 5-fold, 10-fold or higher, indicates that a patient should undergo further testing such as imaging to further assess prognosis and response to therapy. Any decrease, be it 2-fold, 5-fold, 10-fold or higher, in the number of CTCs over time shows disease stabilization and a patient's response to therapy, and is an indicator to not change therapy. For those at risk of cancer, a sudden increase in the number of CTCs detected may provide an early warning that the patient has developed a tumor thus providing an early diagnosis. In one embodiment, the detection of revealed CTCs increases the staging of the cancer.

**[0109]** In any of the methods provided herein, additional analysis may also be performed to characterize CTCs, to provide additional clinical assessment. For example, in addition to image analysis, gene expression analysis and PCR techniques may be employed, such as gene chip analysis and multiplexing with primers specific for particular cancer markers to obtain information such as the type of tumor, from which the CTCs originated, metastatic state, and degree of malignancy. Additionally, cell size, DNA or RNA analysis, proteome analysis, or metabolome analysis may be performed as a means of assessing additional information regarding characterization of the patient's cancer. In various aspects, analysis includes antibodies directed to or PCR multiplexing using primers specific for one or more of the following markers: EGFR, HER2, ERCC1, CXCR4, EpCAM, E-Cadherin, Mucin-1, Cytokeratin, PSA, PSMA, RRM1, Androgen Receptor, Estrogen Receptor, Progesterone Receptor, IGF1, cMET, EML4, or Leukocyte Associated Receptor (LAR).

**[0110]** For example, the additional analysis may provide data sufficient to make determinations of responsiveness of a subject to a particular therapeutic regime, or for determining the effectiveness of a candidate agent in the treatment of cancer. Accordingly, the present invention provides a method of determining responsiveness of a subject to a particular therapeutic regime or determining the effectiveness of a candidate agent in the treatment of cancer by detecting CTCs of the subject as described herein and analyzing the detected CTCs. For example, once a drug treatment is administered to a patient, it is possible to determine the efficacy of the drug treatment using the methods of the invention. For example, a sample taken from the patient before the drug treatment, as well as one or more

cellular samples taken from the patient concurrently with or subsequent to the drug treatment, may be processed using the methods of the invention. By comparing the results of the analysis of each processed sample, one may determine the efficacy of the drug treatment or the responsiveness of the patient to the agent. In this manner, early identification may be made of failed compounds or early validation may be made of promising compounds.

**[0111]** Four important indicators that provide insight to the clinical activity of candidate compounds include HER2, EGFR, CXCR4, and EphB4 RTK. HER2 provides an indicator of malignancy of a cell by determining mRNA stability and subcellular localization of HER2 transcripts. The resistance of EGFR to acquire mutations, and/or the mutations acquired provides important indicators of the activity of a candidate compound in addition to possible alternative compounds that may be used in combination with the candidate compound. An assessment of the level of DNA repair interference induced with platinum provides insight as to the status of the CXCR4 marker and metastatic condition. Additionally, assessment of the status of EphB4 receptor tyrosine kinase provides insight as to the metastatic potential of the cell. Accordingly, using the methods of the present invention, patients taking such candidate drugs may be monitored by taking frequent samples of blood and determining the number of circulating epithelial cells, for example CTCs, in each sample as a function of time. A further analysis of the Her2, EGFR, CXCR4, and EphB4 RTK indicators provides information as to pathology of the cancer and efficacy of the candidate drug. Similarly, ERRC1, Cytokeratin, PSA, PSMA, RRM1, Androgen Receptor, Estrogen Receptor, Progesterone Receptor, IGF1, cMET, EML4 and others provide insight into the clinical activity of candidate compounds. The analysis of these indicators of clinical activity may be through analysis of detectable markers as discussed herein (*e.g.*, immunohistochemistry and fluorescent in situ hybridization (FISH)) or further analysis via techniques such as sequencing, genotyping, gene expression or other molecular analytical technique.

**[0112]** Analysis of CTCs provide a method of determining candidate subjects for a particular clinical trial. For example, the detected CTCs of a candidate may be analyzed to determine whether specific markers exist in order to determine whether the particular therapeutic regime of the clinical trail may be potentially successful. Accordingly in another embodiment, the invention provides a method for determining a candidate subject for a clinical trial. The method includes detecting CTCs of the subject as described herein. The

CTCs may then be analyzed to determine whether the candidate subject is suitable for the particular clinical trial.

[0113] Analysis of CTCs during a clinical trial will provide information on whether the patient is responding or not responding to the experimental drug, where no substantial change or a decrease in revealed CTCs indicates response and an increase in revealed CTCs indicates poor response. The increase or decrease may be 2-fold, 10-fold or higher. This information is an early indicator of the drug's effectiveness and may be used by the investigators as a secondary endpoint in the clinical trial.

[0114] The following examples are intended to illustrate but not limit the invention.

### **EXAMPLE 1**

#### **CTC ASSAY AND IDENTIFICATION OF CTC SUBPOPULATION**

[0115] The data presented here demonstrate the methodology of the present invention as applied to CTCs and subpopulations of CTCs, such as HD-CTCs as defined herein. The assay is performed via a controlled prospective protocol to address the reliability and robustness of the assay as well as a split sample comparison with the Cellsearch®. After this technical validation, the assay was used to investigate the incidence and prevalence of CTCs and specific CTC subpopulations in patients with metastatic breast, prostate, and pancreatic cancers as well as normal controls. The specific subpopulation of CTCs targeted by the assay requires that the cell(s) have an intact nucleus, express cytokeratin and not CD45, are morphologically distinct from surrounding white blood cells (WBCs) and have cytologic features consistent with intact malignant epithelial cells suitable for downstream analysis.

[0116] The following methods and protocols were utilized.

[0117] Patients and Blood Sample Collection was performed as follows. Samples were collected from metastatic cancer patients in anti-coagulated blood tubes at Scripps Clinic, University of California, San Diego, Billings Clinic, and University of California, San Francisco under IRB approved protocols. Samples from non-local sites (UCSF, Billings Clinic) were shipped overnight so that the sample was received and processed within 24 hours. Samples from local sites (Scripps Clinic and UCSD) were held at room temperature

for 16-24 hours to mimic samples coming from non-local sites. Blood specimens were also drawn from normal controls from the TSRI Normal Blood Donor Service.

**[0118]** Blood sample processing for HD-CTC detection was performed as follows. Blood specimens were rocked for 5 minutes before a white blood cell (WBC) count was measured using the Hemocue<sup>TM</sup> white blood cell system (HemoCue, Sweden). Based upon the WBC count, a volume of blood was subjected to erythrocyte lysis (ammonium chloride solution). After centrifugation, nucleated cells were re-suspended in PBS and attached as a monolayer on custom made glass slides (Marienfeld, Germany). The glass slides are the same size as standard microscopy slides but have a proprietary coating that allows maximal retention of live cells. Each slide can hold approximately 3 million nucleated cells, thus the number of cells plated per slide depended on the patients WBC count. Enough blood was lysed to produce 15 slides and the cells were subsequently dried onto the slides after a cell preservative was added. All 15 slides were stored at -80°C for at least 24 hours.

**[0119]** For HD-CTC detection in cancer patients for this investigation, 4 slides were used as a test. The remaining slides created for each patient were stored at -80°C for future experiments. Four slides were thawed from each patient, then cells were fixed with 2% paraformaldehyde, permeabilized with cold methanol, and non-specific binding sites were blocked with goat serum. Slides were subsequently incubated with monoclonal anti-pan cytokeratin antibody (Sigma) and CD45-Alexa 647 (Serotec). After PBS washes, slides were incubated with Alexa Fluor 555 goat anti-mouse antibody (Invitrogen). Cells were counterstained with DAPI and mounted with an aqueous mounting media.

**[0120]** Imaging and technical analysis was performed as follows. All four slides from each patient were scanned using a custom made fluorescent scanning microscope which has been developed and optimized for fast, reliable scanning. One scanning instrument was used for all patient samples in this report to standardize results. Additionally, the light source was calibrated weekly and an algorithm was developed to standardize the exposures of each fluorophore on each patient slide during the scan. Each slide was scanned entirely at 10X magnification in 3 colors and produced over 6900 images. The resulting images were fed to an analysis algorithm that identifies likely candidate HD-CTCs based upon numerous measures, including cytokeratin intensity, CD45 intensity, as well as nuclear and cytoplasmic



shape and size. A technical analyst then goes through algorithm generated likely candidates and removes hits that are obviously not cells, such as dye aggregates.

[0121] Professional analysis and interpretation was performed as follows. All likely candidate CTCs are presented to a hematopathologist for analysis and interpretation through a web based report where the pathologist is able to include or exclude each candidate cell as an HD-CTC. Cells are classified as HD-CTCs if they are cytokeratin positive, CD45 negative, contained an intact DAPI nucleus without identifiable apoptotic changes (blebbing, degenerated appearance) or a disrupted appearance, and are morphologically distinct from surrounding white blood cells (usually a shape based feature, although occasionally purely size based. They must have cytoplasm that is clearly circumferential and within which the entire nucleus is contained. The cytoplasm may show apoptotic changes such as blebbing and irregular density or mild disruption at the peripheral cytoplasmic boundary, but must not be so disrupted that its association with the nucleus is in question. The images are presented as a digital image, with individual fluorescent channel viewing capability as well as a composite image. Each cell image is annotated with ancillary statistical data regarding relative nuclear size, fluorescent intensities, and comparative fluorescent intensities. Each HD-CTC candidate is presented in a field of view with sufficient surrounding WBCs to allow for contextual comparison between cytomorphologic features of the cell in question versus the background white blood cells.

[0122] Cell line experiments were performed as follows. Four aliquots from the donor (2 ml each) were spiked with varying numbers of SKBR-3 cells to produce 4 slides with approximately 300, 100, 30 and 10 cancer cells per a slide. The 16 slides were then processed and analyzed by a single operator according to the HD-CTC sample preparation protocol. A single instrument was used to image all 16 slides.

[0123] **HD-CTC Classification:** HD-CTCs were defined as cells that are a) cytokeratin positive (intensity >6 times that of nucleated white blood cells' cytokeratin intensity); b) CD45 negative (intensity in lowest 2% of all cellular events); c) include an intact non-apoptotic appearing nucleus by DAPI imaging; and d) are morphologically distinct from normal WBCs.

[0124] Inclusion requirements for the morphological assessment of HD-CTC include 1) a nuclear size 30% greater than the average surrounding WBC nuclei, and 2) circumferential cytokeratin positive cytoplasm with an average intensity 600% greater than surrounding nucleated WBCs. Common, although not required, features of HD-CTCs include quite large nuclei up to five times the average size of surrounding WBC nuclei, nuclear contours distinct from surrounding WBC nuclei including elongation, large cytoplasmic domain with a frequently eccentric distribution and/or polygonal or elongated cytoplasmic shape, and doublets and clusters of 3 or more HD-CTCs. Other cell-like objects that are cytokeratin positive, CD45 negative, and contain a nucleus but do not meet the inclusion criteria, for example are the same size as WBCs or have cytoplasm that is not circumferential, are not counted as HD-CTCs but are tracked by the assay. The purpose of this approach is to have strict inclusion criteria for a specific phenotype of CTCs, while retaining data about events that fulfill only some of the requirements, but which might still be clinically meaningful, such as apoptotic tumor cells or tumor cell fragments or cells undergoing epithelial to mesenchymal transition.

[0125] **Results.**

[0126] **Assay Linearity and Sensitivity using Spike-in experiments:** To test assay linearity and sensitivity, specific dilutions ranging from 10 to 300 breast cancer cell line SKBR3s were spiked into normal control blood. Experiments were performed in quadruplicates and processed and analyzed according to the HD-CTC assay as explained in the Methods Section. The mean observed SKBR3s is plotted against expected SKBR3s and displays a correlation coefficient ( $R^2$ ) of 0.9997 (Figure 1).

[0127] **Assay Robustness of HD-CTC Counts in Patients with Carcinomas:** Assay robustness of the HD-CTC assay was tested against multiple processors and split samples. Duplicate tests were performed by two separate processors on 9 different patient samples. A comparison of HD-CTC/mL counts between two processors using split samples has a correlation coefficient ( $R^2$ ) of 0.979 (Figure 3). All data were analyzed by a single operator blinded to the experiment.

[0128] **Assay Specificity in samples from Normal Controls:** Fifteen healthy donors from an institutional healthy donor pool were evaluated as a control population consisting of

8 females and 7 males with an age range of 24 to 62 years. In all but one healthy control, the number of such events when corrected for volume was 1 HD-CTC/ml or less. The outlier was a healthy female donor with an HD-CTC count of 4/ml. Upon explicit review of her cells, about one third of them strongly met all inclusion criteria, while the remaining two thirds fulfilled all criteria but were near the lower limit for inclusion by one or more criteria. Four other healthy donors had 1 HD-CTC/ml. Explicit review of these cells revealed a similar pattern, in that about one third strongly met all criteria, while the remaining two thirds of the cells fulfilled criteria, but were near the lower limit for inclusion by one or more criteria. Examples of included events that are near the lower limit for inclusion are cells that measure 30% larger than surrounding WBCs but don't appear significantly larger by morphologic evaluation, and cells that are slightly out of focus and might have apoptotic nuclear changes that are not detectable by eye, and finally, occasional cells that have objective cytokeratin intensity measurements above the cutoff but subjectively don't appear significantly brighter than surrounding WBCs by single channel fluorescent review.

**[0129] Table 1: Comparison of HD-CTC Assay to CellSearch®**

<b>Cancer Type</b>	<b>HD-CTCs/mL</b>	<b>CellSearch/mL</b>
Breast #1	49.3	0.1
Breast #2	87.0	0.0
Breast #3	33.4	0.1
Breast #4	199.3	0.1
Breast #5	5.0	3.1
Prostate #1	2.3	0.0
Prostate #2	8.4	0.4
Prostate #3	107.3	2.8
Prostate #4	1.3	0.0
Prostate #5	150.5	0.1
Prostate #6	0.0	0.0
Prostate #7	1.4	0.5
Prostate #8	1.5	0.1
Prostate #9	145.3	0.8
Prostate #10	57.6	0.0

(Extrapolated to CTCs/ mL)

**[0130] Comparison of HD-CTC assay to CellSearch®:** A total of 15 patients, 5 metastatic breast cancer and 10 metastatic prostate cancer, were evaluated for CTCs with both Cellsearch® and the HD-CTC assays. Two tubes of blood were collected from each patient. One tube of 7.5 mL of blood was collected in CellSave™ tubes (Veridex, Raritan NJ) and sent to Quest Diagnostics (San Juan Capistrano, CA) for enumeration of CTCs using the Cellsearch® assay. A second tube of blood was collected from each patient and processed according to the HD-CTC protocol 24 hours after the blood draw, consistent with the standard HD-CTC process in order to mimic the timing at which samples were processed at Quest Diagnostics. The CellSearch® assay detected 2 or more CTCs per 7.5mL of blood in 5/15 patients tested. In contrast, the HD-CTC assay detected significantly higher numbers of CTCs in significantly more patients (HD-CTCs were identified in 14/15 patients tested, Table 1).

**[0131] Incidence of HD-CTCs in patients with metastatic cancer:** HD-CTCs were also enumerated in an additional cohort of 30 metastatic breast cancer patients, 20 metastatic prostate cancer patients, 18 metastatic pancreatic cancer patients, and 15 normal controls. Using this approach,  $\geq 5$  HD-CTCs/mL were found in 80% of the prostate cancer patients (mean = 92.2), 70% of the breast cancer patients (mean = 56.8), 50% of the pancreatic cancer patients (mean = 15.8), and 0% of normal controls (mean = 0.6) (Table 2).

**[0132] Table 2: Percentage of patients with HD-CTCs/mL of blood obtained from cohort.**

	N	$\geq 2$	$\geq 5$	$\geq 10$	$\geq 50$
Prostate	20	90%	80%	65%	40%
Breast	30	80%	70%	60%	27%
Pancreatic	18	61%	44%	44%	11%
Normal	15	7%	0%	0%	0%

**[0133] Morphology of HD-CTCs:** A heterogeneous population of CTCs within and across patients was observed. CTCs had various shapes, sizes, and cytokeratin intensities. In some cases, distinctive cytologic features such as large size or polygonal cytoplasmic shape, were quite distinctive and monotonous within the patient's sample. In other cases, there was cytomorphologic variability between HD-CTCs within a single sample. Cell size also

varied; many patient samples had HD-CTCs with nuclei uniformly three or four times the size of neighboring WBC nuclei, while other patients had cells with nuclei only 1.3 times the size of neighboring WBC nuclei. Some patients had a range of sizes. A lower limit for HD-CTC nuclear size of 1.3 times the average WBC nucleus was selected based on evaluation of the largest nuclear size of cells we identified as WBCs showing false nonspecific staining with cytokeratin, for instance, CD45 positive and cytokeratin positive.

**[0134]** Interestingly, using this platform that allows for detailed morphologic evaluation, HD-CTC clusters were identified in the majority of the cancer patients (88%) in this cohort, ranging from clusters of 2 HD-CTCs to greater than 30 HD-CTCs (data not shown). Each HD-CTC was cytokeratin positive, CD45 negative, contained a DAPI nucleus, and was morphologically distinct from surrounding nucleated cells.

**[0135]** In addition to counting HD-CTCs, a number of different categories of cells were tracked including cells that had nuclei displaying apoptosis, cells that didn't have circumferential cytokeratin, other cells that were the same size or smaller than surrounding WBC, and cells that were cytokeratin dim or negative (images not shown). Specifically, some CTCs were excluded because they lacked various morphologic or morphometric inclusion criteria: including one or more of: a) cytokeratin intensity too dim; b) nuclear size too small; c) cytokeratin insufficiently circumferential (surrounds less than 2/3 of nucleus); d) cytokeratin too dim, although appears to be a cluster of two very large cells; e) nucleus shows apoptotic disintegration changes; f) nucleus too small and cytoplasm insufficiently circumferential; appears to be a cell in late apoptosis; g) nucleus too small (same size as surrounding WBC nuclei); h) cytokeratin present, but not circumferential; and i) cytoplasm insufficiently circumferential, nucleus too small.

**[0136]** Although many of these events may in fact represent circulating malignant epithelial cells in various stages of anoikis or disruption secondary to even the minimal processing utilized in the platform, the goal is to identify a 'pure' population of cells with a very high likelihood of representing intact circulating functional malignant potentially metastasizing epithelial cells that are suitable for downstream analysis by secondary methodologies. Fragmented, disrupted, pyknotic or otherwise damaged carcinoma cells are not considered evaluable in standard diagnostic pathology, and thus they are excluded in this fluid phase biopsy platform as well. They are enumerated and tracked, as it is recognized

that their presence likely correlates overall with the tumor biology in the patient, either by reflecting overall tumor burden or by reflecting some as yet ill-understood complex equation involving tumor burden and tumor vascularity and efficiency of intravascular immune surveillance; however, they are not useful for secondary analysis, and thus they are not designated as HD-CTCs.

[0137] Many patients, in addition to having HD-CTCs, had a substantial number of cells that had nuclei that were morphologically distinct from surrounding WBC, resembled the nuclei of the HD-CTCs within that sample, and were CD45 negative, but were also cytokeratin dim or negative (data not shown). Representative types of CTCs found in a single prostate cancer patient included CTCs that were negative for cytokeratin and CD45, but exhibited a large nucleus similar to other CTCs found in this patient, typical CTCs that were cytokeratin positive, CD45 negative, and which had a DAPI nucleus, and CTC clusters of 4 cells.

[0138] In light of the extensive current debate about the possible existence of carcinoma cells undergoing epithelial-to-mesenchymal transition, the appearance and protein expression pattern of these cells identifies them as possible candidates for such a cell type.

[0139] **Discussion.**

[0140] The robustness of the HD-CTC platform was evaluated with both cell lines and patient samples. Despite little automation, and complete manual wetlab processing, of the current HD-CTC assay used for this cohort, the reproducibility of the assay is impressive with an average CV of less than 9% for 9 different samples processed by two independent technicians.

[0141] The criteria used to define a CTC are different across different technologies. The currently disclosed identified a subpopulation that has the highest likelihood of being *bonafide* tumor cells. Even with strict criteria, the incidence of CTCs using our assay is much higher than many technologies and in the same range as reported by the CTC-chip. Additionally, a head on comparison to CellSearch® showed significantly more CTCs in a higher proportion of patients. Whereas other technologies have observed occasional doublets and clusters of CTCs, clusters of CTCs were observed in most patients.

[0142] It is noteworthy that in the small cohort tested, the frequency of detection, and relative concentration, of CTCs among different tumor types using the methodologies of the invention (prostate > breast > pancreatic) parallels the findings observed using other methods such as CellSearch®.

[0143] Observationally, CTCs track over the clinical course of a small subset of prostate cancer patients in which serial draws were performed (Figure 4). Serial HD-CTC detection may be embedded into therapeutic clinical trials. This is expected to allow study of patients with uniform clinical characteristics who are treated similarly and in whom long-term clinical follow-up will be performed. In addition to correlating these cells with patient outcomes to determine their prognostic and monitoring value, these HD-CTCs are expected to serve as a pharmacodynamic tool for assessing on-target effects at a molecular level of drugs of interest.

[0144] In summary, the instant example provides data that the HD-CTC assay (i) finds significant number of CTCs in most patients with metastatic cancer, (ii) has improved sensitivity over the Cellsearch® System, (iii) provides HD-CTCs in an ideal format for downstream characterization, (iii), enables the prospective collection of samples that can be stored frozen for long periods of time, and then retrospectively analyzed as new assays or markers become available.

## EXAMPLE 2

### CTC ASSAY AND IDENTIFICATION OF RARE CELL POPULATIONS

[0145] The data presented here demonstrate identification of putative rare cell populations. Using the methodology described herein, a putative rare cell population was identified. Sample processing and imaging was performed as disclosed in Example 1. Additionally, HD-CTCs were identified and defined as in Example 1.

[0146] In performing the assay, no CTCs were assumed to be cytokeratin positive. A putative rare cell population was identified having the following characteristics: a) cytokeratin dim or negative; b) CD45 negative; and c) intact non-apoptotic appearing nucleus by DAPI imaging. Figure 5 displays the incidence rate of the putative rare cell population across patients relative to identified HD-CTCs.

[0147] Although the invention has been described, it will be understood that modifications and variations are encompassed within the spirit and scope of the invention. Accordingly, the invention is limited only by the following claims.



What is claimed is:

1. A method for detecting a cell in a sample from a subject comprising:
  - a) providing a sample suspected of having at least one rare cell and at least one cell that is present at a concentration that is at least 10 times that of the rare cell;
  - b) contacting the sample with at least one detectable agent;
  - c) performing cell imaging on the sample of (b) to generate a cell image; and
  - d) detecting the at least one rare cell as compared with other cells in the sample by analyzing the cell from the image of (c), thereby detecting the rare cell in the sample.
2. The method of claim 1, further comprising (a'), wherein the at least one rare cell and at least one non-rare cell are affixed to a slide.
3. The method of claim 1, wherein (d) is performed on a computer comprising executable code for performing the analysis.
4. The method of claim 3, further comprising (e), wherein the results of (d) are presented on a graphical user interface.
5. The method of claim 1, wherein the cell imaging of (c) further comprises optimizing an exposure limit for the at least one detectable agent.
6. The method of claim 5, wherein the exposure limit for the at least one detectable agent is determined using a signal from the at least one non-rare.
7. The method of claim 5, wherein the at least one detectable agent binds a cell marker and is selected from a positive marker, negative marker, nuclear marker, content marker, and any combination thereof.
8. The method of claim 5, wherein the exposure limit is optimized before generating the image.
9. The method of claim 5, wherein the exposure limit is optimized after generating the image.

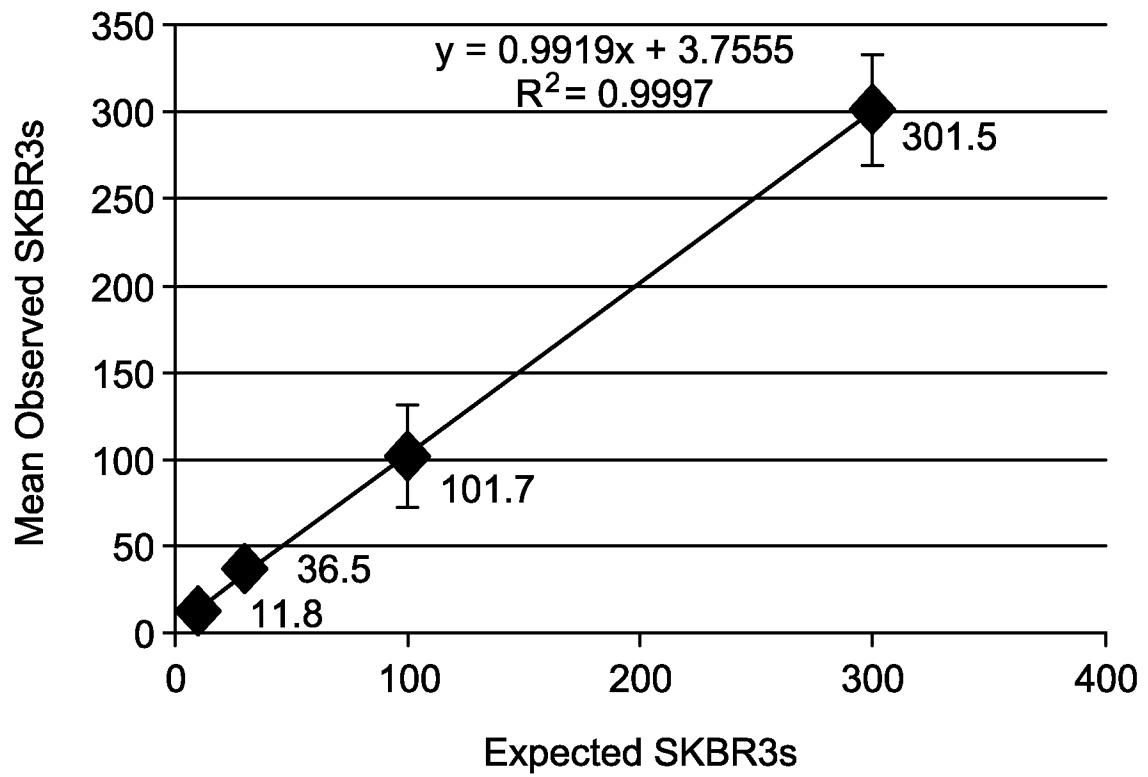
10. The method of claim 5, further comprising (c'), wherein the sample is re-imaged prior to performing (d).
11. The method of claim 1, wherein the cell imaging of (c) further comprises optimizing the exposure time to decrease data collection time.
12. The method of claim 1, wherein a signal from the at least one cell is used to determine a quality control parameter.
13. The method of claim 12, wherein the quality control parameter is distribution of the at least one cell on the slide.
14. The method of claim 12, wherein the quality control parameter is alignment of multiple cell images.
15. The method of claim 14, wherein the at least one detectable agent is a negative marker and a nuclear marker and the alignment is determined by aligning a signal from the nuclear marker with a signal from the negative marker, each present on different images.
16. The method of claim 12, wherein the quality control parameter is quality of cell staining.
17. The method of claim 16, wherein the at least one detectable agent is a negative marker and a nuclear marker and the quality of cell staining is determined by calculating a ratio of signals from the negative markers to signals from the nuclear markers.
18. The method of claim 17, wherein the ratio is less than 1.2.
19. The method of claim 12, wherein the at least one detectable agent is a negative marker and the quality control parameter is distribution of the negative marker throughout the non-rare cells.
20. The method of claim 19, wherein the distribution is calculated using the mean, standard deviation, coefficient of variation, or any combination thereof.
21. The method of claim 12, wherein the at least one detectable agent is a positive marker and the quality control parameter is distribution of the positive marker throughout the non-rare cells.

22. The method of claim 21, wherein the distribution is calculated using the mean, standard deviation, coefficient of variation, or any combination thereof.
23. The method of claim 12, wherein the quality control parameter is cell loss.
24. The method of claim 23, wherein the at least one detectable agent is a nuclear marker or a negative marker and cell loss is determined by calculating a ratio of signals from the nuclear markers or signals from the negative markers to the number of non-rare cells on the slide.
25. The method of claim 1, wherein the cell imaging of (c) further comprises optimizing intensity limits to distinguish rare cells from other cells.
26. The method of claim 25, wherein the at least one detectable agent is a positive marker and the intensity limits are determined using mean, standard deviation, coefficient of variation, or any combination thereof, for a background signal of the positive marker.
27. The method of claim 26, wherein the mean, standard deviation, or coefficient of variation are determined on a global slide basis, single slide basis, region of a single slide, or any combination thereof.
28. The method of claim 25, wherein the at least one detectable agent is a positive marker and the intensity limits are determined within a single image by identifying a highest signal of the positive marker and comparing the highest signal to a standard deviation calculated from signals of additional positive marker.
29. The method of claim 26, wherein the intensity limit is equal to a factor of between 5 and 20 multiplied by the standard deviation and added to the mean.
30. The method of claim 29, wherein the intensity limit is equal to a factor of 12.5 multiplied by the standard deviation and added to the mean.
31. The method of claim 30, wherein the intensity limit is equal to a factor of 15 multiplied by the standard deviation and added to the mean.
32. The method of claim 29, wherein the intensity limit is calculated for each sample image.

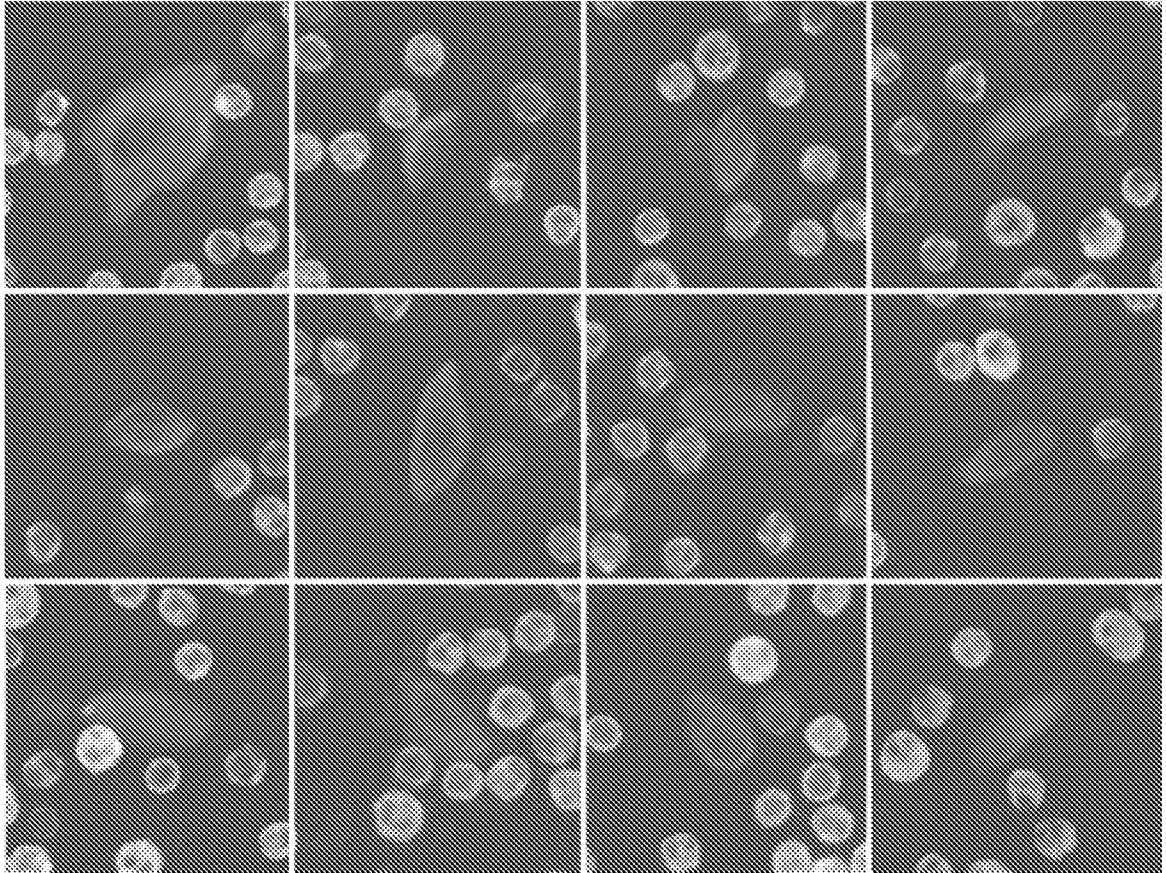
33. The method of claim 25, wherein the at least one detectable agent is a negative marker and the intensity limit for the negative marker is determined using mean and standard deviation of signals from the negative markers from non-rare cells.
34. The method of claim 33, wherein the mean and standard deviation are determined on a global slide basis, single slide basis, region of a single slide, or any combination thereof.
35. The method of claim 1, wherein (d) further comprises measuring cellular sizes for the at least one cell, distribution of cellular sizes for the at least one cell, or combination thereof, and comparing the cellular sizes or distribution of cellular sizes for the at least one cell, to known cellular sizes and distributions.
36. The method of claim 1, wherein the analysis of (d) further comprises measuring nuclear sizes for the at least one cell, distribution of nuclear sizes for the at least one cell, contour patterns for the at least one cell, or combination thereof, and comparing the nuclear sizes, distribution of nuclear sizes or contour patterns to a putative rare cell to identify the suspected rare cell as a rare cell.
37. The method of claim 1, wherein the analysis of (d) further comprises determining the concentration of rare cells in a fluid of the subject from which the sample was taken.
38. The method of claim 1, wherein the detectable agent is a content marker and the analysis of (d) further comprises determining an expression level of the content marker.
39. The method of claim 36, wherein the expression level of the content marker is determined in the at least one rare cell and in the at least one cell and compared to the expression level of the content marker in a patient population.
40. The method of claim 1, wherein the rare cell is a circulating tumor cell (CTC).
41. The method of claim 40, wherein the CTC expresses a positive marker, has an intact nucleus, and is morphologically distinct from a normal WBCs, wherein the CTC is not positive for a negative marker.
42. The method of claim 41, wherein the positive marker is cytokeratin or EpCAM.
43. The method of claim 41, wherein the negative marker is CD45.

44. The method of claim 40, further comprising providing a diagnosis or prognosis to the subject.
45. The method of claim 40, wherein the subject is known to have cancer and is undergoing cancer therapy.
46. The method of claim 45, wherein the therapy is chemotherapy.
47. The method of claim 46, wherein a content marker is used to determine a chemotherapeutic agent.
48. The method of claim 44, wherein the subject is being administered a candidate agent.
49. The method of claim 44, further comprising determining the responsiveness of the subject to the cancer therapy.
50. A method for diagnosing or prognosing cancer in a subject comprising:
  - a) performing the method of claim 40; and
  - b) analyzing the detected CTC to provide a diagnosis or prognosis, thereby diagnosing or prognosing cancer in a subject.
51. A method for determining responsiveness of a subject to a therapeutic regime comprising:
  - a) performing the method of claim 40; and
  - b) analyzing the detected CTC, thereby determining the responsiveness of the subject to a therapeutic regime.
52. The method of claim 51, wherein the therapeutic regime is changed based on the results of b).
53. A method for determining a candidate subject for a clinical trial comprising:
  - a) performing the method of claim 40; and

b) analyzing the detected CTC, thereby determining a candidate subject for a clinical trial.

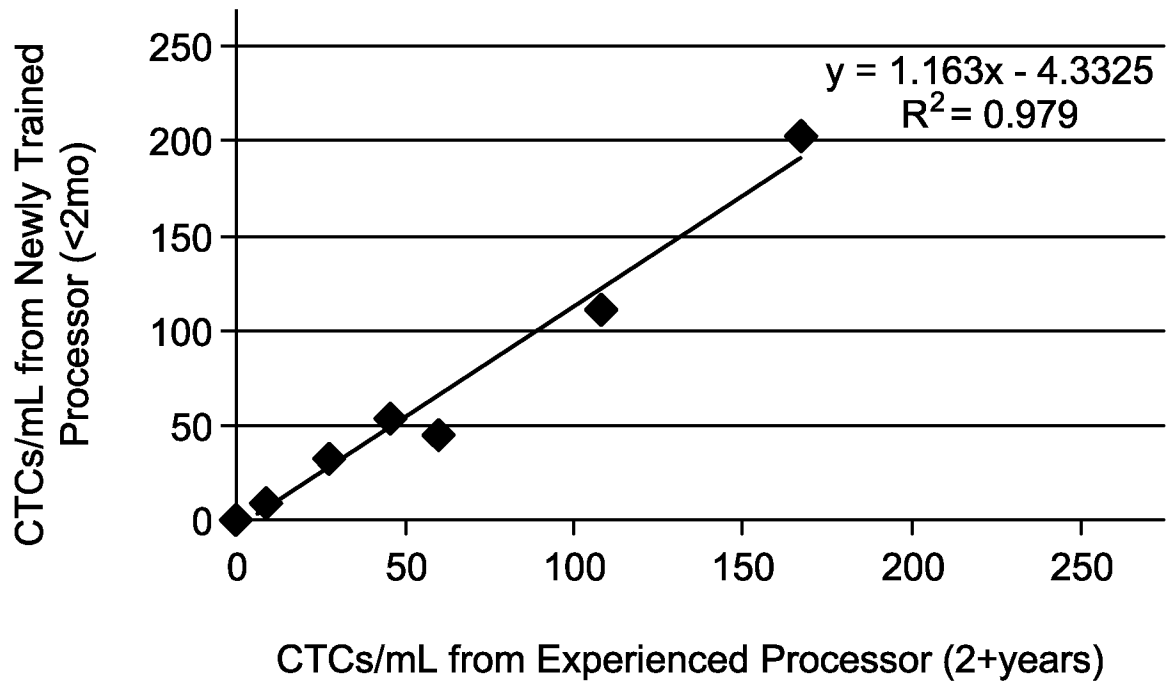


**FIG. 1**

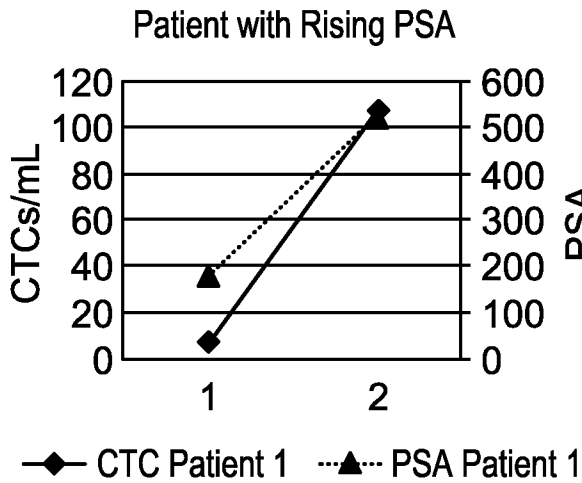


**FIG. 2**

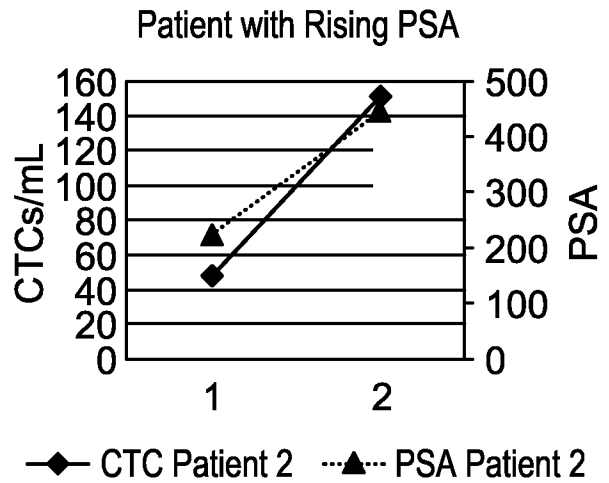




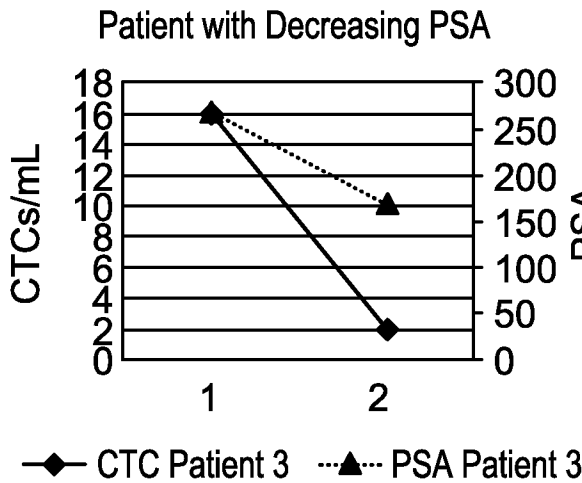
**FIG. 3**



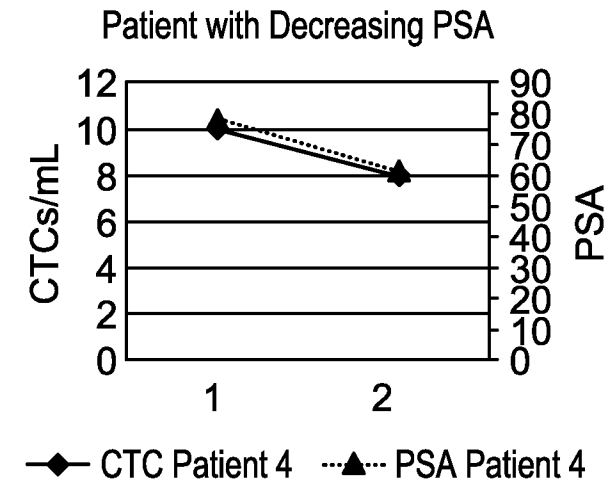
**FIG. 4A**



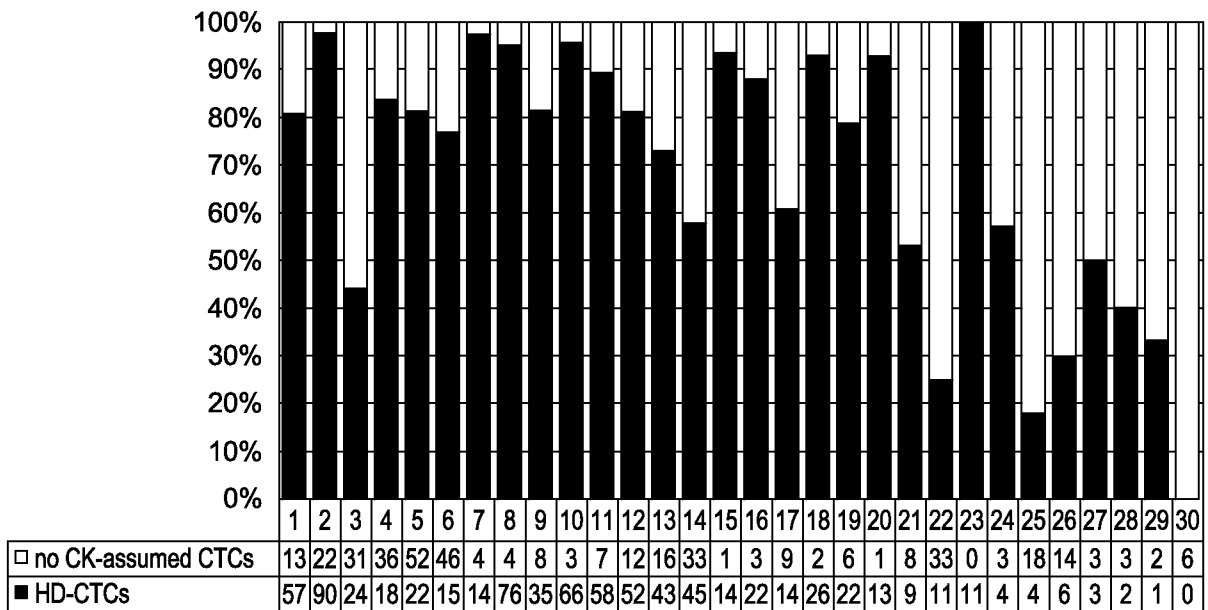
**FIG. 4B**



**FIG. 4C**



**FIG. 4D**



**FIG. 5**

**INTERNATIONAL SEARCH REPORT**

International application No.  
PCT/US 10/53431

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(8) - G01N 33/574 (2010.01)

USPC - 702/21

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

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

USPC: 702/21

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

USPC: 702/1.1, 21, 22, 29; 700/1, 90; 435/6, 41, 91.2; 600/300 (text search - see terms below)

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

PubWEST(USPT,PGPB,EPAB,JPAB); Google

Search Terms: circulating, tumor, ctc, cell, image, negative, nuclear, content, exposure, time, optimize, re-image, quality, control, nucleus, intact, unharmed, rare cell, non rare, chemotherapy

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X — Y	US 2004/0029213 A1 (CALLAHAN et al.) 12 February 2004 (12.02.2004), entire document especially paras [0011]-[0014], [0022], [0025], [0027], [0088],[0108], [0115], [0212], [0233], [0288]	1-6, 8-10, 25, 35, 37-38 ----- 7, 11-24, 26-34, 36, 39-53
Y	US 2009/0105963 A1 (LAURSEN et al.) 23 April 2009 (23.04.2009), entire document especially para [0080]	7, 19-22, 26-27, 29-34
Y	HSIEH et al., High speed detection of circulating tumor cells, Biosensors and Bioelectronics 21 (2006) 1893-1899 [online], 07 February 2006 (07.02.2006), [retrieved on 07 December 2010 (07.12.2010)]. Retrieved from the Internet:<URL: http://kuhn.scripps.edu/PDF/cbrp_highSpeed_biosensors2006.pdf>, pg 1189	11, 53
Y	US 2009/0072171 A1 (SO et al.) 19 March 2009 (19.03.2009), entire document especially paras [0079]-[0080]	12-24
Y	US 2006/0024756 A1 (TIBBE et al.) 02 February 2006 (02.02.2006), entire document especially para [0118]	28
Y	US 2005/0003464 A1 (TIBBE et al.) 06 January 2005 (01.06.2005), entire document especially para [0003]	36, 39
Y	US 2009/0081688 A1 (LUO et al.) 26 March 2009 (26.03.2009), entire document especially paras [0335]-[0336]	40-53

Further documents are listed in the continuation of Box C.

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

Date of the actual completion of the international search 08 December 2010 (08.12.2010)	Date of mailing of the international search report <b>21 DEC 2010</b>
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young  PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 10/53431

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	SCHREUDER, Laser Image Cytometer for Analysis of Circulating Tumor Cells, Thesis [online], 01 February 2008 (01.02.2008) [retrieved on 07 December 2010 (07.12.2010)]. Retrieved from the Internet:<URL: <a href="http://doc.utwente.nl/58400/1/thesis_Schreuder.pdf">http://doc.utwente.nl/58400/1/thesis_Schreuder.pdf</a> > pg 125	15
Y	US 2008/0113358 A1 (KAPUR et al.) 15 May 2008 (15.05.2008), entire document especially para [0126]	17-18, 24
Y	US 2008/0009019 A1 (HAIZLIP et al.) 10 January 2008 (10.01.2008), entire document especially the Abstract; Claim 13	47-48
Y	US 2007/0212736 A1 (CHEN-KIANG et al.) 13 September 2007 (13.09.2007), entire document especially para [0043])	52
	--	
X	KRIVACIC et al., A rare-cell detector for cancer, 07 June 2004 (07.06.2004) [retrieved on 07 December 2010 (07.12.2010)]. Retrieved from the Internet:<URL: <a href="http://www.pnas.org/content/101/29/10501.full.pdf+html">http://www.pnas.org/content/101/29/10501.full.pdf+html</a> > Abstract, pgs 1-2	1
P, Y	US 2009/0317836 A1 (KUHN et al.) 24 December 2009 (24.12.2009), entire document especially the Abstract; paras [0006], [0009], [0011], [0016], [0109]-[0110]	40-53
A	US 2007/0212698 A1 (BENDELE et al.) 13 September 2007 (13.09.2007), entire document	1-53