



(51) International Patent Classification:

G06F 17/00 (2006.01) *G06F 19/10* (2011.01)
G06F 19/00 (2011.01) *G06F 19/24* (2011.01)

(21) International Application Number:

PCT/US2016/024446

(22) International Filing Date:

28 March 2016 (28.03.2016)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/138,695 26 March 2015 (26.03.2015) US

(71) Applicants: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US). **CHILDREN'S MEDICAL CENTER CORPORATION** [US/US]; Children's Hospital Boston, LO 350, 300 Longwood Avenue, Boston, MA 02115 (US).

(72) Inventors: **HENNEK, Jonathan, W.**; 40 Elmwood Street, Somerville, MA 02144 (US). **KUMAR, Ashok, A.**; 27 Medford Street, Medford, MA 02153 (US). **WHITESIDES, George, M.**; 124 Grasmere Street, Newton, MA 02458 (US). **ADAMS, Ryan, P.**; 9B Pitman Street, Somerville, MA 02143 (US). **WILTSCSKO, Alexander, B.**; 22 Linden Place, Brookline, MA 02445 (US). **BRUGNARA, Carlo**; 300 Longwood Avenue, Boston, MA 02115 (US).

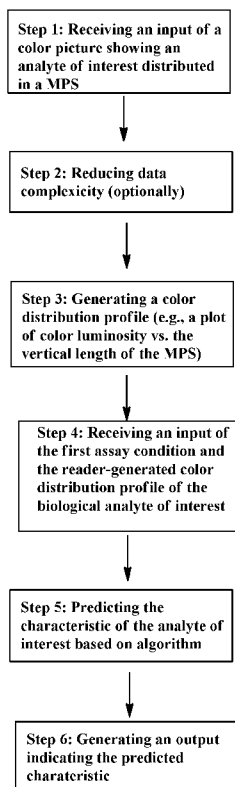
(74) Agents: **SCOZZAFAVA, Mary Rose** et al.; Wilmer Cutler Pickering Hale and Dorr LLP, 60 State Street, Boston, MA 02109 (US).

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

[Continued on next page]

(54) Title: METHODS FOR BIOLOGICAL ANALYTES SEPARATION AND IDENTIFICATION

FIG. 1



(57) Abstract: A system is described, including: a reader for generating a color distribution profile of a biological analyte of interest in a phase-separated multi-phase system under a first assay condition; a memory for storing one or more algorithms and one or more assay conditions, wherein each algorithm is associated with an assay condition and configured to predict a characteristic of the biological analyte of interest based on its color distribution profile under the assay condition; and wherein at least one of the assay conditions is the first assay condition; a computer processor coupled to the reader and the memory, the computer processor is configured to: receive an input of the first assay condition and the reader-generated color distribution profile of the biological analyte of interest; predict a characteristic of the biological analyte of interest using the algorithm associated with the first condition; and provide an output identifying the same.



SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every
kind of regional protection available*): ARIPO (BW, GH,
GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ,
TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU,
TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,

DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT,
LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE,
SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA,
GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

Methods for Biological Analytes Separation and Identification

Incorporation by Reference

[0001] All patents, patent applications and publications cited herein are hereby incorporated by reference in their entirety in order to more fully describe the state of the art as known to those skilled therein as of the date of the invention described herein.

Related Application

[0002] The present application claims the priority and benefit of U.S. Provisional Application No. 62/138,695, filed on March 26, 2015, the content of which is incorporated by reference herein in its entirety. The present application is generally related to PCT/US14/35697, filed on April 28, 2014, the content of which is incorporated by reference herein in its entirety.

Background

[0003] Aqueous mixtures of two polymers such as poly(ethylene glycol) (PEG) and dextran can separate spontaneously into two aqueous phases, called aqueous two-phase systems. Phase separation in aqueous solutions of polymers is an extraordinary and underexplored phenomenon. When two aqueous solutions of polymers are mixed, the resulting system is often not homogeneous; rather, two discrete phases, or layers, form. These layers are ordered according to density and arise from the limited interaction of the polymers for one another. In these systems, each phase predominantly consists of water (upwards of 70 – 90% (w/v)), while the polymer component is present in concentrations ranging from micromolar to millimolar. A low interfacial tension and rapid mass transfer of water-soluble molecules across the boundary characterize the interface between layers.

[0004] Iron deficiency anemia (IDA) is anemia due to insufficient amount of iron. The WHO estimates that IDA is responsible for ~270,000 deaths and 19.7 million disability-adjusted life years lost annually. Over one billion people are estimated to suffer from iron deficiency anemia (IDA). As a result of depleted iron stores in the body, adults may experience chronic fatigue, among other symptoms. IDA during pregnancy has been shown to increase the risk of preterm birth and low birth weight; infants with untreated IDA can have permanent cognitive impairments and delayed physical development.

[0005] Iron supplements provide a simple intervention to treat IDA, but the use of iron supplements when IDA is not present can result in iron overload. The proper diagnosis of IDA is important to connect patients to effective care. Simple interventions, such as oral iron supplements, exist for treating IDA. Supplements, however, should be used only when a diagnosis is available in order to avoid possible side effects. These side effects include iron overload, impaired growth in children, and increased risk of severe illness and death in malaria endemic areas. In developed countries, IDA is easily diagnosed in a central laboratory by a complete blood count and measurement of serum ferritin concentration. In LMICs, however, a lack of instrumentation, trained personnel, and consistent electricity prohibits effective diagnosis. A rapid, low-cost, and simple to use platform to diagnose IDA is needed. While current clinical capabilities can effectively diagnose IDA in the developed world, many countries lack the expensive instrumentation necessary to detect IDA, especially at the point-of-care.

[0006] Red blood indices—measurements of the properties and numbers of red blood cells—are commonly used for the diagnosis of IDA, because they (in contrast to serum iron or ferritin) respond quickly to changes in the iron level in the body, and require a less painful and less invasive procedure for the patient than the gold standard measurement (iron in bone marrow). Red blood cell indices measured by a complete blood count require a hematology analyzer (a flow cytometer, typically with impedance, photometry, and chemical staining capabilities). A hematology analyzer, however, is expensive (\$20,000-\$50,000) and requires highly trained personnel and significant technical maintenance. An inexpensive, rapid, and simple method that approaches the specificity and sensitivity provided by a hematology analyzer could find widespread clinical use.

[0007] An inexpensive and point-of-care tool for the diagnosis of IDA is especially needed in resource-limited countries where the rate of IDA is often very high—affecting nearly 20 % of the population—and where hematology analyzers are only available in major hospitals.

[0008] "Anemia" is defined as a condition in which the patient has a low hemoglobin concentration (HGB) in the blood. Various methods have been developed to diagnose anemia in low-resource settings, either by measuring the number of red blood cells (RBCs) per unit volume through spun hematocrit (HCT), or by measuring HGB directly. Anemia, both chronic and acute, can, however, have many causes, and a diagnosis limited to "anemia"

with no further detailed cellular and/or molecular description does not necessarily provide enough information for the effective treatment of a patient.

[0009] Anemia associated with microcytic (*i.e.*, smaller cells than normal) and hypochromic (*i.e.*, lower concentration of hemoglobin per cell than normal) cells, on the other hand, is mostly a result of IDA or thalassemia trait (α or β -thalassemias). IDA affects > 10 times more people globally than does β -thalassemia trait. Due to the dominance of IDA among other conditions causing microcytic, hypochromic (micro/hypo) red blood cells, several studies have shown good diagnostic accuracy for IDA by measuring the number of hypochromic red blood cells. Micro/hypo anemias are also associated with a reduction in the mass density of red blood cells.

[0010] A tool to distinguish micro/hypo anemia, and thus IDA, quickly from normal blood and other forms of anemia would improve the effectiveness of healthcare, and promote a better use of resources at the level of primary healthcare, in resource-limited countries.

Summary

[0011] Described herein are computer/machine-aided systems and methods for determining and/or predicting the characteristic of a biological analyte of interest. In certain embodiments, the biological analyte of interest having a recognizable color is separated or distributed in a multi-phase system described herein. Based on its properties (*e.g.*, density, size, shape, and/or mass), the biological analyte spreads across the vertical length of the multi-phase system and a color distribution profile of the biological analyte along the vertical length of the multi-phase system is generated. An Algorithm using a computer/machine is then used to predict one or more characteristics of the biological analyte of interest based on the color distribution profile of the biological analyte of interest.

[0012] Machine learning attempts to build learning algorithms that learn the associations between properties of data that are taken to always be available (inputs) and properties of data that are taken to not always be available (outputs). As used herein, “regression” refers to the issue of predicting continuously-varying outcomes from data by using a computer. As used herein, prediction is the act of producing, guessing, imputing, or computing an ordinarily unavailable property of data by the machine, given available properties of data. As used herein, “classification” refers to discrete outcomes that don’t necessarily have a prescribed ordering.

[0013] As used herein, MPS refers to a multi-phase system. In certain embodiments, each of the phases contains a solvent and a phase component which is selected from the group consisting of polymers and surfactants. When two or more solutions containing a phase component are mixed, the resulting system is not homogeneous; rather, two or more discrete phases, or layers, form. These layers are ordered according to density and arise from the exhibit limited interaction of the phase components with one another. The two or more phases or solutions exhibit limited interaction and form distinct, stable phase boundaries between adjacent phases. Each phase can be aqueous or non-aqueous. The non-aqueous phase comprises an organic liquid or an organic solvent.

[0014] In some embodiments, MPS as described herein are used to separate/distribute analytes when the analytes migrate to phases characteristic of their properties, *e.g.*, densities, shape, size, mass, or a combination thereof. In certain embodiments, the analyte contacts each phase of the multi-phase system sequentially. As used herein, “sequential contact” means that the analyte contacts and interacts with only one phase (and its phase component) at a time except at the interface where the analyte contacts and interacts with two adjacent phases simultaneously. That is, the interaction of the analyte with the MPS occurs when the MPS has already phase separated and not during the process of phase separation. In other embodiments, biological analytes of interest are deposited into a formed MPS and the sedimentation profile of the analyte can be studied. The sedimentation rate of the biological analyte can be affected by its density, size, shape, and mass. In still other embodiments, biological analytes of interest are mixed with the components of the MPS and the formation of the MPS and the separation of the analyte are accomplished in one step.

[0015] The phase “combination” refers to the combination of a polymer and a surfactant, a combination of two or more polymers, a combination of two or more surfactants, or a combination of any number of polymers and any number of surfactants.

[0016] As used herein, the use of the phrase “polymer” includes, but is not limited to, the homopolymer, copolymer, terpolymer, random copolymer, and block copolymer. Block copolymers include, but are not limited to, block, graft, dendrimer, and star polymers. As used herein, copolymer refers to a polymer derived from two monomeric species; similarly, a terpolymer refers to a polymer derived from three monomeric species. The polymer also includes various morphologies, including, but not limited to, linear polymer, branched polymer, random polymer, crosslinked polymer, and dendrimer systems. As an example,

polyacrylamide polymer refers to any polymer including polyacrylamide, *e.g.*, a homopolymer, copolymer, terpolymer, random copolymer, block copolymer or terpolymer of polyacrylamide. Polyacrylamide can be a linear polymer, branched polymer, random polymer, crosslinked polymer, or a dendrimer of polyacrylamide.

[0017] As used herein, MPS refers to any one of the multi-phase systems described herein. AMPS refers to any one of the aqueous multi-phase systems described herein (*i.e.*, the solvent used in the MPS is water). As used herein, ATPS refers to an aqueous two-phase polymer system.

[0018] As used herein, the MPSs described herein may be used for analysis of mammalian blood and/or separation of biological analytes from the mammalian blood. In some embodiments, the mammal is human.

[0019] In some embodiments, the phase component is a polymer or a combination of two or more polymers. In some embodiments, the aqueous multi-phase polymer system can be combined with one or more immiscible organic phases to form a multi-phase system.

[0020] As used herein, the phrase “mixture” refers to the combination of two components, which may be mixed or layered one on top of another.

[0021] As used herein, when specific values are disclosed, the ranges bounded by any of the specific values are also contemplated.

[0022] As used herein, the phrase “at the interface” of the adjacent phases of the MPS includes the situation where the biological analytes of interest is between the two adjacent phases or close to the border of one of the two adjacent phases.

[0023] In one aspect, a system for determining a characteristic of a biological analyte of interest is described, including:

a reader for generating a color distribution profile of a biological analyte of interest in a phase-separated multi-phase system under a first assay condition;

a memory for storing one or more algorithms and one or more assay conditions, wherein each algorithm is associated with an assay condition and configured to predict a characteristic of the biological analyte of interest based on its color

distribution profile in a phase-separated multi-phase system under the assay condition;
and wherein at least one of the assay conditions is the first assay condition;

a computer processor coupled to the reader and the memory, the computer processor is configured to:

receive an input of the first assay condition and the reader-generated color distribution profile of the biological analyte of interest;

based on the reader-generated color distribution profile of the biological analyte, predict a characteristic of the biological analyte of interest using the algorithm associated with the first condition; and

provide an output identifying the predicted characteristic of the biological analyte of interest.

[0024] In any one of the embodiments described herein, the algorithm is built by machine-learning. In any one of the embodiments described herein, the machine-learning comprises a process comprising creating, training, validating and/or testing the algorithm using a plurality of biological analytes with known characteristics.

[0025] In any one of the embodiments described herein, at least one of the algorithms is configured to make continuously-varying prediction. In any one of the embodiments described herein, at least one of the algorithms is configured to make discrete prediction.

[0026] In any one of the embodiments described herein, at least one of the algorithms is configured to predict the characteristic of the biological analyte based on comparing and/or matching one or more color distribution profiles of known biological analytes associated with the first assay condition with the reader-generated color distribution profile of the biological analyte.

[0027] In any one of the preceding embodiments, the biological analyte has a recognizable color or is dyed with a recognizable color.

[0028] In any one of the preceding embodiments, the characteristic of the biological analyte is a disease state or a biological index of the biological analyte.

[0029] In any one of the preceding embodiments, the biological analyte is selected from the group consisting of multicellular organisms, cells, organelles, cell fragments, cell

membranes, cell membrane fragments, viruses, virus-like particles, bacteriophage, cytosolic proteins, secreted proteins, signaling molecules, embedded proteins, nucleic acid/protein complexes, organelles, minicells, nucleic acid precipitants, chromosomes, nuclei, mitochondria, chloroplasts, flagella, biominerals, protein complexes, protein aggregates, and combinations thereof.

[0030] In any one of the preceding embodiments, the biological analyte is red blood cell or a population of red blood cell.

[0031] In any one of the preceding embodiments, the characteristic of the red blood cell or the population of red blood cell is one or more indexes selected from the group consisting of the average size of a red blood cell (MCV), the average amount of hemoglobin per red blood cell (MCH), the average amount of hemoglobin per red blood cell (MCHC), the red blood cell distribution width (RDW), percentage of hypochromic red blood cells (%Hypo), hemoglobin concentration (HGB), corpuscular hemoglobin concentration (CH), per unit volume through spun hematocrit (HCT), hemoglobin distribution width (HDW), the number of red blood cells (RBCs), the percentage of red blood cells that are microcytic (%Micro), %Micro/%Hypo, the percentage of cells that are hyperchromic red blood cells (%Hyper), and the percentage of cells that are microcytic red blood cells (%Macro).

[0032] In any one of the preceding embodiments, the output is a print out or a file or image displayed on a smartphone, a PC, or a monitor.

[0033] In any one of the preceding embodiments, the computer processor is configured to predict the characteristic of the biological analyte based on comparing and/or matching one or more color distribution profiles of known biological analytes associated with the first assay condition with the reader-generated color distribution profile of the biological analyte.

[0034] In any one of the preceding embodiments, the computer processor is configured to identify one or more stored color distribution profiles of known biological analytes associated with the first assay condition which is similar to the reader-generated color distribution profile of the biological analyte.

[0035] In any one of the preceding embodiments, the system further comprises a separation unit comprising the multi-phase system.

[0036] In any one of the preceding embodiments, the multi-phase system comprises at least adjacent first and second phase-separated phases, wherein

the first phase comprises a first phase component predominantly dissolved in the solvent of the first phase; and the second phase comprises a second phase component predominantly dissolved in the solvent of the second phase; wherein

the solvents of the first and second phases are the same;

the first phase component is different from the second phase component;

each of the first and second components is selected from the group consisting of a polymer, a surfactant and combinations thereof; and

at least one of the first and second phase components comprises a polymer;

each of the first and second phases has a different density and the first and second phases, taken together, represent a density gradient; and

the first and second phases have a stable interface in-between.

[0037] In any one of the preceding embodiments, the first and second phase components are each selected from the group consisting of Carboxy-polyacrylamide, Dextran, Ficoll, N,N-dimethyldodecylamine N-oxide, poly(2-ethyl-2-oxazoline), poly(acrylic acid), poly(ethylene glycol), poly(methacrylic acid), poly(vinyl alcohol), polyacrylamide, polyethyleneimine, hydroxyethyl cellulose, poly(2-acrylamido-2-methyl-1-propanesulfonic acid), polyvinylpyrrolidone, Nonyl, polyallylamine, (hydroxypropyl)methyl cellulose, diethylaminoethyl-dextran, nonylphenol polyoxyethylene 20, copolymer, terpolymer, block copolymer, random polymer, linear polymer, branched polymer, crosslinked polymer, and dendrimer system thereof.

[0038] In any one of the preceding embodiments, the solvent is water.

[0039] In any one of the preceding embodiments, the assay condition is one or more conditions selected from the group consisting of the composition of the multi-phase system and the distribution condition of the biological analyte in the multi-phase system.

[0040] In any one of the preceding embodiments, the distribution condition of the biological analyte in the multi-phase system comprises the separation time of the biological analyte in the multi-phase system and/or the centrifuge force used for the separation of the biological analyte in the multi-phase system.

[0041] In any one of the preceding embodiments, the reader is a scanner, a camera, or smartphone camera.

[0042] In any one of the preceding embodiments, the memory is selected from the group consisting of a hard drive, a thumb drive, a magnetic disk, an optical disk, and magnetic tape.

[0043] In any one of the preceding embodiments, the color distribution profile comprises a distribution of the biological analyte's color luminosity along the vertical length of the multi-phase system.

[0044] In another aspect, a method for determining a characteristic of a biological analyte of interest is described, including:

generating a color distribution profile of a biological analyte of interest in a phase-separated multi-phase system under a first assay condition;

generating a database and storing the database in a memory, the database comprising one or more algorithms and one or more assay conditions, wherein each algorithm is associated with an assay condition and configured to predict a characteristic of the biological analyte of interest based on its color distribution profile in a phase-separated multi-phase system under the assay condition; and wherein at least one of the one or more assay conditions is the first assay condition; and

based on the reader-generated color distribution profile of the biological analyte, using a computer to predict a characteristic of the biological analyte of interest using the algorithms associated with the first condition.

[0045] In any one of the embodiments described herein, the algorithm is built by machine-learning. In any one of the embodiments described herein, the machine-learning comprises a process comprising creating, training, validating and/or testing the algorithm using a plurality of biological analytes with known characteristics.

[0046] In any one of the embodiments described herein, at least one of the algorithms is configured to make continuously-varying prediction. In any one of the embodiments described herein, at least one of the algorithms is configured to make discrete prediction.

[0047] In any one of the embodiments described herein, at least one of the algorithms is configured to predict the characteristic of the biological analyte based on comparing and/or matching one or more color distribution profiles of known biological analytes associated with the first assay condition with the reader-generated color distribution profile of the biological analyte.

[0048] In any one of the preceding embodiments, the biological analyte has a recognizable color or is dyed with a recognizable color.

[0049] In any one of the preceding embodiments, the characteristic of the biological analyte is a disease state or a biological index of the biological analyte.

[0050] In any one of the preceding embodiments, the biological analyte is selected from the group consisting of multicellular organisms, cells, organelles, cell fragments, cell membranes, cell membrane fragments, viruses, virus-like particles, bacteriophage, cytosolic proteins, secreted proteins, signaling molecules, embedded proteins, nucleic acid/protein complexes, organelles, minicells, nucleic acid precipitants, chromosomes, nuclei, mitochondria, chloroplasts, flagella, biominerals, protein complexes, protein aggregates, and combinations thereof.

[0051] In any one of the preceding embodiments, the biological analyte is a red blood cell or a population of red blood cell.

[0052] In any one of the preceding embodiments, the characteristic of the red blood cell or the population of red blood cell is one or more indexes selected from the group consisting of the average size of a red blood cell (MCV), the average amount of hemoglobin per red blood cell (MCH), the average amount of hemoglobin per red blood cell (MCHC), the red blood cell distribution width (RDW), percentage of hypochromic red blood cells (%Hypo), hemoglobin concentration (HGB), corpuscular hemoglobin concentration (CH), per unit volume through spun hematocrit (HCT), hemoglobin distribution width (HDW), the number of red blood cells (RBCs), the percentage of red blood cells that are microcytic

(%Micro), %Micro/%Hypo, the percentage of cells that are hyperchromic red blood cells (%Hyper), and the percentage of cells that are microcytic red blood cells (%Macro).

[0053] In any one of the preceding embodiments, one or more stored color distribution profiles of known biological analytes associated with the first assay condition is compared and/or matched with the generated color distribution profile of the biological analyte to predict the characteristic of the biological analyte.

[0054] In any one of the preceding embodiments, the method further comprises separating the biological analyte of interest in the multi-phase system.

[0055] In any one of the preceding embodiments, multi-phase system comprises at least adjacent first and second phase-separated phases, wherein

the first phase comprises a first phase component predominantly dissolved in the solvent of the first phase; and the second phase comprises a second phase component predominantly dissolved in the solvent of the second phase; wherein

the solvents of the first and second phases are the same;

the first phase component is different from the second phase component;

each of the first and second components is selected from the group consisting of a polymer, a surfactant and combinations thereof; and

at least one of the first and second phase components comprises a polymer;

each of the first and second phases has a different density and the first and second phases, taken together, represent a density gradient; and

the first and second phases have a stable interface in-between.

[0056] In any one of the preceding embodiments, the first and second phase components are each selected from the group consisting of Carboxy-polyacrylamide, Dextran, Ficoll, N,N-dimethyldodecylamine N-oxide, poly(2-ethyl-2-oxazoline), poly(acrylic acid), poly(ethylene glycol), poly(methacrylic acid), poly(vinyl alcohol), polyacrylamide, polyethyleneimine, hydroxyethyl cellulose, poly(2-acrylamido-2-methyl-1-propanesulfonic acid), polyvinylpyrrolidone, Nonyl, polyallylamine, (hydroxypropyl)methyl cellulose,

diethylaminoethyl-dextran, nonylphenol polyoxyethylene 20, copolymer, terpolymer, block copolymer, random polymer, linear polymer, branched polymer, crosslinked polymer, and dendrimer system thereof.

[0057] In any one of the preceding embodiments, the assay condition is one or more conditions selected from the group consisting of the composition of the multi-phase system and the distribution condition of the biological analyte in the multi-phase system.

[0058] In any one of the preceding embodiments, the distribution condition of the biological analyte in the multi-phase system comprises the separation time of the biological analyte in the multi-phase system and/or the centrifuge force used for the separation of the biological analyte in the multi-phase system.

[0059] In any one of the preceding embodiments, the color distribution profile comprises a distribution of the biological analyte's color luminosity along the vertical length of the multi-phase system.

[0060] As used herein, the "luminosity" of a color and the "intensity" color may be used interchangeably.

[0061] Any embodiment or aspect described herein may be properly combined with another embodiment or aspect described herein. The combination of any two or more embodiments or aspects described herein is expressly contemplated.

Brief Description of the Drawings

[0062] The invention is described with reference to the following figures, which are presented for the purpose of illustration only and are not intended to be limiting. In the Drawings:

[0063] **Fig. 1** illustrates a flow diagram of the method and/or system described herein for determining/predicting the characteristics of a biological analyte of interest, according to one or more embodiments.

[0064] **Fig. 2A** illustrates a design of IDA-AMPS rapid test loaded with blood before and after centrifugation for a representative IDA and Normal sample, according to one or more embodiments.

[0065] **Fig. 2B** illustrates a schematic of the analysis of the quantity and location of red blood cells in an AMPS test using a digital scanner and a custom computer program, according to one or more embodiments.

[0066] **Fig. 3A** shows an example of IDA-AMPS tests after 2 minutes of centrifugation for a representative normal sample, where an image of the tube and its corresponding image with pixels converted to S/V, 1-D red intensity trace, and the first derivative of the trace, according to one or more embodiments.

[0067] **Fig. 3B** shows an example of IDA-AMPS tests after 2 minutes of centrifugation for a representative IDA sample, where an image of the tube and its corresponding image with pixels converted to S/V, 1-D red intensity trace, and the first derivative of the trace, according to one or more embodiments.

[0068] **Fig. 4A** shows example of micro/hypo sample (laid on its side) after 2 minutes of centrifugation, according to one or more embodiments; and **Fig. 4B** shows red intensity versus distance plots averaged for 152 samples showing discrimination between normal (solid blue) and micro/hypo anemic (dashed red) samples at 2, 6, and 10 minutes centrifugation, according to one or more embodiments.

[0069] **Fig. 5A** illustrates receiver operating characteristic (ROC) curves for hypochromia having different threshold values for the percentage of hypochromic red blood cells (%Hypo) as determined by visual evaluation of the IDA-AMPS test, according to one or more embodiments. **Fig. 5B** illustrates receiver operating characteristic (ROC) curves for diagnosis of hypochromia (%Hypo > 3.9%), micro/hypo anemia, and IDA as determined by visual evaluation of the IDA-AMPS test, according to one or more embodiments.

[0070] **Fig. 6A** illustrates the area under the curve (AUC) values for classifying micro/hypo anemia at 2 – 10 minutes centrifugation time determined by machine learning (n = 152), according to one or more embodiments; and **Fig. 6B** illustrates receiver operating characteristic (ROC) curves for diagnosis of hypochromia (%Hypo > 3.9%, solid black line), micro/hypo anemia (dashed red line), and IDA (dotted blue line) as determined by machine learning, according to one or more embodiments.

[0071] **Fig. 7A** illustrates Machine learning prediction results for %Hypo (Predicted %Hypo) compared to a hematology analyzer (True %Hypo), according to one or

more embodiments. **Fig. 7B** shows a Bland-Altman plot showing good agreement between true and predicted %Hypo (n = 152), according to one or more embodiments.

[0072] **Fig. 8** illustrates a reader training guide used to assign redness score to IDA-AMPS tests, according to one or more embodiments.

[0073] **Fig. 9** is a flow chart illustrating the classification of the diagnosis of hypochromia, micro/hypo anemia, iron deficiency anemia, and β -thalassemia trait used in this study based on hematological indices measured by a hematology analyzer (Advia 2120, Siemens), according to one or more embodiments.

Detailed Description

[0074] Multi-phase systems have been described previously and used for a variety of applications, *e.g.*, separation of biological analytes. *See, e.g.*, PCT/US14/35697, filed on April 28, 2014, WO2012/024688, filed on August 22, 2011, WO2012/024693, filed on August 22, 2011, WO2012/024690, filed on August 22, 2011, and WO2012/024691, filed on August 22, 2011, all of which are hereby incorporated by reference herein in their entirety. However, existing multi-phase separation systems often require human users' visual observation to detect whether a biological analyte is present or not at certain locations of the multiphase system, *e.g.*, at the boundary of two adjacent phases. Such a determination is often subjective and is prone to human error. Furthermore, the dynamic/thermodynamic distribution profile of a biological analyte in a multi-phase system is often affected by more than one characteristics of the analyte (*e.g.*, density, mass, shape, size, volume, chemical compositions), thus such distribution profile can provide information-rich data which cannot be interpreted by human observation.

System for determining a characteristic of a biological analyte

[0075] In one aspect, computer/machine-aided systems for determining/predicting the characteristic of a biological analyte of interest through algorithm, *e.g.*, machine-assisted data regression/classification are described. In certain embodiments, the biological analyte of interest has a recognizable color and is separated or distributed in a multi-phase system described herein. Based on its properties (*e.g.*, density, size, shape, and/or mass), the biological analyte spread across the vertical length (height) of the multi-phase system and a color distribution profile of the biological analyte along the vertical length of the multi-phase system is generated. Computer-aided algorithm, *e.g.*, data regression or classification, is then

used to predict a characteristic of the biological analyte of interest. In certain embodiments, the characteristic of the biological analyte is a disease state or a biological index of the biological analyte.

[0076] In one aspect, a system for determining a characteristic of a biological analyte of interest is described, comprising:

a reader for generating a color distribution profile of a biological analyte of interest in a phase-separated multi-phase system under a first assay condition;

a memory for storing one or more algorithms and one or more assay conditions, wherein each algorithm is associated with an assay condition and configured to predict a characteristic of the biological analyte of interest based on its color distribution profile in a phase-separated multi-phase system under the assay condition; and wherein at least one of the assay conditions is the first assay condition;

a computer processor coupled to the reader and the memory, the computer processor is configured to:

receive an input of the first assay condition and the reader-generated color distribution profile of the biological analyte of interest;

based on the reader-generated color distribution profile of the biological analyte, predict a characteristic of the biological analyte of interest using the algorithm associated with the first condition; and

provide an output identifying the predicted characteristic of the biological analyte of interest.

[0077] In some embodiments, the biological analyte of interest has a color (*e.g.*, red blood cells) recognizable to a machine (*e.g.*, a scanner). In other embodiments, the biological analyte of interest is dyed with a recognizable color. In some embodiments, stains or other reactants can be included in the MPS prior to mixing with the analyte. In some embodiments, the analyte can be pre-stained. In some embodiments, the dyed/stained biological analyte maintains the same characteristics of the undyed biological analyte. In some specific embodiments, the white blood cells (leukocytes) are stained/dyed with Acridine orange to differentiate subpopulations. In some specific embodiments, *E.coli* is stained with Fluorescein isothiocyanate (FITC) labeled antibodies. In some specific embodiments,

mitochondria in living cells is stained using rhodamine 123. In some specific embodiments, circulating tumor cells are stained by Hematoxylin and Eosin (H&E). In some specific embodiments, platelets can be stained to have a visible color.

[0078] In some embodiments, the instant invention is further described with reference to **Fig. 1**.

[0079] As used herein, multi-phase systems include two or more phase-separated phases each containing a phase component and these phases are arranged by density and form a density gradient. In some embodiments, the biological analyte of interest is separated by a MPS according to its density or according to its settlement rate in the MPS. As described herein, the settlement rate of the analyte can be affected by many factors such as its mass, volume, chemical composition, and shape. In some embodiments, under a set of assay conditions (explained in further detail below), a biological analyte of interest is separated in the multiphase system and a reader is used to record a color distribution profile of the biological analyte spreading through the vertical length of the multi-phase system. As shown in Step 1 of **Fig. 1**, such a color picture showing an analyte of interest distributed in a MPS may be inputted into the system.

[0080] Non-limiting examples of the reader include a scanner, a smart phone having a scanner or camera, and a camera. In some embodiments, more than one reader can be used.

[0081] In some specific embodiments, the color distribution profile is optionally further simplified or processed to reduce the data complexity (Step 2 of **Fig. 1**). In certain specific embodiments, a color picture of the biological analyte separated in the multiphase system is taken by a scanner or a camera. The background of the picture is then removed. In certain embodiments, the color intensity values for each row of pixels in the picture are summed and a one-dimensional plot of "color luminosity" versus distance of the multiphase system is compiled. These process steps can further simplify the computational load by reducing the data's dimensions, and/or reducing the redundancy in the dataset by grouping together adjacent pixels that are co-varying. A color distribution profile representing a plot of color luminosity vs. the vertical length of the MPS is then generated (Step 3 of **Fig. 1**).

[0082] In some specific embodiments, the computed/machine-aided system comprises a memory storing a database. In some embodiment, the memory (also referred to as "memory device") may be removable. Some exemplary memory devices include hard drive, thumb

drive, magnetic disk, an optical disk, and magnetic tape. The memory device can further store reference data that can be used as a baseline for comparison. The memory device may also be used for storing software, computer algorithms, and temporary files created by the computer processor during analysis of the data from the reader. The data can be stored in the form of images, time-evolution spectra, static or time-dependent photodetector signals, or color intensities along the length of the multiphase system.

[0083] In some specific embodiments, the database includes one or more algorithms and one or more assay conditions, where each algorithm is associated with an assay condition and configured to predict a characteristic of the biological analyte of interest based on its color distribution profile in a phase-separated multi-phase system under the assay condition; and where at least one of the assay conditions is the first assay condition. The characteristic of the biological analyte may be a disease state or a biological index of the biological analyte. Non-limiting examples of properties of the analyte include healthy or diseased state of cell, color, porosity, tendency to swell, size, shape, chemical composition, or distribution of cell. Thus, each of the data set corresponds to the color distribution profile of a known biological analyte with a known characteristic under an assay condition. In some embodiments, the assay condition, as described above, includes the composition of the multiphase system and the settlement conditions of the biological analyte. Thus, the machine/computer aided system as described herein use algorithms, built by using a plurality of color distribution profiles of known analytes, to predict the characteristic of an unknown biological analyte of interest. In certain embodiments, once the reader generates the color distribution profile of a biological analyte of interest in a phase-separated multi-phase system under a first assay condition, the computer processor of the system is capable of receiving an input of the first assay condition and the reader-generated color distribution profile of the biological analyte of interest (Step 4 of **Fig. 1**). The computer processor may be coupled to the reader and the memory by wire or wirelessly.

[0084] In some specific embodiments, the computer processor is configured to, based on the reader-generated color distribution profile of the biological analyte, predict a characteristic of the biological analyte of interest using the algorithm associated with the first condition (Step 5 of **Fig. 1**). In other embodiments, the computer processor is configured to identify one or more stored color distribution profiles of known biological analytes associated with the first assay condition which is similar to the reader-generated color distribution

profile of the biological analyte. In some embodiments, the computer processor is configured to reduce the computational load by reducing the data complexity. In some embodiments, any combination or manipulation of the color/color space could be used. In some embodiments, the reader is configured to detect absorbance or fluorescents. For example, the "color luminosity" may be generated by summing the RGB values. In other embodiments, the colorspace is converted to HSV (hue, saturation and value—a cylindrical-coordinate representations of points in an RGB color model) and S/V is used as the "color luminosity" value. In certain specific embodiments, a color picture of the biological analyte separated in the multiphase system is taken by a scanner or a camera and the color intensity values for each row of pixels in the picture are summed and a one-dimensional plot of "color luminosity" versus distance of the multiphase system is compiled. These process steps can further simplify the computational load by reducing the data's dimensions, and/or reducing the redundancy in the dataset by grouping together adjacent pixels that are co-varying. In some specific embodiments, a particular manipulation is chosen which matches human's visual intuition of "how red the tube should be" at any given point along its length. It is contemplated that any manipulation of the raw pixel values can be part of a machine learning algorithm.

[0085] In some specific embodiments, using the color distribution profile of the biological analyte of interest, the computer processor uses one or more algorithms to predict the characteristic of the biological analyte. "Machine learning" refers to the process of creating, training, validating and testing algorithms that learn from and adapt to data. In certain embodiments, machine learning is used to build algorithms that can predict properties of new analytes by being trained on a set of labeled analytes. These labels may either be continuously varying, in the case of predicting blood parameters such as percentage of hypochromic red blood cells (%Hypo), or may be discrete classes, in the case of predicting whether a patient has anemia. The case of making continuously-varying predictions is called 'regression', and the case of making discrete predictions is called 'classification'. In certain embodiments, in order to train an algorithm, a set of known, labeled analytes are first divided into "train", "validation" and "test" sets. The computer algorithm is configured to accurately predict the labels of the "train" set. This configuration is then tested on the "validation" set. In certain embodiments, the process of reconfiguring the algorithm, training, and validating is repeated until satisfactory performance on the validation set has been achieved. In certain embodiments, to guard against over-tuning the computer algorithm to the particulars of the

data in the training and validation sets, the performance of the computer algorithm is evaluated on the final data set, the “test” set, which contains analytes that the computer algorithm has not processed before. The computer algorithms that can be used for producing continuously-varying predictions include, but are not limited to linear regression, support vector regression, random forest regression, neural networks and Gaussian process regression. The computer algorithms which can be used for producing predictions of the class of analyte include, but are not limited to, logistic regression, support vector classification, neural networks, random forest classifier, Gaussian process classifier, boosting and naive Bayes. In certain specific embodiments, logistic regression is used in comparing the stored color distribution profile of known analyte with that of the biological analyte of interest. In other specific embodiments, support vector regression (SVR) with a radial basis function kernel is used in comparing the stored color distribution profile of known analyte with that of the biological analyte of interest (see, *e.g.*, Step 5 of **Fig. 1**).

[0086] In some other embodiments, the reader and/or the computer processor can further carry out additional data manipulation before and/or after receiving the data from reader. For example, the reader and/or the computer processor can carry out image cropping, edge detection, thresholding, area detection, and/or the like. In some other embodiments, the computer processor can determine tiebreakers for color and/or edge determinations. In some embodiments, the processing unit further includes software such as, image analysis, statistical analysis, comparison with stored calibration curves etc., can be used to analyze the results and rapidly print out a meaningful response to the user of the system.

[0087] In certain embodiments, the diseased biological analyte differs from the healthy analyte in shape, size, mass, density, or a combination thereof. Thus, the separation of the diseased analyte in multi-phase systems, under similar condition, results in a color distribution profile characteristic of its diseased state. In certain embodiments, the system described herein stores one or more algorithms built by machine-learning using one or more known color distribution profiles of analytes with known characteristics. By using computer algorithm built from analyzing known biological analytes of interest, the system will predict the characteristic of the unknown biological analyte (Step 5 of **Fig. 1**). In certain embodiments, using a computer algorithm that has been previously trained and validated on a set of labeled, known analytes, a property of a new analyte is predicted. In the process of training a computer algorithm, the algorithm has accumulated knowledge of the relationships

of all of the input values of the color profile distribution of the analyte. The computer algorithm exploits knowledge of these relationships in order to make predictions of the properties of the analyte. For example, the computer algorithm may learn that a more intense color in a particular physical location of the analyte may indicate with high probability that the analyte has a certain class label, or is drawn from a patient with a particular disease. The different types of computer algorithms outlined above all learn to relate varying configurations of input data (here, the color profile distribution of an analyte) with analyte properties, and are distinguished in the specific manner in which they build and store these relationships. All computer algorithms outlined above share the property that they are able to learn to predict labeled properties, given input data. The processor will then generate an output of the prediction of the characteristic of the biological analyte of interest (Step 6 of **Fig. 1**). The output may be in the form of a print out, a file or image displayed on a smartphone, a PC, or generally a monitor.

[0088] In certain embodiments, the biological analyte is deposited into a pre-formed multi-phase system, which, under certain assay conditions, generates a color distribution profile of the biological analyte through the vertical length of the multiphase system. In other embodiments, the biological analytes are mixed with the components (*e.g.*, solvent, phase components) for the multi-phase system to form the multi-phase system with the biological analyte distributed therein in one step.

[0089] In certain embodiments, the multi-phase system is unit independent from the computer/machine-aided systems disclosed herein. In other embodiments, the computer/machine-aided system further comprises a separation unit comprising the multi-phase system.

[0090] Types of biological analytes that can be analyzed include, without limitation, cells, cancer cells, stem cells, cell extracts, tissue extracts, cell organelles, cell fragments, cell membranes, cell membrane fragments, viruses, virus-like particles, bacteriophage, cytosolic proteins, secreted proteins, signaling molecules, embedded proteins, nucleic acid/protein complexes, nucleic acid precipitants, chromosomes, nuclei, mitochondria, chloroplasts, flagella, biominerals, protein complexes, phage, minicells, and protein aggregates, tissues, organisms, small molecules, large-sized molecules, *e.g.*, biomolecules including proteins, and particles. In one or more aspects, the types of cells used in the disclosed methods include mammalian cells selected from the group consisting of gland cells (*e.g.*, exocrine secretory

epithelial cells, salivary gland mucous cells, salivary gland serous cells, Von Ebner's gland cells, mammary gland cells, lacrimal gland cells, ceruminous gland cells, eccrine sweat gland dark cells, eccrine sweat gland clear cells, apocrine sweat gland cells, gland of Moll cells, sebaceous gland cells, Bowman's gland cells, Brunner's gland cells, seminal vesicle cells, prostate gland cells, bulbourethral gland cells, Bartholin's gland cells, gland of Littre cells, uterine endometrial cells, isolated goblet cells, stomach lining mucous cells, gastric gland zymogenic cells, gastric gland oxyntic cells, pancreatic acinar cells, Paneth cells, type II pneumocyte cells, and Clara cells), hormone secreting cells (*e.g.*, anterior pituitary cells, intermediate pituitary cells, magnocellular neurosecretory cells, gut and respiratory tract cells, thyroid gland cells, parathyroid gland cells, adrenal gland cells, chromaffin cells, Leydig theca interna cells, corpus luteum cells, granulosa lutein cells, theca lutein cells, juxtaglomerular cells, racula densa cells, peripolar cells, and mesangial cells), epithelial cells lining closed internal body cavities (*e.g.*, blood vessel and lymphatic vascular endothelial fenestrated cells, blood vessel and lymphatic vascular endothelial continuous cells, blood vessel and lymphatic vascular endothelial splenic cells, synovial cells, serosal cells, squamous cells, columnar cells, dark cells, vestibular membrane cells, stria vascularis basal cells, stria vascularis marginal cells, Claudius cells, Boettcher cells, choroid plexus cells, pia-arachnoid squamous cells, pigmented and non-pigmented ciliary epithelial cells, corneal endothelial cells, and peg cells), ciliated cells of the respiratory tract cells, oviduct cells, uterine endometrium cells, rete testis cells, and ductulus efferens cells, ciliated ependymal cells of central nervous system, keratinizing epithelial cells (*e.g.*, epidermal keratinocyte, epidermal basal cells, keratinocytes, nail bed basal cells, medullary hair shaft cells, cortical hair shaft cells, cuticular hair shaft cells, cuticular hair root sheath cells, hair root sheath cell of Huxley's layer, hair root sheath cell of Henle's layer, external hair root sheath cells, and hair matrix cells), wet stratified barrier epithelial cells (*e.g.*, surface epithelial cell of stratified squamous epithelium of cornea, tongue, oral cavity, esophagus, anal canal, distal urethra and vagina; basal cell of epithelia of the cornea, tongue, oral cavity, esophagus, anal canal, distal urethra, and vagina; and urinary epithelium cells), cells of the nervous system (*e.g.*, sensory transducer cells, auditory inner hair cell of organ of Corti, auditory outer hair cell of organ of Corti, basal cell of olfactory epithelium, cold-sensitive primary sensory neurons, heat-sensitive primary sensory neurons, Merkel cell of epidermis, olfactory receptor neurons, pain-sensitive primary sensory neurons, photoreceptor cells of the retina, proprioceptive primary sensory neurons, touch-sensitive primary sensory neurons, cholinergic neurons, adrenergic neurons, peptidergic neural cells, inner and outer pillar cells, inner and outer phalangeal cells,

border cells, hensen cells, taste bud supporting cells, olfactory epithelium supporting cells, Schwann cells, satellite cells, enteric glial cells, central nervous system neural and glial cells, and lens cells), hepatocyte, adipocytes, liver lipocytes, kidney cells (*e.g.*, glomerulus parietal cells, glomerulus podocyte cells, proximal tubule brush border loop of Henle thin segment cells, distal tubule cells, and collecting duct cells), lung cells, Type I pneumocytes, pancreatic duct cells, nonstriated duct cells, principal cells, intercalated cells, duct cells, intestinal brush border cells, exocrine gland striated duct cells, gall bladder epithelial cells, ductulus efferens nonciliated cells, epididymal principal cells, epididymal basal cells, extracellular matrix cells, ameloblast epithelial cells, planum semilunatum epithelial cells, loose connective tissue fibroblasts, corneal fibroblasts, tendon fibroblasts, bone marrow reticular tissue fibroblasts, nucleus pulposus cells, cementoblast/cementocytes, odontoblast/odontocytes, hyaline cartilage chondrocytes, fibrocartilage chondrocytes, fibroblast cartilage chondrocytes, osteoblast/osteocytes, osteoprogenitor cells, hyalocytes of vitreous body of eye, stellate cells of perilymphatic space of ear, hepatic stellate cells, pancreatic stele cells, contractile cells, skeletal muscle cells, heart muscle cells, smooth muscle cells, blood and immune cells (*e.g.*, erythrocyte, megakaryocyte, monocyte, connective tissue macrophage, epidermal langerhans, osteoclast, dendritic cell, microglial cell, neutrophil granulocyte, eosinophil granulocyte, basophil granulocyte, mast cell, T cell, suppressor T cell, cytotoxic T cell, natural killer T cell, B cell, and reticulocyte), Stem cells and committed progenitors for the blood and immune system (*e.g.*, pigment cells, melanocytes, and retinal pigmented epithelial cells), germ cells (*e.g.*, oocyte, spermatid, spermatocyte, spermatogonium cell, and spermatozoon, nurse cells (*e.g.*, ovarian follicle cell, and sertoli cells, and thymus epithelial cells), interstitial cells, and combinations thereof. In certain embodiments, the biological analyte is selected from the group consisting of cells, organelles, cell fragments, cell membranes, cell membrane fragments, viruses, virus-like particles, bacteriophage, cytosolic proteins, secreted proteins, signaling molecules, embedded proteins, nucleic acid/protein complexes, organelles, minicells, nucleic acid precipitants, chromosomes, nuclei, mitochondria, chloroplasts, flagella, biominerals, protein complexes, protein aggregates, urine, saliva, feces, bacteria or algae in water (ground water, streams, oceans, etc.), bacteria (*e.g.*, *ecoli*) in food, and combinations thereof. In certain embodiments, the biological analyte is selected from the group consisting of bacteria, circulation tumor cells, and parasites, and combinations thereof.

[0091] In some specific embodiments, the analytes of interest are healthy red blood cell, diseased red blood cell, or sickle cell. In some specific embodiments, the biological analyte

is selected from the group consisting of normal erythrocyte with hemoglobin Hb AA, Hb CC, and Hb AS, sickle cell erythrocyte with hemoglobin Hb SS, Hb SC, HbSbeta⁺, HbSD, HbSE and HbSO, reticulocyte, predominantly hypochromic red blood cells (*e.g.*, iron deficiency anemia (IDA)), predominantly microcytic red blood cells (*e.g.*, β -thalassemia trait (β -TT)), α -thalassemia minor, hemoglobin H disease, Bart's hydrops fetalis, other α -thalassemias, normal red blood cells, and white blood cells.

[0092] In some embodiments, the characteristic of the biological analyte is a disease state or a biological index of the biological analyte. In some embodiments, non-limiting examples of the index of the biological analytes include shape, size, density, mass, width, volume of the biological analyte. In some embodiments, the characteristic of the red blood cell is one or more indices selected from the group consisting of the average size of a red blood cell (mean corpuscular volume, MCV), the average amount of hemoglobin per red blood cell (mean corpuscular hemoglobin, MCH), the average amount of hemoglobin per volume of red blood cells (mean corpuscular hemoglobin concentration, MCHC), and the red blood cell distribution width (RDW). In certain specific embodiments, the disease state of the biological analyte is anemia or the sickle state of the cells. Non-limiting examples of anemia include microcytic anemia, hypochromic anemia, iron deficiency anemia, and β -thalassemia trait. *See*, Blood 2014, 123, 615-624, for additional examples of anemia.

[0093] Any multi-phase system described herein may be used. In some embodiments, the multi-phase system comprises at least adjacent first and second phase-separated phases, wherein

the first phase comprises a first phase component predominantly dissolved in the solvent of the first phase; and the second phase comprises a second phase component predominantly dissolved in the solvent of the second phase; wherein

the solvents of the first and second phases are the same;

the first phase component is different from the second phase component;

each of the first and second components is selected from the group consisting of a polymer, a surfactant and combinations thereof; and

at least one of the first and second phase components comprises a polymer;

each of the first and second phases has a different density and the first and second phases, taken together, represent a density gradient; and

the first and second phases have a stable interface in-between.

[0094] In certain embodiments, the first and second phase components are each selected from the group consisting of Carboxy-polyacrylamide, Dextran, Ficoll, N,N-dimethyldodecylamine N-oxide, poly(2-ethyl-2-oxazoline), poly(acrylic acid), poly(ethylene glycol), poly(methacrylic acid), poly(vinyl alcohol), polyacrylamide, polyethyleneimine, hydroxyethyl cellulose, poly(2-acrylamido-2-methyl-1-propanesulfonic acid), polyvinylpyrrolidone, Nonyl, polyallylamine, (hydroxypropyl)methyl cellulose, diethylaminoethyl-dextran, nonylphenol polyoxyethylene 20, copolymer, terpolymer, block copolymer, random polymer, linear polymer, branched polymer, crosslinked polymer, and dendrimer system thereof.

[0095] The biological analyte is distributed in the multi-phase system to generate a color distribution profile under certain assay conditions. In some embodiments, the assay condition is one or more conditions selected from the group consisting of the composition of the multi-phase system (*e.g.*, phase component and solvent, and amounts thereof) and the distribution condition of the biological analyte in the multi-phase system. In some embodiments, the assay condition is one or more conditions selected from the group consisting of the force used for the settlement of the biological analyte in the multi-phase system, and the time allowed for the settlement of the biological analyte in the multi-phase system.

[0096] In some embodiments, the phase components are any of the phase components described herein. In certain embodiments, the solvent used in the multi-phase is water. In certain embodiments, the solvent used in the multi-phase is an organic solvent. In some embodiments, the amount of the solvent used is 1 μ l, 2 μ l, 5 μ l, 10 μ l, 20 μ l, 50 μ l, 100 μ l, or 1 ml. Any other suitable amount for the solvent is contemplated. In some embodiments, the multiphase system is placed on top of the biological analyte. In some embodiments, the biological analyte is deposited at the top of the multiphase system. In some embodiments, the biological analyte is allowed to settle in the multi-phase system under gravity. In other embodiments, the biological analyte is allowed to settle in the multi-phase system under centrifugation. In some embodiments, the centrifugation time is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, or 20 minutes. Any known centrifugation force may be used. Non-limiting examples of the

centrifugation force include 100, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000, 13000, 13500, 14000, 15000, 20000, 50000, 100000 g, or the centrifugation force is in any ranges limited by any two of the values disclosed herein.

[0097] In some embodiments, the assay conditions, as described herein, also include thermodynamic conditions or dynamic conditions. In some embodiments, the biological analyte is separated by the multi-phase systems under thermodynamic conditions. As used herein, the term thermodynamic condition refers to the scenario that the biological analyte reaches its location in the multi-phase system characteristic of its density. The thermodynamic location of the analyte is determined by the density gradient of the multi-phase system and the density of the analyte. Once reaching its thermodynamic location in the multi-phase system, the biological analyte will not further change its location therein even upon further settlement under centrifugation.

[0098] In other embodiments, the biological analyte is separated by the multi-phase systems under dynamic conditions. As used herein, the term dynamic condition refers to the scenario that the biological analyte has not reached its thermodynamic location in the multi-phase system. Various other properties of the analyte, *e.g.*, mass, volume, width, chemical composition, shape, may affect the settlement rate of the analyte. If allowed to settle for sufficient time, the biological analyte will eventually reach its thermodynamic location in the multiphase system. The color distribution profile under dynamic conditions provides an information-rich data set for the analyte, because it provides information not only related to the density of the analyte, but also its mass, volume, width, chemical composition, shape, or a combination of any two or more of these factors. In other embodiments, the color distribution profile is a recorded video over time of the biological analyte settling and moving through the MPS. In these embodiments, the video can be recorded by the use of a high speed camera, the use of a strobe light, the use of scanning optics (*e.g.*, CD/DVD drive), or the use of a linear CCD array.

Method for determining a characteristic of a biological analyte of interest

[0099] In another aspect, a method for determining a characteristic of a biological analyte of interest is described, comprising:

generating a color distribution profile of a biological analyte of interest in a phase-separated multi-phase system under a first assay condition;

generating a database and storing the database in a memory, the database comprising one or more algorithms and one or more assay conditions, wherein each algorithm is associated with an assay condition and configured to predict a characteristic of the biological analyte of interest based on its color distribution profile in a phase-separated multi-phase system under the assay condition; and wherein at least one of the one or more assay conditions is the first assay condition; and

based on the reader-generated color distribution profile of the biological analyte, using a computer to predict a characteristic of the biological analyte of interest using the algorithms associated with the first condition.

[0100] In some embodiments, the algorithm is built by machine-learning. In any one of the embodiments described herein, the machine-learning comprises a process comprising creating, training, validating and/or testing the algorithm using a plurality of biological analytes with known characteristics.

[0101] The biological analytes, the multi-phase system, the algorithm, the color distribution profile, and the assay conditions are as described herein. In some embodiments, the method further comprises separating the biological analyte of interest in the multi-phase system.

Multi-phase system

[0102] Described herein are multi-phase systems including two or more phase-separated phases each containing a phase component. The MPS can be a two- or three-phase system as disclosed herein. However, MPSs containing more than three phases are also contemplated. The MPSs described herein have important biological applications, including, but not limited to, enrichment of reticulocytes, diagnosis of iron deficiency anemia and β -thalassemia trait, and diagnosis of sickle cell disease and its subtypes.

[0103] In some embodiments, MPSs as described herein are used to separate analytes (e.g., red blood cells, reticulocytes, erythrocytes, etc.) from each other or from impurities and other objects in the sample. The analytes migrate to phases characteristic of their densities, and in so doing, contact each phase of the multi-phase system sequentially. As used herein, “sequential contact” means that the analyte contacts and interacts with only one phase (and its phase component) at a time except at the interface between two phases. That is, the interaction of the analyte with the MPS occurs when the MPS has already phase separated

and not during the process of phase separation. In some embodiments, the pH, osmolality, and the polymer used in the preparation of the phase separated components are selected to be compatible with the cells to be analyzed or separated.

[0104] The concept of the multi-phase system (MPS) is further explained herein. When two or more solutions each containing a phase component are mixed, the resulting system is not homogeneous; rather, two or more discrete phases, or layers, form. These layers are ordered according to density and arise from the exhibited limited interaction of the phase components with one another. The two or more phases or solutions thus exhibit limited interaction and form distinct phase boundaries between adjacent phases. The two adjacent phases have rapid material exchange and reach a thermodynamic, stable equilibrium. Thus, the phase separation and phase boundary are stable and not easily disturbed. Each phase can be aqueous or non-aqueous. The non-aqueous phase comprises an organic liquid or an organic solvent. When the solvent used for the MPS is water, the MPS is also called an aqueous multi-phase system (AMPS).

[0105] The multi-phase systems disclosed herein comprise two or more zones or regions that are phase-separated from each other, wherein each of the two or more phases comprises a phase component. The phase component is a polymer or a combination of two or more polymers.

[0106] Non-limiting examples of polymer used in the formation of a phase include dextran, polysucrose (herein referred to by the trade name “Ficoll”), poly(vinyl alcohol), poly(2-ethyl-2-oxazoline), poly(methacrylic acid), poly(ethylene glycol), polyacrylamide, polyethyleneimine, hydroxyethyl cellulose, polyvinylpyrrolidone, carboxy-polyacrylamide, poly(acrylic acid), poly(2-acrylamido-2-methyl-1-propanesulfonic acid), dextran sulfate, diethylaminoethyl-dextran, chondroitin sulfate A, poly(2-vinylpyridine-N-oxide), poly(diallyldimethyl ammonium chloride), poly(styrene sulfonic acid), polyallylamine, alginic acid, nonylphenol polyoxyethylene, poly(bisphenol A carbonate), polydimethylsiloxane, polystyrene, poly(4-vinylpyridine), polycaprolactone, polysulfone, poly(methyl methacrylate-co-methacrylic acid), poly(methyl methacrylate), poly(tetrahydrofuran), poly(propylene glycol), poly(vinyl acetate), copolymer, terpolymer, block copolymer, random polymer, linear polymer, branched polymer, crosslinked polymer, and dendrimer system thereof. As used herein, a polymer includes its homopolymer,

copolymer, terpolymer, block copolymer, random polymer, linear polymer, branched polymer, crosslinked polymer, and/or dendrimer system.

[0107] The phase components are selected so that the resulting phases are phase-separated from each other. As used herein, phase-separation refers to the phenomena where two or more solutions, each comprising a phase component, when mixed together, form the same number of distinct phases where each phase has clear boundaries and is separated from other phases. Each phase component used in the solution is selected to be soluble in the solvent of the phase, so that each resulting phase is a distinct solution of the phase component and each phase is phase-separated from other adjacent phase(s). When the multi-phase polymer system is designed, each phase component is selected to predominantly reside in one particular phase of the multi-phase system. It should be noted that in the resulting multi-phase system, every phase can contain varying amounts of other phase components from other phases in the MPS, in addition to the selected desired phase component in that phase. Unless otherwise specified, the phase component composition in each phase of the multi-phase system recited herein generally refers to the starting phase component composition of each phase, or to the predominant phase component composition of each phase. The boundary between every two adjacent phases is also called the interface between the two phases. In some embodiments, the MPS is placed in a container and there is also an interface formed between the bottom phase and the container.

[0108] In some embodiments, the MPS described herein to analyze the biological analyte is a two-phase aqueous system. Non-limiting examples of the two phase systems include aqueous two-phase systems where the phase component combination of the two phases is selected from the group consisting of:

Phase component combinations		
1	poly(2-ethyl-2-oxazoline)	poly(methacrylic acid)
2	poly(2-ethyl-2-oxazoline)	poly(vinyl alcohol)
3	poly(ethylene glycol)	poly(methacrylic acid)
6	poly(ethylene glycol)	poly(2-ethyl-2-oxazoline)
8	dextran	poly(2-ethyl-2-oxazoline)
10	Ficoll	poly(methacrylic acid)
11	Ficoll	poly(vinyl alcohol)
12	Ficoll	poly(2-ethyl-2-oxazoline)
15	polyacrylamide	poly(methacrylic acid)

16	polyacrylamide	poly(acrylic acid)
18	polyacrylamide	poly(2-ethyl-2-oxazoline)
19	polyacrylamide	poly(ethylene glycol)
20	poly(diallyldimethyl ammonium chloride)	poly(methacrylic acid)
21	poly(diallyldimethyl ammonium chloride)	poly(acrylic acid)
22	poly(diallyldimethyl ammonium chloride)	poly(vinyl alcohol)
23	poly(diallyldimethyl ammonium chloride)	poly(2-ethyl-2-oxazoline)
24	poly(diallyldimethyl ammonium chloride)	poly(ethylene glycol)
25	dextran sulfate	poly(vinyl alcohol)
26	dextran sulfate	poly(2-ethyl-2-oxazoline)
28	chondroitin sulfate A	poly(methacrylic acid)
29	chondroitin sulfate A	poly(vinyl alcohol)
30	chondroitin sulfate A	poly(2-ethyl-2-oxazoline)
31	polyethyleneimine	poly(methacrylic acid)
32	polyethyleneimine	poly(2-ethyl-2-oxazoline)
33	polyethyleneimine	poly(ethylene glycol)
34	polyethyleneimine	Ficoll
35	polyethyleneimine	polyacrylamide
36	polyvinylpyrrolidone	poly(methacrylic acid)
39	poly(propylene glycol)	poly(methacrylic acid)
41	poly(propylene glycol)	polyacrylamide
42	poly(2-acrylamido-2-methyl-1-propanesulfonic acid)	dextran
43	poly(2-acrylamido-2-methyl-1-propanesulfonic acid)	polyvinylpyrrolidone
44	poly(styrene sulfonic acid)	poly(2-ethyl-2-oxazoline)
45	poly(styrene sulfonic acid)	dextran sulfate
46	diethylaminoethyl-dextran	poly(acrylic acid)
47	polyallylamine	dextran sulfate
48	alginate acid	poly(acrylic acid)
49	alginate acid	poly(propylene glycol)
50	(hydroxypropyl)methyl cellulose	poly(diallyldimethyl ammonium chloride)
51	(hydroxypropyl)methyl cellulose	poly(propylene glycol)
52	carboxy-polyacrylamide	poly(methacrylic acid)
53	carboxy-polyacrylamide	poly(vinyl alcohol)
54	carboxy-polyacrylamide	polyethyleneimine
55	hydroxyethyl cellulose	dextran
56	hydroxyethyl cellulose	Ficoll
57	methyl cellulose	Ficoll
58	Zonyl	poly(methacrylic acid)
59	Zonyl	dextran

60	Zonyl	polyacrylamide
61	Brij	poly(2-ethyl-2-oxazoline)
62	Brij	Ficoll
63	Brij	polyallylamine
64	Tween	poly(methacrylic acid)
65	Tween	poly(vinyl alcohol)
66	Tween	poly(2-ethyl-2-oxazoline)
69	Tween	Ficoll
70	Tween	polyacrylamide
71	Tween	polyallylamine
72	Tween	hydroxyethyl cellulose
73	Triton	poly(methacrylic acid)
74	Triton	poly(acrylic acid)
75	Triton	poly(2-ethyl-2-oxazoline)
77	Triton	Ficoll
78	Triton	polyacrylamide
79	Triton	polyallylamine
81	nonylphenol polyoxyethylene	poly(methacrylic acid)
82	nonylphenol polyoxyethylene	dextran
83	1-O-Octyl-B-D-glucopyranoside	poly(methacrylic acid)
84	1-O-Octyl-B-D-glucopyranoside	poly(2-ethyl-2-oxazoline)
86	1-O-Octyl-B-D-glucopyranoside	polyethyleneimine
87	Pluronic	poly(methacrylic acid)
88	Pluronic	poly(vinyl alcohol)
89	Pluronic	poly(2-ethyl-2-oxazoline)
90	Pluronic	dextran
91	Pluronic	Ficoll
92	Pluronic	polyacrylamide
93	Pluronic	polyethyleneimine
94	sodium dodecyl sulfate	poly(acrylic acid)
95	sodium cholate	poly(methacrylic acid)
96	sodium cholate	dextran sulfate
97	N,N-dimethyldodecylamine N-oxide	poly(methacrylic acid)
98	N,N-dimethyldodecylamine N-oxide	polyacrylamide
99	CHAPS	poly(methacrylic acid)
100	CHAPS	poly(2-ethyl-2-oxazoline)
101	CHAPS	poly(ethylene glycol)
102	CHAPS	dextran
103	CHAPS	Ficoll

104	CHAPS	polyacrylamide
105	CHAPS	polyethyleneimine
106	CHAPS	Pluronic
107	PVPNO	PA
108	PVPNO	PMAA
111	PVPNO	PEOZ
112	PVPNO	PEG
116	PVPNO	PEI
117	PVPNO	Tween
118	Ficoll	Dextran
119	Poly(ethylene glycol)	Ficoll

[0109] In some embodiments, the MPS described herein to analyze the biological analyte is a three-phase aqueous system. Non-limiting examples of the three phase systems include aqueous three-phase systems where wherein the phase component combination of the three phases is selected from the group consisting of:

Number	Phase component combinations		
1	poly(methacrylic acid)	poly(2-ethyl-2-oxazoline)	poly(ethylene glycol)
2	poly(methacrylic acid)	poly(2-ethyl-2-oxazoline)	Ficoll
3	poly(methacrylic acid)	poly(2-ethyl-2-oxazoline)	polyacrylamide
4	poly(methacrylic acid)	poly(2-ethyl-2-oxazoline)	poly(diallyldimethyl ammonium chloride)
5	poly(methacrylic acid)	poly(2-ethyl-2-oxazoline)	chondroitin sulfate A
6	poly(methacrylic acid)	poly(2-ethyl-2-oxazoline)	polyethyleneimine
7	poly(methacrylic acid)	poly(2-ethyl-2-oxazoline)	Tween
8	poly(methacrylic acid)	poly(2-ethyl-2-oxazoline)	Triton
9	poly(methacrylic acid)	poly(2-ethyl-2-oxazoline)	1-O-Octyl-B-D-glucopyranoside
10	poly(methacrylic acid)	poly(2-ethyl-2-oxazoline)	Pluronic
11	poly(methacrylic acid)	poly(2-ethyl-2-oxazoline)	CHAPS
12	poly(methacrylic acid)	poly(ethylene glycol)	Ficoll
13	poly(methacrylic acid)	poly(ethylene glycol)	polyacrylamide
14	poly(methacrylic acid)	poly(ethylene glycol)	poly(diallyldimethyl ammonium chloride)
15	poly(methacrylic acid)	poly(ethylene glycol)	polyethyleneimine
16	poly(methacrylic acid)	poly(ethylene glycol)	polyvinylpyrrolidone
17	poly(methacrylic acid)	poly(ethylene glycol)	Tween 20
18	poly(methacrylic acid)	poly(ethylene glycol)	1-O-Octyl-B-D-glucopyranoside
19	poly(methacrylic acid)	poly(ethylene glycol)	CHAPS
20	poly(methacrylic acid)	Ficoll	polyethyleneimine
21	poly(methacrylic acid)	Ficoll	Tween
22	poly(methacrylic acid)	Ficoll	Triton
23	poly(methacrylic acid)	Ficoll	Pluronic
24	poly(methacrylic acid)	Ficoll	CHAPS

25	poly(methacrylic acid)	polyacrylamide	polyethyleneimine
26	poly(methacrylic acid)	polyacrylamide	poly(propylene glycol)
27	poly(methacrylic acid)	polyacrylamide	Zonyl
28	poly(methacrylic acid)	polyacrylamide	Tween
29	poly(methacrylic acid)	polyacrylamide	Triton
30	poly(methacrylic acid)	polyacrylamide	Pluronic
31	poly(methacrylic acid)	polyacrylamide	N,N-dimethyldodecylamine N-oxide
32	poly(methacrylic acid)	polyacrylamide	CHAPS
33	poly(methacrylic acid)	polyethyleneimine	carboxy-polyacrylamide
34	poly(methacrylic acid)	polyethyleneimine	1-O-Octyl-B-D-glucopyranoside
35	poly(methacrylic acid)	polyethyleneimine	Pluronic
36	poly(methacrylic acid)	polyethyleneimine	CHAPS
37	poly(methacrylic acid)	Pluronic F68	CHAPS
38	poly(acrylic acid)	poly(ethylene glycol)	polyacrylamide
39	poly(acrylic acid)	poly(ethylene glycol)	poly(diallyldimethyl ammonium chloride)
40	poly(acrylic acid)	polyacrylamide	Triton
41	poly(vinyl alcohol)	poly(2-ethyl-2-oxazoline)	poly(ethylene glycol)
42	poly(vinyl alcohol)	poly(2-ethyl-2-oxazoline)	dextran
43	poly(vinyl alcohol)	poly(2-ethyl-2-oxazoline)	Ficoll
44	poly(vinyl alcohol)	poly(2-ethyl-2-oxazoline)	polyacrylamide
45	poly(vinyl alcohol)	poly(2-ethyl-2-oxazoline)	poly(diallyldimethyl ammonium chloride)
46	poly(vinyl alcohol)	poly(2-ethyl-2-oxazoline)	dextran sulfate
47	poly(vinyl alcohol)	poly(2-ethyl-2-oxazoline)	chondroitin sulfate A
48	poly(vinyl alcohol)	poly(2-ethyl-2-oxazoline)	Tween
49	poly(vinyl alcohol)	poly(2-ethyl-2-oxazoline)	Pluronic
50	poly(vinyl alcohol)	poly(ethylene glycol)	dextran
51	poly(vinyl alcohol)	poly(ethylene glycol)	Ficoll
52	poly(vinyl alcohol)	poly(ethylene glycol)	polyacrylamide
53	poly(vinyl alcohol)	poly(ethylene glycol)	poly(diallyldimethyl ammonium chloride)
54	poly(vinyl alcohol)	poly(ethylene glycol)	dextran sulfate
55	poly(vinyl alcohol)	poly(ethylene glycol)	Tween
56	poly(vinyl alcohol)	dextran	Ficoll
57	poly(vinyl alcohol)	dextran	Tween
58	poly(vinyl alcohol)	dextran	Pluronic
59	poly(vinyl alcohol)	Ficoll	Tween
60	poly(vinyl alcohol)	Ficoll	Pluronic
61	poly(vinyl alcohol)	polyacrylamide	Tween
62	poly(vinyl alcohol)	polyacrylamide	Pluronic
63	poly(2-ethyl-2-oxazoline)	poly(ethylene glycol)	dextran
64	poly(2-ethyl-2-oxazoline)	poly(ethylene glycol)	Ficoll
65	poly(2-ethyl-2-oxazoline)	poly(ethylene glycol)	polyacrylamide
66	poly(2-ethyl-2-oxazoline)	poly(ethylene glycol)	poly(diallyldimethyl ammonium chloride)
67	poly(2-ethyl-2-oxazoline)	poly(ethylene glycol)	dextran sulfate
68	poly(2-ethyl-2-oxazoline)	poly(ethylene glycol)	polyethyleneimine

69	poly(2-ethyl-2-oxazoline)	poly(ethylene glycol)	Tween
70	poly(2-ethyl-2-oxazoline)	poly(ethylene glycol)	1-O-Octyl-B-D-glucopyranoside
71	poly(2-ethyl-2-oxazoline)	poly(ethylene glycol)	CHAPS
72	poly(2-ethyl-2-oxazoline)	dextran	Ficoll
73	poly(2-ethyl-2-oxazoline)	dextran	Tween
74	poly(2-ethyl-2-oxazoline)	dextran	Triton
75	poly(2-ethyl-2-oxazoline)	dextran	Pluronic
76	poly(2-ethyl-2-oxazoline)	dextran	CHAPS
77	poly(2-ethyl-2-oxazoline)	Ficoll	polyethyleneimine
78	poly(2-ethyl-2-oxazoline)	Ficoll	Brij
79	poly(2-ethyl-2-oxazoline)	Ficoll	Tween
80	poly(2-ethyl-2-oxazoline)	Ficoll	Triton
81	poly(2-ethyl-2-oxazoline)	Ficoll	Pluronic
82	poly(2-ethyl-2-oxazoline)	Ficoll	CHAPS
83	poly(2-ethyl-2-oxazoline)	polyacrylamide	polyethyleneimine
84	poly(2-ethyl-2-oxazoline)	polyacrylamide	Tween
85	poly(2-ethyl-2-oxazoline)	polyacrylamide	Triton
86	poly(2-ethyl-2-oxazoline)	polyacrylamide	Pluronic
87	poly(2-ethyl-2-oxazoline)	polyacrylamide	CHAPS
88	poly(2-ethyl-2-oxazoline)	dextran sulfate	poly(styrene sulfonic acid)
89	poly(2-ethyl-2-oxazoline)	polyethyleneimine	1-O-Octyl-B-D-glucopyranoside
90	poly(2-ethyl-2-oxazoline)	polyethyleneimine	Pluronic
91	poly(2-ethyl-2-oxazoline)	polyethyleneimine	CHAPS
92	poly(2-ethyl-2-oxazoline)	Pluronic F68	CHAPS
93	poly(ethylene glycol)	dextran	Ficoll
94	poly(ethylene glycol)	dextran	polyvinylpyrrolidone
95	poly(ethylene glycol)	dextran	Tween
96	poly(ethylene glycol)	dextran	CHAPS
97	poly(ethylene glycol)	Ficoll	polyethyleneimine
98	poly(ethylene glycol)	Ficoll	Tween
99	poly(ethylene glycol)	Ficoll	CHAPS
100	poly(ethylene glycol)	polyacrylamide	polyethyleneimine
101	poly(ethylene glycol)	polyacrylamide	Tween
102	poly(ethylene glycol)	polyacrylamide	CHAPS
103	poly(ethylene glycol)	polyethyleneimine	1-O-Octyl-B-D-glucopyranoside
104	poly(ethylene glycol)	polyethyleneimine	CHAPS
105	dextran	Ficoll	hydroxyethyl cellulose
106	dextran	Ficoll	Tween
107	dextran	Ficoll	Triton
108	dextran	Ficoll	Pluronic
109	dextran	Ficoll	CHAPS
110	dextran	polyvinylpyrrolidone	poly(2-acrylamido-2-methyl-1-propanesulfonic acid)
111	dextran	hydroxyethyl cellulose	Tween
112	dextran	hydroxyethyl cellulose	Triton

113	dextran	Pluronic F68	CHAPS
114	Ficoll	polyethyleneimine	Pluronic
115	Ficoll	polyethyleneimine	CHAPS
116	Ficoll	hydroxyethyl cellulose	Tween
117	Ficoll	hydroxyethyl cellulose	Triton
118	Ficoll	Pluronic F68	CHAPS
119	polyacrylamide	polyethyleneimine	Pluronic
120	polyacrylamide	polyethyleneimine	CHAPS
121	polyacrylamide	Pluronic F68	CHAPS
122	polyethyleneimine	Pluronic F68	CHAPS
123	PEOZ	PEG	PVPNO
124	PEOZ	PEI	PVPNO
125	PEOZ	PA	PVPNO
126	PEOZ	PMAA	PVPNO
127	PEG	PEI	PVPNO
128	PEG	PMAA	PVPNO
129	PEG	PA	PVPNO
130	PEI	PA	PVPNO
131	PEI	PMAA	PVPNO
132	PA	PMAA	PVPNO
133	PEOZ	PEG	PVPNO
134	PEOZ	TWEEN	PVPNO
135	PEOZ	PA	PVPNO
136	PEOZ	PMAA	PVPNO
137	PEG	TWEEN	PVPNO
138	TWEEN	PA	PVPNO
139	TWEEN	PMAA	PVPNO
140	PA	PMAA	PVPNO
141	PEG	PA	PVPNO
142	PEG	PMAA	PVPNO

[0110] A list of some of the abbreviations for polymer used in this study are as follows:

poly(2-vinylpyridine-N-oxide) – PVPNO;

poly(methacrylic acid) – PMAA;

poly(acrylic acid) – PAA;

polyacrylamide – PA;

poly(vinyl alcohol) – PVA;

poly(2-ethyl-2-oxazoline) – PEOZ;

poly(ethylene glycol) – PEG;

hydroxyethylcellulose – HEC;

polyethyleneimine – PEI; and

polyvinylpyrrolidone – PVP.

Further examples of MPS

[0111] In some embodiments, the MPS used herein are further described below.

[0112] The concentration of the phase component in each phase is selected so that the resulting density of each phase will fall in the range of density as described herein. In some embodiments, the concentration of the first phase component in the first phase or the concentration of the second phase component in the second phase is between about 1-40 % (w/v). In some specific embodiments, the concentration of the first phase component in the first phase and the concentration of the second phase component in the second phase are each independently about 9.0%, 9.3%, 9.5%, 10.0%, 10.1%, 10.3%, 10.5%, 10.6%, 10.8%, 11.0%, 11.1%, 11.4%, 11.6%, or 12.0% (w/v). Ranges bounded by any of the specific values noted above are also contemplated. In some specific embodiments, the concentration of the first phase component in the first phase or the concentration of the second phase component in the second phase is about 9.0%-12.0% (w/v).

[0113] In some embodiments, the first and second phase components for the aqueous two-phase system for enrichment of reticulocytes are each selected from the group consisting of Caboxy-polyacrylamide, Dextran, Ficoll, N,N-dimethyldodecylamine N-oxide, poly(2-ethyl-2-oxazoline), poly(acrylic acid), poly(ethylene glycol), poly(methacrylic acid), poly(vinyl alcohol), polyacrylamide, polyethyleneimine, hydroxyethyl cellulose, poly(2-acrylamido-2-methyl-1-propanesulfonic acid), polyvinylpyrrolidone, Nonyl, polyallylamine, (hydroxypropyl)methyl cellulose, diethylaminoethyl-dextran, and nonylphenol polyoxyethylene 20.

[0114] In one specific embodiment, the aqueous two-phase system for enrichment of reticulocytes has dextran and Ficoll as its the first and second phase components, respectively. In some embodiments, Ficoll having a molecular weight of 70 K Da or 400 Da is used. In some embodiments, dextran having a molecular weight of 500 K Da is used. In other embodiments, Ficoll or dextran with other molecular weight known in the art can be used.

[0115] The differences in the densities of the phases of MPSs provide a means to perform density-based separations. The interfaces between phases mark discontinuities (on the molecular scale) is between continuous fluid phases of different density. The densities (ρ_A and ρ_B) of the phases above and below the interface establish the range of densities for

components (ρ_C) that will localize at the interface ($\rho_A > \rho_C > \rho_B$). The interfacial surface energy between the phases of a MPS is astonishingly low (from nJ m^{-2} to mJ m^{-2}); a low interfacial surface energy reduces the mechanical stress on cells as they pass through the interface.

[0116] Compared to layered gradients in density (*e.g.*, Percoll, Optiprep, or Nycodenz), the MPSs described herein offer several advantages: i) they are thermodynamically stable, ii) they self-assemble rapidly ($t \sim 15$ minutes, 2000 g) on centrifugation or slowly ($t \sim 24$ hours) on settling in a gravitational field, iii) they can differentiate remarkably small differences in density ($\Delta\rho < 0.001 \text{ g cm}^{-3}$), and iv) they provide well-defined interfaces that facilitate both the identification and extraction of sub-populations of cells by concentrating them to quasi-two-dimensional surfaces.

[0117] In some embodiments, the various components of the blood sample naturally settle in the MPS to their thermodynamically stable states. In some embodiments, during settling the reticulocytes contact one or more of the two phases sequentially. As a result, the enriched reticulocytes will settle to a location in the MPS characteristic of its density, *e.g.*, at the interface between phases of lower and higher density than the reticulocytes. In other embodiments, the multi-phase system containing the blood sample can be centrifuged. The use of centrifuge facilitates the settlement process, by speeding up the migration of the biological analyte, *e.g.*, reticulocyte, to a location in the MPS characteristic of its density. In some embodiments, the multi-phase system and the human blood sample (placed on top of the MPS) is centrifuged for about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, or 30 minutes. When the centrifuging process is conducted only for a short period of time, *e.g.*, about 1, 2, 3, 4, or 5 minutes, the analyte in the blood sample may not have reached its thermodynamic state. Ranges bounded by any of the specific values noted above are also contemplated. In some embodiments, the centrifuging process is stopped while the reticulocyte is still migrating through the phases in the MPS. Such shorter centrifuge can reveal the size or shape profile of the reticulocytes, as reticulocytes of the same density but with different sizes or shapes can have different settlement rates in the MPS (given sufficient time, it is expected that all reticulocytes with the same density will occupy the same location – regardless of differences in sedimentation rates).

[0118] In some embodiments, the volume ratio of the human blood sample to the multiphase system is about 4:1, 2:1, 1:1, 1:2, 1:3, 1:4, 1:5, or 1:6, or 0.15-4:1, or 0.2-2:1, or

0.3-1:1. Ranges bounded by any of the specific values noted above are also contemplated. In some embodiments, the volume ratio of the human blood sample to the multiphase system is about 1:1.

[0119] The tonicity of a MPS system is a colligative property that depends primarily on the number of dissolved particles in solution. Thus, the tonicity of the MPS can be adjusted by using a tonicity adjusting agent. Non-limiting examples of tonicity adjusting agent include dextrose, glycerin, mannitol, KCl, and NaCl.

[0120] The biological analyte, *e.g.*, blood cells, can change its size or shape in response to the change of the tonicity or pH of the MPS. As a result, changing tonicity or pH will affect the biological analyte's size and thereby its density and/or migration speed in the MPS. Additionally, changing tonicity or pH will affect the biological analyte's shape and thereby its migration speed through the MPS phases. The tonicity or pH can affect different cells differently. For instance, tonicity or pH may affect reticulocytes and erythrocytes to different extents. Therefore, changing tonicity or pH may provide another parameter to improve the separation and enrichment of the biological analyte, *e.g.*, reticulocytes. In some embodiments, Applicants have surprisingly found that hypertonic MPS provides superior enrichment results for reticulocytes.

[0121] Additionally, the addition of certain tonicity-adjusting agents may also change the density of all the phases in the MPS. For instance, the addition of NaCl or KCl will increase the density and tonicity of each of the MPS phases. This provides another way of fine-tuning the density ranges of the MPS phases. In other embodiments, Nycodenz can be added to the phases in the MPSs to adjust the density alone without affecting the tonicity of the phases.

[0122] In some embodiments, the enriched reticulocytes are collected with a yield of more than 0.5%, 1%, 1.5%, 2%, 3%, 4%, 5%, 10%, or 20% using the method described herein.

[0123] In some embodiments, the method described herein can be adopted to various scales, *e.g.*, the microliter scale, the milliliter scale, or the multi-liter scale.

[0124] In some embodiments, the MPS is a two-phase system and the density ranges of the two phases are described herein and selected to allow the iron deficiency microcytic and/or hypochromic red blood cells characteristic of iron deficiency anemia or the α/β -

thasassemias to be easily identified. For instance, the densities of the two phases may be selected so that iron deficiency anemia red blood cells or β -thalassemia trait red blood cells will settle and reside in the interface of the first and second phases to allow easy identification of the IDA or β -TT. In a specific embodiment, the MPS is an aqueous two-phase system having the first and second densities at about 1.0784 g/cm^3 and 1.0810 g/cm^3 , respectively. In one embodiment, the first and second phase components are dextran and Ficoll, respectively.

[0125] In some other embodiments, the aqueous multi-phase system further comprises: a third aqueous phase comprising a third phase component and having a third density between about 1.073 g/cm^3 and about 1.093 g/cm^3 ; wherein the third density is higher than the first density but lower than the second density; and the third phase component comprises at least one polymer.

[0126] In some embodiments, the third density is less than about 0.002 g/cm^3 , 0.0019 g/cm^3 , 0.0018 g/cm^3 , 0.0017 g/cm^3 , 0.0016 g/cm^3 , 0.0015 g/cm^3 , 0.0014 g/cm^3 , 0.0013 g/cm^3 , 0.0012 g/cm^3 , 0.0011 g/cm^3 , 0.0010 g/cm^3 , 0.0009 g/cm^3 , 0.0008 g/cm^3 , 0.0007 g/cm^3 , 0.0006 g/cm^3 , 0.0005 g/cm^3 , 0.0004 g/cm^3 , 0.0003 g/cm^3 , 0.0002 g/cm^3 , or 0.0001 g/cm^3 lower than the second density. Ranges bounded by any of the specific values noted above are also contemplated.

[0127] In some embodiments, the first, third, and second densities are about 1.040 - 1.055 g/cm^3 , 1.075 - 1.085 g/cm^3 , and 1.080 - 1.085 g/cm^3 , respectively. In one specific embodiment, the first, third, and second densities are about 1.0505 g/cm^3 , 1.0810 g/cm^3 , and 1.0817 g/cm^3 , respectively. In one specific embodiment, the first, third, and second phase components are PVA, dextran and Ficoll, respectively.

[0128] In some embodiments, the first and second phase components are selected so that the resulting two phases phase separate to form the two-phase system. Similarly, in other embodiments, the first, third, and second phase components are selected so that the resulting three phases phase separate to form the three-phase system. In some embodiments, the first, second, and third phase components are each selected from the group consisting of Carboxy-polyacrylamide, Dextran, Ficoll, N,N-dimethyldodecylamine N-oxide, poly(2-ethyl-2-oxazoline), poly(acrylic acid), poly(ethylene glycol), poly(methacrylic acid), poly(vinyl alcohol), polyacrylamide, polyethyleneimine, hydroxyethyl cellulose, poly(2-acrylamido-2-

methyl-1-propanesulfonic acid), polyvinylpyrrolidone, Nonyl, polyallylamine, (hydroxypropyl)methyl cellulose, diethylaminoethyl-dextran, and nonylphenol polyoxyethylene 20, copolymer, terpolymer, block copolymer, random polymer, linear polymer, branched polymer, crosslinked polymer, and dendrimer system thereof. Generally, any two- or three- phase system described herein can be used, provided that the density of each phase falls in the ranges of the phase densities described herein. In other embodiments, a single phase is used and a color distribution profile of the analyte in the single phase is generated. In certain embodiments, the single phase is viscous.

[0129] The concentration of the phase component in each phase also can be fine-tuned to adjust the density of each phase so that the density of each phase falls in the ranges of the phase densities described herein. In other embodiments, the density ranges of the phases can also be achieved by adding additives such as Nycodenz. In some embodiments, the concentration of the first phase component in the first phase or the concentration of the second phase component in the second phase is between about 1-40 % (w/v). In some specific embodiments, the concentration of the first phase component in the first phase or the concentration of the second phase component in the second phase is about 5%, 10%, 15%, 20%, or 25% (w/v). In some specific embodiments, the concentration of the third phase component in the third phase is between about 1-40 % (w/v) or about 5%, 10%, 15%, 20%, or 25% (w/v). In some specific embodiments, the concentration of the first phase component in the first phase, the concentration of the third phase component in the third phase, or the concentration of the second phase component in the second phase is about 5%-25%, 10%-20%, or 15%-20% (w/v).

[0130] As described herein, the tonicity of a MPS system can be adjusted using a tonicity adjusting agent including, but are not limited to, dextrose, glycerin, mannitol, NaH_2PO_4 (or its hydrate form), KH_2PO_4 , KCl, and NaCl. In some embodiments, the aqueous multi-phase system for the diagnosis of iron deficiency anemia and/or β -thalassemia trait is isotonic.

[0131] The use of other additives are contemplated. In some embodiments, cells are tagged with NPs specific to certain markers (*e.g.*, CD4). In other embodiments, additives are used to specifically destroy certain cells (RBC lysis with saponin) or to create aggregation (adding prothrombin to aggregate platelets).

[0132] Further details and examples of the multi-phase systems can be found in PCT/US14/35697, filed on April 28, 2014, WO2012/024688, filed on August 22, 2011, WO2012/024693, filed on August 22, 2011, WO2012/024690, filed on August 22, 2011, and WO2012/024691, filed on August 22, 2011, all of which are hereby incorporated by reference herein in their entirety.

Detection of Iron Deficiency Anemia (IDA) and Prediction of Red Blood Cell Index

[0133] In one specific embodiment, the machine/computer-aided system and/or method described herein are used for the detection of iron deficiency anemia (IDA) and/or the prediction of red blood cell index.

[0134] Aqueous multiphase systems (AMPS) are aqueous solutions of polymers and surfactants that spontaneously phase segregate and form discrete, immiscible layers. Between each phase is an interface with a molecularly sharp step in density; these steps in density can be used to separate subpopulations of cells by density. The phases of an AMPS can be tuned to have very small steps in density ($\Delta\rho < 0.001 \text{ g/cm}^3$), can be made biocompatible, are thermodynamically stable, and reform if shaken. We previously used AMPS as a tool to enrich reticulocytes from whole blood, and to detect sickle cell disease.

[0135] In some embodiments, we demonstrate the use of AMPS to diagnose IDA, by exploiting the fact that RBCs in patients with micro/hypo anemia have lower density than those of healthy patients. In certain specific embodiments, using only a drop of blood (a volume easily obtainable from a finger prick), we can detect, by eye, low density RBCs and diagnose IDA in under three minutes; this method had a true positive rate (sensitivity) of 84%, with a 95% confidence interval (CI) of 72-93%, and a true negative rate (specificity) of 78% (CI = 68-86%).

[0136] In certain embodiments, the diagnostic accuracy of the system disclosed herein is improved by imaging each AMPS test with a digital scanner and analyzing the distribution of red color—corresponding to the RBCs—found in the tube. In some embodiments, using standard machine learning protocols, we are able to diagnose IDA with a sensitivity of 90% (83-96%) and a specificity of 77% (64-87%), and were able to detect hypochromic RBCs above a threshold of 3.9% with a sensitivity of 96% (CI = 88-99%) and a specificity of 92% (CI = 84-97%). Thus, in certain embodiments, a simple optical reader paired with appropriate algorithms provides rapid, reader-insensitive diagnosis.

[0137] In certain embodiments, using machine learning, many of the important values measured during a complete blood count (namely, values pertaining to red blood cells or "red blood cell indices") can also be predicted. Red blood cell indices are used to diagnose many diseases and, therefore, predicting their values quickly and simply may be clinically useful.

Experimental Design

The sedimentation rate of red blood cells is related to important red-cell indices

[0138] The sedimentation rate of red blood cells through a fluid is a function of several physical characteristics of the cells: mass, volume, size, shape, deformability, and density (mass per unit volume). These characteristics are related, directly or indirectly, to a number of red blood cell indices, including, mean corpuscular volume (MCV, fL) or the average size of a red blood cell, mean corpuscular hemoglobin (MCH, pg/cell) or the average amount of hemoglobin per cell, mean corpuscular hemoglobin concentration (MCHC, g/dL) or the average amount of hemoglobin per volume of blood, red blood cell distribution width (RDW, %) or the distribution in volume of the RBCs. These characteristics, in addition to the hematocrit (HCT)—the ratio of the volume of the RBCs to the total volume of blood—can be used to derive the total number of RBCs (#RBCs) and the total hemoglobin concentration in the blood (HGB, g/dL).

[0139] Many hematology analyzers use these indices to bin red blood cells. The percentage of red blood cells that are microcytic (%Micro) is defined as the fraction of cells below a specific MCV. The percentage of red blood cells that are hypochromic (%Hypo) is defined as the fraction of cells below a specific MCHC.

[0140] The hematology analyzer used in this study (ADVIA 2120, Siemens) defines %Micro as the percentage of RBCs of MCV < 60 fL and %Hypo as the percentage of RBCs with MCHC < 28 g/dL. IDA corresponds to a decrease in MCV, MCH, MCHC, and HGB, and an increase in RDW, %Hypo, and %Micro. Several other hemoglobinopathies have been shown to affect the density of RBCs and could affect the performance of a density-based test. Sick cell disease, and spherocytosis are known to increase the density of some or all of the population of RBCs, while β -thalassemia, α -thalassemia, and malaria decrease RBC density.

Classifying Blood Samples Using Hematological Indices

[0141] Possible markers used to make a diagnosis of IDA include transferrin saturation, and ferritin. These methods, however, are time consuming and impractical in many settings; extensive research has focused on using red blood cell indices to diagnose IDA.

[0142] Here, we discuss several conditions with overlapping population. Using hematological parameters (*i.e.*, red blood cell indices), we define four different states: 1) hypochromia—the condition of having hypochromic RBCs— as %hypo $\geq 3.9\%$, 2) micro/hypo anemia—the condition of having hypochromic RBCs and low HGB—as %hypo $\geq 3.9\%$ and when HGB < 12.0 g/dL for females over 15 yrs, < 13.0 g/dL for males over 15 yrs, < 11.0 g/dL for children under 5 yrs, and < 11.5 g/dL for children 5 to 15 yrs,⁸ 3) IDA as micro/hypo anemia when %micro/%hypo ≤ 1.5 , and 4) β -thalassemia. **Fig. 9** is a flow chart illustrating the classification, *i.e.*, the diagnosis of hypochromia, micro/hypo anemia, iron deficiency anemia, and β -thalassemia trait used in this study based on hematological indices measured by a hematology analyzer (Advia 2120, Siemens).

Designing AMPS to Identify Micro/Hypo RBCs

[0143] An AMPS with n total phases will contain $n+1$ interfaces (*e.g.*, there are 3 interfaces in a two phase system: air/phase-1, phase-1/phase-2, and phase-2/container).

[0144] In order to detect the presence of microcytic and hypochromic red blood cells, a properly designed AMPS may: i) have a top layer with density greater than that of plasma and its components (≥ 1.025 g cm⁻³) in order to minimize dilution of the AMPS, ii) have a bottom layer less dense than the average red blood cell density (which are represented by a Gaussian distribution where mature erythrocytes have a density of 1.095 g cm⁻³ and immature erythrocytes (*i.e.*, reticulocytes) of 1.086 g cm⁻³) such that normal blood will pack at the bottom of the tube, iii) maintain biocompatibility by tuning the pH (7.4) and osmolality (290 mOsm/kg) to match blood, and iv) undergo phase separation in a short amount of time (≤ 5 minutes) under centrifugation (*e.g.*, 13,700 g, the speed of the StatSpin CritSpin centrifuge used in this study).

[0145] We designed two different AMPS: 1) A simple two-phase AMPS (IDA-AMPS-2) to diagnose microcytic and hypochromic anemia and IDA by the presence of a band or streak of redness above the packed hematocrit; 2) A three-phase AMPS (IDA-AMPS-3) to capture

microcytic and hypochromic RBCs at two liquid/liquid interfaces and to provide additional information about the density distribution of the RBCs of a patient.

[0146] IDA-AMPS-2 was designed to have a top layer density $\rho_{\text{top}} = 1.0784 \text{ g cm}^{-3}$ and bottom layer density $\rho_{\text{bot}} = 1.0810 \text{ g cm}^{-3}$ and was composed of 10.2% (w/v) dextran (MW ~500 kD), 10.2% (w/v) Ficoll (MW ~400 kD), and 0.7% (w/v) Nycodenz. IDA-AMPS-3 contained 10.2% (w/v) partially hydrolyzed poly(vinyl alcohol) (containing 78% hydroxyl and 22% acetate groups) (MW ~6 kD), 5.6% (w/v) dextran (MW ~500 kD), and 7.4% (w/v) Ficoll (MW ~400 kD). The density of the phases were $\rho_{\text{top}} = 1.0505 \text{ g cm}^{-3}$, $\rho_{\text{mid}} = 1.0810 \text{ g cm}^{-3}$, $\rho_{\text{bot}} = 1.0817 \text{ g cm}^{-3}$ as measured by a U-tube oscillating densitometer (DMA 35, Anton-Paar).

An AMPS diagnostic system easy to use, rapid, and fieldable

[0147] We previously demonstrated the use of a point-of-care assay for sickle cell disease using AMPS. A similar strategy is employed here. Briefly, a plastic microhematocrit tube is preloaded with 15 μl of AMPS solution that has been sealed at the bottom with a white vinyl-based sealant (Critoseal, Leica), and centrifuged for 2 minutes at 13,700 g in a hematocrit centrifuge (CritSpin, Iris Sample Processing) in order to separate the phases.

[0148] A drop (5 μl) of blood is loaded at the top of the tube through capillary action enabled by a small hole in the side of the tube; the hole allows the blood to enter the tube up to and not beyond the hole (by capillary wicking). We previously demonstrated that blood can be loaded in this manner with a coefficient of variance (CV) in the volume loaded < 4%. A elastomeric silicone sleeve is then slid over the hole to prevent the blood leaking during centrifugation. Up to 12 tubes can then be loaded into the hematocrit centrifuge and spun for the desired time. In the current study we used a centrifuge that cost ~\$1,600 (CritSpin, Iris Sample Processing). Alternatively, a more portable centrifuge manufactured by HWLab was used that provides similar performance and costs \$150 (\$60 each for orders > 400 units).

[0149] The cost of the materials and reagents necessary to fabricate a test at the laboratory scale is ~\$0.20.

[0150] The total time needed to perform this assay is less than ten minutes (it can be done in as little as three minutes), and all of the components, including a battery to power the centrifuge, can fit into a backpack. In certain embodiments, a lead-acid 12V car battery is

chosen because it is widely available, has a long life cycle, is relatively low cost, and can be charged by nearly every car and truck in the world as well as by solar panels). In other embodiments, 4 lithium ion cells (*e.g.*, 18650 cells) or 9 primary alkaline batteries are used.

Visual analysis of IDA-AMPS tests after centrifugation provides a simple diagnostic test for micro/hypo anemia

[0151] The diagnostic readout of an IDA-AMPS test is designed to be done readily with the naked eye by visualizing the presence or absence of red color above the packed hematocrit at the bottom of the tube. IDA-AMPS provides three bins of density in which red blood cells can collect: 1) the top/middle (T/M) interface ($\text{RBCs} \leq 1.081 \text{ g cm}^{-3}$), 2) the middle/bottom (M/B) interface ($\text{RBCs} > 1.081 \text{ g cm}^{-3}$ and $\leq 1.0817 \text{ g cm}^{-3}$), and 3) the bottom/seal (B/S) interface ($\text{RBCs} \geq 1.0817 \text{ g cm}^{-3}$) (**Fig. 2A**). Blood is loaded into the top of the tube, from a finger prick, using capillary action provided by a hole in the side of the tube. A silicone sleeve is used to cover the hole to prevent leakage during centrifugation. Normal blood packs at the bottom of the tube, while less dense RBCs can be seen packing at the interfaces between the phases and inside of the phase of the AMPS. Normal and IDA blood can be differentiated, by eye, after only 2 minutes of centrifugation. White blood cells (leukocytes) collect at the Top/Middle interface. In some cases white blood cells can agglomerate with RBCs, resulting in a slight red color at the Top/Middle interface, even in a normal sample.

[0152] Discarded blood samples were obtained from Children's Hospital Boston ($n = 152$, see Table 5 for a summary of the populations of interest) along with complete blood counts from a hematology analyzer (ADVIA 2120, Siemens).

[0153] For the purpose of understanding the optimum timing for a test, the assay was performed by stopping centrifugation every two minutes and imaging the tubes. After ten minutes of centrifugation at 13,700 g nearly all of the RBCs reach their equilibrium position; at lower centrifugation times red color is found throughout the phases of the AMPS in samples having micro/hypo RBCs. The time-dependence of the distributions at short centrifugation times provides additional information regarding the size and density distribution of red blood cells. For this reason—and because a rapid test is desirable—we chose results from $t = 2$ minutes and evaluated the ability of blinded readers to diagnose hypochromia, micro/hypo anemia, and IDA.

[0154] Three readers were trained using a guide (**Fig. 8**) to classify the amount of red color above the packed hematocrit as 1) none or nearly none, 2) some, 3) moderate, 4) strong, and 5) very strong. In some of the cases red cells were more prevalent at the interfaces, while in others, the red color was highly visible in the phases of the AMPS. The guide was available to readers during each reading for reference. An average score was determined based on concordance between at least two of the readers.

Results and Discussion

[0155] Using aqueous multiphase systems (AMPS), we have created a simple and low-cost method to detect microcytic and hypochromic red blood cells, and hence, IDA. After two minutes in a centrifuge that can be powered by battery, the AMPS test can be evaluated, by eye, and used to diagnose IDA with an AUC of 0.88. Using a desktop scanner to image the tests, we evaluate the images of the IDA-AMPS tests and use standard machine learning protocols to diagnose IDA (AUC = 0.90)—a computer aided diagnosis may be desirable for a fielded device in order to reduce possible user variability. The performance of the IDA-AMPS test is comparable to previous studies using reticulocyte hemoglobin concentration to diagnose IDA (AUC = 0.91) and, therefore, may have a high enough performance to find clinical use.

[0156] To our knowledge, there are currently no direct methods of measuring serum ferritin concentration at the POC. Several methods for measuring hemoglobin—providing a diagnosis for anemia, but not necessarily iron deficiency anemia—are available at the POC. These methods include: 1) the cyanmethemoglobin method using a photoelectric colorimeter, 2) spectrophotometrically using the azidemethoglobin method (*e.g.*, the HemoCue system), 3) colorimetrically (both by eye and a smartphone app) using a redox reaction, 4) paper-based devices, 5) the hematocrit estimate method, and 6) the WHO Hemoglobin Color Scale. Since IDA is a nutritional disorder, molecular diagnostics are not useful for diagnosis, except for a rare hereditary form of IDA called “iron refractory IDA”.

[0157] In certain embodiments, the IDA-AMPS test described herein is able to detect microcytic and hypochromic RBCs and diagnose IDA with an AUC comparable to other metrics that have found clinical use, suggesting that it could find widespread use as a screening tool for IDA. In particular, because the equipment needed to run the test is portable, this method may find use in rural clinics where large fractions of the population at risk for

IDA, such as children and pregnant women, seek care in LMICs. Ultimately, the performance of this test is to be validated in such settings to demonstrate feasibility of using and interpreting the assay.

[0158] Using machine learning analysis of digital images, we demonstrate an algorithm that is as good as visual interpretation at identifying IDA; the algorithm determined by the machine learning can be readily implemented into a smartphone application (app). Mobile health—or mHealth, the general term given to portable technologies used to diagnose disease that can transmit data over mobile phone networks—is becoming increasingly widespread in Sub-Saharan Africa. By integrating algorithms determined by machine learning into a smartphone app—eliminating the need for visual interpretation and potential bias from users—our test might be used by minimally trained healthcare workers in LMICs. Interestingly, this test may also find use in veterinary medicine. IDA in livestock, especially pigs, is increasingly common due to modern rearing facilities that eliminate the animals' exposure to iron-containing soil; IDA in pigs can cause weight loss, retarded growth, and an increased susceptibility to infection.

[0159] A simple method to perform a complete blood count without the need to draw large volumes of blood and send that sample to a central laboratory—so called, point-of-care hematology—has been a major goal of the diagnostic community for several decades. We use machine learning to analyze images of IDA-AMPS tests to predict several red blood cell indices, a first step towards POC hematology. We found, in the best case, a Pearson's r of 0.94 for predicting the number of hypochromic RBCs (%Hypo). HemoCue, the most widely used portable test used to measure hemoglobin concentration, in comparison, has been shown to correlate nearly perfectly with a hematology analyzer (Pearson's $r = 0.99$) when operated by trained laboratory staff. When the device was used by clinical staff, however, the correlation was much poorer (Pearson's $r = 0.66$). The lessons from this study suggest that the design and use of a POC hematology system needs to be simple enough for a technician with minimal training to operate.

[0160] The IDA-AMPS test described herein is a new approach to diagnosing IDA and, using machine learning algorithms, to predict red blood cell indices. Instead of directly measuring a biological marker such as concentration of hemoglobin or serum ferritin, our method relies on observing the way in which red blood cells move through a viscous media

(a function of their density as well as size and shape) to make a diagnosis. This approach may be applied to other diseases or biological applications.

Centrifugation of blood through IDA-AMPS provides a clear diagnostic for micro/hypo anemia

[0161] IDA-AMPS-2 provides two bins of density in which blood can collect: 1) blood of low density ($< 1.081 \text{ g cm}^{-3}$) at the interface between the top and bottom phases (T/B), and 2) normal blood ($\geq 1.085 \text{ g cm}^{-3}$) at the bottom of the tube above the white sealing clay. IDA-AMPS-3 provides three bins of density: the T/M interface ($< 1.081 \text{ g cm}^{-3}$), the M/B interface ($> 1.081 \text{ g cm}^{-3}$ and $< 1.0817 \text{ g cm}^{-3}$), and normal blood at the bottom of the tube.

[0162] The transient nature of the systems at short centrifugation times provides additional information regarding the size and density distribution of red blood cells.

[0163] We chose the best results from digital analysis (IDA-AMPS-3, $t = 2$ minutes) and evaluated the ability of blinded readers to diagnose hypochromia, micro/hypo anemia, and IDA.

Digital analysis of IDA-AMPS improves diagnostic performance

[0164] We sought to improve our ability to diagnose IDA or provide reader-insensitive and automated method to make this diagnosis, by analyzing the images obtained using a flatbed scanner. Digital analysis of the AMPS tests was performed using the following steps (**Fig. 2B**): i) a flatbed scanner in transmission mode imaged up to 12 tests simultaneously (Epson Perfection V330 Photo), ii) using a custom program written in Python (iPython Notebook) individual capillary tubes were detected and cropped, and the tube image was converted to hue-saturation-value (HSV) colorspace, iii) the HSV value of each pixel was converted to its corresponding S/V value, and iv) a one dimensional plot of “red intensity” versus distance above the (cropped) seal was compiled by summing the S/V values for each row of pixels and saved for later analysis. Further details can be found below.

[0165] **Fig. 3A** and **3B** show examples of IDA-AMPS tests after 2 minutes of centrifugation for a representative normal and IDA sample, respectively, where an image of the tube and its corresponding image with pixels converted to S/V, 1-D red intensity trace, and the first derivative of the trace. Specifically, **Figs. 3A-B** show, for a representative

normal (3A) and IDA (3B) sample, i) a scanned test image, ii) its corresponding red intensity image where each pixel was converted to S/V, iii) 1-dimensional red intensity trace, and iv) the first derivative of the 1-dimensional red intensity trace. Digital analysis of images of the IDA-AMPS tests enables the direct comparison of a large number of samples. Normal RBCs packs at the bottom of the tube, similarly to a hematocrit, while less dense RBCs can be seen packing at the interfaces between the phases and inside of the phases of the AMPS. White blood cells (leukocytes) pack at the T/M interface and can sometimes agglomerate with RBCs, even in normal blood, causing a slight red color at the T/M interface. For clarity the top (loading port) of the tubes is not shown.

[0166] In **Figs. 4A-B**, the average red intensity for all normal and micro/hypo samples is plotted as a function of distance from the sealed (bottom) end of the tube for different centrifugation times; the shaded region represents the 99% confidence intervals. The red intensity is highest at distance = 0 cm where the hematocrit packs at the bottom of the tube (the white plastic seal is excluded during analysis and the loading end of the tube is excluded from **Figs. 4A-B** for clarity). After 2 minutes of centrifugation, the red intensity difference between the normal and micro/hypo anemic samples in the majority of the tube is high; most of the red color is spread throughout the phases. As the centrifugation time increases, the signal decreases in the phases and increases at the interfaces as red blood cells reach their equilibrium position based on their density.

[0167] Receiver operating characteristic (ROC) curves were generated for visual analysis of IDA-AMPS-3 (**Figs. 5A-5B**) using the 1-5 redness threshold for hypochromia, micro/hypo anemia, and IDA. The area under the curve (AUC) is highest for hypochromia (0.98, CI = 0.96-1.00) (**Figs. 5A-B**); the test is excellent at detecting the presence of hypochromic RBCs. Perfect diagnostic accuracy (*i.e.*, no false positives or false negatives) would result in an AUC = 1.00.

[0168] As shown in **Fig. 5B**, the ability to predict micro/hypo anemia and IDA for the IDA-AMPS test is lower, with an AUC of 0.89 (CI = 0.83-0.94) and 0.88 (CI = 0.81-0.94). For IDA, this corresponds to a sensitivity of 84% (CI = 72-93%) and a specificity of 78% (CI = 68-86%) at a maximum efficiency cutoff threshold of redness > 2 (some red above the packed hematocrit).

[0169] As a diagnostic for IDA, the performance of IDA-AMPS (AUC = 0.89) exceeds that of using only hemoglobin concentration (AUC = 0.73) (often the only metric available in low-resource settings). The AUC, sensitivity, and specificity of IDA-AMPS is also comparable to that of a test for IDA using the reticulocyte hemoglobin concentration (CHr)—a red blood cell parameter measured by a hematology analyzer (AUC of 0.91, sensitivity of 93.2% and a specificity of 83.2%). Although not perfect, this performance for CHr has been high enough to gain popularity in clinical use. The similar AUC of our test for diagnosing IDA (0.89 vs 0.913) suggests that it could be clinically useful as well, especially in LMICs where a hematology analyzer is often unavailable.

[0170] We analyzed the concordance between blinded readers and found excellent intra-reader agreement for duplicates of the sample blood sample. On average the three readers showed a Lin's concordance correlation coefficient, ρ_c of 0.99 (a ρ_c of 1.00 is perfect concordance). Inter-reader agreement was slightly lower; we found a ρ_c of 0.91 between the three readers (Table 1). These results suggest that 1) the IDA-AMPS tests are highly reproducible for the same samples, and 2) stringent training for readers may be necessary to ensure inter-reader concordance.

Table 1. Lin's concordance correlation coefficients assessing inter- and intra- reader concordance for the visual analysis of the IDA-AMPS test after two minutes centrifugation. Inter-reader ρ_c was assessed by averaging the ρ_c between each pair of readers.

Lin's Concordance Correlation Coefficient	
	ρ_c (95% Confidence Interval)
Intra-reader - Reader 1	0.995 (0.994 - 0.997)
Intra-reader - Reader 2	0.985 (0.979 - 0.989)
Intra-reader - Reader 3	0.984 (0.978 - 0.988)
Inter-reader - Replicate 1	0.910 (0.879 - 0.934)
Inter-reader - Replicate 2	0.913 (0.882 - 0.936)

Machine learning providing a method to predict blood parameters and diagnose IDA as an alternative to blinded readers

[0171] Machine learning is a powerful approach for finding an efficient way to make predictions or decisions from data. The general problem of predicting continuously-varying outcomes from data is called “regression”, and predicting classes, or labels, from data is called “classification”. Here we apply standard machine learning techniques to 1) the classification problem of distinguishing micro/hypo anemic samples from normal samples

and 2) the regression problem of predicting continuously-varying red blood cell indices from images of the IDA-AMPS test.

[0172] First, we use principle component analysis (PCA) on the red luminosity versus distance plots to reduce the number of dimensions, which has the effect of a) reducing the computational load by reducing the data from 180 to 30 dimensions, b) reducing the redundancy in the dataset by grouping together adjacent pixels that are co-varying, and c) examining the first derivative of the data.

[0173] We then divide the data into 5 equally-sized stratified groups, or “folds”, taking care that each fold contains a similar proportion of the patient population (normal, micro/hypo anemia, etc). The fifth fold is held out as a “test” dataset, which is never used for training.

[0174] To distinguish normal from micro/hypo anemic patients, we use logistic regression. To predict red blood cell indices, we use support vector regression (SVR) with a radial basis function kernel. We use repeated random sub-sampling validation to guard against over fitting. A training set and validation set are randomly sampled 500 times and the average performed across all sampled validation sets is tabulated. We optimize parameters of the algorithm (*e.g.*, the number of dimensions to use in PCA, the “C” parameter of SVR) using the average R^2 value of the true blood parameters (as measured by a hematology analyzer) vs. predicted blood parameters across each of the four validation sets.

[0175] Using the red intensity traces as input data, we trained a machine-learning algorithm (logistic regression) to discriminate micro/hypo anemic from normal samples; each sample was given a score based on its difference from an average normal sample. Using these scores, receiver operating characteristic (ROC) curves were generated for IDA-AMPS for $t = 2, 4, 6, 8$, and 10 min by changing the decision threshold for micro/hypo anemia using the assigned score. **Fig. 6A** shows the area under the curve (AUC) values obtained from these ROC curves. Perfect diagnostic accuracy (compared to classification by a hematology analyzer) would result in an $AUC = 1.00$. The algorithm randomly samples from the dataset and optimizes the hyperparameters using a validation set of data, and then repeats the process many times. Once the algorithm has been optimized using the training and validation data sets, it analyzes the test data set only one time. For this reason, the results for the AUC calculation are presented without error bars. At short centrifugation times, the test provides

excellent discrimination for micro/hypo anemia; the AUC for IDA-AMPS diminishes after 6 minutes of centrifugation. These data suggest that the optimum centrifugation time for the assay is 2 minutes. For the IDA-AMPS test, the best discrimination between normal and micro/hypo anemic samples is, therefore, at 2 or 4 minutes centrifugation.

[0176] Using the machine learning algorithm, we are able to distinguish hypochromia from normal samples with an AUC of 0.98. For micro/hypo anemia we found an AUC of 0.93, and for IDA we found an AUC of 0.90, corresponding to a sensitivity of 90% (CI = 83-96%) and a specificity of 77% (CI = 64-87%) (**Fig. 6B**). Table 2 provides a comparison of AUC values for visual and digital (machine learning) evaluation of several important subpopulations. These results indicate that: 1) Both digital and visual analysis are excellent at detecting the presence of hypochromic RBCs (*i.e.*, hypochromia); 2) In all cases, the machine learning results are either slightly better or similar to visual evaluation. Depending on the use case, this suggests that the IDA-AMPS test might be best used alongside a low-cost optical detector (*e.g.*, a desktop scanner) or read by the naked eye, with some trade-off between cost and diagnostic accuracy; and 3) Interestingly, our test is able predict hypochromia, micro/hypo anemia, and IDA in women better than men. The AUC for all women in our data set ($n = 74$) is 0.95 compared to 0.86 for men ($n = 78$) and, impressively, is a perfect 1.00 for women ≥ 15 yrs ($n = 47$). This difference may be because the normal range of HGB and MCV for women is lower than for men and the current density of the phases of the AMPS used here is closer to the density of RBCs in female blood. An AMPS with a slightly denser bottom phase density might improve the performance in diagnosing the male population (though with a possible tradeoff in performance for women). This difference may be because the normal range of HGB and MCV for women is lower than for men and the current density of the phases of the AMPS used here is closer to the density of RBCs in female blood. An AMPS with a slightly denser bottom phase density might improve the performance in diagnosing the male population (though with a possible tradeoff in women).

Table 2. Area under the curve (AUC) and 95% confidence interval (CI) results for diagnosing hypochromia, micro/hypo anemia, and IDA using visual and digital analysis of the IDA-AMPS-3 system after 2 minutes centrifugation.

Population	Visual			Digital		
	Hypochromia	Micro/Hypo Anemia	IDA	Hypochromia	Micro/Hypo Anemia	IDA
	AUC (CI)	AUC (CI)	AUC (CI)	AUC (CI)	AUC (CI)	AUC (CI)
General (n = 152)	0.98 (0.96 - 1.00)	0.89 (0.83 - 0.94)	0.88 (0.81 - 0.94)	0.98	0.93	0.90
M (n = 78)	0.95 (0.90 - 1.00)	0.87 (0.79 - 0.96)	0.80 (0.70 - 0.91)	0.97	0.91	0.86
F (n = 74)	1.00 (0.99 - 1.00)	0.91 (0.82 - 0.99)	0.91 (0.83 - 0.99)	0.99	0.95	0.95
Age $\geq$ 15 yrs (n = 47)	0.98 (0.94 - 1.00)	0.97 (0.91 - 1.00)	0.97 (0.91 - 1.00)	1.00	1.00	1.00
Age $\geq$ 5 yrs, < 15 yrs (n = 40)	0.95 (0.86 - 1.00)	0.92 (0.82 - 1.00)	0.88 (0.76 - 1.00)	0.95	0.91	0.91
Age < 5 yrs (n = 65)	0.97 (0.93 - 1.00)	0.83 (0.72 - 0.93)	0.82 (0.70 - 0.93)	0.98	0.86	0.86

[0177] One potential confounding factor for a diagnostic that evaluates the presence of low-density RBCs is other hemoglobinopathies. Beta-thalassemia minor (*i.e.*, β -thalassemia trait, β -TT) and α -thalassemia trait are benign genetic disorders that can present a confounding diagnosis to IDA because both conditions result in microcytic and hypochromic red blood cells. Identification of thalassemic trait is desired to aid (through genetic counseling) in prevention of β -thalassemia major, HbH disease, and Hemoglobin Bart's. Several RBC indices have been shown to provide discrimination between β -TT from IDA. Our initial results indicates that testing on a larger population that includes patients with β -TT (and other thalassemias) might be needed for our test to be implemented in regions with a high prevalence of β -TT; many Mediterranean countries have a prevalence approaching 10%. Many countries in Sub-Saharan Africa and parts of India, however, have a prevalence of β -TT < 3% and some level of uncertainty in differentiating β -TT and IDA might be acceptable.

Predicting Red Blood Cell Indices using Machine Learning

[0178] In the process of identifying micro/hypo anemia, we noticed that the characteristic curves of red luminosity provided an information-rich picture of the dynamics of red blood cells moving through AMPS. Several red blood cell indices should have an impact on the distribution and movement of cells in a gradient.

[0179] The way in which an object moves through an AMPS is related to the density, shape, and size of that object. Many of the parameters measured by a hematology analyzer—so called red blood cell indices—should be related to the distribution and movement of red blood cells in an AMPS. Given the ability of our machine learning approach to identify micro/hypo anemia as well as a trained human user, we tested the ability to use the images of blood moving through the IDA-AMPS tests to predict common red blood cell indices. A rapid and inexpensive test that could predict red blood cell indices could have important clinical implications.

[0180] Using standard machine learning techniques for this “regression” problem, we were able to predict red blood cell indices from the 1D representation of the output of the IDA-AMPS system (see SI for details). We guarded against over-fitting using repeated random sub-sampling validation, in which we randomly sampled a training set and a validation set 500 times, and averaged the performance across all validation sets. For each blood parameter we wished to predict, we independently repeated our cross-validated training approach. We included blood parameters we believed would yield good regression performance (those related to red blood cells, %Hypo, HGB) and as negative controls, those that the IDA-AMPS system would not be able to detect (those related to colorless cells outside of the density range of our system, WBC, PLT).

[0181] True blood parameters, as measured by a hematology analyzer, are compared with predicted parameters determined by machine learning in Figs. 7A-7B and summarized in Table 3. **Fig. 7A** illustrates Machine learning prediction results for %Hypo (Predicted %Hypo) compared to a hematology analyzer (True %Hypo). **Fig. 7B** shows a Bland-Altman plot showing good agreement between true and predicted %Hypo (n = 152). In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting. %Hypo showed the best correlation with a Pearson's r of 0.94 while the other blood cell indices show a lower correlation. As a comparison, other point-of-care tests used to measure HGB (some commercially available) have been found to have r = 0.85 - 0.96. A Pearson's r of 1.00 would represent perfect correlation between the machine learning predictions and the values measured by the hematology analyzer. The ability of a machine learning algorithm to predict any variable in a regression problem is related to the total size of the data set. While the number of patients tested here are substantial for a prototype POC

device, the predictive ability of the algorithm could likely be improved by increasing the size of the data set.

Table 3. Pearson product-moment correlation coefficient (Pearson's r) for the predictive ability of scikit-learn for the IDA-AMPS test.

Blood Parameter	Pearson's r
%Hypo	0.94
MCHC	0.80
CH	0.80
HGB	0.78
HCT	0.76
MCH	0.73
RDW	0.71
HDW	0.68
%Micro	0.65
%Micro/%Hypo	0.63
RBC	0.60
%Hyper	0.50
MCV	0.49
%Macro	0.30

[0182] Table 4 illustrates hemoglobin concentration thresholds used to define anemia in the study. Table 5 illustrates populations of interest for the patients involved in the assessment of the IDA-AMPS test.

Table 4. Hemoglobin concentration thresholds used to define anemia in our study.

Hemoglobin Concentration (HGB)	
Population	g/dl
Female ≥ 15 years	< 12.0
Male ≥ 15 years	< 13.0
Children < 5 years	< 11.0
Children $\geq 5, < 15$ years	< 11.5

Table 5. Populations of interest for the patients involved in the assessment of the IDA-AMPS test.

Population	Male			Female		
	Normal	IDA	β-TT	Normal	IDA	β-TT
< 4.99 years	18	15	3	19	9	1
5 to 14.99 years	12	9	1	13	5	0
> 15 years	13	7	0	15	12	0
Total	43	31	4	47	26	1

[0183] One risk of machine learning is over-fitting. To guard against this we evaluated the tests performance for negative controls that should not correlate to images being evaluated (WBC and PLT) and found, as expected, a very low correlation ($r < 0.2$). Owing to the density of the current test, the algorithm is also unable to predict %Macro or %Hyper; an AMPS with increased density of the phases might instead be used to identify macrocytosis.

Statistical Methods

[0184] We define sensitivity (*i.e.*, the true positive rate) as (number of true positives)/(number of true positives + number of false negatives) and specificity (*i.e.*, the true negative rate) as (number of true negatives)/(number of true negatives + number of false positives) and defined their corresponding 95% Exact binomial confidence intervals. Receiver operating characteristic (ROC) curves and their corresponding area under the curve (AUC) and 95% confidence intervals were calculated in MatLab. Lin's concordance correlation coefficient was calculated using an open-license tool from the National Institute of Water and Atmospheric research of New Zealand (<https://www.niwa.co.nz/node/104318/concordance>).

Choice of Read Guide vs. Color Bar

[0185] Many point-of-care diagnostic tests use a color bar to guide readers to make a quantitative (or semiquantitative) determination. Here, we instead use a "read guide". Using the read guide, the readers are doing pattern recognition over the length of the tube rather than evaluating "redness" at any specific point in the tube. Using a color bar, therefore, is difficult as the patterns are more complicated than a simple shift in color. Two patients with IDA might have very different distributions in the low density of their RBCs, giving very different looking results. For example: Patient A may have 30% hypochromic RBCs, but those RBCs might be only slightly under the threshold of low MCHC (how %Hypo are defined). The IDA-AMPS test for this Patient A might have a strong band of red only in the bottom phase of the AMPS with some RBCs settled at the M/B interface. Patient B may have 15% hypochromic RBCs, but those RBCs might have a larger distribution in MCHC (some very low, some only slightly below the threshold). The IDA-AMPS test for this patient might appear to have a strong red streak in both the bottom and middle phases and a small number of RBCs settled at the T/M and M/B interfaces. In both cases, the patients have red cells

above the bottom packed cells that are visible by eye and would be classified as IDA, even though the distribution of the red cells is different.

Chemicals

[0186] AMPS solutions were prepared using the following reagents: poly(vinyl alcohol) containing 78% hydroxyl and 22% acetate groups (MW = 6 kD, Acros Organics), dextran (MW = 500 kD, Spectrum Chemicals), Ficoll (MW = 400 kD, Sigma-Aldrich), ethylenediaminetetra-acetic acid disodium salt (EDTA, Sigma-Aldrich), sodium phosphate dibasic (Mallinckrodt), potassium phosphate monobasic (EMD), and sodium chloride (EMD). All chemicals were used as received from the suppliers.

Preparation of IDA-AMPS

[0187] IDA-AMPS was prepared mixing, in a volumetric flask 10.2% (w/v) partially hydrolyzed poly(vinyl alcohol) (containing 78% hydroxyl and 22% acetate groups) (MW ~6 kD), 5.6% (w/v) dextran (MW ~500 kD), 7.4% (w/v) Ficoll (MW ~400 kD), 5 mM EDTA (to prevent coagulation), 9.4 mM sodium phosphate dibasic, and 3.0 mM potassium phosphate monobasic. The solution was brought to volume and the pH was brought to 7.40 ± 0.01 (Orion 2 Star, Thermo Scientific) using sodium hydroxide and hydrochloric acid. The osmolality was measured to 290 ± 15 using a vapor pressure osmometer (Vapro 5500, Wescor). We measured density with an oscillating U-tube densitometer (DMA35 Anton Paar). Rapid tests were prepared as described previously.

Automate Sedimentation Analysis with Machine Learning

[0188] In another specific embodiment, machine/computer-aided system and method as described herein can be used to analyze sedimentation data and complete blood count. Supervised learning approach may be used to map characteristics of sedimentation data to common hematological parameters.

[0189] Machine learning provides a method to interpret medical data that is inherently complex and contains multiple dimensions. When applied to medical images, techniques from machine learning are used to perform computer assisted diagnosis in mammograms. When used on complex temporal data, such as electrocardiograms, these methods also aid in the identification of pathologies. The sedimentation of blood in AMPS provides an opportunity to apply machine learning to images of capillaries that evolve over time. Scans

of the sedimentation of blood in AMPS over time provide an information-rich set of data with signals that are related to the dynamics of the cells moving through the fluid phases.

Interpreting tests by eye or with simple algorithms leaves significant amounts of data underutilized. Machine learning enables this underutilized data to be identified and accessed so that all useful signals are available for interpretation. Supervised learning approaches are developed to identify hematological parameters from sedimentation data.

[0190] Experimental Strategy: A large ($n \sim 1,000$) set of discarded blood samples is evaluated to create a database of sedimentation scans along with basic hematology parameters from CBCs. Digital scans of blood moving through AMPS systems will provide the raw data to analyze; until the analytical micro-centrifuge of SA 1 is finished, scans will be obtained using a digital scanner to image tubes after different increments of centrifugation time. The sedimentation scans provide a source of high-quality data. Using machine learning, patterns are examined in images of sedimentation that may correlate with specific parameters from CBCs. For example, blood with a high red blood cell distribution width (RDW) may have a wider band of red color than blood with a normal RDW. Combinations of different parameters will lead to a complex pattern in the image analysis that I will develop algorithms to decipher by taking advantage of recent advances in machine learning techniques to interpret complex images robust to environmental variables, such as viewpoint.

[0191] Preliminary Data: Preliminary results indicated differences in common hematological parameters and correlate with differences in the sedimentation of blood cells. **Fig. 9** shows typical results for line scans of the red luminosity of capillary tubes for blood with different levels of hypochromic red blood cells.

[0192] Specific Methods: Using the analytical micro-centrifuge developed in SA 1 and the AMPS developed in SA 2, a database is built of line scans and CBC data using discarded blood samples ($n \sim 1,000$) from Children's Hospital Boston. This large dataset will serve as a learning set.

[0193] Standard, off-the-shelf machine learning tools can be used for the predictive task of converting an image of blood in a capillary tube filled with AMPS into common blood indices. A simple and robust image preprocessing pipeline was developed to convert a 2D color image of a sedimentation test into a 1D vector indicating the amount of blood in a tube along the length of the tube (**Fig. 9**). Light colored regions depict the standard deviation for

over 100 samples. We dimensionally reduce this data using principal component analysis (PCA) to remove redundant pixels, and increase the separability between data points. We then use support vector regression (SVR) to learn the best transfer function from the principal components of the data from the blood to a single blood parameter output, such as the percent of cells that are hypochromic. We tune the parameters of the SVR algorithm and pipeline (*e.g.*, the amount that we downsample the relatively high-resolution scans) using Bayesian Optimization and 4-fold cross-validation (see below). All core code methods are implemented in the peer-reviewed Python machine learning library “scikit-learn.”

[0194] Before SA 1 is completed, initial data will come from a digital scanner with a transmission mode. After spinning blood for short intervals of time in a standard microhematocrit centrifuge, scans are collected to create a record of the sedimentation through the tubes over time. Although this information will have lower time resolution than a fully functional analytical micro-centrifuge, this alternative provides a contingency plan to begin analyzing the dynamics of sedimentation to identify parameters that correlate with the measurements of a CBC. Several AMPS will be developed in SA 2, but a number of AMPS from previous work and preliminary work can be used while other systems are in development. Systems to separate reticulocytes and to diagnose sickle cell disease will give a starting point to characterize blood flow in different systems.

[0195] Validation: A common pitfall in applying learning algorithms to rich datasets is over-fitting—fitting a function so tightly to the learning data that natural variations result in poor performance when testing the function on actual test data. Due to the risk of over-fitting in machine learning generally, there are very standard techniques and approaches to avoid it. A key first step is to divide available labeled data into three broad categories: 1) training data, 2) validation data, and 3) test data. The algorithm will be fit on training data, and evaluated on validation data in order to improve the specific parameter choices (every algorithm has “knobs” that need to be set correctly in order to achieve good predictive performance, such as the number of dimensions to reduce image data via PCA). During the process of fitting (with the training data) and tuning (with the validation data) the supervised learning algorithm, the algorithm is not shown in the test data. After satisfactorily maximizing performance on the validation data, the generalized performance of the algorithm is tested by applying it to the test data (only once). A useful extension of this process is called “K-fold validation”. This process subdivides the training data into a number “K” evenly-sized subsets, called “folds,”

and trains over permutations of the folds. The parameters are tuned to maximize the average performance of the all “K” iterations. Finally, the resulting algorithm is tested on the reserved test data. For the proposed analysis, a 4-fold cross validation is used.

[0196] Upon review of the description and embodiments described above, those skilled in the art will understand that modifications and equivalent substitutions may be performed in carrying out the invention without departing from the essence of the invention. Thus, the invention is not meant to be limiting by the embodiments described explicitly above.

Claims

We claim:

1. A system for determining a characteristic of a biological analyte of interest, comprising:

a reader for generating a color distribution profile of a biological analyte of interest in a phase-separated multi-phase system under a first assay condition;

a memory for storing one or more algorithms and one or more assay conditions, wherein each algorithm is associated with an assay condition and configured to predict a characteristic of the biological analyte of interest based on its color distribution profile in a phase-separated multi-phase system under the assay condition; and wherein at least one of the assay conditions is the first assay condition;

a computer processor coupled to the reader and the memory, the computer processor is configured to:

receive an input of the first assay condition and the reader-generated color distribution profile of the biological analyte of interest;

based on the reader-generated color distribution profile of the biological analyte, predict a characteristic of the biological analyte of interest using the algorithm associated with the first condition; and

provide an output identifying the predicted characteristic of the biological analyte of interest.

2. The system of claim 1, wherein the algorithm is built by machine-learning.
3. The system of claim 2, wherein the machine-learning comprises a process comprising creating, training, validating and/or testing the algorithm using a plurality of biological analytes with known characteristics.
4. The system of any one of the preceding claims, wherein at least one of the algorithms is configured to make continuously-varying prediction.
5. The system of any one of the preceding claims, wherein at least one of the algorithms is configured to make discrete prediction.
6. The system of any one of the preceding claims, wherein at least one of the algorithms is configured to predict the characteristic of the biological analyte based on comparing

and/or matching one or more color distribution profiles of known biological analytes associated with the first assay condition with the reader-generated color distribution profile of the biological analyte.

7. The system of any one of the preceding claims, wherein the biological analyte has a recognizable color or is dyed with a recognizable color.
8. The system of any one of the preceding claims, wherein the characteristic of the biological analyte is a disease state or a biological index of the biological analyte.
9. The system of any one of the preceding claims, wherein the biological analyte is selected from the group consisting of multicellular organisms, cells, organelles, cell fragments, cell membranes, cell membrane fragments, viruses, virus-like particles, bacteriophage, cytosolic proteins, secreted proteins, signaling molecules, embedded proteins, nucleic acid/protein complexes, organelles, minicells, nucleic acid precipitants, chromosomes, nuclei, mitochondria, chloroplasts, flagella, biominerals, protein complexes, protein aggregates, and combinations thereof.
10. The system of any one of the preceding claims, wherein the biological analyte is red blood cell or a population of red blood cell.
11. The system of claim 10, wherein the characteristic of the red blood cell or the population of red blood cell is one or more indexes selected from the group consisting of the average size of a red blood cell (MCV), the average amount of hemoglobin per red blood cell (MCH), the average amount of hemoglobin per red blood cell (MCHC), the red blood cell distribution width (RDW), percentage of hypochromic red blood cells (%Hypo), hemoglobin concentration (HGB), corpuscular hemoglobin concentration (CH), per unit volume through spun hematocrit (HCT), hemoglobin distribution width (HDW), the number of red blood cells (RBCs), the percentage of red blood cells that are microcytic (%Micro), %Micro/%Hypo, the percentage of cells that are hyperchromic red blood cells (%Hyper), and the percentage of cells that are microcytic red blood cells (%Macro).
12. The system of any one of the preceding claims, wherein the output is a print out or a file or image displayed on a smartphone, a PC, or a monitor.
13. The system of any one of the preceding claims, wherein the system further comprises a separation unit comprising the multi-phase system.

14. The system of any one of the preceding claims, wherein the multi-phase system comprises at least adjacent first and second phase-separated phases, wherein
- the first phase comprises a first phase component predominantly dissolved in the solvent of the first phase; and the second phase comprises a second phase component predominantly dissolved in the solvent of the second phase; wherein
 - the solvents of the first and second phases are the same;
 - the first phase component is different from the second phase component;
 - each of the first and second components is selected from the group consisting of a polymer, a surfactant and combinations thereof; and
 - at least one of the first and second phase components comprises a polymer;
 - each of the first and second phases has a different density and the first and second phases, taken together, represent a density gradient; and
 - the first and second phases have a stable interface in-between.
15. The system of claim 14, wherein the first and second phase components are each selected from the group consisting of Carboxy-polyacrylamide, Dextran, Ficoll, N,N-dimethyldodecylamine N-oxide, poly(2-ethyl-2-oxazoline), poly(acrylic acid), poly(ethylene glycol), poly(methacrylic acid), poly(vinyl alcohol), polyacrylamide, polyethyleneimine, hydroxyethyl cellulose, poly(2-acrylamido-2-methyl-1-propanesulfonic acid), polyvinylpyrrolidone, Nonyl, polyallylamine, (hydroxypropyl)methyl cellulose, diethylaminoethyl-dextran, nonylphenol polyoxyethylene 20, copolymer, terpolymer, block copolymer, random polymer, linear polymer, branched polymer, crosslinked polymer, and dendrimer system thereof.
16. The system of claim 14 or 15, wherein the solvent is water.
17. The system of any one of the preceding claims, wherein the assay condition is one or more conditions selected from the group consisting of the composition of the multi-phase system and the distribution condition of the biological analyte in the multi-phase system.

18. The system of claim 17, wherein the distribution condition of the biological analyte in the multi-phase system comprises the separation time of the biological analyte in the multi-phase system and/or the centrifuge force used for the separation of the biological analyte in the multi-phase system.
19. The system of any one of the preceding claims, wherein the reader is a scanner, a camera, or smartphone camera.
20. The system of any one of the preceding claims, wherein the memory is selected from the group consisting of a hard drive, a thumb drive, a magnetic disk, an optical disk, and magnetic tape.
21. The system of any one of the preceding claims, wherein the color distribution profile comprises a distribution of the biological analyte's color luminosity along the vertical length of the multi-phase system.
22. A method for determining a characteristic of a biological analyte of interest, comprising:
 - generating a color distribution profile of a biological analyte of interest in a phase-separated multi-phase system under a first assay condition;
 - generating a database and storing the database in a memory, the database comprising one or more algorithms and one or more assay conditions, wherein each algorithm is associated with an assay condition and configured to predict a characteristic of the biological analyte of interest based on its color distribution profile in a phase-separated multi-phase system under the assay condition; and wherein at least one of the one or more assay conditions is the first assay condition; and
 - based on the reader-generated color distribution profile of the biological analyte, using a computer to predict a characteristic of the biological analyte of interest using the algorithms associated with the first condition.
23. The method of claim 22, further comprising building the algorithm by machine-learning.
24. The method of claim 23, wherein building the algorithm comprises creating, training, validating and/or testing the algorithm using a plurality of biological analytes with known characteristics.

25. The method of any one of claims 22-24, wherein at least one of the algorithms is configured to make continuously-varying prediction.
26. The method of any one of claims 22-25, wherein at least one of the algorithms is configured to make discrete prediction.
27. The method of any one of claims 22-26, wherein at least one of the algorithms is configured to predict the characteristic of the biological analyte based on comparing and/or matching one or more color distribution profiles of known biological analytes associated with the first assay condition with the reader-generated color distribution profile of the biological analyte.
28. The method of any one of claims 22-27, wherein the biological analyte has a recognizable color or is dyed with a recognizable color.
29. The method of any one of claims 22-28, wherein the characteristic of the biological analyte is a disease state or a biological index of the biological analyte.
30. The method of any one of claims 22-29, herein the biological analyte is selected from the group consisting of multicellular organisms, cells, organelles, cell fragments, cell membranes, cell membrane fragments, viruses, virus-like particles, bacteriophage, cytosolic proteins, secreted proteins, signaling molecules, embedded proteins, nucleic acid/protein complexes, organelles, minicells, nucleic acid precipitants, chromosomes, nuclei, mitochondria, chloroplasts, flagella, biominerals, protein complexes, protein aggregates, and combinations thereof.
31. The method of any one of claims 22-30, wherein the biological analyte is a red blood cell or a population of red blood cell.
32. The method of any one of claims 22-31, wherein the characteristic of the red blood cell or the population of red blood cell is one or more indexes selected from the group consisting of the average size of a red blood cell (MCV), the average amount of hemoglobin per red blood cell (MCH), the average amount of hemoglobin per red blood cell (MCHC), the red blood cell distribution width (RDW), percentage of hypochromic red blood cells (%Hypo), hemoglobin concentration (HGB), corpuscular hemoglobin concentration (CH), per unit volume through spun hematocrit (HCT), hemoglobin distribution width (HDW), the number of red blood cells (RBCs), the percentage of red blood cells that are microcytic (%Micro), %Micro/%Hypo, the

percentage of cells that are hyperchromic red blood cells (%Hyper), and the percentage of cells that are microcytic red blood cells (%Macro).

33. The method of any one of claims 22-32, wherein the method further comprises separating the biological analyte of interest in the multi-phase system.

34. The method of any one of claims 22-33, wherein multi-phase system comprises at least adjacent first and second phase-separated phases, wherein

the first phase comprises a first phase component predominantly dissolved in the solvent of the first phase; and the second phase comprises a second phase component predominantly dissolved in the solvent of the second phase; wherein

the solvents of the first and second phases are the same;

the first phase component is different from the second phase component;

each of the first and second components is selected from the group consisting of a polymer, a surfactant and combinations thereof; and

at least one of the first and second phase components comprises a polymer;

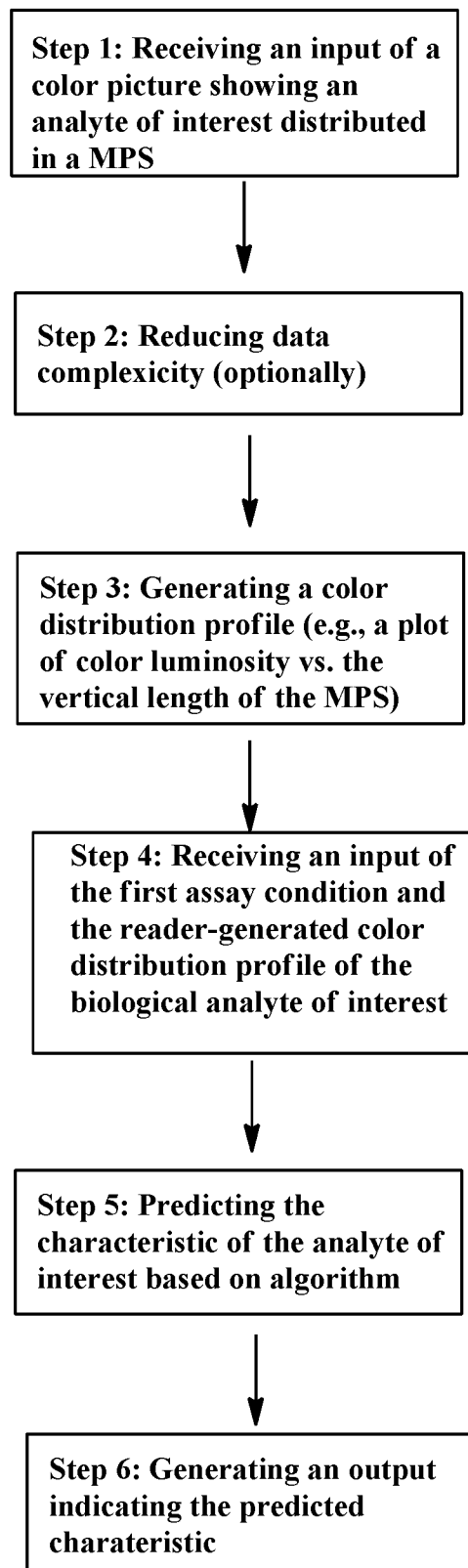
each of the first and second phases has a different density and the first and second phases, taken together, represent a density gradient; and

the first and second phases have a stable interface in-between.

35. The method of claim 34, wherein the first and second phase components are each selected from the group consisting of Carboxy-polyacrylamide, Dextran, Ficoll, N,N-dimethyldodecylamine N-oxide, poly(2-ethyl-2-oxazoline), poly(acrylic acid), poly(ethylene glycol), poly(methacrylic acid), poly(vinyl alcohol), polyacrylamide, polyethyleneimine, hydroxyethyl cellulose, poly(2-acrylamido-2-methyl-1-propanesulfonic acid), polyvinylpyrrolidone, Nonyl, polyallylamine, (hydroxypropyl)methyl cellulose, diethylaminoethyl-dextran, nonylphenol polyoxyethylene 20, copolymer, terpolymer, block copolymer, random polymer, linear polymer, branched polymer, crosslinked polymer, and dendrimer system thereof.

36. The method of any one of claims 22-35, wherein the assay condition is one or more conditions selected from the group consisting of the composition of the multi-phase system and the distribution condition of the biological analyte in the multi-phase system.
37. The method of claim 36, wherein the distribution condition of the biological analyte in the multi-phase system comprises the separation time of the biological analyte in the multi-phase system and/or the centrifuge force used for the separation of the biological analyte in the multi-phase system.
38. The method of any one of claims 22-37, wherein the color distribution profile comprises a distribution of the biological analyte's color luminosity along the vertical length of the multi-phase system.

FIG. 1



2/15

FIG. 2A

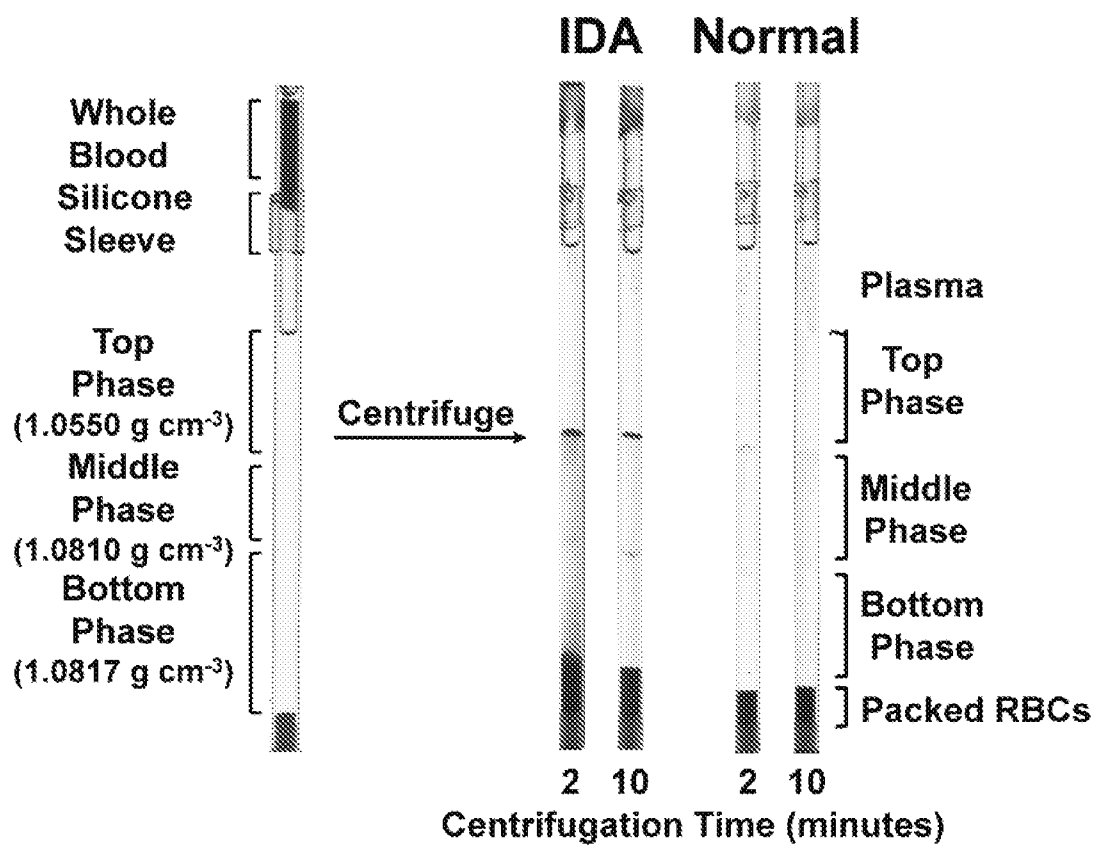


FIG. 2B

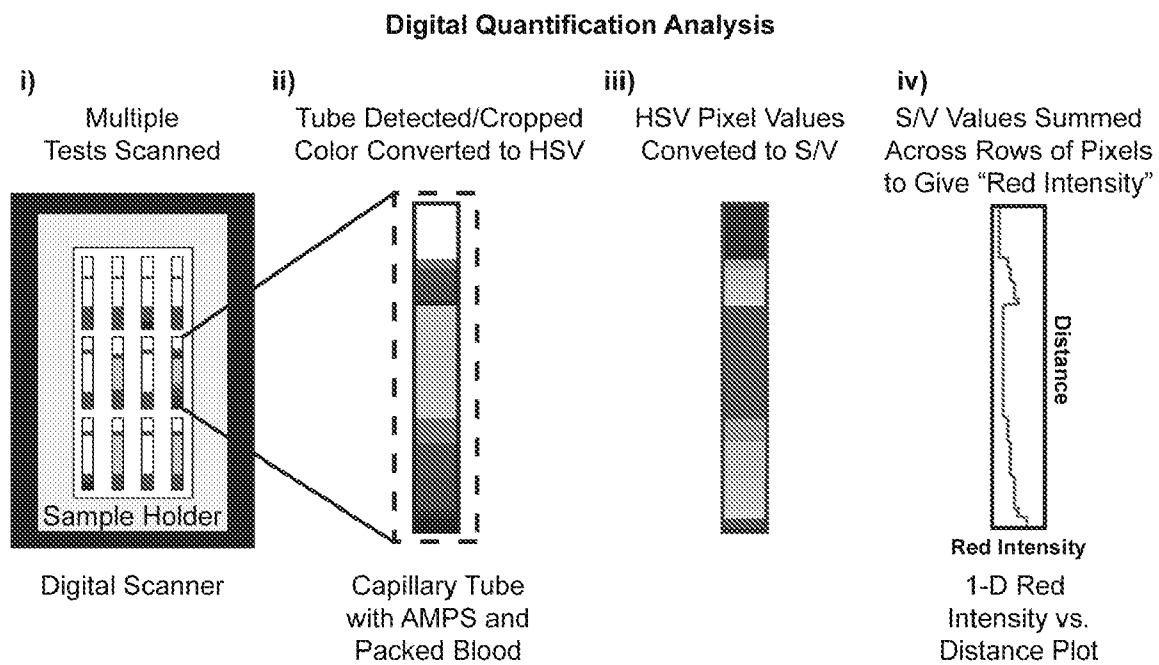


FIG. 3A

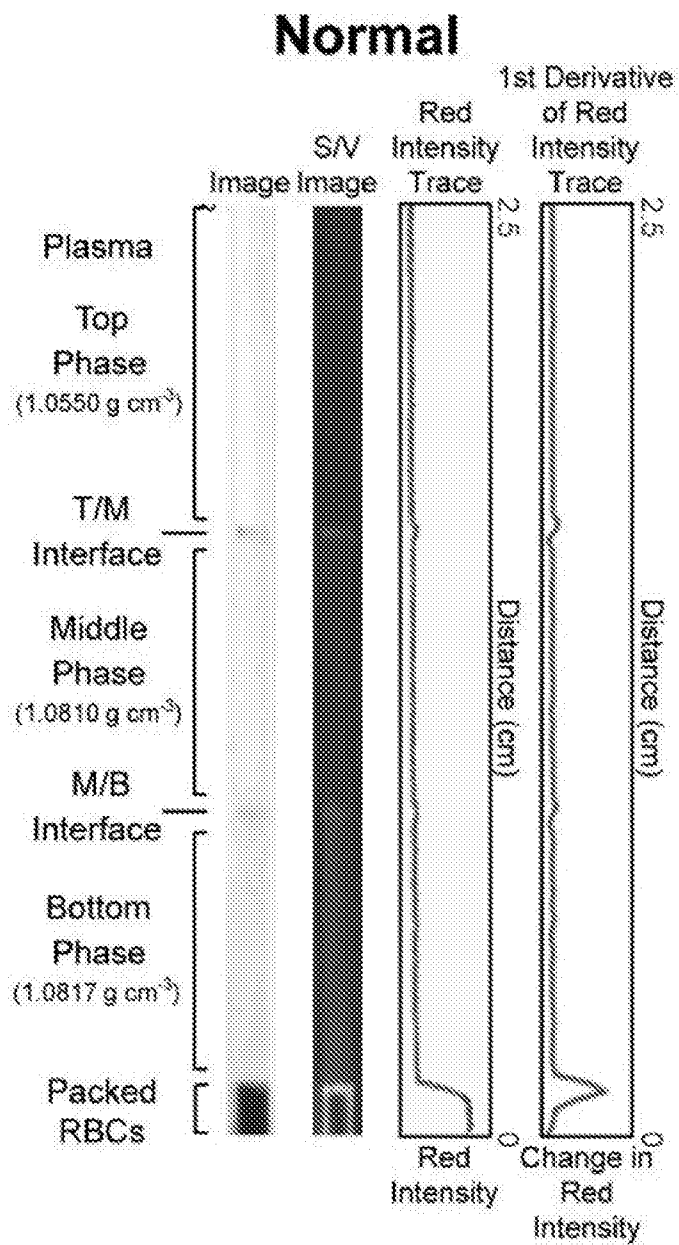


FIG. 3B

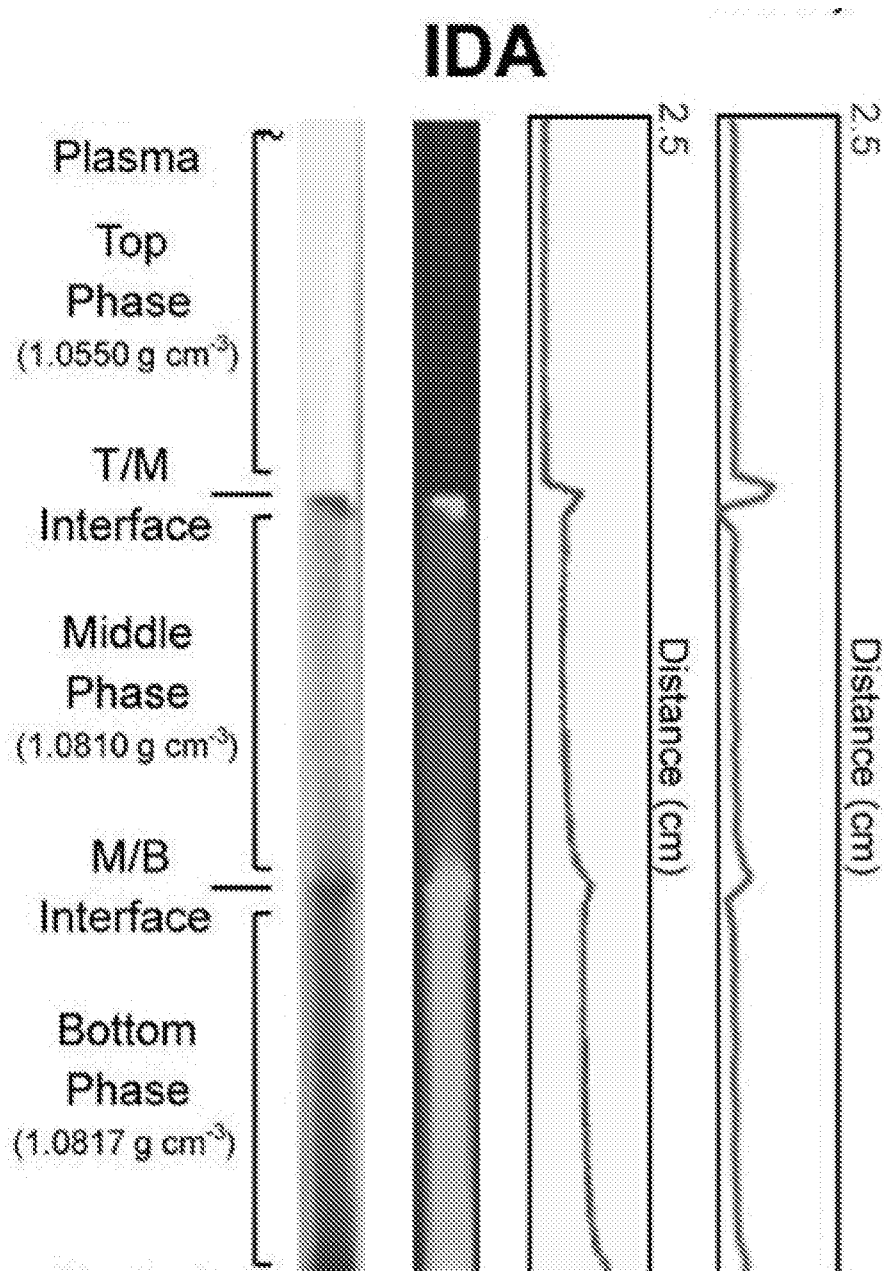
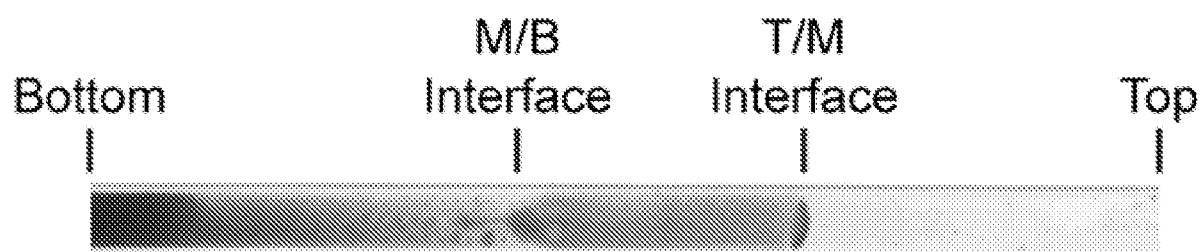


FIG. 4A

Example of Micro/Hypo Sample



7/15

FIG. 4B

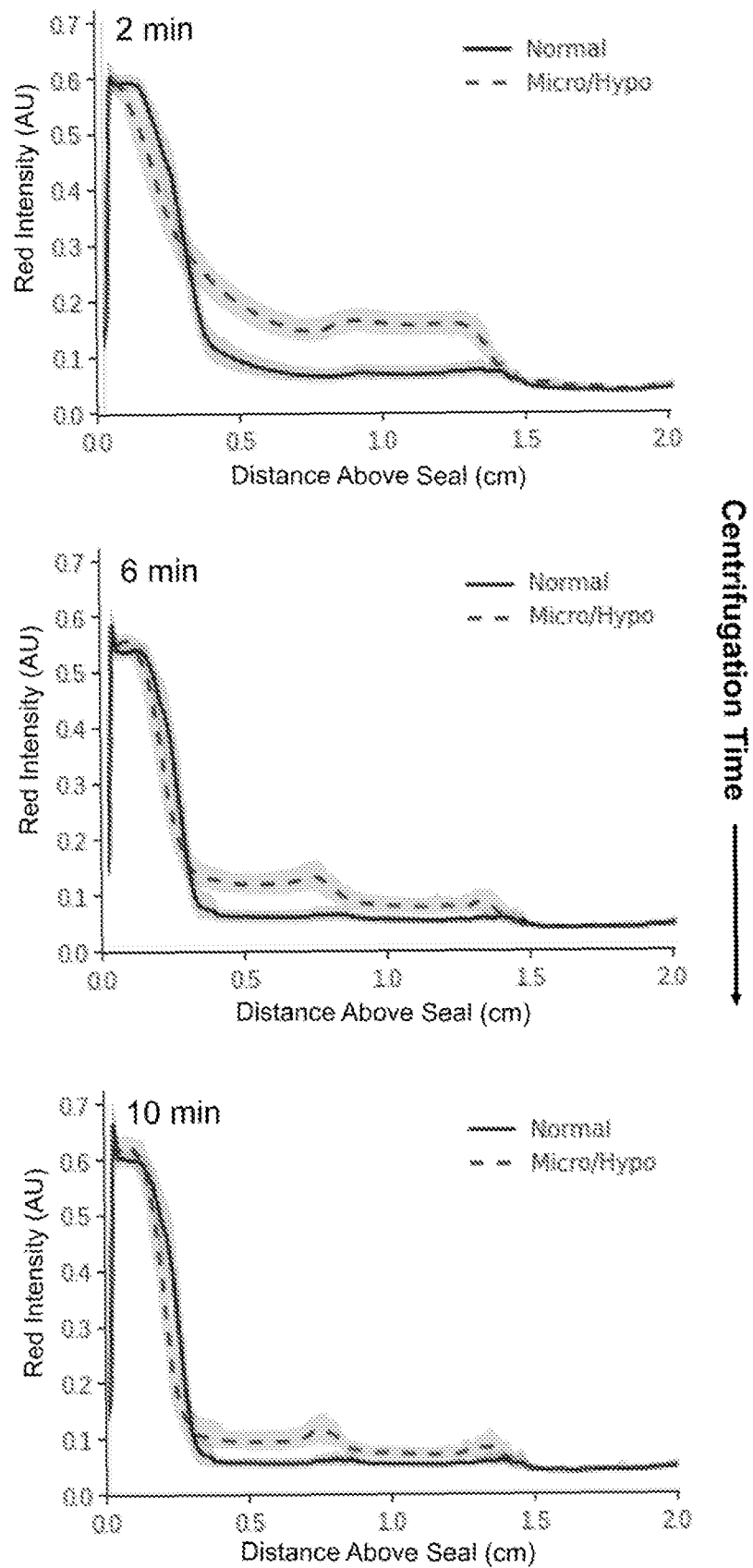


FIG. 5A

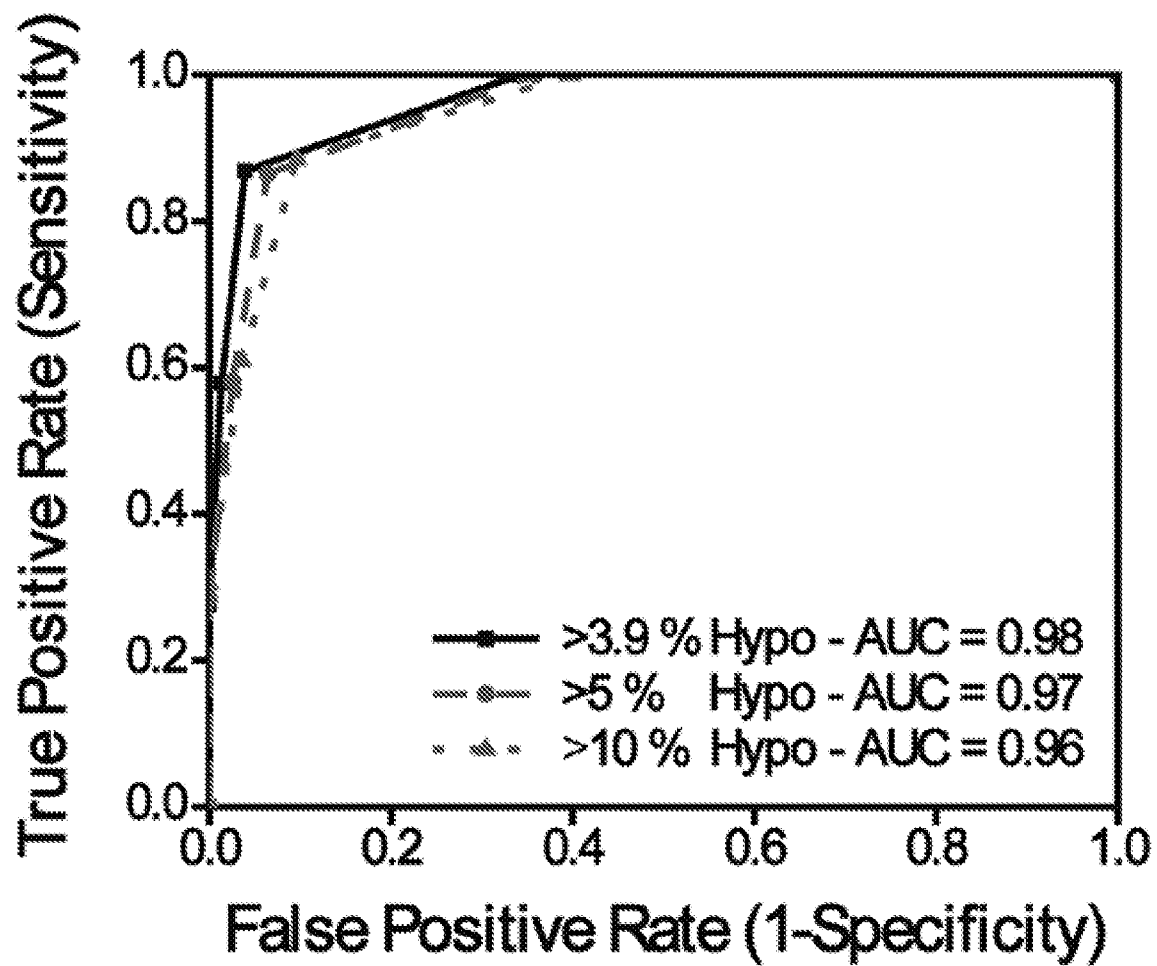
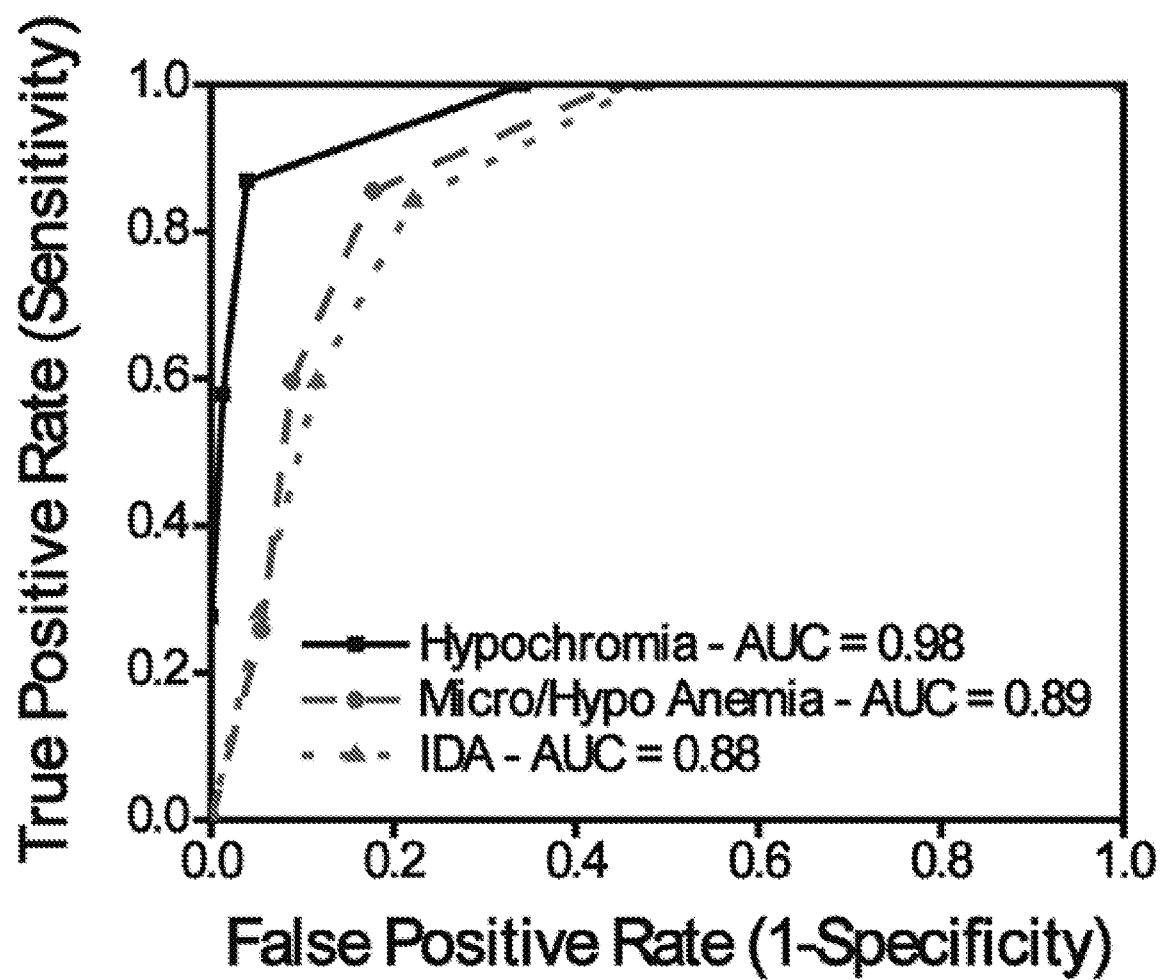
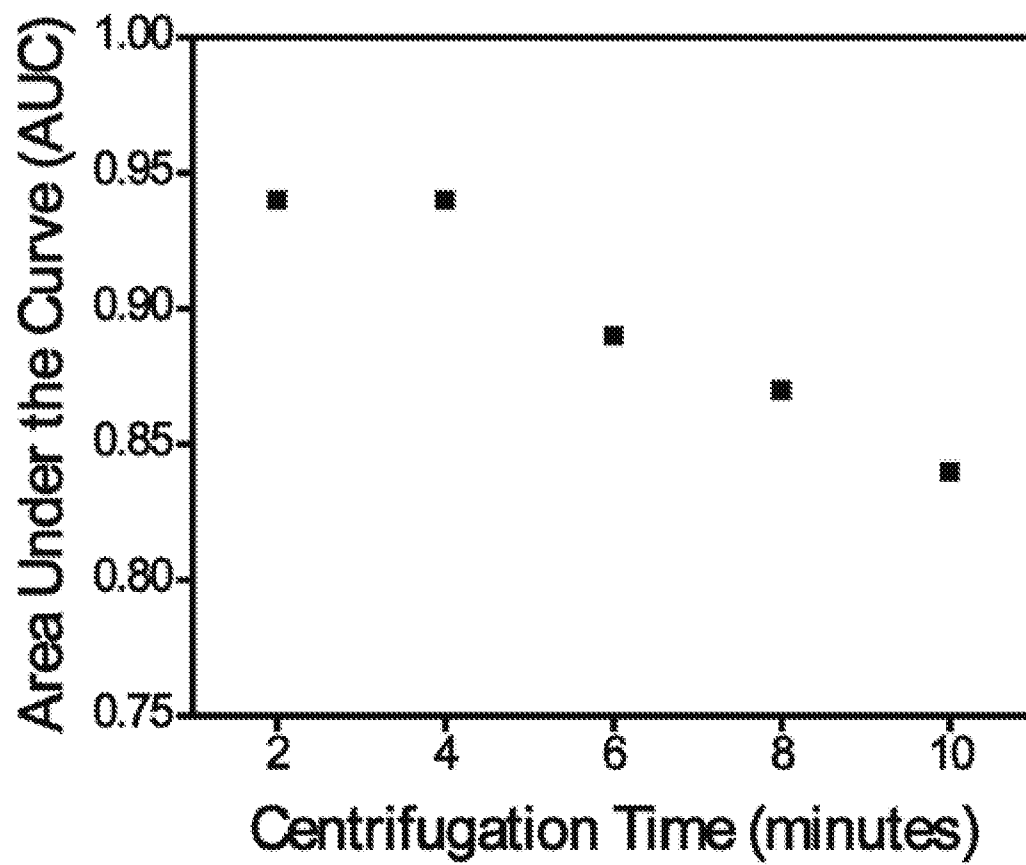


FIG. 5B



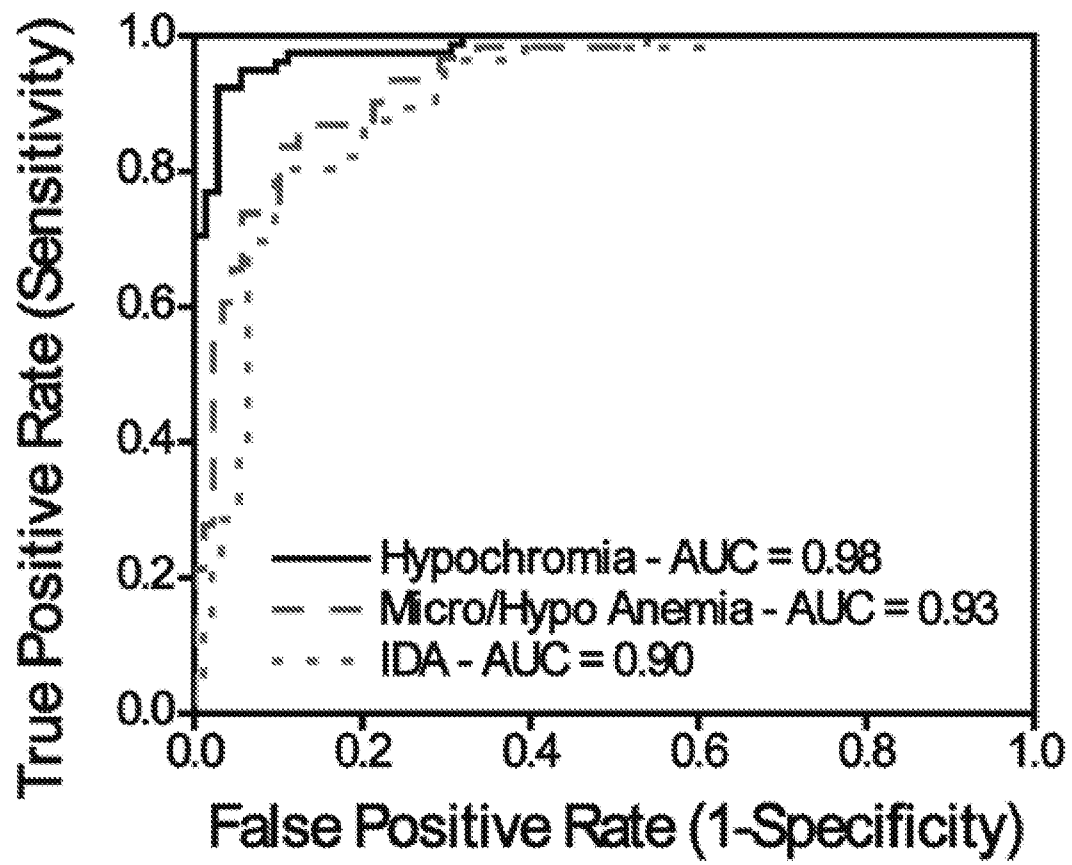
10/15

FIG. 6A



11/15

FIG. 6B



12/15

FIG. 7A

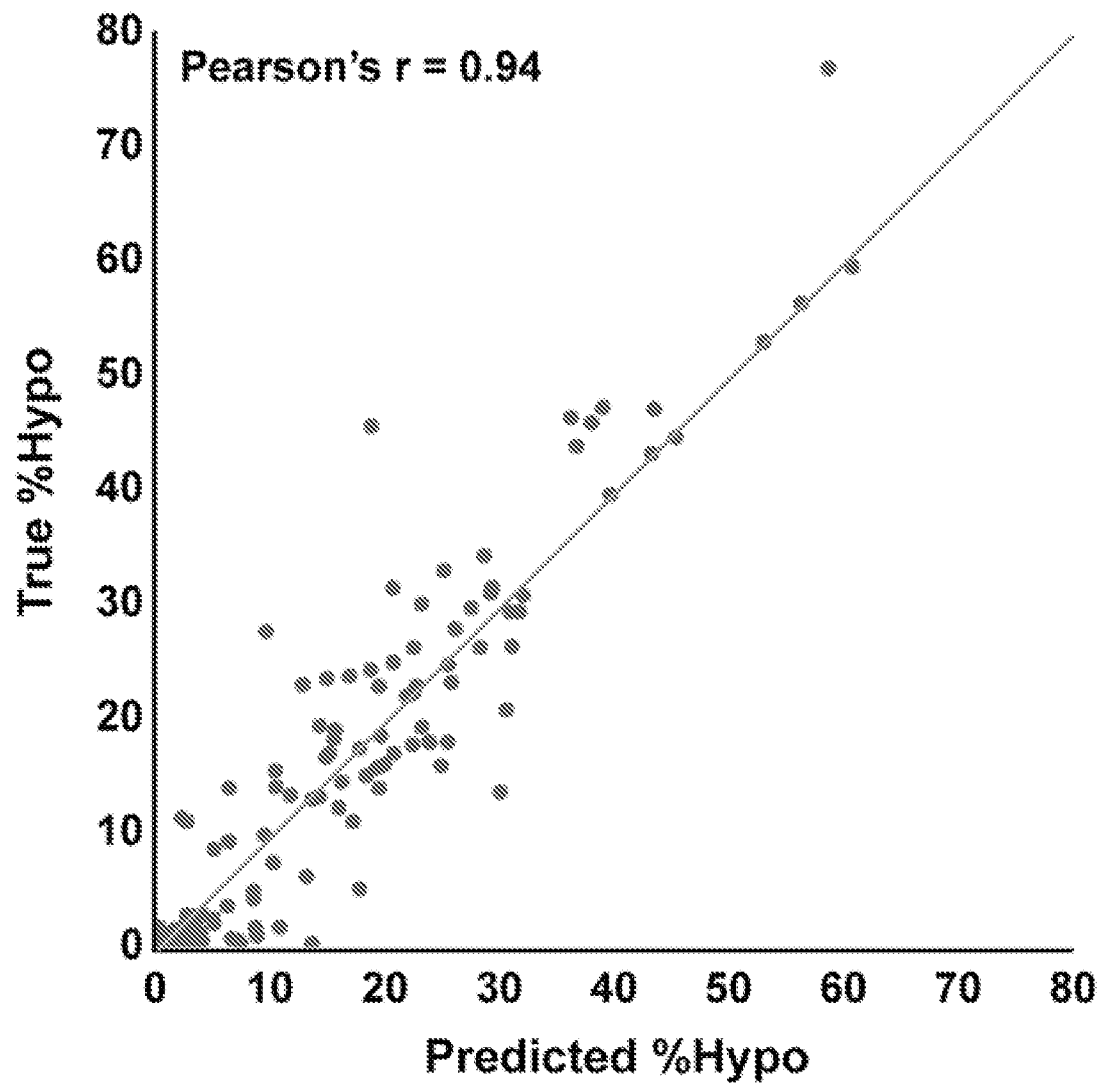


FIG. 7B

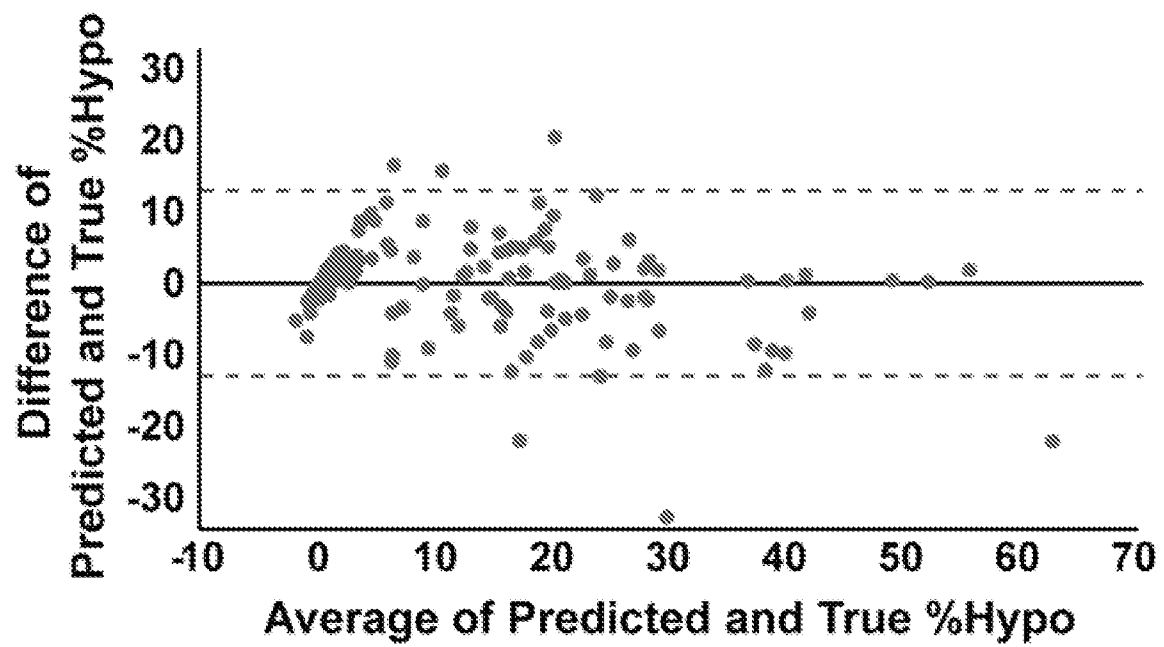


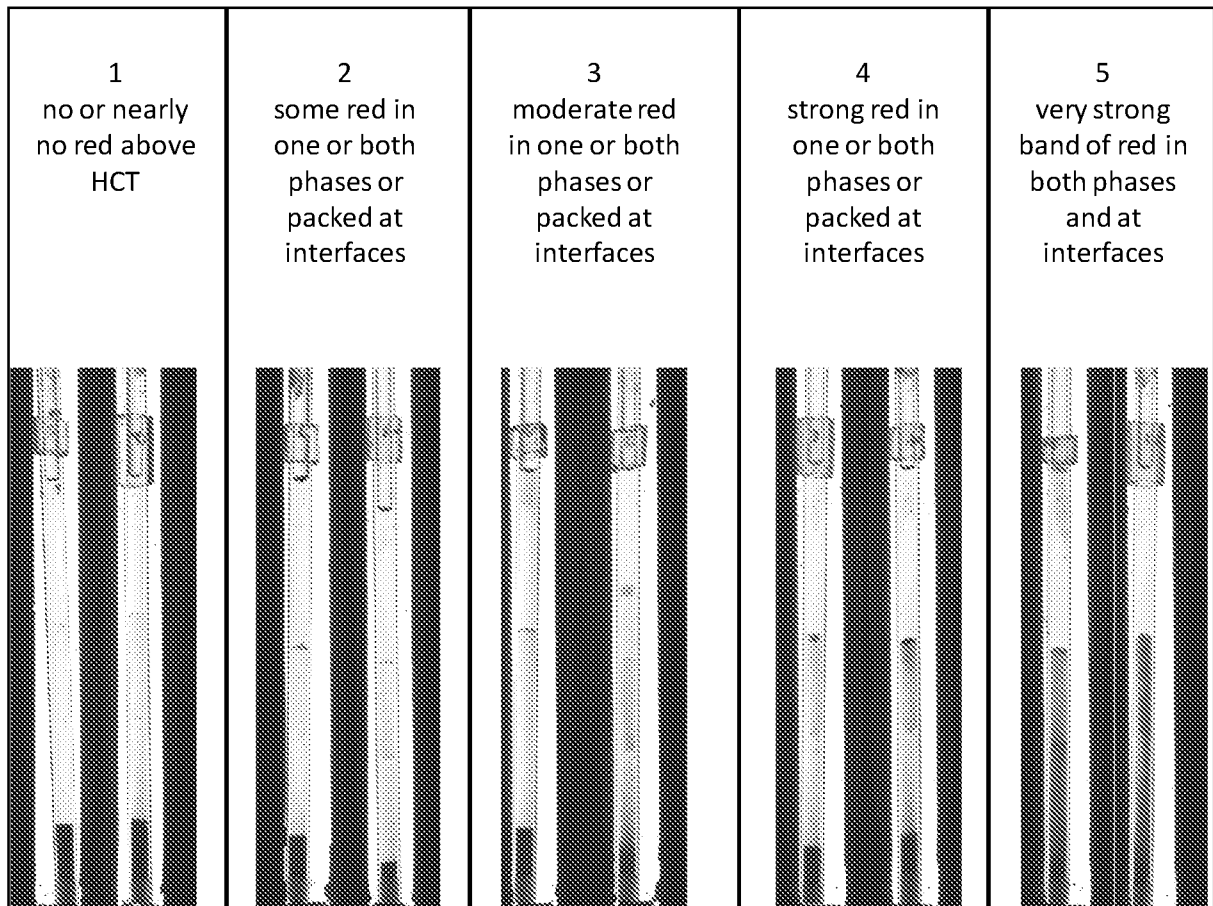
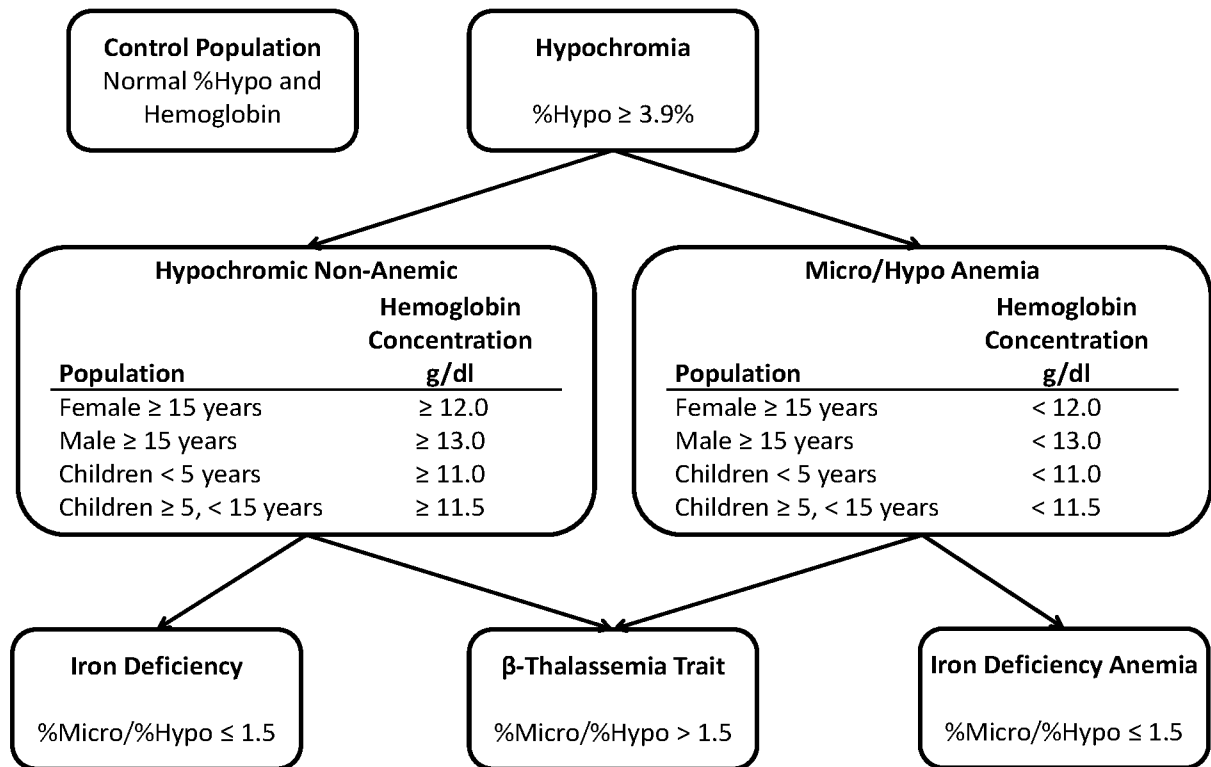
FIG. 8

FIG. 9



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/024446

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 17/00; G06F 19/00; G06F 19/10; G06F 19/24 (2016.01)

CPC - G06F 17/00; G06F 19/00; G06F 19/10; G06F 19/24 (2016.05)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC(8) - G06F 17/00; G06F 19/00; G06F 19/10; G06F 19/24 (2016.01)

CPC - G06F 17/00; G06F 19/00; G06F 19/10; G06F 19/24 (2016.05)

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

U.S. Classes - 702/1; 702/19; 702/22; 702/23 (keyword delimited)

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

Orbit, Google Patents, PubMed, Google.

Search terms used: system determine characteristic multi-phase biological analyte reader generate color distribution profile memory store algorithms assay conditions prediction machine learning

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2005/0151976 A1 (TOMA) 14 July 2005 (14.07.2005) entire document	1-4, 22-25
A	US 2003/0191376 A1 (SAMUELS et al) 09 October 2003 (09.10.2003) entire document	1-4, 22-25
A	US 6,872,298 B2 (KERMANI) 29 March 2005 (29.03.2005) entire document	1-4, 22-25
A	US 2002/0081232 A1 (LEWIS et al) 27 June 2002 (27.06.2002) entire document	1-4, 22-25
A	US 7,174,198 B2 (TROFIMOV et al) 06 February 2007 (06.02.2007) entire document	1-4, 22-25

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

31 May 2016

Date of mailing of the international search report

01 JUL 2016

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2016/024446

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 5-21, 26-38
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.