# Supporting Information for:

# Diagnosis of Iron Deficiency Anemia Using Density-based Fractionation of Red Blood Cells

Jonathan W. Hennek[1], Ashok A. Kumar[1], Alex B. Wiltschko,[2,3] Matthew Patton[1], Si Yi Ryan Lee[1], Carlo Brugnara[4], Ryan P. Adams[2], and George M. Whitesides[1,5*]

[1]Department of Chemistry and Chemical Biology

[2]School of Engineering and Applied Sciences

[3]Harvard Medical School, Department of Neurobiology

[4]Department of Laboratory Medicine, Boston Children's Hospital and Department of Pathology, Harvard Medical School

[5]Wyss Institute for Biologically Inspired Engineering

**Additional details of image analysis and machine learning**

The general problem of predicting continuously-varying outcomes from data is called "regression", and predicting classes, or labels, from data is called "classification".[1,2] Here we apply computer vision to extract summary data from scanned images of IDA-AMPS tests, and standard machine learning techniques to the classification problem of distinguishing micro/hypo anemic samples from normal samples and the regression problem of predicting continuously-varying red blood cell indices from images of the IDA-AMPS test.

The scanned images of groups of 12 IDA-AMPS tests were first segmented and aligned from the raw color scan image, resulting in cropped images of individual tubes. We then standardized the data by calculating the red intensity per pixel of each image, and then calculated the mean across the width of the tube, reducing each image into a single vector of 180 dimensions, representing the red intensity of the test at a 180 sequential distances from the test's plug. All computer vision was performed using the Python programming language, using the SciPy and NumPy libraries.

The resulting 180-dimensional red intensity summary of each test's original image was further dimensionally reduced to 30 dimensions using principal components analysis (PCA) to make further analysis more tractable, and to simultaneously remove redundant information (e.g. nearby pixels that were highly correlated).

To automate distinguishing normal from micro/hypo anemic patients based on the IDA-AMPS test, we use logistic regression[1,3] as a machine learning algorithm to discriminate between tests belonging to different patient types. Logistic regression is an algorithm which, given a training set of input data (here, our 30-dimensional representation of each test image), and associated labels (here, whether the patient was anemic or normal), learns a function mapping the input data into a probability of the test sample having come from a patient from either the anemic or normal group. We fit the logistic regression model using L2-

regularization using scikit-learn, guarding against over-fitting by repeated random resampling with replacement of our training set, and fully holding out a test set during all cross-validation.[4] A training set and validation set are randomly sampled 500 times and the average performance across all sampled validation sets is tabulated. In the case of classification, performance is defined as the AUC of the resulting discriminator.

To predict red blood cell indices, we use a different, but equally standard, algorithm called support vector regression (SVR) with a radial basis function kernel.[3] The goal of SVR is to produce a mapping from our 30-dimensional representation of an IDA-AMPS test to one continuous red blood cell index, which we fit in the Python language using the scikit-learn package. We define performance in the regression task using the same cross-validation approach as described above, but with performance defined as 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. The Python code used in this manuscript can be obtained from a GitHub repository by contacting the first author.

## Differentiating IDA from β-Thalassemia Trait (β-TT)

IDA is characterized by microcytic hypochromic red blood cells caused by low levels of iron in the blood and bone marrow. Diagnostic biomarkers for IDA include serum iron, transferrin saturation, and ferritin. In several clinical conditions, however, these biomarkers do not change rapidly enough to reflect actual iron concentration.[2] Recently, a variety of red blood cell indices have found favor in clinical diagnoses including mean corpuscular volume (MCV), mean corpuscular hemoglobin concentration (MCHC), red blood cell distribution width (RDW), and reticulocyte hemoglobin content (CHr).[5] Relative to healthy patients, patients with IDA have low values of MCV, MCHC, and CHr, and a larger RDW. Beta-thalassemia minor (i.e. β-thalassemia trait: β-TT) is a benign genetic disorder that presents a potentially confounding diagnosis to IDA because both conditions result in microcytic and

hypochromic red blood cells. Identification of β-TT is vital to aid in prevention of β-thalessemia major through genetic counseling; transmission of β-thalessemia is autosomal recessive and so heterozygous carriers can attempt to mitigate the risk of a homozygous offspring. Due to the characteristically smaller (bimodal) RDW and larger total red blood cell count for patients with β-TT, it can be distinguished from IDA.[5] A population of cells with a larger RDW will likely have a larger distribution in the density of red blood cells. A diagnostic designed to quantify the density and distribution in density of red blood cells should, in principle, be able to differentiate IDA.

We evaluated the predictive ability of our test to differentiate β-TT and IDA using a small data set from IDA (n = 5 and 57) and found, using digital analysis, an AUC of 0.41, effectively a random guess. This analysis suggests that our test may not be able to distinguish between β-TT and IDA (although a larger sample size of β-TT patients is required for a more definitive conclusion). The inability of a test to differentiate β-TT and IDA could limit its value in regions where the prevalence of β-TT is highest; India, Thailand, and Indonesia account for nearly 50% of the global burden of β-TT.[6] Although we have no specifically identified samples with alpha-thalassemia minor in our study, a similar concern applies, and would be relevant in South East Asia and China.

A subset of our patients classified as IDA based on %Hypo and HGB, had MCV and MCH values in the normal range (80 - 100 $\mu m^3$, and 26 - 34 pg/cell). We evaluated the impact of these samples on the diagnostic capabilities of the IDA-AMPS test using visual analysis from blinded readers. If the normal MCV/MCH samples were excluded from the analysis (14 samples out of 57), the AUC was nearly unchanged—0.89 versus 0.88 when these samples are not excluded.

**Statistical Methods**

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 (Cardillo G. [2008] ROC curve: compute a Receiver Operating Characteristic curve). Lin's concordance correlation coefficient[7] 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**

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" (Figure SX). 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**

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-Alrich), 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**

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.[8]


**The bottle of PVA we use (lot # A0324406) is labeled as "MW = 2 kD, 75% hydrolized."
We discovered, however, that this bottle was mislabeled by the supplier and is actually MW
= 6 kD, 78% hydrolized.

**Figure S1.** Flow chart illustrating 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).

**Control Population**
Normal %Hypo and Hemoglobin

**Hypochromia**
%Hypo ≥ 3.9%

**Hypochromic Non-Anemic**

| Population | Hemoglobin Concentration g/dl |
| --- | --- |
| Female ≥ 15 years | ≥ 12.0 |
| Male ≥ 15 years | ≥ 13.0 |
| Children < 5 years | ≥ 11.0 |
| Children ≥ 5, < 15 years | ≥ 11.5 |

**Micro/Hypo Anemia**

| Population | Hemoglobin Concentration g/dl |
| --- | --- |
| Female ≥ 15 years | < 12.0 |
| Male ≥ 15 years | < 13.0 |
| Children < 5 years | < 11.0 |
| Children ≥ 5, < 15 years | < 11.5 |

**Iron Deficiency**
%Micro/%Hypo ≤ 1.5

**β-Thalassemia Trait**
%Micro/%Hypo > 1.5

**Iron Deficiency Anemia**
%Micro/%Hypo ≤ 1.5

**Table S1.** Hemoglobin concentration thresholds used to define anemia in our study.[9]

| Population | Hemoglobin Concentration (HGB) g/dL |
|---|---|
| Female ≥ 15 years | < 12.0 |
| Male ≥ 15 years | < 13.0 |
| Children < 5 years | < 11.0 |
| Children ≥ 5, < 15 years | < 11.5 |

**Table S2.** 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 |

**Table S3.** Lin's concordance correlation coefficients assessing inter- and intra- reader concordance for the visual analysis of the IDA-AMPS test after two minutes centrifugation.[7] Inter-reader $\rho_c$ was assessed by averaging the $\rho_c$ between each pair of readers.

| | Lin's Concordance Correlation Coefficient |
| --- | --- |
| | $\rho_c$ (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) |

**Figure S2.** A. Machine learning results for mean corpuscular hemoglobin concentration (MCHC) compared to a hematology analyzer and B. Bland-Altman plot showing the agreement between true and predicted MCHC (n = 152). In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting. Error bars represent the 99% bootstrap confidence intervals for 500 validation sets.

## A.
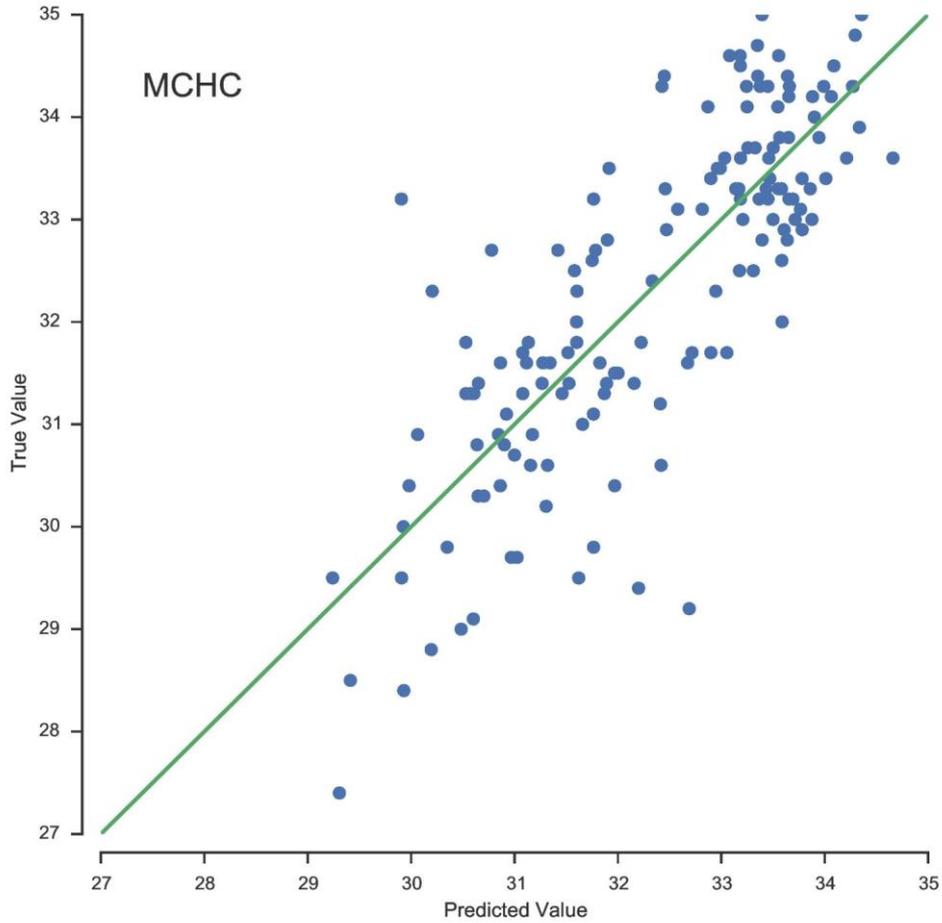


## B.
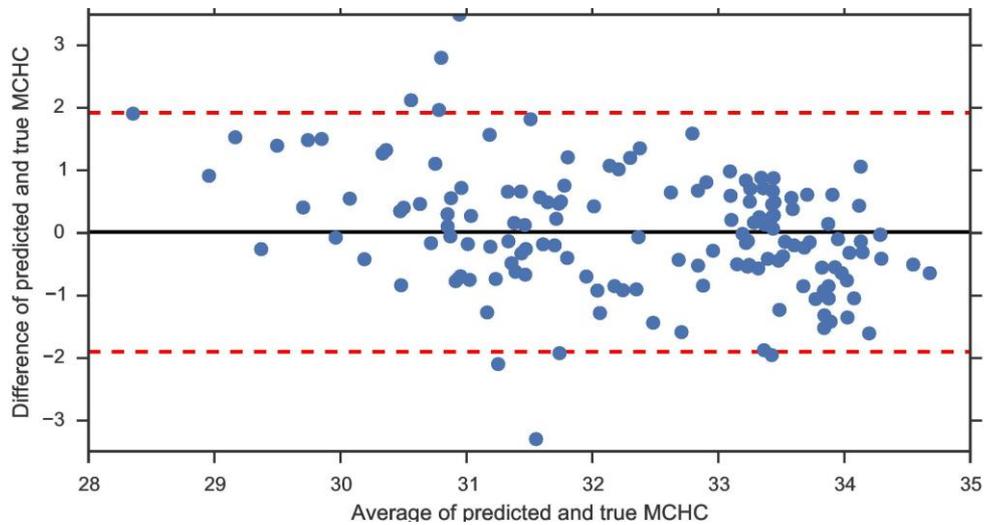
**Figure S3.** A. Machine learning results for corpuscular hemoglobin concentration (CH) compared to a hematology analyzer and B. Bland-Altman plot showing agreement between true and predicted CH (n = 152). In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting.

**A.**
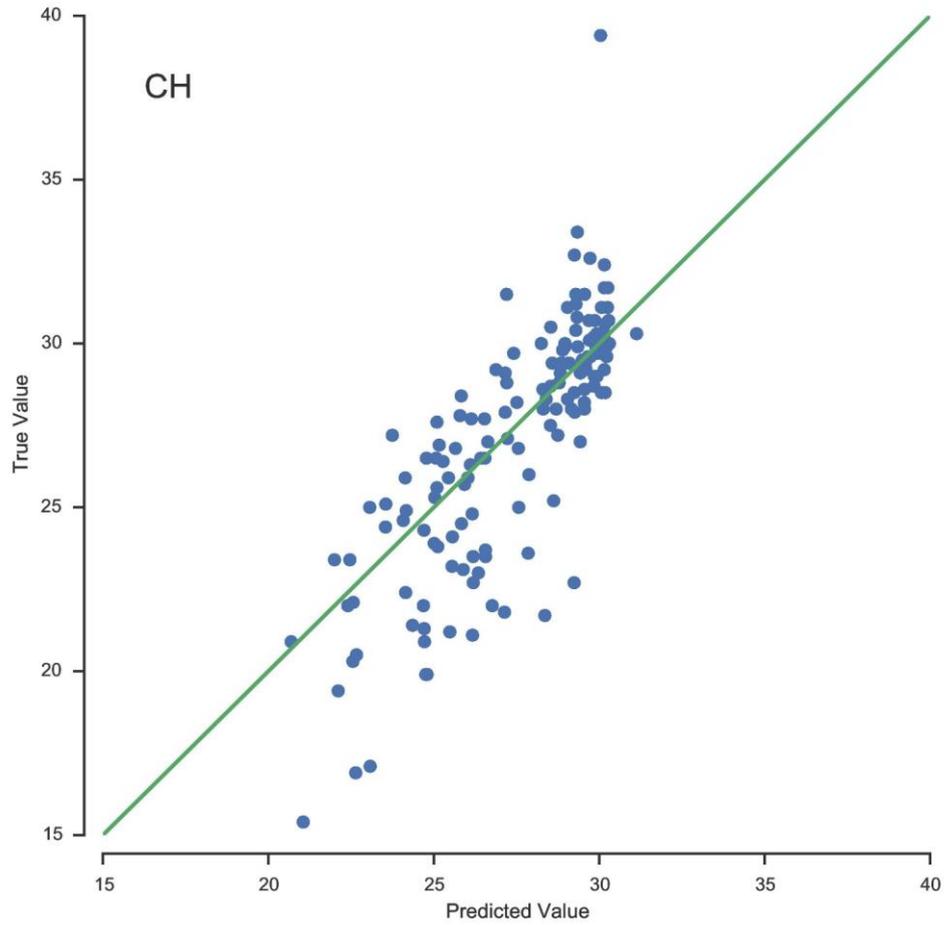


**B.**
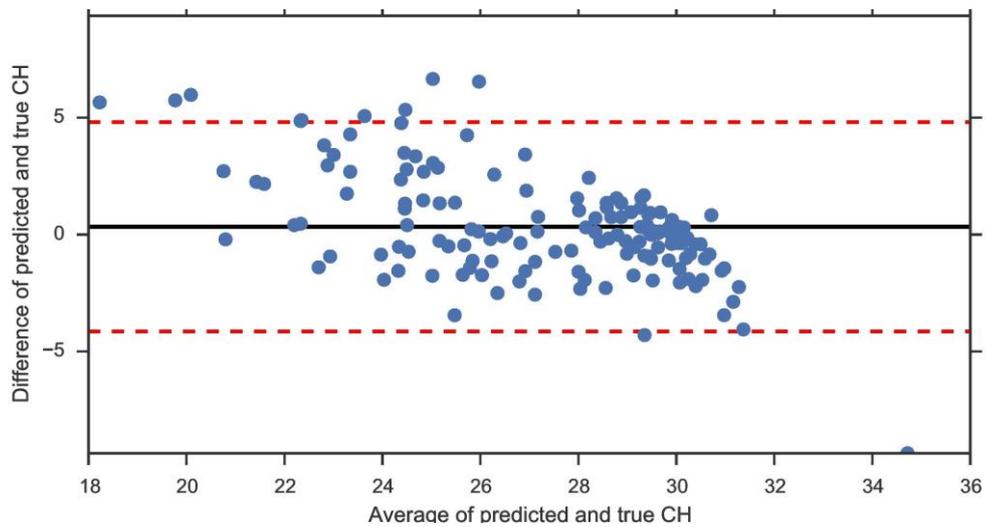
**Figure S4.** A. Machine learning results for hemoglobin concentration (HGB) compared to a hematology analyzer and B. Bland-Altman plot showing the agreement between true and predicted HGB (n = 152). In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting.

**A.**
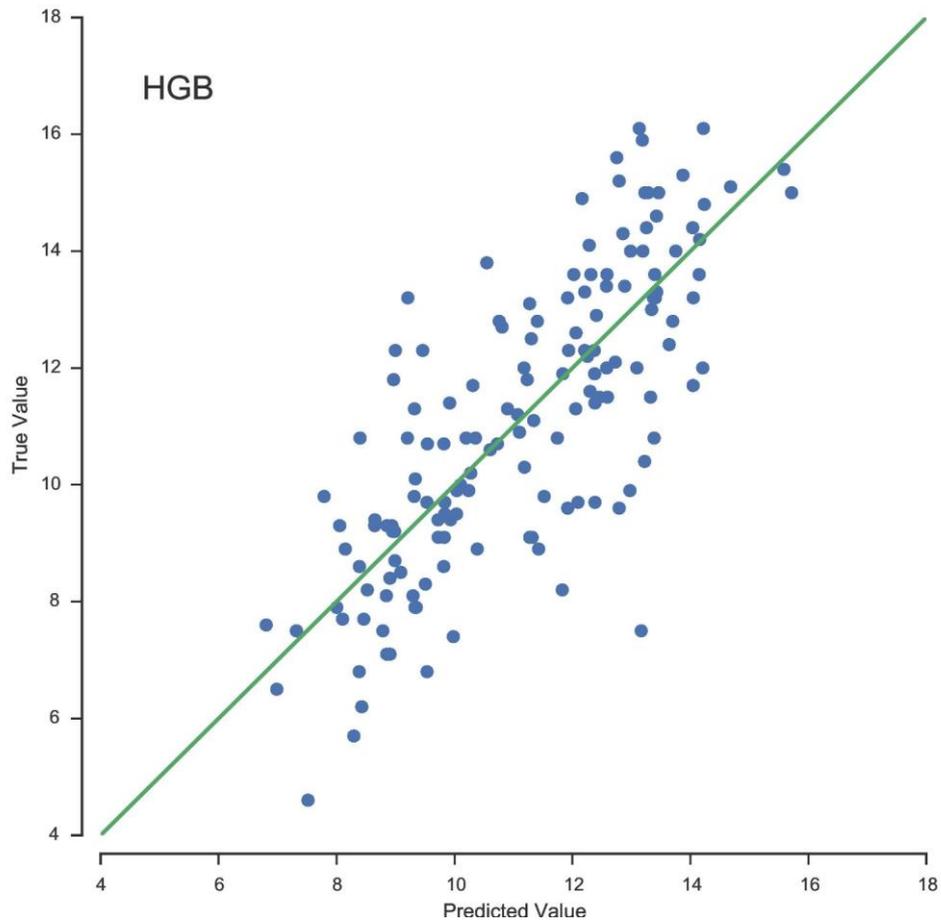


**B.**
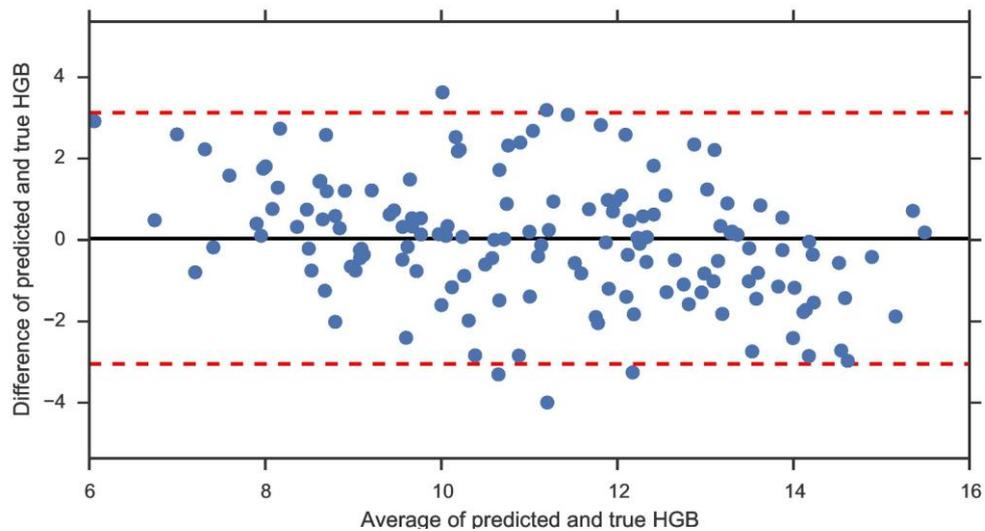
**Figure S5.** A. Machine learning results for hematocrit (HCT) compared to a hematology analyzer and B. Bland-Altman plot showing the agreement between true and predicted HCT (n = 152). In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting.
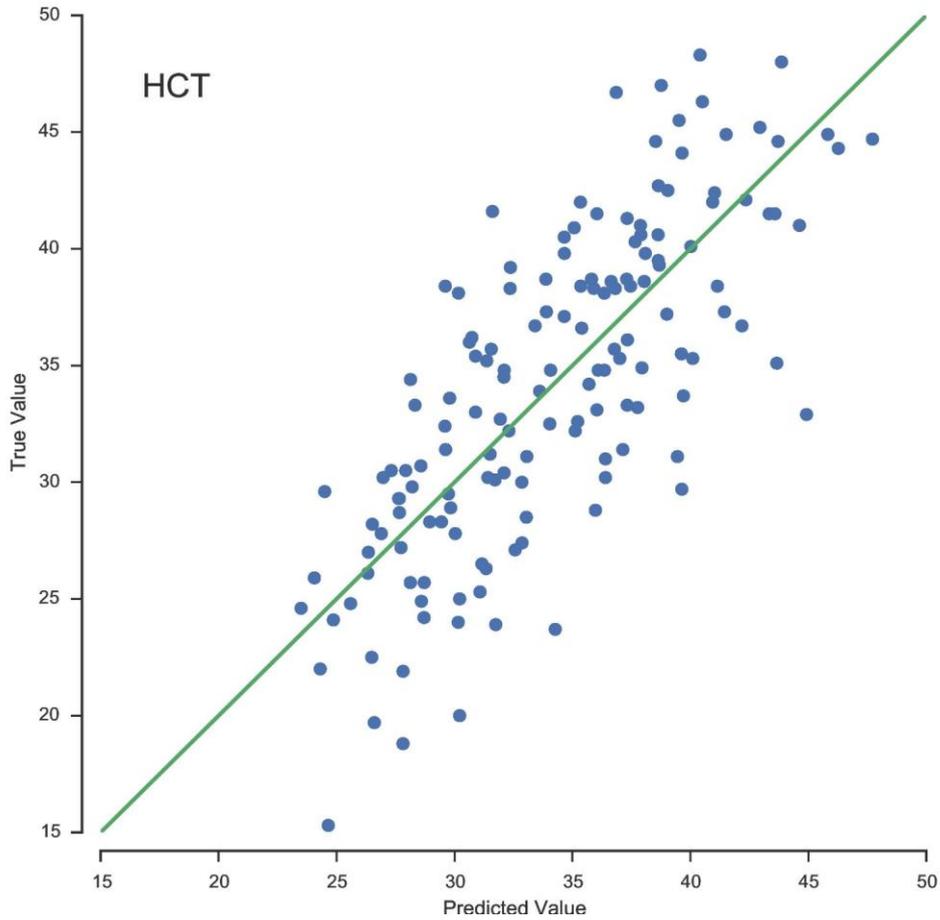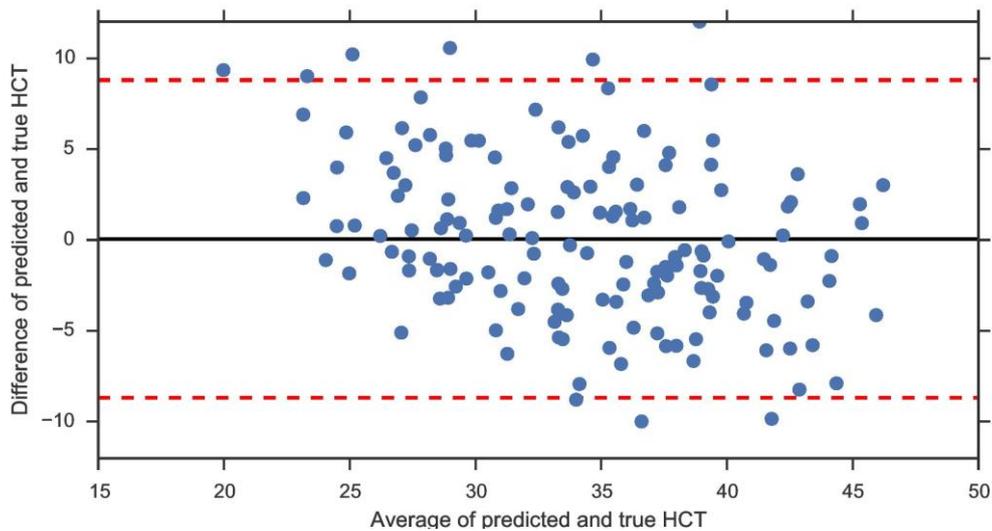
**Figure S6.** A. Machine learning results for corpuscular hemoglobin (MCH) compared to a hematology analyzer and B. Bland-Altman plot showing the agreement between true and predicted MCH (n = 152). In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting.

**A.**
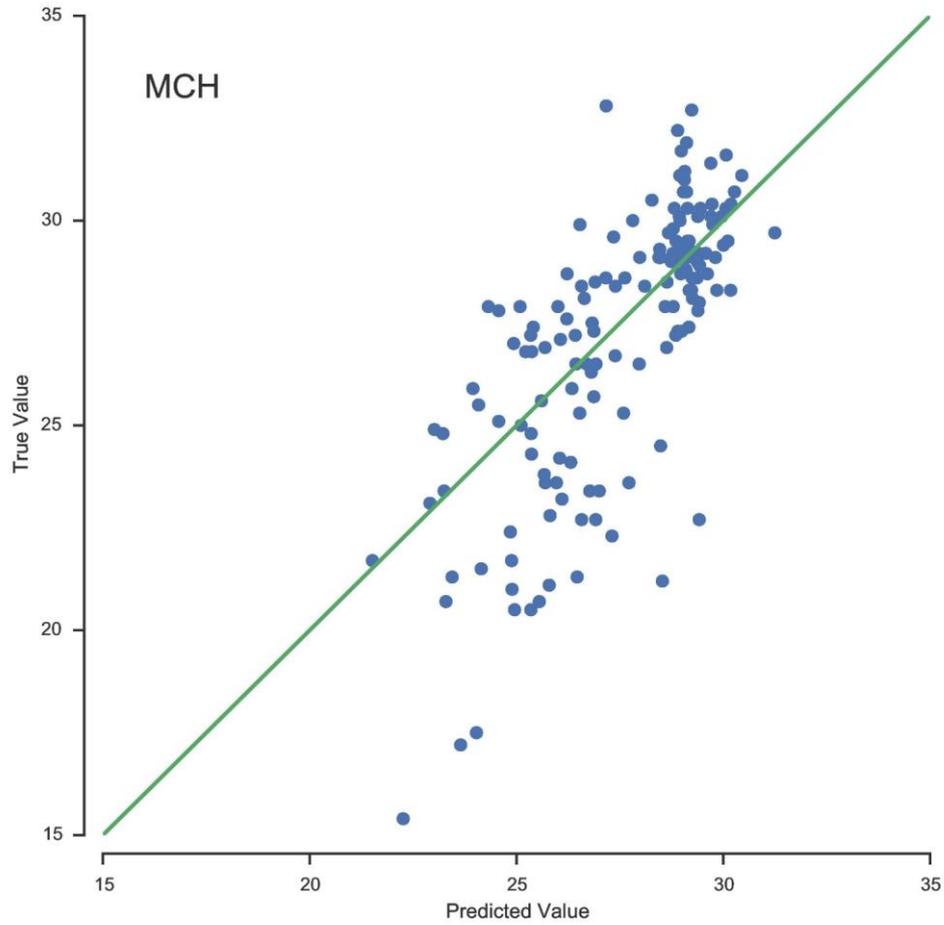


**B.**
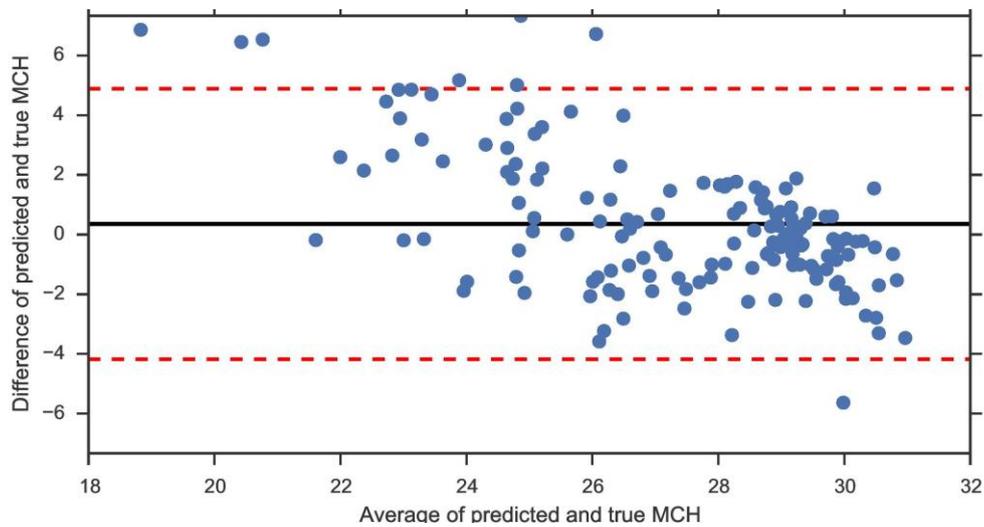
**Figure S7.** A. Machine learning results for red blood cell distribution width (RDW) compared to a hematology analyzer and B. Bland-Altman plot showing the agreement between true and predicted RDW (n = 152).  In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting.

**A.**
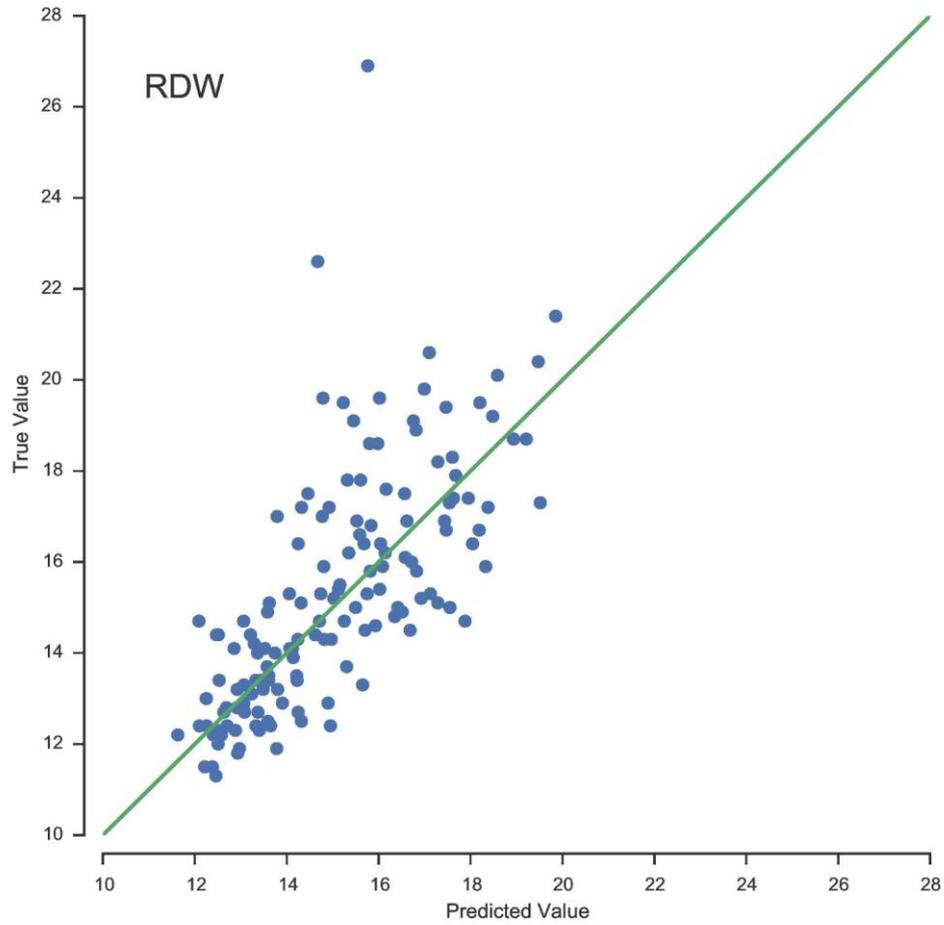


**B.**
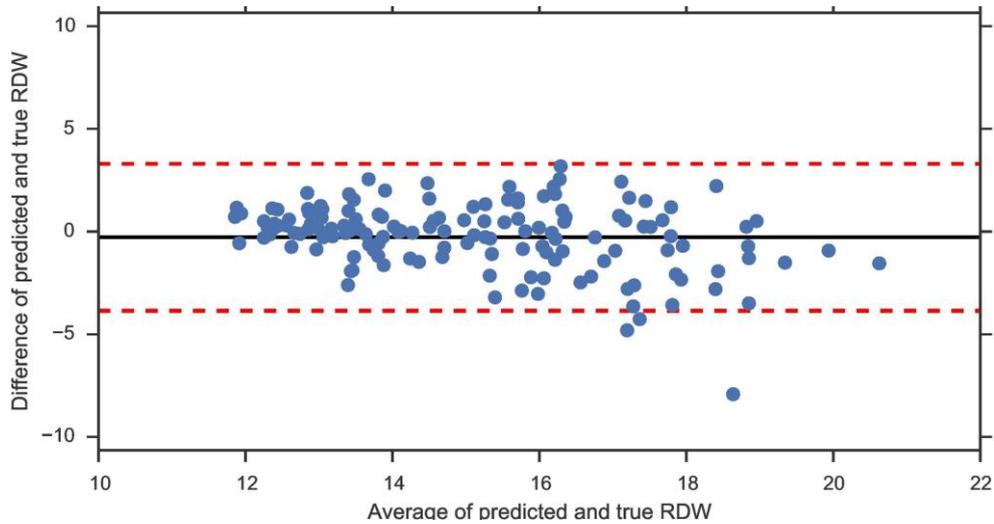
**Figure S8.** A. Machine learning results for hemoglobin distribution width (HDW) compared to a hematology analyzer and B. Bland-Altman plot showing the agreement between true and predicted HDW (n = 152). In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting.

**A.**
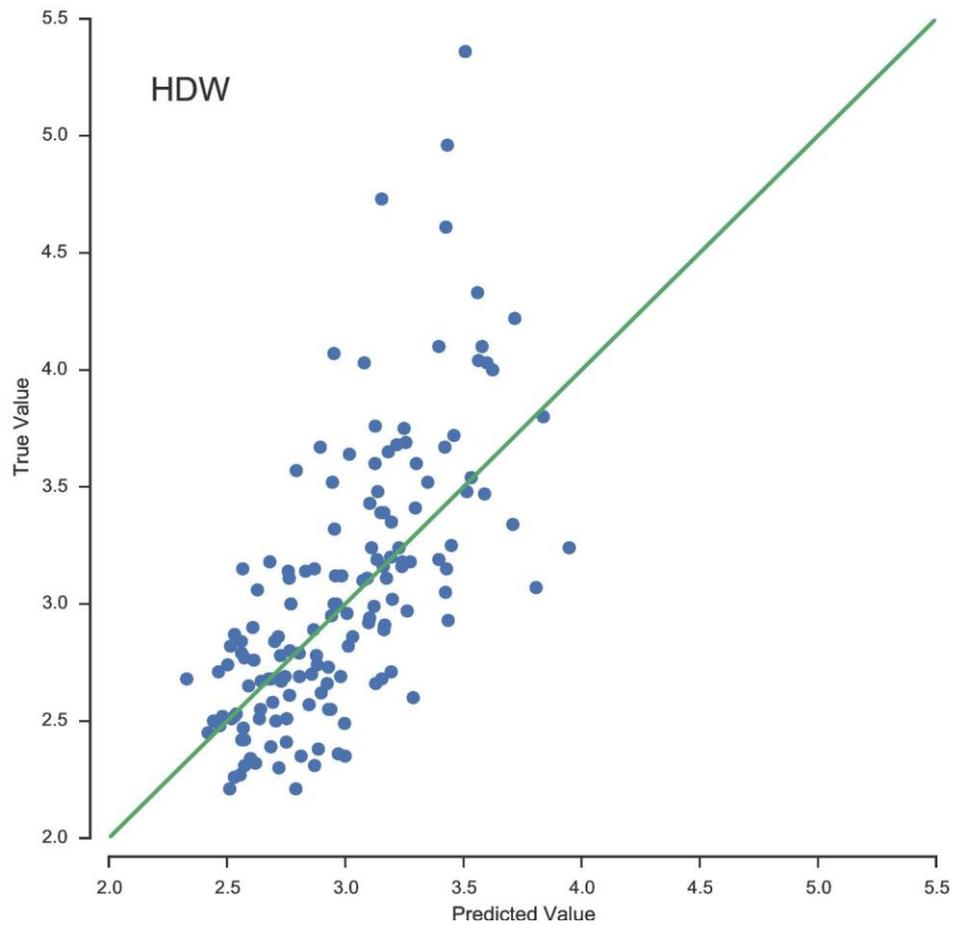


**B.**
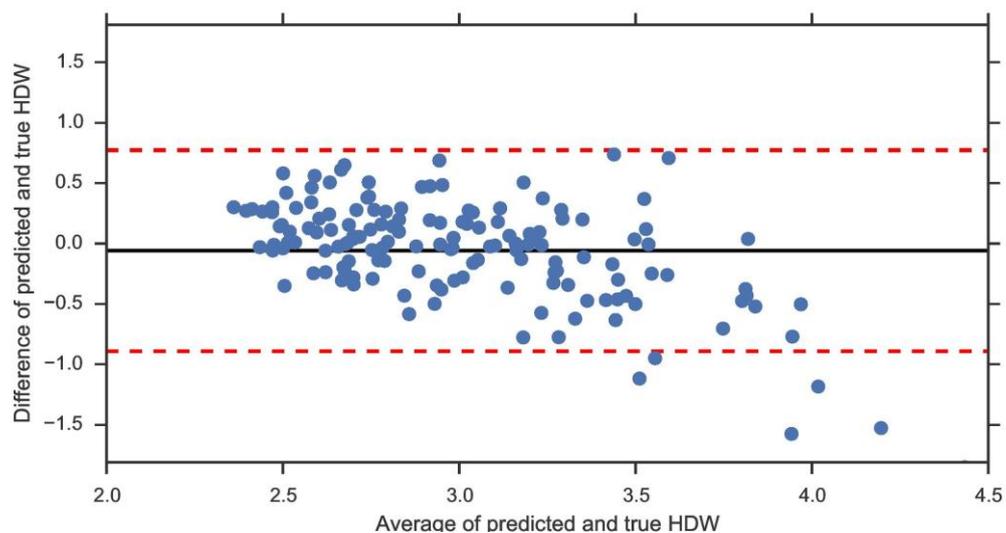
**Figure S9.** A. Machine learning results for percent microcytic red blood cells (%Micro) compared to a hematology analyzer and B. Bland-Altman plot showing the agreement between true and predicted %Micro (n = 152).  In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting.
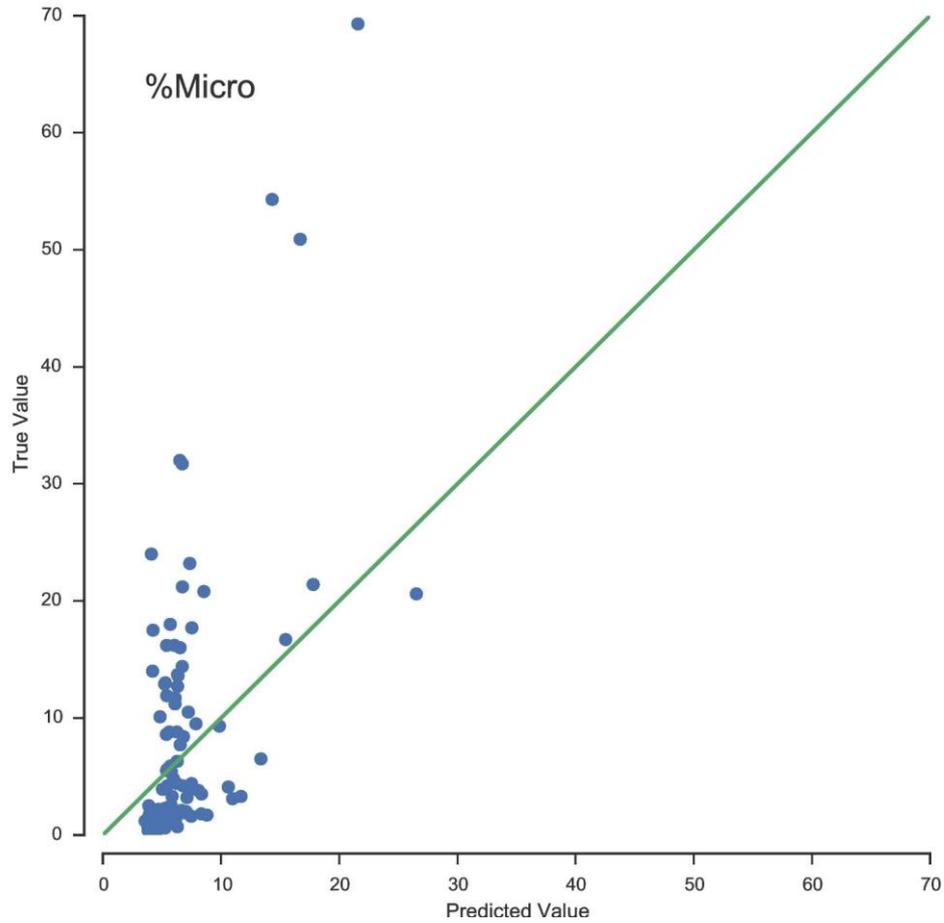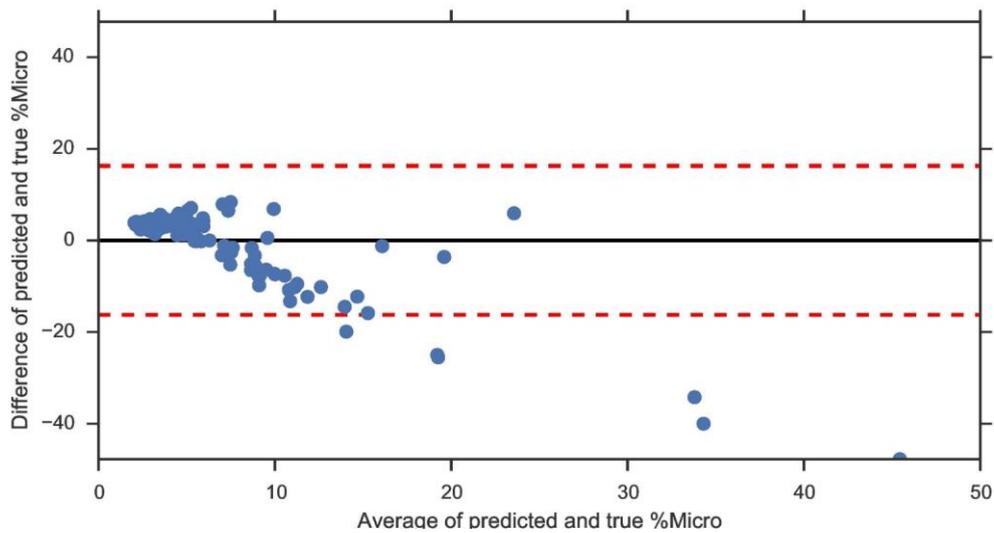
**Figure S10.** A. Machine learning results for percent microcytic versus hypochromic red blood cells (%Micro/%Hypo) compared to a hematology analyzer and B. Bland-Altman plot showing the agreement between true and predicted %Micro/%Hypo)(n = 152). In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting. .
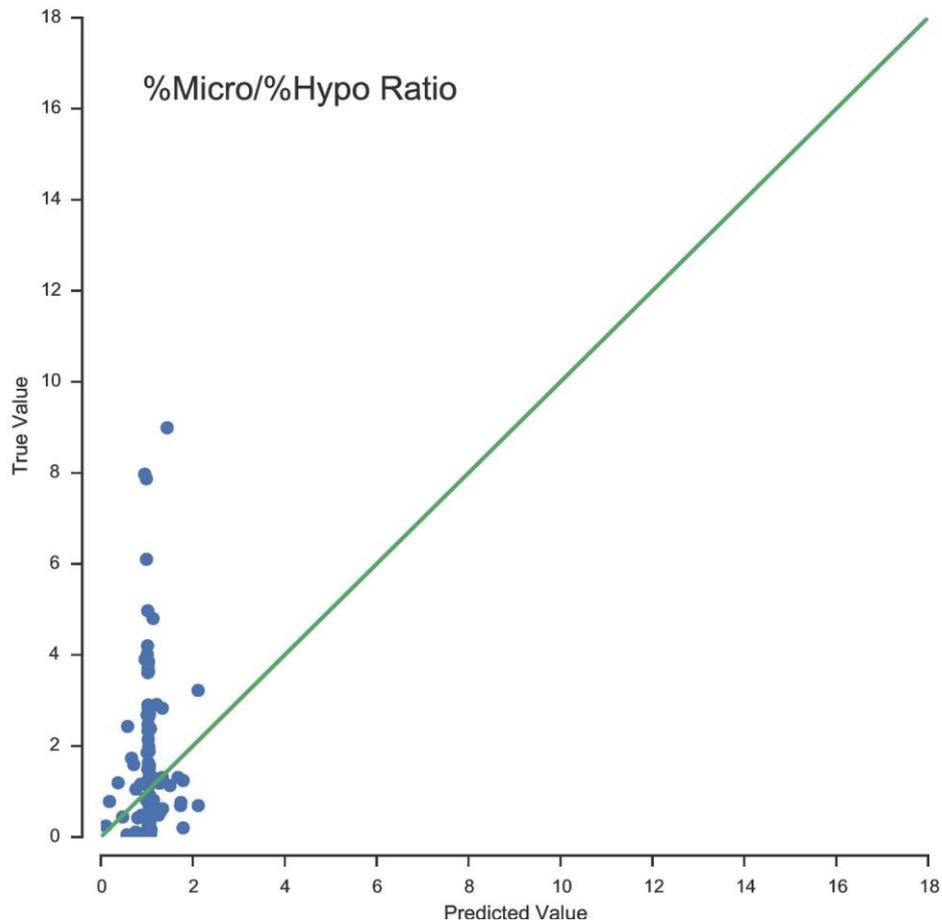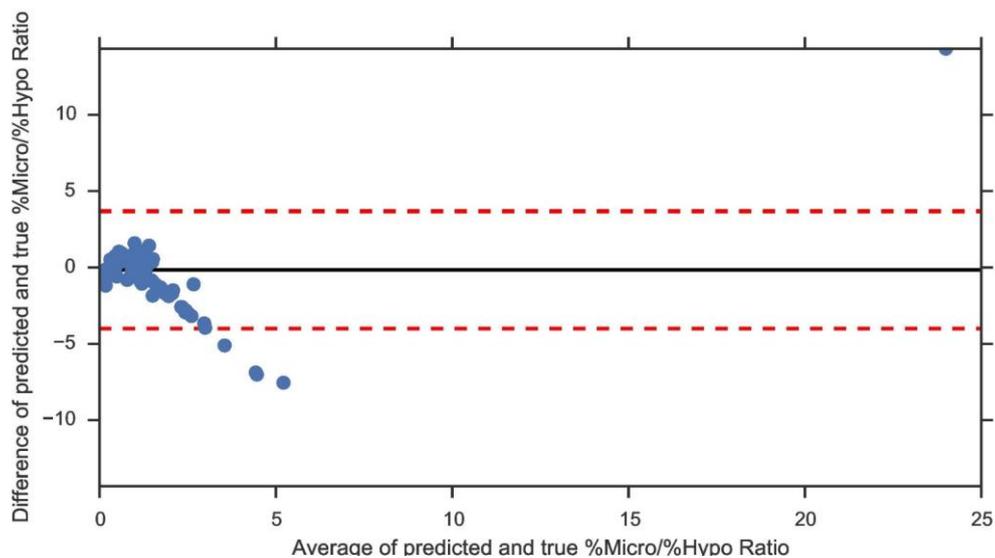
**Figure S11.** A. Machine learning results for the number of red blood cells (RBC) compared to a hematology analyzer and B. Bland-Altman plot showing the agreement between true and predicted RBC (n = 152). In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting.

**A.**
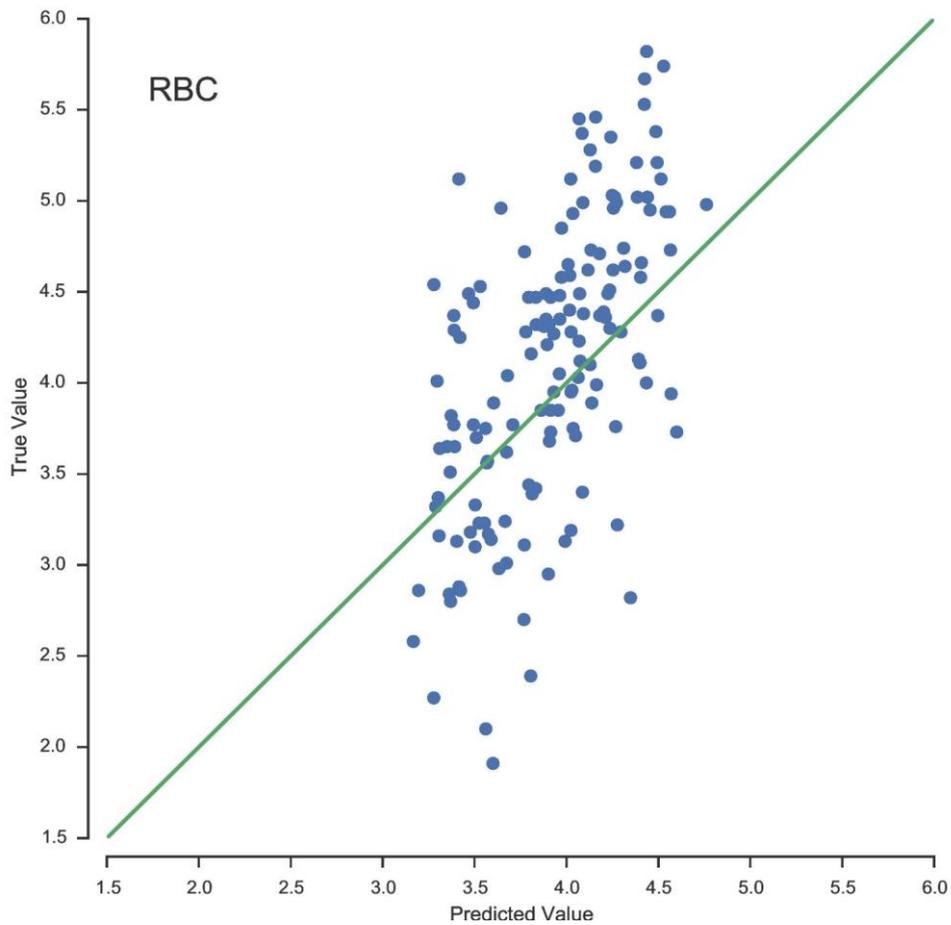


**B.**
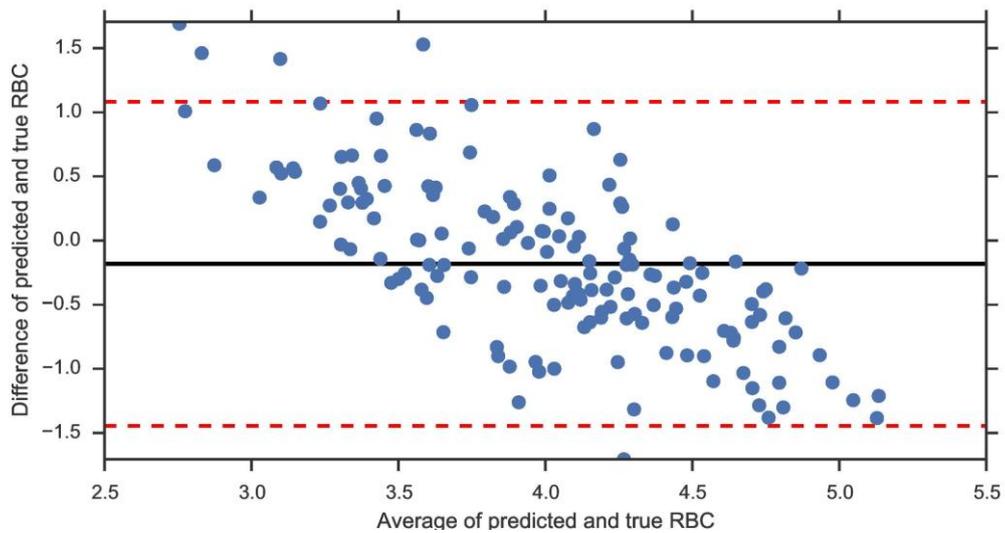
**Figure S12.** A. Machine learning results for percent hypochromic red blood cells (%Hypo) compared to a hematology analyzer and B. Bland-Altman plot showing the agreement between true and predicted %Hypo (n = 152). In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting.

**A.**
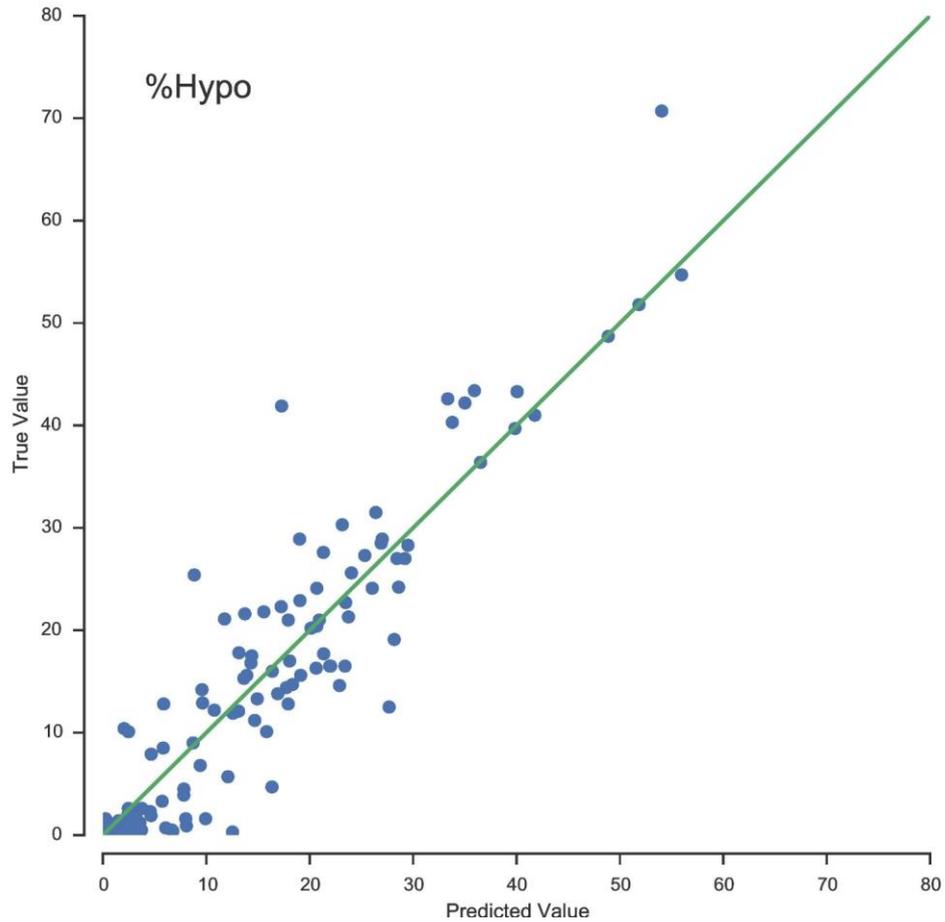


**B.**
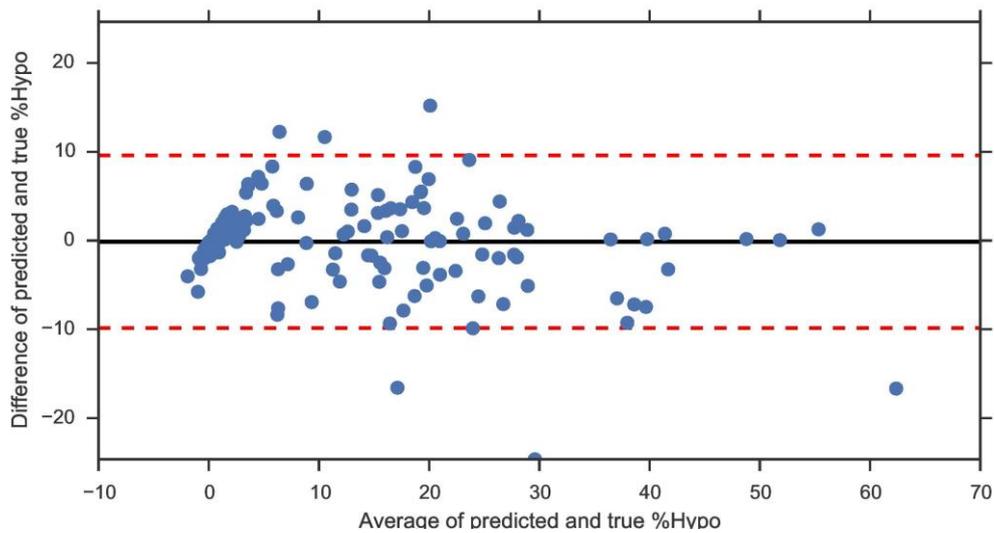
**Figure S13.** A. Machine learning results for mean corpuscular volume (MCV) compared to a hematology analyzer and B. Bland-Altman plot showing the agreement between true and predicted MCV (n = 152). In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting.

**A.**
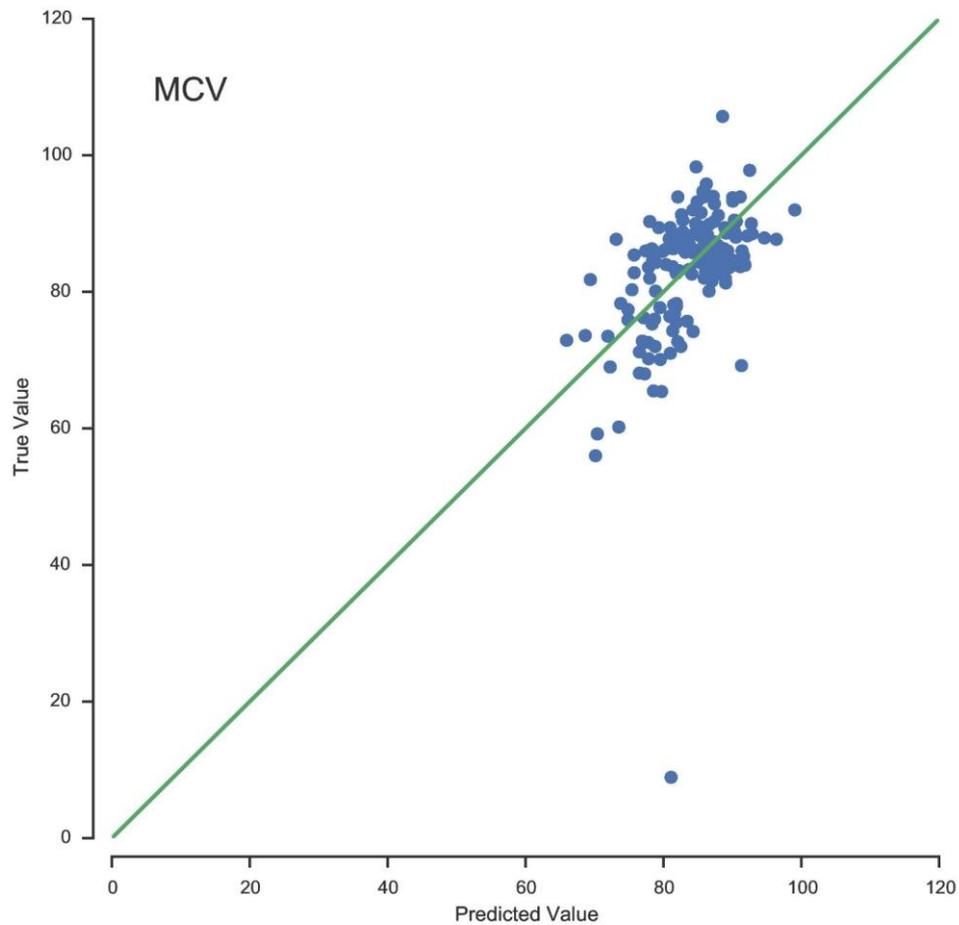


**B.**
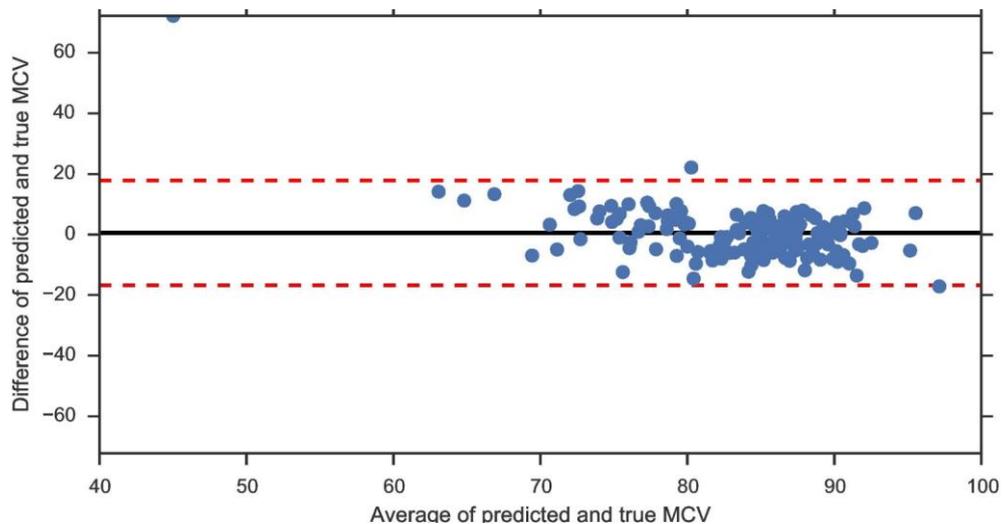
**Figure S14.** A. Machine learning results for percent macrocytic red blood cells (%Macro) compared to a hematology analyzer and B. Bland-Altman plot showing the agreement between true and predicted %Macro (n = 152). In both cases repeated random sub-sampling validation (n = 500) was used to guard against over-fitting.

**A.**
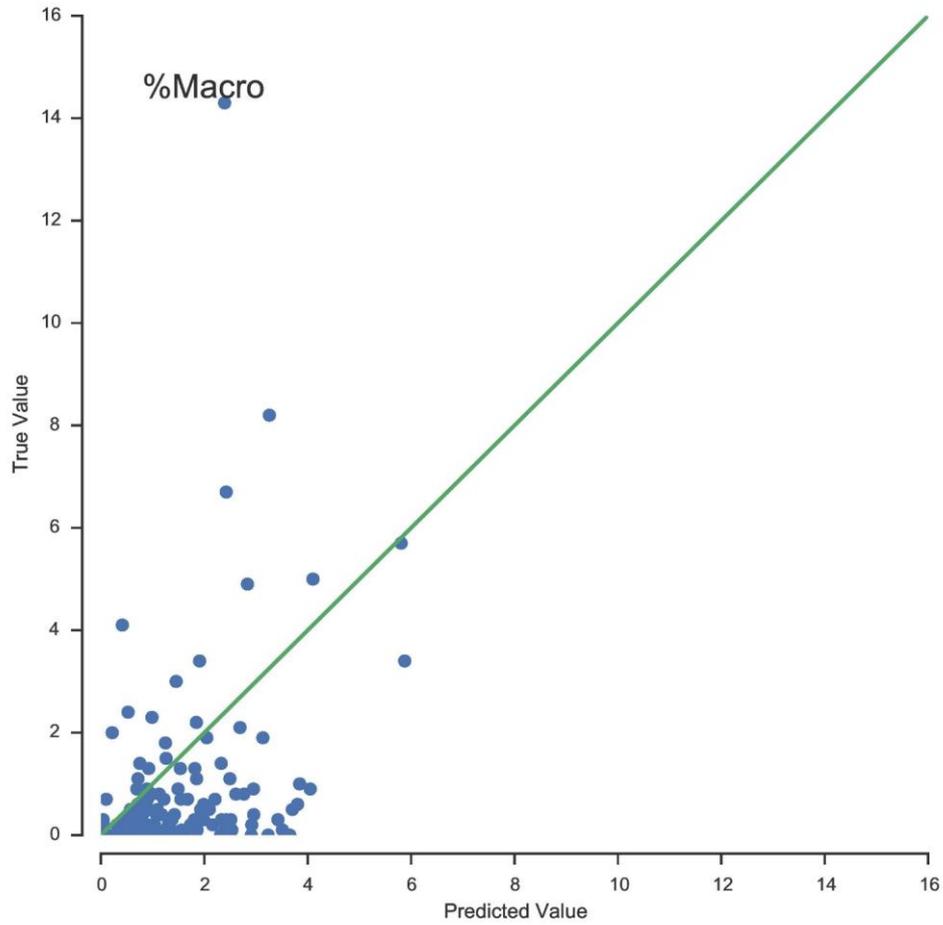


**B.**
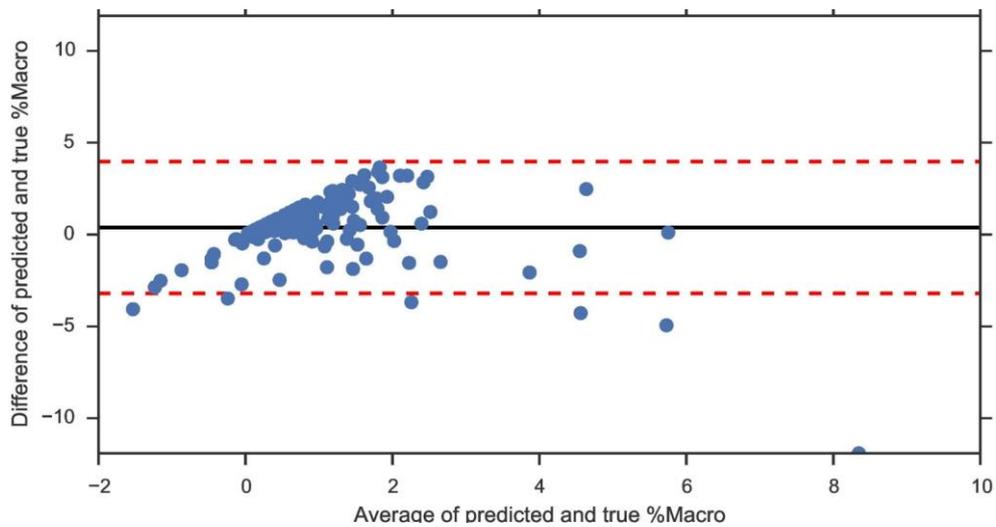
**Figure S15.** Reader training guide used to assign redness score to IDA-AMPS tests.

| 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| no or nearly no red above HCT | some red in one or both phases or packed at interfaces | moderate red in one or both phases or packed at interfaces | strong red in one or both phases or packed at interfaces | very strong band of red in both phases and at interfaces |
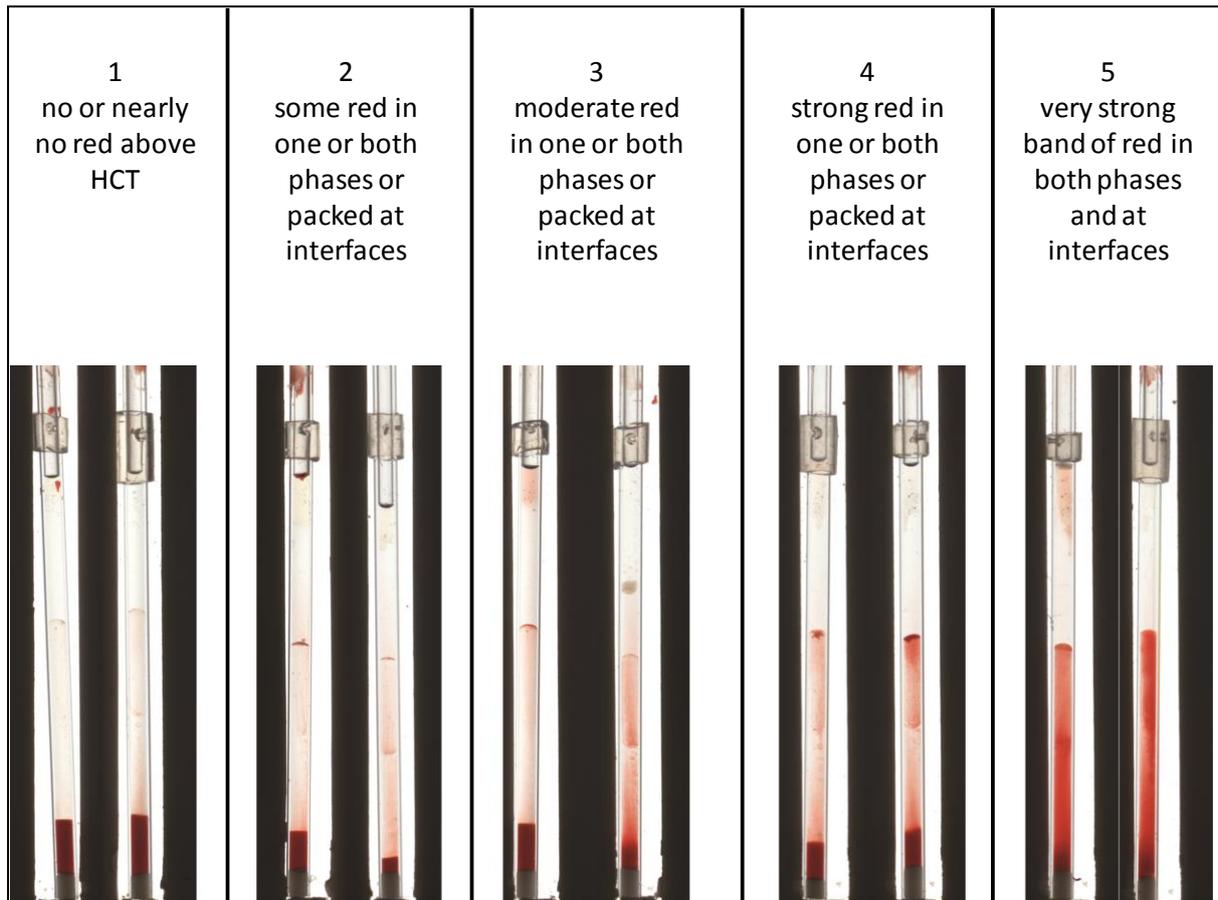
# References

1. C. M. Bishop, *Pattern recognition and machine learning*, Springer, 2006.

2. K. P. Murphy, *Machine learning: a probabilistic perspective*, MIT press, 2012.

3. H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. N. Vapnik, *'Support Vector Regression Machines' in Advances in Neural Information Processing Systems 9: NIPS 1996*, MIT Press, 1997.

4. R. Kohavi, in *Ijcai*, 1995, vol. 14, pp. 1137–1145.

5. A. Demir, N. Yarali, T. Fisgin, F. Duru, and A. Kara, *Pediatr. Int.*, 2002, **44**, 612–616.

6. R. Colah, A. Gorakshakar, and A. Nadkarni, *Expert Rev. Hematol.*, 2010, **3**, 103–17.

7. L. I. Lin, *Biometrics*, 1989, **45**, 255–68.

8. A. A. Kumar, M. R. Patton, J. W. Hennek, S. Y. R. Lee, G. D'Alesio-Spina, X. Yang, J. Kanter, S. S. Shevkoplyas, C. Brugnara, and G. M. Whitesides, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 14864–9.

9. WHO/CDC, *Worldwide prevalance of anaemia 1993-2005*, Geneva, 2008.