scientific reports

Check for updates

OPEN

Artificial intelligence-driven label-free detection of chronic myeloid leukemia cells using ghost cytometry

Kohjin Suzuki[1,2], Naoki Watanabe[1], Yutaka Tsukune[1], Tadaaki Inano[1], Shintaro Kinoshita[1], Sayuri Tomoda[2], Kohei Yamada[2], Yusuke Konishi[2], Takuya Kuwana[2], Takeshi Sugiyama[2], Kenji Fukada[2], Kazuhiro Yamada[2], Miki Ando[1] & Tomoiku Takaku[1,3]✉

Early diagnosis and treatment initiation of chronic myeloid leukemia (CML) are considered to increase the rate of deep molecular response. However, the early diagnosis of CML is challenging due to the absence of clinical symptoms and peripheral blood test anomaly, especially at the timing of peripheral white blood cell count is within a normal range. This study explored the possibility of artificial intelligence (AI)-based quantitative detection of CML cells using ghost cytometry (GC) technology. We created pre-trained AI models, using the morphological information data of the peripheral blood leukocytes obtained from patients newly diagnosed with CML and healthy individuals. We applied these models to peripheral blood samples from CML patients after initiating tyrosine kinase inhibitor treatment at various time points, which contains smaller amounts of tumor cells. The AI model accurately detected CML cells and a strong correlation between AI-detected CML cells and actual *BCR::ABL1*^{IS} mRNA levels was observed. We concluded that the multidimensional morphological information of single cells obtained using GC, combined with machine learning algorithms, enables the quantitative detection of label-free CML cells. Our finding may enable the development of a screening test that can identify early-stage patients with CML before numerical abnormalities appear in blood tests.

The Philadelphia chromosome arises from a translocation between chromosomes 9 and 22, leading to the formation of the *BCR::ABL1* fusion gene[1,2]. ABL1 is a tyrosine kinase that plays a role in cellular proliferation signaling. The *BCR::ABL1* fusion gene results in the constitutive activation of ABL1, thereby driving abnormal cellular proliferation[3,4]. The *BCR::ABL1* fusion gene is detected in chronic myeloid leukemia (CML), and is considered central to CML pathogenesis. This condition is characterized by the monoclonal neoplastic proliferation of hematopoietic stem cells that retain their ability to mature and is currently the only therapeutic target[5,6]. Although CML progresses from the chronic phase to the advanced phase, tyrosine kinase inhibitors (TKIs) are the primary treatment option for all disease phases[7]. However, despite significant improvements in the prognosis of CML due to TKIs, about half of the patients fail to achieve treatment-free remission (TFR) because of TKI-resistant *BCR::ABL1* mutations[8,9] and residual CML stem cells[10,11]. As the long-term use of TKIs is associated with various adverse events, such as fetal cardiovascular complications, edema, rashes, fatigue, and pleural effusions[12], it is important to increase the TFR achievement rate. Currently, most patients with CML are diagnosed when their white blood cell counts significantly increase, indicating a substantial replacement of normal cells with tumor cells. In addition, factors of the commonly used prognostic scoring systems for CML, such as splenomegaly, peripheral blood blasts, and platelet count, worsen and increase over time[13,14]. Given that a higher rate of deep molecular response (DMR) has been reported in patients who achieve early molecular response[15,16], we hypothesized that early diagnosis and treatment intervention would increase the number of patients achieving DMR and improve TFR rates.

To achieve early diagnosis of CML, we focused on a ghost cytometry (GC)[17]. GC is a recently developed technology based on a combination of artificial intelligence (AI) and flow cytometry that utilizes micropatterned

[1]Department of Hematology, Juntendo University Graduate School of Medicine, 3-1-3, Hongo, Bunkyo-ku, Tokyo 113-8421, Japan. [2]Technology Innovation, Sysmex Corporation, 4-4-4, Takatsukadai, Nishi-ku, Kobe 651-2271, Japan. [3]Department of Hematology, Saitama Medical University, 38 Moro-Hongo, Moroyama-machi, Iruma-gun, Saitama 350-0495, Japan. ✉email: ttakaku@saitama-med.ac.jp

illumination to capture temporal waveform signals that reflect subtle structural and morphological changes in cells[18,19]. Notably, this machine learning-based analysis of signals can accurately identify various cell types, including peripheral blood leukocytes, comparable to surface marker-based classification[20]. Furthermore, GC enables the label-free identification of blast cells in bone marrow samples from acute myeloid leukemia and acute lymphoblastic leukemia[21]. However, it has not yet been reported whether GC can quantitatively detect leukemia cells from actual clinical specimens.

We previously reported that the *BCR::ABL1* oncogene played a significant role in promoting mitochondrial fragmentation by activating dynamin-related protein 1 through the mitogen-activated protein kinase pathway, which can be a potential biomarker for the early diagnosis of CML using AI-driven GC[22]. Currently, automated hematology analyzers are a cost-effective and convenient method for identifying numerical abnormalities[23], raising the suspicion of hematological disease including CML[24]. However, they are not suitable for the highly sensitive and specific detection of CML cells. Measurement of *BCR::ABL1*^{IS} mRNA by reverse transcription-quantitative polymerase chain reaction (RT-qPCR) is essential for detecting a low percentage of CML cells with high sensitivity[25,26] and is crucial for assessing treatment efficacy and monitoring disease progression[27,28]. However, RT-qPCR is not commonly performed as a screening test, particularly in the early stages of CML, due to the absence of symptoms and normal white blood cell counts. In this study, we aimed to explore AI-based quantitative label-free detection of CML cells to identify before numerical abnormalities in blood cells appear in blood tests for early CML diagnosis.

Results

Label-free discrimination between chronic myeloid leukemia and healthy leukocytes using ghost cytometry

To detect CML cells using GC, we first evaluated the ability of GC to distinguish between CML and healthy cells (Fig. 1A). Leukocytes measured using GC were classified into white blood cells (WBCs), granulocytes, and lymphocytes based on the forward-scatter-side-scatter scattergram (Fig. 1B). Using AI, the waveform signals obtained from the healthy and CML patient cells were classified for each cell population. As a result, the average discrimination scores (F1 scores) for WBCs, granulocytes, and lymphocytes were 0.79 (0.66–0.88), 0.73 (0.56–0.86), and 0.83 (0.75–0.91), respectively (Fig. 1C, n = 8 for all), indicating approximately 80% accuracy in identifying CML cells from healthy cells.

Quantitative label-free detection using ghost cytometry

Having demonstrated the potential of GC to discriminate between the leukocytes of healthy individuals and patients with CML, we evaluated the feasibility of label-free quantitative detection using GC. Quantitative detection involves acquiring waveform signals from the cells of interest beforehand and pre-training the AI models using these signals. Subsequently, AI models were used to calculate the cell ratios within the samples using the waveform signals obtained from samples with varying cell composition ratios (Fig. 2A).

Initially, we performed validation using two different cell lines: K562 and UT-7/erythropoietin (EPO) to verify the feasibility of the concept of quantitative detection using GC. These two cell types were successfully discriminated with an F1 score of 0.89 (0.85–0.92, n = 3) using GC. Next, we calculated the K562 cell ratio using samples consisting of a mixture of these two cell types at ratios ranging from 10 to 90%. The AI-predicted K562 cell ratios showed a highly positive correlation (r = 0.99) with the actual K562 cell ratios, indicating strong consistency between the predicted and actual values (Fig. 2B).

Next, we conducted a similar validation using peripheral blood leukocytes obtained from patients with CML before treatment and from healthy individuals. The F1 scores against healthy donors for WBCs, granulocytes, and lymphocytes of three different patients were 0.76 (0.68–0.82), 0.76 (0.65–0.85), and 0.76 (0.66–0.83), respectively, demonstrating high positive correlations with correlation coefficients of r = 0.99 (0.98–0.99), 0.98 (0.97–0.99), and 0.86 (0.68–0.96), respectively (Fig. 2C).

Based on these results, the positive correlation between the AI-predicted and the actual ratios of CML cells in the samples indicated the potential for label-free quantitative detection of CML cells using GC and pre-trained AI.

Model of early-stage chronic myeloid leukemia (CML) patient samples

In daily clinical practice, the ratio of *BCR::ABL1*-positive cells in peripheral blood at diagnosis is typically over 80%, as determined by fluorescence in situ hybridization analysis. Therefore, if GC can detect small amounts of CML cells, it could facilitate early screening. Hence, we investigated the feasibility of utilizing samples from patients with CML who had already initiated TKI treatment and every month of follow-up as a model for early-stage CML patient samples.

A comparison of blood test results between pre-treatment CML patient samples (n = 6) and post-treatment CML patient samples (n = 11) revealed significant improvements in WBC counts, including neutrophils and basophils, and in immature granulocyte (IG) counts following treatment ($p < 0.0001$ for each). Consequently, the *BCR::ABL1*^{IS} values were significantly lower ($p < 0.0005$), with a median of 55.3% (34.0–70.3%) in post-treatment samples (Table 1). In automated blood count testing, when specific numerical abnormalities are detected, flagging is performed to alert healthcare providers. In pre-treatment CML patient samples, the positive flagging was observed in response to increased peripheral blood WBC counts, including "WBC Abn Scg" (100%), "Neutro+" (100%), "Lympho+" (100%), "Mono+" (100%), "Eo+" (83.3%), "Baso+" (100%), and "Leuko+" (100%) and in response to increased immature cells, such as "IG present" (100%), "Bl/Abn Ly?" (100%), "and Left Shift?" (100%). In contrast, post-treatment CML patient samples exhibited a significant disappearance of this flagging, including "WBC Abn Scg "(9.1%), "Neutro+" (0%), "Lympho+" (0%), "Mono+" (0%), "Eo+" (9.1%), "Baso+" (0%), "Leuko+" (0%), "IG present" (9.1%), "Bl/Abn Ly?" (27.3%), and "Left Shift?" (0%) (Fig. 3).

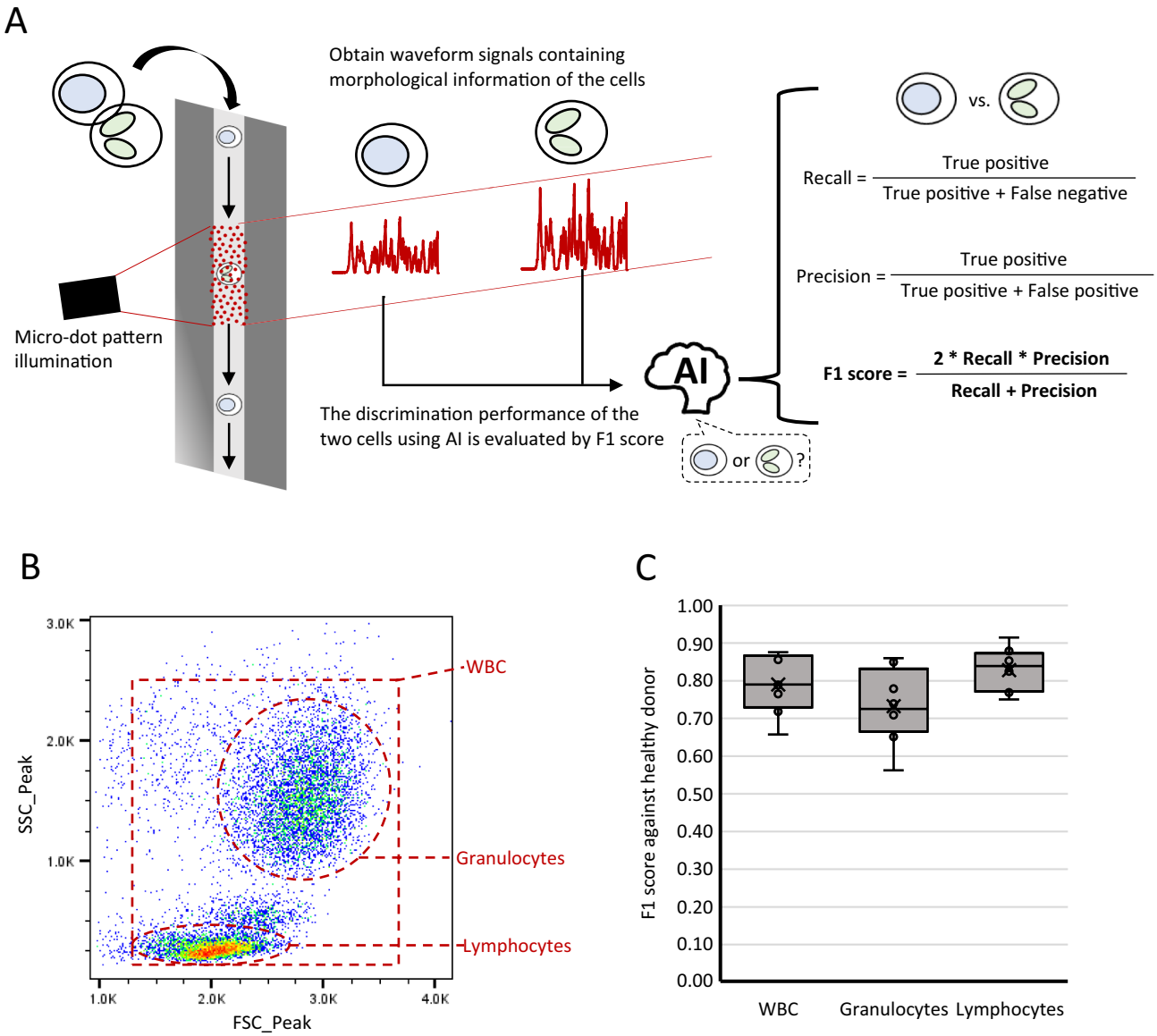


Fig. 1. Discriminations of leukocytes from patient with CML by ghost cytometry (GC). (**A**) Principle of the AI-driven GC was illustrated. Waveform signals containing the morphological information of the cells obtained by GC were analyzed using AI to distinguish two kinds of cells. The classification accuracy was evaluated by F1 score, harmonic means of precision and recall. (**B**) Leukocytes obtained from CML patients and healthy individual were divided into WBCs, granulocytes, and lymphocytes from SSC-FSC scattergram. (**C**) The classification accuracy of F1 score between WBC, Granulocytes, and Lymphocytes from CML patients at diagnosis and those from healthy individual (n = 8). *CML* chronic myeloid leukemia, *AI* artificial intelligence, *GC* ghost cytometry, *WBCs* white blood cells, *SSC* side scatter, *FSC* forward scatter.

Our results indicated that the specimens obtained 1–3 months after the initiation of treatment showed improvement in the numerical abnormality of peripheral leukocytes and the resolution of abnormal flagging, despite the persistence of approximately 50% *BCR::ABL1*-positive cells.

Quantitative label-free detection using ghost cytometry from early-stage CML patient model samples

The next step involved investigating the potential for label-free CML cell detection using GC from these specimens. Therefore, we utilized specimens from untreated CML patients for pre-training the AI model (n = 6, average F1 score of 0.84 against healthy control). Subsequently, waveform signals obtained from the leukocytes of healthy individuals (n = 5) and patients with CML during treatment (n = 11) were analyzed using a pre-trained AI model, and the percentage of CML cells within each specimen was calculated (Fig. 4A).

As a result, a significant prediction of CML cells was achieved in the samples from patients with CML during treatment, compared with the specimens from healthy individuals, in each fraction of WBCs ($p = 0.0009$),

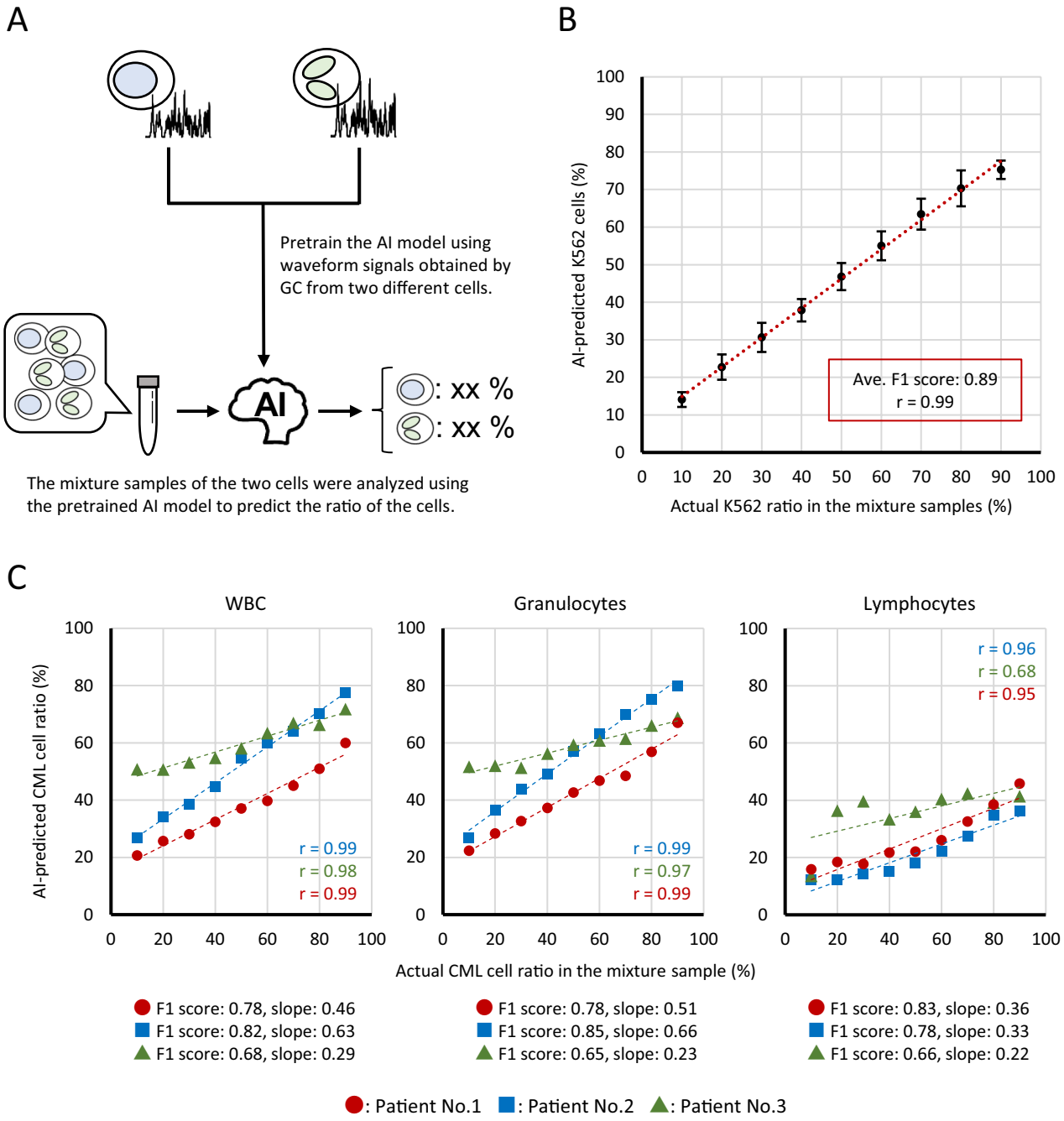


Fig. 2. Quantitative label-free detection of CML cells. (**A**) The experimental method of the quantitative label-free detection by GC was illustrated. The classification AI model was pre-trained by the waveform signals obtained from cells using GC. The mixture samples of the two cells were then evaluated by the pre-trained AI model to predict the ratio of each cell in the sample. (**B**) The correlation between the AI-predicted K562 ratio and actual mixing ratio (10–90%) of K562 in each sample. Three different experiments were performed to obtain the training and testing data, respectively. The data were shown in average ± SD. The r value was evaluated by Spearman's rank correlation coefficient. (**C**) Quantitative label-free detection of CML patient leukocytes by GC. Mixture samples of the leukocytes from CML patients and healthy individual (10–90%) were evaluated by the pretrained AI model to calculate the ratio of CML cells in the samples. Leukocytes were divided into WBCs, granulocytes, and lymphocytes, and samples from three different patients were evaluated. F1 score indicates the classification accuracy of the healthy or CML leukocytes. The r value was evaluated by Spearman's rank correlation coefficient. *CML* chronic myeloid leukemia, *AI* artificial intelligence, *GC* ghost cytometry, *WBCs* white blood cells.

	Untreated CML patients (n = 6)	CML patients during treatment (n = 11)	
	Median (1Q–3Q)	Median (1Q–3Q)	*p* value
WBC (10^2/uL)	1972.9 (1571.9–2628.9)	49.9 (43.7–83.4)	**0.000**
RBC (10^4/uL)	368 (333.5–400.5)	410 (368–468)	**0.045**
HGB (g/dL)	11.3 (10.5–12.7)	12.6 (11.3–13.8)	0.106
HCT (%)	33.4 (31.4–35.5)	36.5 (34.8–41.8)	0.067
MCV (fL)	93.1 (89.2–95.1)	89.3 (84.7–93.8)	0.325
MCH (pg)	31.4 (30.4–32.6)	30.4 (28.1–30.7)	0.054
MCHC (g/dL)	33.7 (33.2–34.7)	33.3 (32.7–34.3)	0.260
PLT (10^4/uL)	60.6 (39.3–80.3)	13.4 (10.1–18.6)	**0.000**
RDW-SD (fL)	55.1 (50.4–57.1)	47.2 (44.3–52.2)	**0.042**
RDW-CV (%)	16.3 (15.7–16.9)	14.4 (13.2–15.4)	**0.024**
PDW (fL)	11.2 (10.5–12.1)	11.6 (10.2–12.6)	0.471
MPV (fL)	10.1 (9.7–10.5)	10.6 (9.5–11.9)	0.211
P-LCR (%)	25.3 (21.7–28.6)	29.1 (22.8–38.9)	0.206
PCT (%)	0.6 (0.4–0.8)	0.1 (0.1–0.2)	**0.000**
NEUT# (10^2/uL)	1631.7 (1274.9–2437.3)	29.1 (21.7–64.3)	**0.000**
LYMPH# (10^2/uL)	97.9 (72.8–123.4)	12.0 (10.3–23.8)	**0.000**
MONO# (10^2/uL)	51.1 (29.5–76.9)	1.8 (1.5–4.3)	**0.000**
EO# (10^2/uL)	49.7 (26.7–62.8)	1.6 (0.5–2.7)	**0.000**
BASO# (10^2/uL)	75.6 (58.8–92.8)	0.5 (0.3–1.2)	**0.000**
NEUT% (%)	84.6 (80.6–92.3)	57.7 (50.3–72.8)	**0.001**
LYMPH% (%)	4.7 (3.3–8.2)	26.5 (16–36.8)	**0.001**
MONO% (%)	2.3 (1.5–4.2)	4.0 (3.4–6.9)	0.089
EO% (%)	2.3 (1.5–3.4)	2.3 (0.9–5.3)	0.440
BASO% (%)	4.2 (2.2–5.4)	1.0 (0.6–1.9)	**0.011**
IG# (10^2/uL)	666.1 (421.4–916.7)	0.3 (0.2–0.6)	**0.000**
IG% (%)	32.3 (26.7–36.8)	0.4 (0.4–0.7)	**0.000**
RET% (%)	3.1 (2.2–4.5)	1.0 (0.5–1.8)	**0.002**
RET# (10^4/uL)	12.1 (6.6–18.0)	4.7 (2.5–8.3)	**0.008**
IRF (%)	25.7 (18.8–35.9)	5.5 (3.0–8.1)	**0.000**
LFR (%)	74.4 (64.2–81.2)	94.5 (91.9–97.0)	**0.000**
MFR (%)	15.2 (14.1–16.6)	4.7 (3.0–7.5)	**0.000**
HFR (%)	9.1 (4.9–20.5)	0.5 (0.0–0.8)	**0.001**
RET-He (pg)	28.1 (27.5- 28.8)	28.9 (26.3–30.3)	0.628
Treatment duration (month)	0 (0–0)	1 (1–2)	**0.000**
BCR::ABL^{IS} *mRNA (%)*	112.5 (90.4–125.4)	55.3 (34.0–70.3)	**0.000**

Table 1. Blood test results from untreated CML patients and patients during TKI treatment. Significant values are in [bold].

granulocytes ($p = 0.0005$), and lymphocytes ($p = 0.01$) (Fig. 4B). Furthermore, receiver operating characteristic (ROC) analysis revealed that GC could discriminate between samples from healthy individuals and patients with *BCR::ABL1*-positive cells, with areas under the curve (AUC) of 0.98 (95% confidence interval, 0.93–1), 1 (1–1), and 0.92 (0.77–1) for WBCs, granulocytes, and lymphocytes, respectively (Fig. 4C). Finally, the correlation between the AI-predicted CML cell ratio and the values of *BCR::ABL1*^{IS} mRNA, which reflect the actual percentage of CML cells in the sample, was evaluated. The results showed positive correlations with correlation coefficients (r) of 0.87, 0.84, and 0.68 for WBCs, granulocytes, and lymphocytes, respectively (Fig. 4D). These findings suggest the potential of GC for detecting CML cells with high sensitivity in a model of early-stage CML.

Ghost cytometry highly identified CD16^{Low} granulocytes

Because the potential for label-free detection of CML cells using GC has been demonstrated, we further investigated methods to improve the detection accuracy and sensitivity of GC. In previous reports, cluster of differentiation (CD) 16 expression was decreased in leukocytes from patients with CML[29]. Therefore, we focused on the CD16 expression in CML cells. There was a significant increase in cells with lower CD16 expression in granulocytes (CD16^{Low} granulocytes) from patients with CML compared with those from healthy individuals (Fig. 5A,B). After sorting each cell population, Wright–Giemsa staining revealed that mature neutrophils were predominantly observed in higher CD16 expressed (CD16^{High}) granulocytes section from healthy individuals and patients with CML. In contrast, immature neutrophils resembling myelocytes and stab neutrophils were frequently observed in the CD16^{Low} granulocyte sections from patients with CML (Fig. 5C).

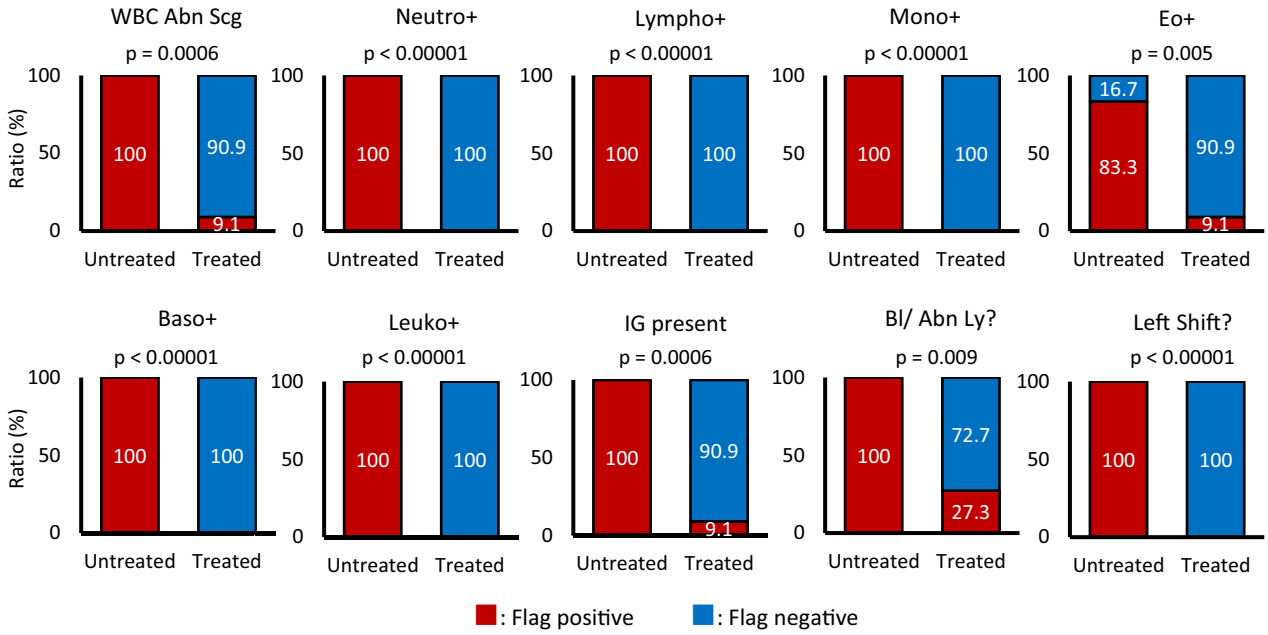


Fig. 3. Flagging information of automated hematological analyzer XN-450 from untreated CML patients (Untreated, n = 6) and patients during TKI treatment (Treated, n = 11). *p* values were evaluated by Fisher's exact test. *CML* chronic myeloid leukemia, *TKI* tyrosine kinase inhibitor, *WBC Abn Scg* white blood cell abnormal scattergram, *Neutro+* increased neutrophil, *Lympho+* increased lymphocytes, *Mono+* increased monocytes, *Eo+* increased eosinophil, *Baso+* increased basophil, *Leuko+* increased leukocytes, *IG present* immature granulocyte present, *Bl/Abn Ly?* blast/abnormal lymphocyte.

We hypothesized that focusing on $CD16^{Low}$ granulocytes in patients with CML would improve the discriminative performance of GC. We compared $CD16^{High/Low}$ granulocytes between patients with CML and healthy individuals. The F1 score between healthy and CML $CD16^{High}$ granulocytes was 0.67 [0.56–0.83], while that between healthy $CD16^{High}$ and CML $CD16^{Low}$ granulocytes was 0.95 [0.90–0.98], which was significantly higher than the former (n = 5, $p = 0.02$, Fig. 5D). Furthermore, the F1 score for $CD16^{Low}$ granulocytes between healthy individuals and patients with CML was 0.87 [0.78–0.95], showing a higher trend compared with the comparison between $CD16^{High}$ granulocytes in healthy individuals and patients with CML ($p = 0.07$, Fig. 5D).

Next, to evaluate label-free detection, we constructed a training model using CD16-labeled cells from untreated patients and attempted to detect CML cells in label-free samples from patients undergoing treatment.

After pre-training the AI using waveform signals derived from $CD16^{Low}$ granulocytes from untreated patients with CML, the percentage of label-free CML cells in patient specimens during treatment (n = 11) was predicted from the granulocyte fraction. In healthy individuals (n = 5), the positivity rate of CML cells was < 20%. However, in CML patient specimens, the positivity rate for CML cells was significantly higher than that in healthy individuals ($p < 0.001$, Fig. 5E). Additionally, ROC analysis revealed that specimens containing label-free CML cells could be distinguished with an AUC of 1.00 (1.00–1.00) by using the AI model pre-trained with $CD16^{Low}$ granulocytes derived from patients with CML (Fig. 5F). Furthermore, the AI-predicted CML cell ratio positively correlated with *BCR::ABL1*^{IS} mRNA, with a correlation coefficient 0.71 (Fig. 5G).

These results suggest that training AI with CML-specific $CD16^{Low}$ granulocytes would improve label-free detection accuracy and sensitivity using GC. These results also indicate that training the model with CML-specific cell populations may contribute to improved specificity and a reduction in false positives.

Discussion

In this study, we aimed to explore the possibility of label-free detection of CML cells using GC. Similar to our previous findings[22], GC could accurately distinguish CML cells in granulocytes, lymphocytes, and total WBC fraction. It has been reported that in CML, *BCR::ABL1* is expressed at the level of hematopoietic stem cells[30–33], and that peripheral blood lymphocytes can also be positive for *BCR::ABL1*[34–36]. Based on this, we hypothesize that the GC was able to achieve high classification performance across various cell populations by capturing subtle morphological changes induced by *BCR::ABL1* expression. The F1 scores of the analysis between healthy and CML patient cells were approximately 0.80. This implies that the accuracy of AI detection is approximately 80%, indicating the minimum detection sensitivity of CML cells using GC may be approximately 20%. In clinical practice, nearly all peripheral blood leukocytes at the time of CML diagnosis are replaced by *BCR::ABL1*-positive cells with increasing peripheral white blood cell count. In pre-treated CML patient specimens used in this study, *BCR::ABL1*^{IS} mRNA was at a median of 112.5%, suggesting that almost all cells were replaced by CML cells at diagnosis. From this, we hypothesized that GC could identify CML cells even if the ratio was approximately 30–70%, which potentially enables earlier diagnosis.

To achieve an early diagnosis of CML using GC, it is necessary to verify whether CML cells can be detected in cases where the disease is still in its early stages, and CML cells have not yet proliferated significantly. However, identifying patients suitable for this testing is challenging because patients with early-stage CML often do not visit the hospital due to the absence of symptoms and normal peripheral white blood cell count. Therefore, we considered using patient samples collected a few months after treatment initiation as they contained a smaller number of CML cells. We considered that the residual *BCR::ABL1*^{IS} mRNA reflects the proportion of remaining CML cells, and that it mimics samples from early-stage CML patients whose blood test values are within the normal range. As results of detecting CML cells from these samples using GC, we achieved a higher detection rate of CML cells in patient samples during treatment compared to those from healthy individuals. Furthermore, the strong positive correlation between the F1 score of CML cell detection by GC and *BCR::ABL1*^{IS} was observed. Since *BCR::ABL1* mRNA levels are considered the gold standard for evaluating treatment response in CML, we believe that correlation with mRNA levels serves as the most clinically relevant indicator for reflecting treatment response and disease progression. Therefore, our study demonstrates that the detection of CML cells using GC may enable the identification of CML in its early stages.

The quality of the training data in machine learning model is considered crucial when utilizing AI[37]. A more accurate AI model can be created by training the AI model using samples with a high discriminative performance against healthy cells. In this study, it was revealed that the low expression of CD16 in granulocytes was one of the specific features of CML, and training AI using these CD16^{Low} granulocytes resulted in a high F1 score and achieved sensitive detection. Thus, by optimizing the types and characteristics of cells used for AI training, it may be possible to achieve higher sensitivity and detect specific population of the cells.

When the peripheral blood CML cell rate exceeds approximately 80%, the detection of numerical abnormalities using an automated hematology analyzer is effective[23,24]. *BCR::ABL1*^{IS} mRNA detection using RT-qPCR remains the standard method for assessing treatment efficacy and monitoring due to its high sensitivity of 10–0.001%[27]. In contrast, introducing a cost-effective and convenient method, such as GC which can detect CML cells in label free manner, may open new possibilities especially for screening and early diagnosis of CML (Fig. 6). Incorporating this screening technology into routine health check-ups may enable the diagnosis of early-stage patients with CML, which is currently highly challenging.

To the best of our knowledge, this is the first study to demonstrate the potential of GC for the quantitative detection of low levels of tumor cells from actual patients specimens using pre-trained AI based on subtle morphological features. Although this study focused exclusively on CML, there is potential to achieve screening tests for various blood disorders by training the model with data obtained from patients with other hematological diseases. Further studies are anticipated to confirm these hypotheses.

This study has several limitations. First, post-treatment samples were used to simulate early-stage CML, however, these samples may differ biologically from actual untreated patients in early-stage CML. The validation with untreated early-stage cases may be required in future study. Second, the number of samples used in this study may be insufficient for robust statistical power. Future studies with larger cohorts are warranted to validate the findings and facilitate clinical implementation. Third, the *BCR::ABL1* mRNA levels and AI detections are shown to correlate in each cell, but there appears to be a variation. Although the *BCR::ABL1* mRNA level is expected to correlate to some extent with the proportion of CML cells, they do not always correspond precisely. This variation may be partly due to the current AI model's insufficient ability to fully detect all CML cells, suggesting that further training is needed to improve its performance. Moreover, incorporating more representative CML cell populations into AI model training may reduce false positives in healthy cells, highlighting the need for further investigation in future studies.

In conclusion, we demonstrated that the multidimensional morphological information of single cells obtained through GC technology, combined with machine learning algorithms, enables the quantitative detection of CML cells in a label free manner. By applying this cost-effective and convenient technology, it may be possible to develop a screening test that can identify early-stage CML patients before numerical abnormalities appear in blood tests, which may potentially contribute to improving the rates of DMR and following TFR achievement rate through early diagnosis and treatment initiation.

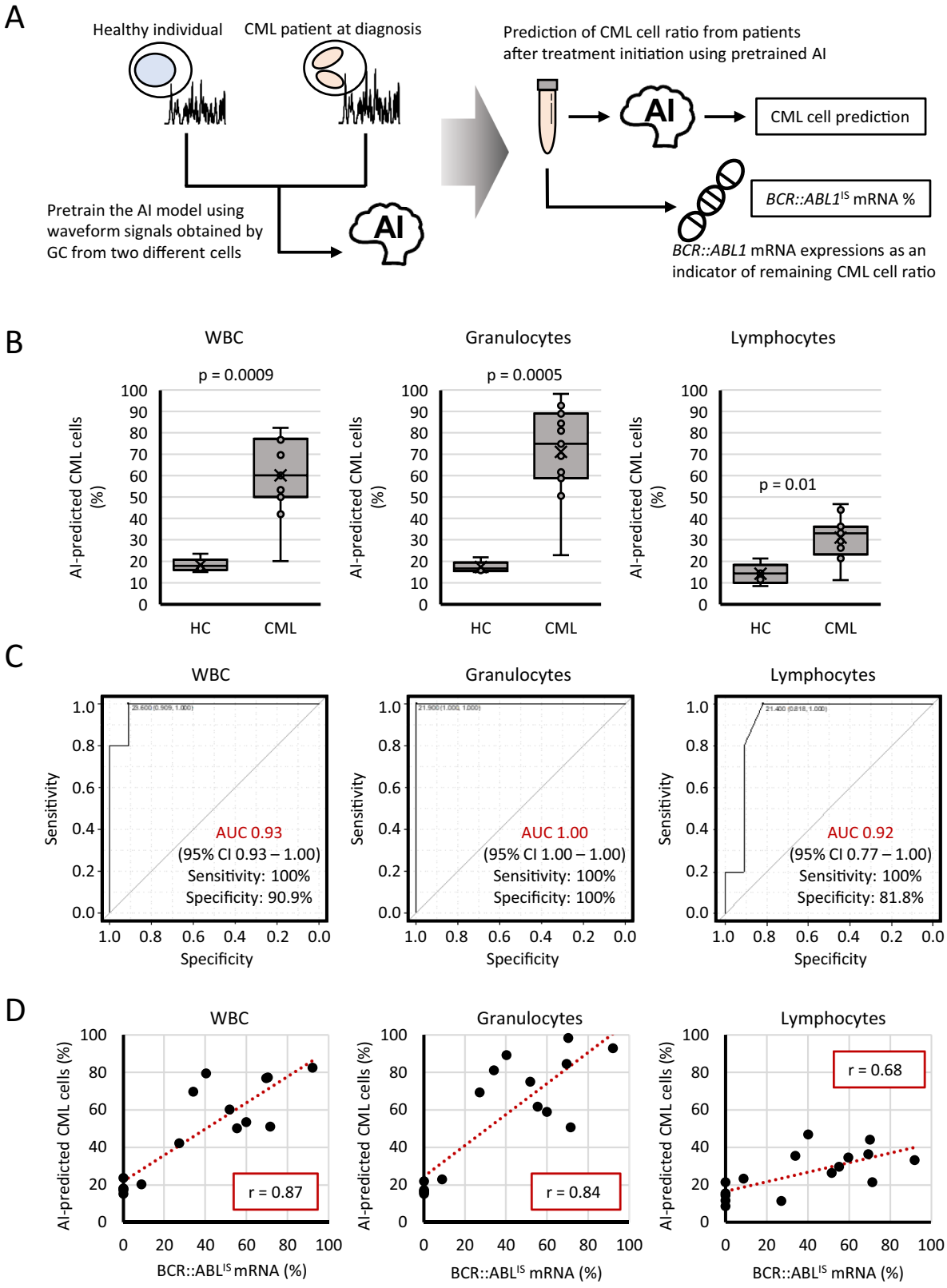
Methods

Clinical samples

Blood samples were obtained from 21 patients with CML at diagnosis, 11 patients with CML under TKI treatment for 1–3 months, and five healthy volunteers. This study was approved by the Ethics Committee of Juntendo University School of Medicine (M12-0895-M01) and Sysmex Corporation (2021-068) and adhered to the Declaration of Helsinki. Patients provided informed consent by accessing the study information published on the website and were free to withdraw their consent at any point during the study, as per the Ethical Guidelines for Medical and Health Research Involving Human Subjects. The ethics committee waived the requirement for written informed consent. Clinical information and data pertaining to clinical tests, including the international scale (IS) of *BCR::ABL1* mRNA, were retrieved from medical records.

Cell culture

CML-derived K562 cells were cultured in Roswell Park Memorial Institute 1640 medium (Nacalai Tesque, Tokyo, Japan) containing 10% heat-inactivated fetal bovine serum (FBS; Nichirei, Tokyo, Japan) and penicillin–streptomycin (100 U/mL; Nacalai Tesque). The cells were incubated in a humidified incubator at 37 °C under 5% CO_2. Acute myeloid leukemia M7-derived UT-7/EPO cells[38] were cultured in Iscove's modified Dulbecco's medium (Nacalai Tesque) containing 10% FBS, 100 U/mL penicillin–streptomycin, and 10 ng/mL recombinant human EPO (Peprotech, Cranbury, NJ, USA) in a humidified incubator maintained at 37 °C under 5% CO_2 conditions.



Complete blood count testing

Blood count data were obtained using an automated hematology analyzer XN-450 (Sysmex, Kobe, Japan). Obtained information is follows: "WBC," white blood cell; "RBC," red blood cell; "HGB," hemoglobin; "HCT," hematocrit; "MCV," mean corpuscular volume; "MCH," mean corpuscular hemoglobin; "MCHC," mean corpuscular hemoglobin concentration; "PLT," platelet; "RDW-SD," standard deviation of red cell distribution

◀**Fig. 4**. Quantitative label-free detection of CML cells from patients during the TKI treatments. (**A**) The experimental method of quantitative label-free detection of CML cells from patients during the TKI treatments was illustrated. The classification AI model was pre-trained by the waveform signals obtained from leukocytes from CML patients at diagnosis and those from healthy individual. Leukocyte samples from CML patients during the TKI treatment and those from healthy individuals were then evaluated using the pre-trained AI model to calculate the ratio of CML cells in the sample. *BCR::ABL1*^{IS} mRNA expressions was used as an indicator of remaining CML cell ratio in the samples. (**B**) Ratios of AI-predicted CML cells in the testing samples. Leukocytes in the samples were divided into WBCs, granulocytes, and lymphocytes from SSC-FSC scattergram, and the ratio of CML cells were calculated using the pre-trained AI model. CML patients at diagnosis for AI training, n = 6; healthy individual for testing (HC), n = 5; CML patients during the treatment for testing (CML), n = 11; *p* value by Mann–Whitney U test. (**C**) Diagnostic performance of CML cell detection by GC was evaluated by ROC analysis of HC and CML by the ratio of AI-predicted CML cells. The results were divided into WBC, granulocytes, and lymphocytes divided from SSC-FSC scattergram. (**D**) The correlation between AI-predicted CML cells and *BCR::ABL1*^{IS} mRNA expressions. The r value was evaluated by Spearman's rank correlation coefficient. *CML* chronic myeloid leukemia, *TKI* tyrosine kinase inhibitor, *AI* artificial intelligence, *GC* ghost cytometry, *WBCs* white blood cells, *SSC* side scatter, *FSC* forward scatter.

width; "RDW-CV," coefficient of variation of red cell distribution width; "PDW," platelet distribution width; "P-LCR," platelet large cell ratio; "PCT," plateletcrit; "NEUT," neutrophil; "LYMPH," lymphocyte; "MONO," monocytes; "EO," eosinophil; "BASO," basophil; "IG," immature granulocyte; "RET," reticulocyte; "IRF," immature reticulocyte fraction; "LFR," low-fluorescence reticulocyte; "MFR," middle-fluorescence reticulocyte; "HFR," high-fluorescence reticulocyte; and "RET-He," reticulocyte hemoglobin equivalent. Obtained flagging information is follows: "WBC Abn Scg," white blood cell abnormal scattergram; "Neutro+," neutrophil increase; "Lympho+," lymphocyte increase; "Mono+," monocyte increase; "Eo+," eosinophil increase; "Baso+," basophil increase; "Leuko+," leukocyte increase; "IG present," immature granulocyte present; "Bl/Abn Ly?," suspicion of blast or abnormal lymphocyte; and "Left Shift?," suspicion of left shift.

Artificial intelligence-driven flow cytometry analysis using ghost cytometry technology

AI-based flow cytometric analysis was performed using a prototype device developed by Sysmex Corporation (Hyogo, Japan) based on GC technology[17].

To collect WBCs from whole blood samples, red blood cells were lysed in five times the amount of red blood cell lysis solution (Qiagen, Hilden, Germany), followed by 15 min incubation at room temperature (24 °C). For GC analysis, one of the two cell groups selected for comparison was labeled, combined, and subjected to analysis using machine learning-integrated GC. Specifically, 1.0×10^6 cells from healthy individuals were incubated at room temperature for 20 min with PC7-labeled anti-CD45 antibody (Clone: J33, Beckman Coulter, Brea, CA, USA) and Alexa Fluor® 488-labeled anti-CD16 antibody (Clone: 3G8, BioLegend, San Diego, CA, USA). The same number of cells from patients with CML were labeled with the same anti-CD16 antibody. These two cell types were mixed and analyzed by GC to obtain waveform signals containing morphological information. The cells were categorized based on their forward scatter (FSC) and side scatter (SSC) scattergrams; SSC-low/FSC-low cells were defined as lymphocytes, and SSC-high/FSC-high cells were defined as granulocytes. We used the PC7-labeled CD45 signal as the ground-truth label to distinguish between the two cell types. Machine learning algorithms, such as convolutional neural networks and light gradient boosting machines, were used to train and classify the waveform signals corresponding to each cell type. The cell discrimination performance was evaluated using the F1 score. The F1 score was calculated as the harmonic mean of the precision and recall. Precision is the ratio of true positives to the sum of true positives and false positives, whereas recall is the ratio of true positives to the sum of true positives and false negatives. The F1 score ranged from 0 to 1, where higher values indicated better discriminative ability between the two cell groups.

In the cell line experiments, one of the two selected cell groups for comparison was stained with MitoTracker Deep Red (10 nM; Thermo Fisher Scientific) as the ground-truth label for 30 min at 37 °C in a humidified incubator, followed by the same experiment described above.

For AI-based cell prediction, machine learning models were pre-trained, as described. The waveform signals were then obtained from the samples for detection without labeling, followed by prediction using the pre-trained AI model. The model classifies the measured waveform signals of cells into one of the two cell types used for pre-training. This allowed the calculation of the cell ratio within the sample.

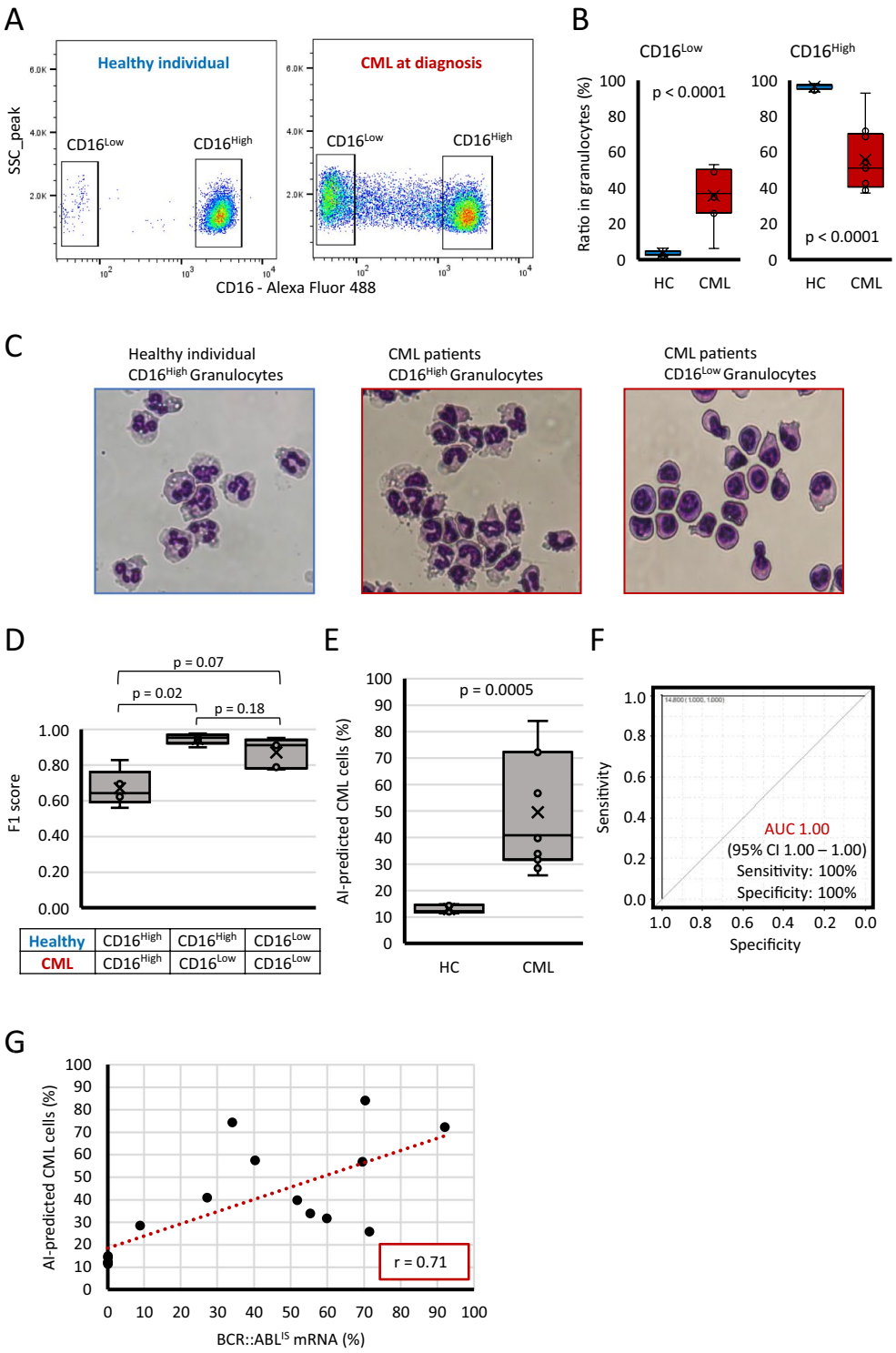
The data related to the flow cytometry measurements obtained in this study were analyzed using the FlowJo software (version 10.8.1, Becton Dickinson, Franklin Lakes, NJ, USA).

Cell sorting

Leukocytes were stained with Alexa Fluor® 488-labeled anti-CD16 antibody (Clone: 3G8, BioLegend), and granulocytes gated from FSC-SSC scattergram were divided into CD16 low and high granulocytes, followed by cell sorting using FACSAria (Becton Dickinson, Franklin Lakes, NJ, USA).

Wright–Giemsa staining

The cells were immobilized on glass slides using Cytospin (Sakura Finetek, Tokyo, Japan), followed by staining with Wright's solution (Muto Pure Chemicals, Tokyo, Japan) containing 50% methanol (Nacalai Tesque) for 5 min. Subsequently, the cells were washed and stained with Giemsa solution (Muto Pure Chemicals) for 20 min. The cells were washed, dried, and observed under a microscope (ECLIPSE TE300; Nikon, Tokyo, Japan).



Statistical analyses

For statistical analyses, the Mann–Whitney U test was used to compare differences between two experimental groups, and the Steel–Dwass test was used to compare differences among more than three experimental groups. Fisher's exact test was used to compare differences in proportions between the two experimental groups. All statistical tests were two-tailed, and statistical significance was set at $p < 0.05$. Spearman's rank correlation coefficient was used to evaluate the correlation between two factors. ROC analysis was used to evaluate diagnostic performance. Statistical analyses were performed using EZR, which is based on R and R commander[39].

◀**Fig. 5**. CML cell detection using pre-trained AI model by $CD16^{Low}$ granulocytes. (**A**) Flow cytometric analysis of CD16 expression in CML patients and healthy individual. Peripheral leukocytes were divided into granulocytes from FSC-SSC scattergram, then CD16 expression were evaluated. (**B**) Percentages of $CD16^{High/Low}$ granulocytes from CML patients (n = 9) and healthy individual (n = 9). *p* value by the Mann–Whitney U test. (**C**) Wright–Giemsa staining of $CD16^{High/Low}$ granulocytes from CML patients and healthy individual. (**D**) The classification accuracy of F1 score between $CD16^{High/Low}$ granulocytes from CML patients or healthy individual (n = 5). (**E**) Ratios of AI-predicted CML cells in the testing samples. Leukocytes in the samples were divided into granulocytes from SSC-FSC scattergram, and the ratio of CML cells were calculated using the pre-trained AI model. CML patients at diagnosis for AI training, n = 6; healthy individual for testing (HC), n = 5; CML patients during the treatment for testing (CML), n = 11; *p* value by the Steel–Dwass test. (**F**) Diagnostic performance of CML cell detection by GC was evaluated by ROC analysis. (**G**) The correlation between AI-predicted CML cells and *BCR::ABL1*^{IS} mRNA expressions. The r value was evaluated by Spearman's rank correlation coefficient. *CML* chronic myeloid leukemia, *AI* artificial intelligence, *GC* ghost cytometry, *WBCs* white blood cells, *SSC* side scatter, *FSC* forward scatter.

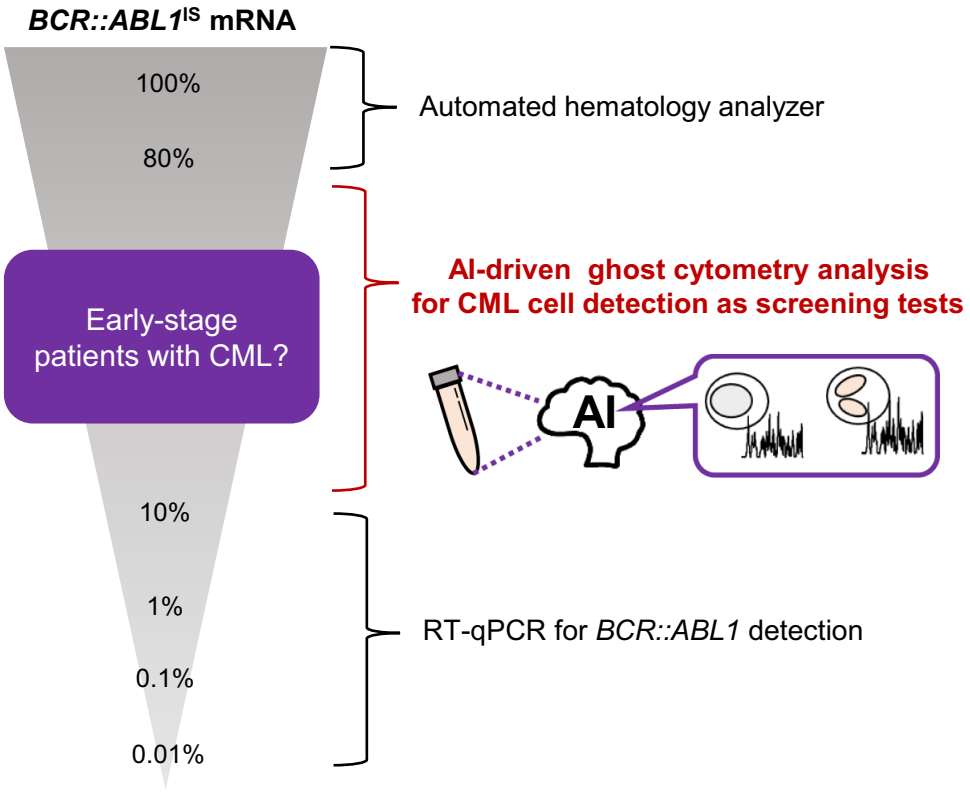


Fig. 6. Overview diagram of CML cell screening using GC. During the diagnosis of CML, numerical abnormalities are detected through blood tests, with *BCR::ABL1*^{IS} mRNA being close to almost 100%. On the other hand, RT-qPCR is useful for evaluating treatment efficacy and disease progression, but it is not suitable as a screening test. GC, however, is a convenient method that can detect CML cells from samples that do not show numerical abnormalities in blood tests. Therefore, it may be possible to utilize GC for early diagnosis through screening of CML. *CML* chronic myeloid leukemia, *GC* ghost cytometry.

Data availability

The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 18 February 2025; Accepted: 10 June 2025
Published online: 01 July 2025

References

1. Rowley, J. D. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**, 290–293 (1973).
2. Nowell, P. C. The minute chromosome (Phl) in chronic granulocytic leukemia. *Blut* **8**, 65–66 (1962).
3. Samad, M. A., Mahboob, E. & Mansoor, H. Chronic myeloid leukemia: A type of MPN. *Blood Res.* **57**, 95–100 (2022).

4. Senapati, J. et al. Management of chronic myeloid leukemia in 2023—Common ground and common sense. *Blood Cancer J.* **13**, 58 (2023).
5. Druker, B. J. et al. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat. Med.* **2**, 561–566 (1996).
6. Bower, H. et al. Life expectancy of patients with chronic myeloid leukemia approaches the life expectancy of the general population. *J. Clin. Oncol.* **34**, 2851–2857 (2016).
7. Hehlmann, R., Saussele, S., Voskanyan, A. & Silver, R. T. Management of CML-blast crisis. *Best Pract. Res. Clin. Haematol.* **29**, 295–307 (2016).
8. Ko, T. K. et al. An integrative model of pathway convergence in genetically heterogeneous blast crisis chronic myeloid leukemia. *Blood* **135**, 2337–2353 (2020).
9. Koyama, D., Kikuchi, J., Kuroda, Y., Ohta, M. & Furukawa, Y. AMP-activated protein kinase activation primes cytoplasmic translocation and autophagic degradation of the BCR-ABL protein in CML cells. *Cancer Sci.* **112**, 194–204 (2021).
10. Tanaka, Y. et al. Eliminating chronic myeloid leukemia stem cells by IRAK1/4 inhibitors. *Nat. Commun.* **13**, 271 (2022).
11. Chen, Y., Zou, J., Cheng, F. & Li, W. Treatment-free remission in chronic myeloid leukemia and new approaches by targeting leukemia stem cells. *Front. Oncol.* **11**, 769730 (2021).
12. Caldemeyer, L., Dugan, M., Edwards, J. & Akard, L. Long-term side effects of tyrosine kinase inhibitors in chronic myeloid leukemia. *Curr. Hematol. Malig. Rep.* **11**, 71–79 (2016).
13. Sokal, J. E. et al. Prognostic discrimination in "good-risk" chronic granulocytic leukemia. *Blood* **63**, 789–799 (1984).
14. Pfirrmann, M. et al. Prognosis of long-term survival considering disease-specific death in patients with chronic myeloid leukemia. *Leukemia* **30**, 48–56 (2016).
15. Shanmuganathan, N. et al. Early BCR-ABL1 kinetics are predictive of subsequent achievement of treatment-free remission in chronic myeloid leukemia. *Blood* **137**, 1196–1207 (2021).
16. Sasaki, K. et al. Prediction for sustained deep molecular response of BCR-ABL1 levels in patients with chronic myeloid leukemia in chronic phase. *Cancer* **124**, 1160–1168 (2018).
17. Ota, S. et al. Ghost cytometry. *Science* **360**, 1246–1251 (2018).
18. Ota, S., Sato, I. & Horisaki, R. Implementing machine learning methods for imaging flow cytometry. *Microscopy (Oxford)* **69**, 61–68 (2020).
19. Adachi, H. et al. Use of ghost cytometry to differentiate cells with similar gross morphologic characteristics. *Cytom. A* **97**, 415–422 (2020).
20. Ugawa, M. et al. In silico-labeled ghost cytometry. *Elife* **10**, e67660 (2021).
21. Kawamura, Y. et al. Label-free cell detection of acute leukemia using ghost cytometry. *Cytom. A* **105**, 196–202 (2024).
22. Suzuki, K. et al. BCR::ABL1-induced mitochondrial morphological alterations as a potential clinical biomarker in chronic myeloid leukemia. *Cancer Sci.* **116**, 673–689 (2025).
23. Furundarena, J. R. et al. Comparison of abnormal cell flagging of the hematology analyzers Sysmex XN and Sysmex XE-5000 in oncohematologic patients. *Int. J. Lab. Hematol.* **39**, 58–67 (2017).
24. Ogasawara, A. et al. A simple screening method for the diagnosis of chronic myeloid leukemia using the parameters of a complete blood count and differentials. *Clin. Chim. Acta* **489**, 249–253 (2019).
25. Nakamae, H. et al. A new diagnostic kit, ODK-1201, for the quantitation of low major BCR-ABL mRNA level in chronic myeloid leukemia: Correlation of quantitation with major BCR-ABL mRNA kits. *Int. J. Hematol.* **102**, 304–311 (2015).
26. Burmeister, T. & Reinhardt, R. A multiplex PCR for improved detection of typical and atypical BCR-ABL fusion transcripts. *Leuk. Res.* **32**, 579–585 (2008).
27. Hughes, T. P. et al. Frequency of major molecular responses to imatinib or interferon alfa plus cytarabine in newly diagnosed chronic myeloid leukemia. *N. Engl. J. Med.* **349**, 1423–1432 (2003).
28. Foroni, L. et al. Guidelines for the measurement of BCR-ABL1 transcripts in chronic myeloid leukaemia. *Br. J. Haematol.* **153**, 179–190 (2011).
29. Kabutomori, O., Iwatani, Y., Koh, T. & Yanagihara, T. CD16 antigen density on neutrophils in chronic myeloproliferative disorders. *Am. J. Clin. Pathol.* **107**, 661–664 (1997).
30. Vetrie, D., Helgason, G. V. & Copland, M. The leukaemia stem cell: Similarities, differences and clinical prospects in CML and AML. *Nat. Rev. Cancer* **20**, 158–173 (2020).
31. Daley, G. Q., Van Etten, R. A. & Baltimore, D. Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science* **247**, 824–830 (1990).
32. Hamilton, A. et al. Chronic myeloid leukemia stem cells are not dependent on Bcr-Abl kinase activity for their survival. *Blood* **119**, 1501–1510 (2012).
33. Koschmieder, S. & Schemionek, M. Mouse models as tools to understand and study BCR-ABL1 diseases. *Am. J. Blood Res.* **1**, 65–75 (2011).
34. Pagani, I. S. et al. Lineage-specific detection of residual disease predicts relapse in patients with chronic myeloid leukemia stopping therapy. *Blood* **142**, 2192–2197 (2023).
35. Takahashi, N., Miura, I., Saitoh, K. & Miura, A. B. Lineage involvement of stem cells bearing the philadelphia chromosome in chronic myeloid leukemia in the chronic phase as shown by a combination of fluorescence-activated cell sorting and fluorescence in situ hybridization. *Blood* **92**, 4758–4763 (1998).
36. Torlakovic, E., Litz, C. E., McClure, J. S. & Brunning, R. D. Direct detection of the Philadelphia chromosome in CD20-positive lymphocytes in chronic myeloid leukemia by tri-color immunophenotyping/FISH. *Leukemia* **8**, 1940–1943 (1994).
37. Rajkomar, A. et al. Scalable and accurate deep learning with electronic health records. *NPJ Digit. Med.* **1**, 18 (2018).
38. Komatsu, N. et al. Establishment and characterization of an erythropoietin-dependent subline, UT-7/Epo, derived from human leukemia cell line, UT-7. *Blood* **82**, 456–464 (1993).
39. Kanda, Y. Investigation of the freely available easy-to-use software "EZR" for medical statistics. *Bone Marrow Transplant.* **48**, 452–458 (2013).

Acknowledgements

We thank the Laboratory of Cell Biology and Molecular and Biochemical Research in Biomedical Research Core Facilities, Juntendo University Graduate School of Medicine, for their technical assistance.

Author contributions

K.S. designed, performed, and analyzed experiments and contributed to writing the manuscript. N.W., Y.T., T.I., and S.K. collected clinical specimens and information. S.T., Kohei Y., Y.K., T.K., T.S., and K.F. developed a prototype of ghost cytometry and provided the methodology. Kazuhiro Y. and M.A. supervised the research and contributed to clinical and scientific discussions. T.T. designed and supervised the research, collected clinical specimens and information, and contributed to writing the manuscript. All authors approved the final manuscript.

Declarations

Competing interests

This work was funded by Sysmex to T.T. as a collaborative research project. K.S. and T.T. have filed patent applications related to the label-free CML cell detection method. K.S., S.T., Kohei Y., Y.K., T.K., T.S., K.F., and Kazuhiro Y. are employees of Sysmex. Other authors declare no competing financial interests regarding this study.

Additional information

Correspondence and requests for materials should be addressed to T.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025