Leukemia Research 69 (2018) 54–59

Contents lists available at ScienceDirect

Leukemia Research

journal homepage: www.elsevier.com/locate/leukres

Research paper

Inter-observer variance and the need for standardization in the morphological classification of myelodysplastic syndrome

Keiko Sasada[a], Noriko Yamamoto[a], Hiroki Masuda[a], Yoko Tanaka[a], Ayako Ishihara[a], Yasushi Takamatsu[b], Yutaka Yatomi[c], Waichiro Katsuda[d], Issei Sato[e], Hirotaka Matsui[a,e,f,*], on behalf of the Kyushu regional department of the Japanese Society of Laboratory Hematology

[a] Department of Laboratory Medicine, Kumamoto University Hospital, Kumamoto University, Japan
[b] Department of Medical Oncology, Hematology and Infectious Diseases, Fukuoka University, Japan
[c] Department of Clinical Laboratory Medicine, Graduate School of Medicine, The University of Tokyo, Japan
[d] ThinkCyte, Inc., Japan
[e] Medical Image Analysis Team, Center for Advanced Intelligence Project, Institute of Physical and Chemical Research (RIKEN), Japan
[f] Department of Molecular Laboratory Medicine, Faculty of Life Sciences, Kumamoto University, Japan

ARTICLE INFO

Keywords:
Myelodysplastic syndrome
Morphological classification
Cytoplasmic hypo–granularity
Pseudo Pelger–Huët anomaly
Inter-observer variance

ABSTRACT

In this era of genome medicine, the sub-classification of myeloid neoplasms, including myelodysplastic syndrome (MDS), is now supported by genetic testing in selected cases. However, as the initial suspicion and primary diagnosis of the disease still largely relies on morphological features and numbers of hematopoietic cells, the establishment of a uniform diagnostic basis, especially for cell morphology, is essential. In this study, we collected nearly 100,000 hematopoietic cell images from 499 peripheral blood smear specimens from patients with MDS and used these to evaluate the standardization of morphological classification by medical technologists. The observers in this study ranged between two to eleven for each image, and the images were classified according to MDS criteria through a web-based system. We found considerable inter-observer variance in the assessment of dysplastic features. Observers did not recognize cytoplasmic hypo–granularity unless almost all granules in neutrophils were absent. Pseudo Pelger–Huët anomalies were also often overlooked, except for cells with a very typical “pince-nez” appearance. Taken together, this study suggests a requirement for further standardization in terms of morphological cell classification, and a need for the development of automatic cell classification–supporting devices for the accurate diagnosis of MDS.

1. Introduction

Myelodysplastic syndrome (MDS) is a hematopoietic malignancy characterized by morphological atypia (dysplasia) of hematopoietic cells, ineffective hematopoiesis, peripheral cytopenia and progression to acute myeloid leukemia (AML). The disease is often found in the elderly: The median age of disease diagnosis is 76 years, with around 30 out of a 100,000 people aged 70 years or older diagnosed with MDS each year, whereas the frequency is 3–4 among people less than 70 years of age [1]. Recently, it was found that the disease becomes obvious when hematopoietic stem cells sequentially acquire several to ten somatic and/or germline gene mutations, and that gene mutations are frequently associated with disease phenotype or prognosis [2]. Based on these recent advances in the understanding of the molecular pathogenesis of the disease, the 2016 version of the WHO classification for myeloid neoplasms widely adopted gene mutation patterns for disease classification [3].

However, needless to say, the initial diagnosis of MDS is still, in principle, based on cell morphology and numbers, in which the disease is diagnosed by the following criteria: the ratio of myeloblasts is less than 5% in the bone marrow and less than 1% in the peripheral blood; and cytopenia or dysplasia is seen in at least one lineage [3], although the patterns and degree of dysplasia in peripheral blood cells, *per se*, seem not to be associated with an MDS subtype [4]. Nevertheless, morphological discrimination between normal and MDS cells is often very difficult, and thus requires substantial experience. For example, though dysplasia of MDS is defined as having a morphological abnormality in 10% or more of hematopoietic cells in a smear sample [3], other diseases or reactive conditions, such as various infections by pathogens or the administration of drugs, including anticancer agents, sometimes display dysplastic features resembling MDS [5,6]. Hematological disorders, including aplastic anemia and primary myelofibrosis,

* Corresponding author at: Department of Molecular Laboratory Medicine, Faculty of Life Sciences, Kumamoto University, 1-1-1 Honjo, Chuo-ku, Kumamoto, 860-8556, Japan.
E-mail address: hmatsui@kumamoto-u.ac.jp (H. Matsui).

https://doi.org/10.1016/j.leukres.2018.04.003
Received 15 February 2018; Received in revised form 23 March 2018; Accepted 3 April 2018
Available online 09 April 2018

also occasionally give rise to cells morphologically close to MDS. Furthermore, a certain number of healthy people exhibit dysplasia in hematopoietic cells [7]. Therefore, subtle changes in cell morphology may dissuade clinicians and medical technologists from a confident judgement [8], possibly giving rise to the variance observed in MDS diagnoses among observers and hospitals [9,10].

In this setting, the board for the standardization of cell morphology in the Japanese Society for Laboratory Hematology (JSLH) is working to standardize the definition of dysplasia. In Japan, Matsuda et al. proposed classifying dysplasia into two categories: (groups A and B), in which group A includes dysplastic features highly characteristic of MDS, including a pseudo Pelger–Huët anomaly, cytoplasmic hypo–granularity (a decrease of cytoplasmic granules by more than 80% compared to normal granular cells), micromegakaryocytes and ring sideroblasts; while group B includes other dysplastic features less specific to MDS [11]. This categorization was also adopted by the study group for idiopathic hematopoietic disorders in the Japanese Ministry of Health, Labour and Welfare (MHLW). Likewise, the International Working Group on the Morphology of MDS (IWGM–MDS) proposes that the presence of a pseudo Pelger–Huët anomaly, cytoplasmic hypo–granularity (decrease of granules by more than two-thirds), abnormal chromatin clumping and macrocytes in granulocytes should be taken into account for dysplasia in MDS [12].

Of course, the quality of smear samples impacts on the judgement of dysplasia. Factors that determine the quality of samples include the application of peripheral blood or bone marrow specimens to glass slides, including their drying and staining, and the camera settings for taking cell images (when stored as digital files and observed through a display monitor). As for sample staining, according to our experience, the Wright–Giemsa staining method tends to stain cytoplasmic granules of neutrophils weakly compared to May–Grünward–Giemsa staining. Thus the former staining may lead to overlooking a decrease in the number of cytoplasmic granules. Collectively, it is very important to standardize the procedure for the preparation of smear samples among institutes/hospitals. The systematic education of clinicians and medical technologists is also crucial so that they are able to evaluate the dysplasia of MDS appropriately. In Japan, medical technologists for hematology are certified by the JSLH, with 1234 medical technologists have been certified as of 2017. However, certified technologists are scarce, especially in regional cities; thus, clinicians and medical technologists who lack relevant experience often have no choice but to evaluate the dysplasia of MDS without input from more experienced personnel.

Under these circumstances, we have developed a cell classification system that can assist in the identification of MDS cells present in peripheral blood. This project is ultimately aimed at the implementation of software that distinguishes subtle dysplasia present in MDS cells utilizing a machine learning concept. Although the automatic identification of dysplastic cells from blood specimens using an automated hematopoietic cell counting system has been attempted previously [13,14], as far as we know, a system that identifies such cells using machine learning is lacking. To this end, we first obtained a total 97,924 cell images from 499 peripheral blood smear samples from patients with MDS (note, however, that the images unavoidably included non-malignant myeloid and non-myeloid cells because they were automatically taken by an imaging device, as described in Section Materials and methods). Each image was subsequently evaluated by at least two medical technologists (2–11, mean 5.67) who majored in hematology, and the data used as a “guide” for machine learning. However, we had previously found variation among observers for about 40% of the images. In this manuscript, we first disclose the evaluations entered by observers and discuss the issues that need to be overcome to realize an automatic diagnostic decision support system.

2. Materials and methods

2.1. Sample preparation and obtaining digital images of peripheral blood cells

A total of 499 peripheral blood smear samples from patients with MDS, which had been prepared for clinical usage at each hospital that participated in this study, were anonymized and collected at the Department of Laboratory Medicine in Kumamoto University Hospital. MDS diagnoses were made at each hospital. In this study, we put our focus on the evaluation of dysplasia in individual hematopoietic cells. Therefore, we collected samples regardless of the MDS subtype, so that we could collect as many samples as possible. Two hundred cell images for each sample, which included all types of nucleated cell lineages, were automatically taken using a CellaVision DM96 digital morphology imaging system (CellaVision AB, Lund, Sweden). The images were stored in JPEG format (360 × 363 pixels, 72 ppi), without having the device apply an automatic classification. Since smear samples that were covered with coverslips could not be screened by CellaVision, the cell images of such samples were manually captured under a light microscope. After the manual exclusion of images that were obviously not suitable for being classified, a final total of 97,924 cell images were initially used for the study.

The use of peripheral blood smear samples for this purpose was approved by the ethics committee at Kumamoto University, and the study was performed in accordance with the Declaration of Helsinki.

2.2. Classification of cells by medical technologists

The entering of data on cell classifications was executed through an original web-based system. A total 71 medical technologists, who mainly belonged to the Kyushu regional department of the JSLH, were recruited as observers in this study; the participants are listed in the Supplemental Table. They were individually issued with an user account to securely access websites to which cell images were uploaded, and were randomly assigned 1–76 (mean 42.5, median 52) smear samples (200 cells in each sample) and blinded to clinical data. The numbers of observers for the 499 MDS samples are shown in Supplemental Fig. S1 (mean observers/cell image is 5.67). The observers were asked to select a cell category from choices as shown in the Table 1. In addition, they were given an opportunity to check and correct their classification for 200 images again before final submission of their decision on each images of a smear specimen. Therefore, they had a chance to compare morphological features in a cell with other cells in the same specimen. Since we used peripheral blood smear samples in this study, it was expected that dysplastic cells seen in the samples were highly biased toward granulocytes (neutrophils and monocytes). Therefore, we put the choices of pseudo Pelger–Huët anomaly and cytoplasmic hypo–granularity in category A, and of micro- or giant-neutrophils, hyper-segmented neutrophils, pseudo Chédiak–Higashi granules and Auer bodies in category B of the classifications proposed by the Japanese MHLW study group for idiopathic hematopoietic disorders [11].

The categorizations entered by observers were statistically analyzed using R 2.8.1 (https://www.r-project.org/) and modified R commander software (http://www.rcommander.com/). Siegel’s kappa coefficient test was applied for the statistical analysis of inter-observer variance.

3. Results

3.1. Overall classification of cell images

Among the 97,924 cell images independently classified by at least two observers, 880 images were classified as “platelet”, “no cell” or “non-typable” by more than 80% of observers. Although we omitted such images before providing them to observers, inappropriate images