

Reproducibility of the WHO histological criteria for the diagnosis of Philadelphia chromosome-negative myeloproliferative neoplasms

Umberto Gianelli¹, Anna Bossi², Ivan Cortinovis², Elena Sabattini³, Claudio Tripodo⁴, Emanuela Boveri⁵, Alessia Moro⁶, Riccardo Valli⁷, Maurilio Ponzoni⁸, Ada M Florena⁴, Giulio F Orcioni⁹, Stefano Ascani¹⁰, Emanuela Bonoldi¹¹, Alessandra Iurlo¹², Luigi Gugliotta¹³ and Vito Franco⁴

¹UOC di Anatomia Patologica, Dipartimento di Fisiopatologia medico-chirurgica e dei trapianti, Università degli Studi di Milano, Fondazione IRCCS Cà Granda—Ospedale Maggiore Policlinico, Milano, Italy; ²Dipartimento di Scienze Cliniche e di Comunità, Università degli Studi di Milano, Milano, Italy; ³Unità Operativa di Emolinfopatia, Azienda Ospedaliero-Universitaria Policlinico S.Orsola—Malpighi, Bologna, Italy; ⁴Dipartimento di Scienze per la Promozione della Salute e Materno-Infantile 'G. D'Alessandro', Sezione di Anatomia Patologica, Università degli Studi di Palermo, Palermo, Italy; ⁵Struttura Complessa di Anatomia Patologica, Fondazione IRCCS Policlinico San Matteo, Pavia, Italy; ⁶UOC di Anatomia Patologica, Azienda Ospedaliera San Paolo, Milano, Italy; ⁷UOC di Anatomia Patologica, Arcispedale Santa Maria Nuova—IRCCS, Reggio Emilia, Italy; ⁸UOC di Anatomia Patologica, Dipartimento di Oncologia, Istituto Scientifico San Raffaele, Milano, Italy; ⁹Anatomia Patologica Ospedaliera ed Universitaria, IRCCS San Martino-IST Genova, Genova, Italy; ¹⁰Istituto di Anatomia Patologica, Università degli Studi di Perugia, Azienda Ospedaliera S. Maria, Terni, Italy; ¹¹UOC di Anatomia Patologica, Ospedale 'A. Manzoni', Lecco, Italy; ¹²U.O.C. di Ematologia, Centro Trapianti di Midollo Osseo, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milano, Italy and ¹³Istituto di Ematologia 'L. e A. Seragnoli', Policlinico S.Orsola—Malpighi, Bologna, Italy

This study, performed on behalf of the Italian Registry of Thrombocythaemias (Registro Italiano Trombocitemie), aimed to test the inter-observer reproducibility of the histological parameters proposed by the WHO classification for the diagnosis of the Philadelphia chromosome-negative myeloproliferative neoplasms. A series of 103 bone marrow biopsy samples of Philadelphia chromosome-negative myeloproliferative neoplasms consecutively collected in 2004 were classified according to the WHO criteria as follows: essential thrombocythaemia ($n=34$), primary myelofibrosis ($n=44$) and polycythaemia vera ($n=25$). Two independent groups of pathologists reviewed the bone marrow biopsies. The first group was asked to reach a collegial 'consensus' diagnosis. The second group reviewed individually all the cases to recognize the main morphological parameters indicated by the WHO classification and report their results in a database. They were subsequently instructed to individually build a 'personal' diagnosis of myeloproliferative neoplasms subtype just assembling the parameters collected in the database. Our results indicate that high levels of agreement ($\geq 70\%$) have been reached for about all of the morphological features. Moreover, among the 18 evaluated histological features, 11 resulted statistically more useful for the differential diagnosis among the different Philadelphia chromosome-negative myeloproliferative neoplasms. Finally, we found a high percentage of agreement (76%) between the 'personal' and 'consensus' diagnosis (Cohen's kappa statistic >0.40). In conclusion, our results support the use of the histological criteria proposed by the WHO

Correspondence: Dr U Gianelli, MD, Division of Pathology, Department of Pathophysiology and Transplantation, University of Milan Medical School, IRCCS Ca' Granda—Maggiore Policlinico Hospital Foundation, Via Francesco Sforza 35, Milan 20122, Italy.

E-mail: umberto.gianelli@unimi.it

Received 11 July 2013; revised 16 September 2013; accepted 16 September 2013; published online 8 November 2013

classification for the Philadelphia chromosome-negative myeloproliferative neoplasms to ensure a more precise and early diagnosis for these patients.

Modern Pathology (2014) **27**, 814–822; doi:10.1038/modpathol.2013.196; published online 8 November 2013

Keywords: essential thrombocythaemia; myeloproliferative neoplasms; primary myelofibrosis; polycythaemia vera; WHO classification

Since the introduction of the WHO classification criteria for the diagnosis of Philadelphia chromosome-negative chronic myeloproliferative disorders in 2001, much emphasis has been put on the bone marrow histological analysis as an irreplaceable tool for the diagnosis of these neoplasms.¹ Such criteria, stemming from the seminal work of leading experts,^{2–5} established the boundaries among the different pathological entities classified as Philadelphia chromosome-negative chronic myeloproliferative disorders, providing indications on the morphological features to be associated with clinical and laboratory data in order to achieve a definite and reproducible diagnosis. The WHO classification criteria, revised in 2008, by means of incorporating recent molecular achievements, underscored the importance of an integrated approach in which the fine morphology has a central role.^{6,7} Prototypical is the example of megakaryocytes, whose degree of atypia, proliferation and aggregation, have been considered the pillars of the differential diagnosis of these disorders. On the other hand, the frequent implementation of blood counts in routine clinical practice implies an increasingly early diagnosis and certainly contributes to the occurrence of Philadelphia chromosome-negative myeloproliferative neoplasms cases with markedly overlapping clinical and pathological features, which could hamper a clear-cut classification. These cases, regarded as ‘unclassifiable’ in the WHO classification, account for 10–15% of all myeloproliferative neoplasms and constitute a diagnostic ‘grey zone’ virtually linking all the Philadelphia chromosome-negative myeloproliferative neoplasms. Most of the diagnosis of myeloproliferative neoplasms ‘unclassifiable’ derive by the impossibility to reach a conclusive diagnosis in the presence of clinical or morphological features that, according to the WHO classification, are borderline with different myeloproliferative neoplasms. Recently, the reproducibility of histopathology of bone marrow according to WHO classification and its usefulness for identifying different myeloproliferative neoplasms has been questioned,^{8–10} and subsequent controversies between supporters and opponents have been raised in this setting.

Provided that the correct classification of myeloproliferative disorders integrates clinical, molecular and pathological data, in this study, through the morphological analysis of bone marrow biopsy samples of a large series of Philadelphia chromo-

some-negative chronic myeloproliferative neoplasms at diagnosis, we aimed to test the inter-observer reproducibility of the histological parameters indicated in the WHO classification and propose a rational approach to the analysis of these variables in the differential diagnosis of myeloproliferative neoplasms.

Materials and methods

A series of 113 patients from four Institutions (Milano, Bologna, Reggio Emilia, Palermo) referring to the Italian Registry of Thrombocythaemias (Registro Italiano Trombocitemie, RIT) were collected for the study. Consecutive patients, with newly diagnosed Philadelphia chromosome-negative myeloproliferative neoplasms, fulfilling the WHO classification criteria and diagnosed in 2004 were included in this study; this choice was made in order to have a reliable clinical follow-up. All of the patients signed their informed consent. As our aim focused on the differential diagnosis among Philadelphia chromosome-negative myeloproliferative neoplasms, cases with advanced bone marrow fibrosis (MF-3 according to the European consensus grading system for marrow fibrosis¹¹) on the basis of the local diagnosis were excluded from the study. For each patient, bone marrow biopsy specimen at diagnosis was available and stained with hematoxylin–eosin, Giemsa and Gomori’s silver impregnation.

Two groups of pathologists reviewed all the cases. The first group composed by eight pathologists (UG, CT, MP, AMF, GFO, SA, EBon, VF) centrally reviewed the cases and reached a ‘consensus’ diagnosis at a multi-headed microscope based on the morphological analysis and the complete knowledge of the clinical data. Only cases for which a 100% agreement was reached were included in this study. A total of 103 out of 113 cases were selected and classified as follows: essential thrombocythaemia: $n = 34$, primary myelofibrosis: $n = 44$ (myelofibrosis grade 0: $n = 15$, myelofibrosis grade 1: $n = 16$, myelofibrosis grade 2: $n = 13$) and polycythaemia vera: $n = 25$ (all in the polycythaemic phase of the disease). Ten cases were excluded from the study because of the inadequacy of the material ($n = 6$) or because they have been classified as myeloproliferative neoplasm ‘unclassifiable’ ($n = 2$) and as prepolycythaemic phase of polycythaemia vera ($n = 2$). Table 1 summarizes the most important clinical data of the study population.

Table 1 Clinical data of the study population

	ET (n = 34)	PMF (n = 44)	PV (n = 25)
Female	50%	60%	44%
Male	50%	40%	56%
Median age	56 years (21–80)	68 years (39–93)	65 years (54–82)
Haemoglobin > 16.5 g/dl	—	—	100%
Haemoglobin < 12 g/dl	10%	34%	—
White blood cells count > 12 × 10 ⁹ /l	5%	15%	10%
Anaemia and leukocytosis	—	5%	—
Increased spleen size	—	62%	30%
Elevated lactate dehydrogenase	10%	61%	70%
Leukoerythroblastic blood picture	—	52%	16%
JAK2 ^{V617F} mutation	54%	55%	92%

The second group of four pathologists (ES, EBov, AM, RV) were asked to review individually all the 103 selected cases, in a blinded fashion, without any knowledge of the ‘consensus’ diagnosis made by the first group and of the clinical data, with the exception of the patient’s age. Their aim was to recognize and register the main morphological parameters included in the WHO classification for each case and report their results in a database. To carry out this task, each reviewer received by mail the same box containing the original slides evaluated by the first panel of pathologists and a histological form reporting the list of the morphological features with minimal explanation on how to compile it, as reported in the following paragraphs.

The following 18 morphological variables were examined: (1) overall bone marrow cellularity in relation to patient’s age; (2) amount of erythropoiesis; (3) left-shifting erythropoiesis; (4) amount of granulopoiesis; (5) left-shifting granulopoiesis; (6) myeloid to erythroid ratio; (7) amount of megakaryocytes; (8) loose clusters of megakaryocytes, defined as aggregation of ≥ 3 megakaryocytes, albeit not in reciprocal contact; (9) dense clusters of megakaryocytes, defined as aggregation of ≥ 3 megakaryocytes in reciprocal contact; (10) pleomorphism of the clusters of megakaryocytes, defined as the presence of a cluster of ≥ 3 megakaryocytes of variable size, ranging from small to giant; (11) small megakaryocytes; (12) giant megakaryocytes (13) megakaryocyte nuclear hyperlobulation (‘stag horn-like’ morphology); (14) megakaryocyte bulbous nuclei (‘cloudy-like’ or ‘balloon-shaped’ morphology); (15) megakaryocyte naked nuclei; (16) megakaryocyte maturation defect defined as nuclear-to-cytoplasmic ratio alteration; (17) megakaryocyte nuclear dysmorphism (hyperchromatic and dysmorphic nuclei); and (18) entity of marrow fibrosis, (determined according to the European consensus grading system for marrow fibrosis¹¹). The four reviewers were instructed to classify the quantitative morphological variables (numbered 1, 2, 4, 6, 7) with three possible categories (‘reduced’, ‘normal’ or ‘increased’) and the qualitative variables (numbered 3, 5, 8–18) with two (‘absent’ or ‘present’).¹² To register a qualitative variable as ‘present’, the reviewer had

to identify that specific morphological features at least two times in the same slide.

Figure 1 represents some of these morphological features.

In order to assess how morphology alone could help to reach the correct diagnosis, the four reviewers were subsequently asked to propose a ‘personal’ diagnosis by just analyzing and assembling the data collected in their databases, without reviewing the case under the microscope.

Statistical Analysis

To investigate the relationship between morphological profile of each case, ‘personal’ and ‘consensus’ diagnosis, we performed a multiple correspondence analysis. This is a data analysis technique for categorical data, used to detect and represent underlying structures in a data set. It achieves the results by representing data as points in a low-dimensional Euclidean space. In this study, we only considered the first two axes. As a final result of multiple correspondence analysis, a two-dimensional map was produced where diagnoses, reviewers and category of each morphological feature were represented as points. The degree of accuracy of the map is given by the percentage of total variance represented on the two axes of the map. In Figure 2, the percentage of variance explained by the two axes is about 91% (horizontal axis: 63.5%; vertical axis: 27.7%).

Cohen’s kappa statistic was used to express the agreement between the ‘personal’ diagnosis made by each reviewer and ‘consensus’ diagnosis.

Asymptotic 95% confidence intervals for kappa statistic were computed according to Fleiss *et al.*¹³

A kappa of 0 means that the agreement is no better than that expected by chance alone, and kappa values of 1.00 indicate perfect agreement. By convention, a kappa between 0.81 and 1.00 is interpreted as indicating excellent agreement. Values of < 0.20, 0.21–0.40, 0.41–0.60 and 0.61–0.80 are interpreted as showing poor, fair, moderate and good agreement, respectively.¹⁴ All data processing and analyses were carried out with the SAS statistical software (version 9.2; SAS Institute, Cary, NC, USA).

Results

Agreement Among the Reviewers on the Morphological Features

Table 2 shows the percentage of patients for whom at least 3 out of 4 reviewers have classified in the same way each category of the 18 morphological variable considered. We found that high levels of crude agreement (≥ 70%) have been reached for all of the morphological features, with the exception of the presence of naked nuclei (65%). Interestingly,

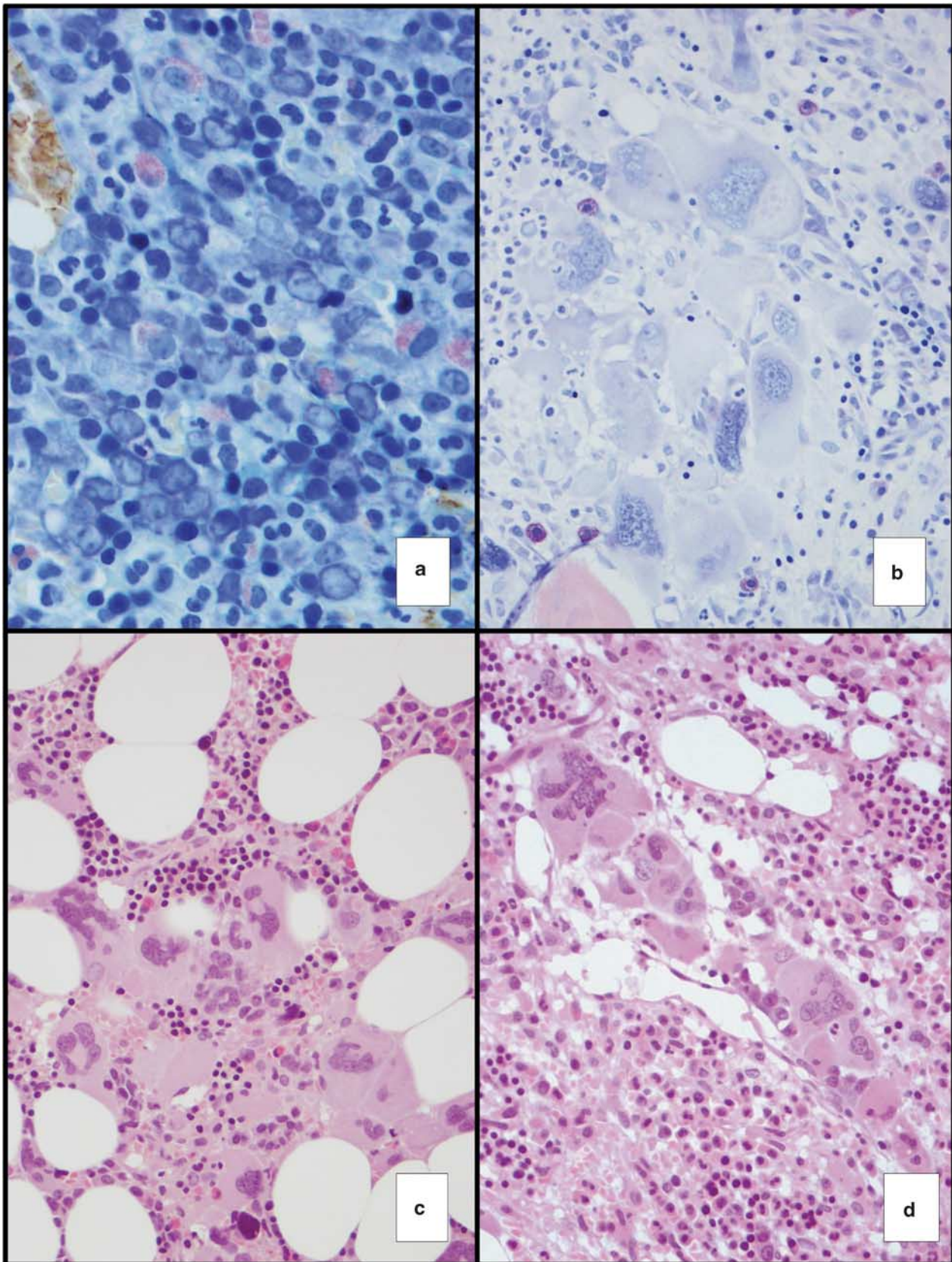


Figure 1 (a) A case of polycythaemia vera with prominent left-shifting erythropoiesis (Giemsa, $\times 40$). (b) Dense cluster of megakaryocytes sometimes with bulbous nuclei are evident in this case of primary myelofibrosis (Giemsa, $\times 20$). (c) Large-to-giant megakaryocytes with hyperlobulated nuclei in a case of essential thrombocythaemia (H&E, $\times 20$). (d) A pleomorphic cluster of megakaryocytes is evident in a case of polycythaemia vera (H&E, $\times 20$).

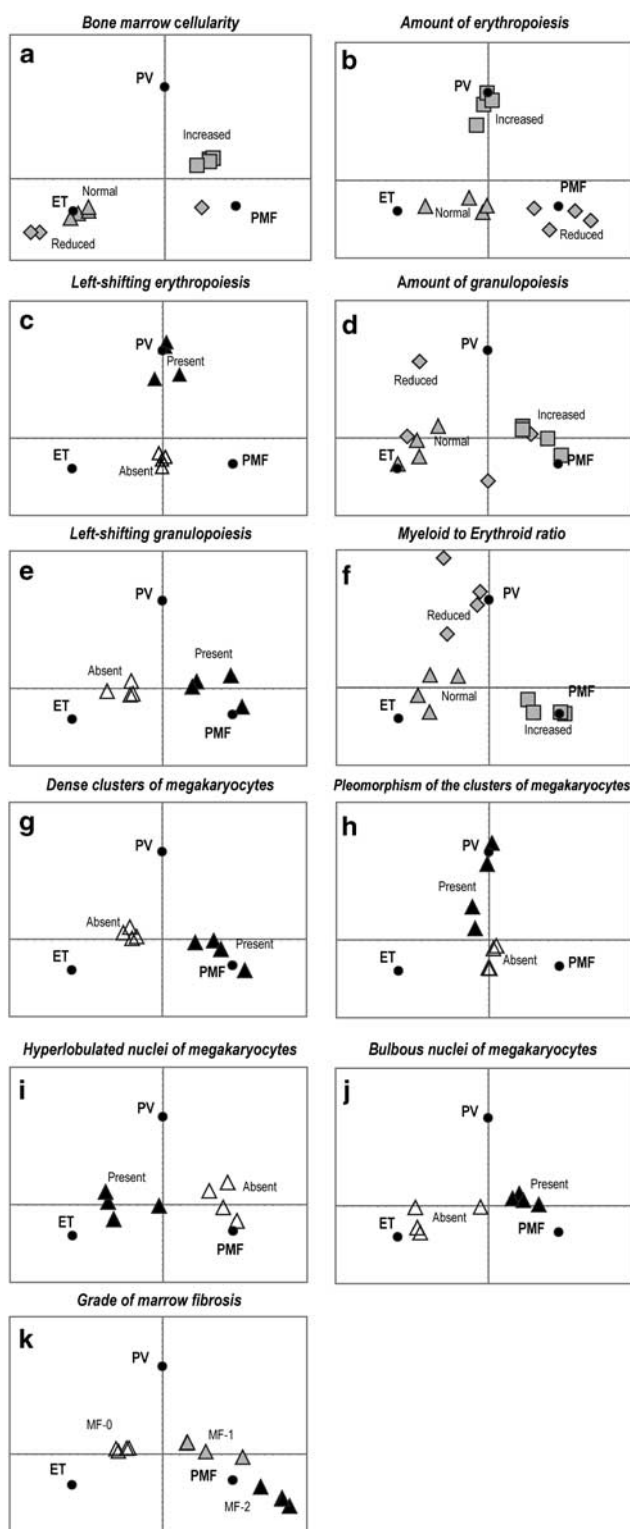


Figure 2 Two dimension graphical representation of Multiple Correspondence Analysis. The horizontal axis contrasts the morphological profiles of essential thrombocythemia (ET) patients vs those with primary myelofibrosis (PMF), while the vertical axis contrasts polycythemia vera (PV) vs ET and PMF. Quantitative variables (◆ = reduced, ▲ = normal, ■ = increased) and qualitative variables (△ = absent, ▲ = present) are represented.

the percentage of crude agreement varied among reviewers in relation to the specific disease: eg, considering the bone marrow cellularity, the percentage of agreement among reviewers was higher in cases of primary myelofibrosis and polycythemia vera (93% and 92%, respectively) than in essential thrombocythemia (83%); focusing on the myeloid-to-erythroid ratio, the percentage of agreement was higher in essential thrombocythemia or primary myelofibrosis (about 80%) than in polycythemia vera (40%). Moreover, agreement among reviewers also varied in relation to the different category of each single morphological variable: eg, considering bone marrow cellularity, the percentage of agreement in primary myelofibrosis ranged from 84% when ‘increased’ to 9% if ‘normal’; focusing on the myeloid-to-erythroid ratio, in cases of essential thrombocythemia the percentage of agreement varied from 71% when normal to 9% if increased.

Relationship and Agreement Between Morphological Features and ‘Consensus’ Diagnosis

To calculate the relationship between the reviewers’ evaluations on each morphological variable and ‘consensus’ diagnosis, we selected a subset of 11 morphological features that had turned out the most statistically useful for the differential diagnosis among the three myeloproliferative neoplasms.

The 11 morphological variables selected for all further analysis were the following: overall bone marrow cellularity, amount and left-shifting erythropoiesis, amount and left-shifting granulopoiesis, myeloid-to-erythroid ratio, dense clusters of megakaryocytes, pleomorphic clusters of megakaryocytes, hyperlobulation or bulbous appearance of the nuclei and the grading of marrow fibrosis.

Figures 2a–k shows the graphs obtained by the multiple correspondence analysis. To allow an easy reading of the graphs, the 11 morphological features were reproduced separately, and the reader should only consider the distance between the points and their positions on the plane identified by the first two axes. The horizontal axis contrasts the morphological profiles of the patients affected by essential thrombocythemia vs those by primary myelofibrosis while the vertical axis contrasts polycythemia vera vs essential thrombocythemia and primary myelofibrosis. The more different are the morphological profiles that characterize the diagnoses, the higher is the distance between the points that represent the three diagnoses. Along with this line, the higher is the agreement between reviewers for each morphological features, the lower is the distance between the points that represent them. If a category of a morphological feature and a particular diagnosis are plotted nearby, this means that such a morphological parameter is typical of that particular diagnosis.

Table 2 Absolute number (*n*) and percentage (%) of patients for whom at least 3 out of 4 reviewers have classified in the same way each category of the 18 morphological variable

<i>Morphological variable</i>	<i>Category</i>	<i>Total</i>	<i>Essential</i>	<i>Primary myelofibrosis</i>	<i>Polycythaemia vera</i>
		(<i>n</i> = 103)	<i>thrombocythaemia</i> (<i>n</i> = 34)	(<i>n</i> = 44)	(<i>n</i> = 25)
		<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)	<i>n</i> (%)
Bone marrow cellularity	Normal	27 (26)	22 (65)	4 (9)	1 (4.0)
	Increased	65 (63)	6 (18)	37 (84)	22 (88.0)
Amount of erythropoiesis	Reduced	5 (5)	—	5 (11)	—
	Normal	49 (48)	28 (82)	19 (43)	2 (8)
Left-shifting erythropoiesis	Increased	19 (18)	—	2 (5)	17 (68)
	Absent	79 (77)	33 (97)	44 (91)	6 (24)
Amount of granulopoiesis	Present	13 (13)	1 (3)	—	12 (48)
	Reduced	0 (0.0)	—	—	—
Left-shifting granulopoiesis	Normal	34 (33)	25 (74)	5 (11)	4 (16)
	Increased	50 (49)	4 (12)	35 (80)	11 (44)
Myeloid /erythroid	Absent	46 (45)	30 (88)	7 (16)	9 (36)
	Present	33 (32)	4 (12)	21 (48)	8 (32)
Amount of megakaryocytes	Reduced	2 (2)	—	—	2 (8)
	Normal	33 (32)	24 (71)	2 (5)	7 (28)
Loose clusters of megakaryocytes	Increased	37 (36)	3 (9)	33 (75)	1 (4)
	Normal	—	—	—	—
Dense clusters of megakaryocytes	Increased	103 (100)	34 (100)	44 (100)	25 (100)
	Absent	1 (1)	—	1 (2)	—
Pleomorphism of the clusters of megakaryocytes	Present	97 (94)	34 (100)	39 (89)	24 (96)
	Absent	59 (57)	29 (85)	12 (27)	18 (72)
Small megakaryocytes	Present	28 (27)	1 (3)	23 (52)	4 (16)
	Absent	74 (72)	30 (88)	39 (87)	5 (20)
Giant megakaryocytes	Present	15 (15)	1 (3)	2 (5)	12 (48)
	Absent	9 (9)	6 (18)	3 (7)	—
Hyperlobulation nuclei of megakaryocytes	Present	69 (67)	14 (41)	36 (82)	19 (76)
	Absent	3 (3)	—	3 (7)	—
Bulbous nuclei of megakaryocytes	Present	89 (86)	34 (100)	34 (77)	21 (84)
	Absent	31 (30)	2 (6)	24 (55)	5 (20)
Naked nuclei	Present	54 (52)	31 (91)	12 (27)	11 (44)
	Absent	30 (29)	26 (77)	1 (2)	3 (12)
Maturation defects	Present	49 (48)	4 (12)	32 (73)	13 (52)
	Absent	33 (32)	16 (47)	7 (16)	10 (40)
Dysmorphic (hyperchromatic) nuclei	Present	34 (33)	4 (12)	26 (59)	4 (16)
	Absent	14 (14)	9 (27)	2 (5)	3 (12)
Grade of marrow fibrosis	Present	61 (59)	12 (35)	36 (82)	13 (52)
	Absent	52 (51)	24 (71)	13 (30)	15 (60)
Grade of marrow fibrosis	Present	26 (25)	3 (9)	20 (46)	3 (12)
	MF-0	58 (56)	30 (88)	13 (30)	15 (60)
	MF-1	16 (16)	3 (9)	9 (21)	4 (16)
	MF-2	11 (11)	—	11 (25)	—

Results reported in Figures 2a–k can be summarized as follows: (1a) bone marrow cellularity: we found a good agreement among reviewers when it was regarded as ‘normal’ or ‘increased’. As expected, the ‘normal’ category was more frequently associated by the four reviewers to essential thrombocythaemia; (1b) amount of erythropoiesis: a good agreement has been found when erythropoiesis increases and this category has been more frequently associated to polycythaemia vera cases; (1c) left-shifting erythropoiesis: a good agreement has been found when this category was ‘absent’ and this modality do not support a diagnosis of polycythaemia vera; (1d) amount of granulopoiesis: a moderate agreement has been reached when ‘normal’ or ‘increased’ and lack agreement when ‘reduced’. Normal granulopoiesis

resulted more frequently associated with essential thrombocythaemia, while increased granulopoiesis with primary myelofibrosis; (1e) left-shifting granulopoiesis: good agreement has been reached when ‘absent’ and this category was more frequently observed in essential thrombocythaemia; (1f) myeloid-to-erythroid ratio: a good agreement has been reached when increased and this appeared more frequent in primary mielofibrosis, while only a moderate agreement when ‘normal’ (more frequently associated with essential thrombocythaemia) or ‘reduced’ (more frequently associated with polycythaemia vera); (1g) dense clusters of megakaryocytes: good agreement was reported when ‘absent’ and moderate-to-good agreement when ‘present’. Presence of dense clusters of megakaryocytes has

Table 3 Crude agreement between 'personal' and 'consensus' diagnosis

	Essential thrombocythaemia (n = 34)	Primary myelofibrosis (n = 44)	Polycythaemia vera (n = 25)	Total (n = 103)
Classified	20 (59%)	39 (89%)	15 (60%)	74 (72%)
Not classified	14 (41%)	5 (11%)	10 (40%)	29 (28%)

A case has been considered classified (agreement reached) when at least 3 out of 4 reviewers made the same 'personal' diagnosis corresponding to the 'consensus' one.

been observed more frequently in primary myelofibrosis; (1h) pleomorphic clusters of megakaryocytes: a good agreement was found when 'absent' and this modality do not support the diagnosis of polycythaemia vera; (1l) hyperlobulated nuclei of megakaryocytes: moderate agreement was reached for both the absent or present categories; (1j) bulbous nuclei of megakaryocytes: good agreement was reached when present (more frequent in primary myelofibrosis); and (1k) grade of marrow fibrosis: good agreement was found in grade 0 and in grade 2, while moderate agreement in grade 1.

Agreement Among 'Personal' and 'Consensus' Diagnosis

Focusing on the 11 selected morphological features, we further investigated if a morphological analysis alone was sufficient *per se* to reach a correct diagnosis of Philadelphia chromosome-negative myeloproliferative neoplasms even in the absence of clinical data.

We calculated the percentage of crude agreement between the 'personal' diagnosis and 'consensus' diagnosis. A case has been considered classified (agreement reached) when at least three out of four reviewers made the same 'personal' diagnosis corresponding to the 'consensus' one. We found that morphology alone allow to correctly classify 72% of the cases (Table 3).

Moreover, we calculated the agreement between 'personal' diagnosis of each of the four reviewers and 'consensus' diagnosis. The results indicate a higher percentage of crude agreement ranging from 65 to 84% (mean value: 76%), with moderate-to-good values of Cohen's kappa statistic (ranging from 0.41 to 0.80) (Table 4).

Discussion

The classification of Philadelphia chromosome-negative myeloproliferative neoplasms is a challenging task because of the potential clinical and pathogenic overlap at onset among the different diseases, which clearly reflects on the lack of pathognomonic morphological features; equally important, most bone marrow biopsies are presently performed earlier than in the past in almost 'asymptomatic' patients, whose abnormalities are mostly restricted to peripheral blood count values, such as thrombocytosis.¹⁵⁻¹⁸ In this study, aimed to

evaluate the inter-observer reproducibility of the morphological parameters considered for the diagnosis of Philadelphia chromosome-negative myeloproliferative neoplasms, we found an overall high level of agreement (>70%) for about all these features. Interestingly, the percentage of agreement ranged among reviewers in relation to the individual pathological entity and to the different categories of each morphological feature. It is important to note that our study is the only one which analyzes the inter-observer reproducibility of the morphological criteria of the WHO classification considering all the three Philadelphia chromosome-negative myeloproliferative neoplasms. We think that this approach could better reproduce the daily clinical practice.

WHO classification of myeloproliferative neoplasms has been considered as poorly reproducible by some groups, while others managed to reach satisfactory interobservers' agreement; even in this latter event, a variable degree of morphologically unclassifiable cases were always recorded.^{8,9,17-19}

This issue is not trivial, especially considering some recent reports, in which the prognostic impact of histology is reached when a central pathology review is asked to experts such as the authors of the WHO chapter on MPN classification.²⁰⁻²³

Although the current classification *per se* enabled to reach inconceivable achievements, the above-reported controversies disclose some drawbacks of this classification scheme of MPN.

In the present study, the lowest scores were obtained when reviewers considered myeloid-to-erythroid ratio (70%) and the presence of naked nuclei (65%), with the former reaching the highest discordance in polycythaemia vera cases. Discordance in the evaluation of the myeloid-to-erythroid ratio, particularly in PV cases, can be understood as this ratio can vary from either normal, if the two lineages increase consistently, or decreased if erythropoiesis is more increased than granulopoiesis. However, as it is relevant for a diagnosis, the use of myeloid- and erythroid-associated markers could be suggested, either always or only in cases of uncertainty. Alternatively, the simultaneous evaluation of bone marrow biopsy and aspirate smears can allow an accurate estimation of the erythroid and myeloid series and their quantification. Moreover, some degree of discordance, and particularly about erythropoiesis, may be attributed to the variable quality of different staining. To this regard, here we like to underline that optimally treated bone marrow

Table 4 Percentage of crude agreement and Cohen's kappa statistic (with 95% confidence interval) between 'personal' diagnosis of each reviewer and 'consensus' diagnosis

Reviewers	<i>Personal diagnosis vs consensus diagnosis</i>		
	(n = 103)	Kappa	IC _{95%}
1	73%	0.58	0.46–0.72
2	84%	0.76	0.65–0.87
3	83%	0.73	0.62–0.84
4	65%	0.46	0.32–0.59
mean	76%		

biopsies and good quality slides and staining are mandatory prerequisites for the best possible diagnoses and that Giemsa staining can help in the evaluation of erythropoiesis and granulopoiesis. At a variance with the other groups,¹⁸ the concordance about the 'loose cluster' parameter is very high (95%), but it did not bear significance in differentiating the three myeloproliferative neoplasms.

Moreover, we selected 11 out of 18 morphological features that resulted the most statistically useful for the differential diagnosis among the three myeloproliferative neoplasms. Our analysis demonstrated that these selected parameters allow to correctly classify 72% of the cases by morphology alone. Interestingly, primary myelofibrosis is the disease for which the highest level (89%) of consensus has been reached.

To further investigate the relationship between morphological profiles identified by four reviewers and 'consensus' diagnosis, we performed a multiple correspondence analysis on the 11 selected variables. Analysis of the results in Figure 2 can help to identify the more useful parameters in the diagnosis of a myeloproliferative neoplasms; (a) cellularity: this parameter, if normal for the patient's age, seems crucial in the diagnostic pathway as all reviewers associated it with a diagnosis of essential thrombocythaemia; (b) increased amount of erythropoiesis and left shifting: when present, these were critical parameters as all reviewers associated them with a diagnosis of polycythaemia vera; (c) increased amount of granulopoiesis and left shifting: their occurrence are useful to rule out a diagnosis of essential thrombocythaemia; and (d) dense clusters of megakaryocytes: this parameter is clearly useful, because its presence is strongly suggestive for primary myelofibrosis. However, as dense clusters are infrequent, particularly in early-onset cases of PMF, it appears advisable to evaluate additional serial sections in order to increase the chance to detect them; (e) bulbous nuclei of megakaryocytes: when evident they suggest a diagnosis of primary myelofibrosis; (f) reticulin fibers: as expected, the presence of significant fibrosis frankly addresses the diagnosis towards primary myelofibrosis.

Although these figures do not allow us to design an independent histological diagnostic flow chart, it

should be taken into account that some morphological parameters (such as cellularity, amount/shifting of erythropoiesis and granulopoiesis) could be more helpful than others in reaching the correct diagnosis.

The putative critical diagnostic role played by megakaryocytes deserves to be slightly re-shaped. In fact, when the WHO classification 2001 was published, the concept that primary myelofibrosis was composed exclusively of bulbous megakaryocytes, while essential thrombocythaemia was represented almost by stag-horn ones was prevalent. Our daily experience actually suggests that megakaryocytes in these diseases are not so monomorphic and such a degree of morphological variability is always encountered, along with the presence of elements with additional morphological properties, including small elements, naked nuclei and maturation defects. The attempt should thus be to discover the typical ones in the whole biopsy setting, and this could be difficult particularly in early stage diseases.

Finally, to assess impact of subjectivity in the 'personal' diagnosis, we compared it with the 'consensus' diagnosis. Our results indicate an overall median percentage of crude agreement of 76% between the 'personal' diagnosis of each reviewers and 'consensus' diagnosis, with moderate-to-good values of Cohen's kappa statistic. These positive results certainly reflect the different experience among the four reviewers, but at the same time they confirm that it is possible to reach high levels of agreement and suggest that better results could probably be obtained by means of specific training. Taken together, these findings suggest that even if morphology alone allows the classification of most of the cases, integration with clinical and molecular data is mandatory for a correct classification of myeloproliferative neoplasms as indicated by the WHO classification.

In conclusion, the present study demonstrates that high levels of agreement could be reached among pathologists in recognizing the morphological features indicated by the WHO classification of Philadelphia chromosome-negative myeloproliferative neoplasms. Although this success reinforces the reproducibility of these histological parameters, about one-third of the cases of myeloproliferative neoplasms could not be efficiently classified by morphological examination alone. Notwithstanding these limitations, bone marrow histology should always be performed in the diagnostic work-up of myeloproliferative neoplasms, as individual parameters (in particular represented by overall cellularity, amount/shifting of erythropoiesis and granulopoiesis and bone marrow fibrosis) may be helpful in the differential diagnosis of Philadelphia chromosome-negative myeloproliferative neoplasms or have prognostic significance.^{24,25} Keeping this in mind, and without entering into the logic of supporting or opposing the WHO classification, we believe that these promising results deserve to be

further improved, possibly with the organization of larger consensus conferences, to ensure a more precise and early diagnosis for patients affected by these dramatic diseases.

Acknowledgments

This study was partially supported by the Registro Italiano Trombocitemie (RIT), which is a project promoted by the GIMEMA Foundation. The authors wish to thank the Beat Leukemia Foundation and ONLUS (www.beat-leukemia.org) for granting funding support.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

- Vardiman JW, Brunning RD, Harris NL. Chronic myeloproliferative diseases, In: Jaffe ES, Harris NL, Stein H, Vardiman JW (eds). WHO Classification of Tumours. Tumours of Haematopoietic and Lymphoid Tissues. IARC Press: Lyon, France; 2001, pp 39–41.
- Thiele J, Kvasnicka HM, Werden C, *et al*. Idiopathic primary osteo-myelofibrosis: a clinico-pathological study on 208 patients with special emphasis on evolution of disease features, differentiation from essential thrombocythemia and variables of prognostic impact. *Leuk Lymphoma* 1996;22:303–317.
- Thiele J, Kvasnicka HM, Boeltken B, *et al*. Initial (prefibrotic) stages of idiopathic (primary) myelofibrosis (MF): a clinicopathological study. *Leukemia* 1999; 13:1741–1748.
- Thiele J, Kvasnicka HM, Zankovich R, *et al*. Relevance of bone marrow features in the differential diagnosis between essential thrombocythemia and early stage idiopathic myelofibrosis. *Haematologica* 2000; 85:1126–1134.
- Thiele J, Kvasnicka HM. Chronic myeloproliferative disorders with thrombocythaemia: a comparative study of two classification systems (PVSG, WHO) on 839 patients. *Ann Hematol* 2003;82:148–152.
- Tefferi A, Thiele J, Orazi A, *et al*. Proposals and rationale for revision of the World Health Organization diagnostic criteria for polycythemia vera, essential thrombocythemia, and primary myelofibrosis: recommendations from an ad hoc international expert panel. *Blood* 2007;110:1092–1097.
- Vardiman JW, Thiele J, Arber DA, *et al*. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* 2009;114:937–951.
- Wilkins BS, Erber WN, Bareford D, *et al*. Bone marrow pathology in essential thrombocythemia: interobserver reliability and utility for identifying disease subtypes. *Blood* 2008;111:60–70.
- Brousseau M, Parot-Schinkel E, Moles MP, *et al*. Practical application and clinical impact of the WHO histopathological criteria on bone marrow biopsy for the diagnosis of essential thrombocythemia *versus* prefibrotic primary myelofibrosis. *Histopathology* 2010;56:758–767.
- Buhr T, Hebeda K, Kaloutsi V, *et al*. European Bone Marrow Working Group trial on reproducibility of World Health Organization criteria to discriminate essential thrombocythemia from prefibrotic primary myelofibrosis. *Haematologica* 2012;97:360–365.
- Thiele J, Kvasnicka HM, Facchetti F, *et al*. European consensus on grading bone marrow fibrosis and assessment of cellularity. *Haematologica* 2005;90:1128–1132.
- Thiele J, Kvasnicka HM, Diehl V. Standardization of bone marrow features—does it work in hematopathology for histological discrimination of different disease patterns? *Histol Histopathol* 2005;20:633–644.
- Fleiss JL, Cohen J, Everitt BS. Large sample standard error of kappa and weighted kappa. *Psychol Bull* 1969;72:323–327.
- Landis JR, Kock GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
- Thiele J, Kvasnicka HM, Diehl V. Initial (latent) polycythaemia vera with thrombocytosis mimicking essential thrombocythaemia. *Acta Haematol* 2005;113: 213–219.
- Gianelli U, Iurlo A, Vener C, *et al*. The significance of bone marrow biopsy and JAK2V617F mutation in the differential diagnosis between ‘early’ pre-polycythemic phase of polycythemia vera and essential thrombocythemia. *Am J Clin Pathol* 2008;30:336–342.
- Florena AM, Tripodo C, Iannitto E, *et al*. Value of bone marrow biopsy in the diagnosis of essential thrombocythemia. *Haematologica* 2004;89:911–919.
- Gianelli U, Vener C, Raviele PR, *et al*. Essential thrombocythemia or chronic idiopathic myelofibrosis? A single-center study based on hematopoietic bone marrow histology. *Leuk Lymphoma* 2006;47:1774–1781.
- Koopmans SM, Bot FJ, Lam KH, *et al*. Reproducibility of histologic classification in nonfibrotic myeloproliferative neoplasia. *Am J Clin Pathol* 2011;136:618–624.
- Barbui T, Thiele J, Passamonti F, *et al*. Survival and disease progression in essential thrombocythemia are significantly influenced by accurate morphologic diagnosis: an international study. *J Clin Oncol* 2011;29: 3179–3184.
- Barbui T, Thiele J, Carobbio A, *et al*. Disease characteristics and clinical outcome in young adults with essential thrombocythemia *versus* early/prefibrotic primary myelofibrosis. *Blood* 2012;120:569–571.
- Passamonti F, Thiele J, Girodon F, *et al*. A prognostic model to predict survival in 867 World Health Organization-defined essential thrombocythemia at diagnosis: a study by the International Working Group on Myelofibrosis Research and Treatment. *Blood* 2012;120:1197–1201.
- Barbui T, Finazzi G, Carobbio A, *et al*. Development and validation of an International Prognostic Score of thrombosis in World Health Organization-essential thrombocythemia (IPSET-thrombosis). *Blood* 2012;120: 5128–5133.
- Vener C, Fracchiolla NS, Gianelli U, *et al*. Prognostic implications of the European consensus for grading of bone marrow fibrosis in chronic idiopathic myelofibrosis. *Blood* 2008;111:1862–1865.
- Gianelli U, Vener C, Bossi A, *et al*. The European Consensus on grading of bone marrow fibrosis allows a better prognostication of patients with primary myelofibrosis. *Mod Pathol* 2012;25:1193–1202.