Detection of Neuroblastoma Cells in Bone Marrow and Peripheral Blood by Different Techniques: Accuracy and Relationship with Clinical Features of Patients

Maria Valeria Corrias,1 Lawrence B. Faulkner,5 Angela Pistorio,2 Cristina Rosanda,3 Francesco Callea,6 Maria Serena Lo Piccolo,3 Paola Scaruffi,6 Cinzia Marchi,5 Laura Lacitignola,5 Marzia Occhino,1 Claudio Gambini,4 Gian Paolo Tonini,7 Riccardo Haupt,2 Bruno De Bernardi,3 Vito Pistoia,1 and Alberto Garaventa3

1Laboratory of Oncology, 2Service of Epidemiology and Biostatistics, 3Department of Hematology-Oncology, and 4Service of Pathology, Istituto Gaslini, Genoa, Italy; 5Azienda Ospedaliera Meyer, Florence, Italy; 6Service of Pathology, Ospedali Civili di Brescia, Brescia, Italy; 7Laboratory of Tumor Genetics, Istituto Nazionale per la Ricerca sul Cancro, Genoa, Italy

ABSTRACT

Purpose: Detection of metastatic tumor cells in bone marrow (BM) and peripheral blood (PB) of children with neuroblastoma is crucial for prognosis and planning of therapy. Aims of this large descriptive repeated survey were to evaluate the diagnostic accuracy of different techniques in diagnostic samples obtained at several disease course time points and to correlate positive results with patient clinical features and outcome.

Experimental Design: BM aspirates, trephine biopsies, PB, and peripheral blood stem cell (PBSC) samples from Italian children with neuroblastoma were analyzed by morphological and histotechniques, as well as by immunocytochemistry (IC) for disialoganglioside GD2 and reverse transcription-PCRs (RT-PCRs) for tyrosine hydroxylase (TH) and pgp9.5 genes. The diagnostic odd ratio (DOR) was used to measure the accuracy of the different techniques.

Results: A total of 2,247 evaluations were done on 561 BM, 265 PB, and 69 PBSC samples from 247 patients. IC showed the best accuracy. Whereas TH RT-PCR accuracy was satisfactory, that of pgp9.5 was very low. Positive results obtained by IC in BM and PB samples at diagnosis from stage 1, 2, and 3 patients correlated with unfavourable outcome. No correlation was found between positive results obtained by IC or TH RT-PCR in BM, PB, and PBSC samples from stage 4 patients and their outcome.

Conclusions: Because of its elevated diagnostic accuracy, IC may represent a useful adjunct to conventional morphological techniques, especially in view of its potential prognostic role in patients with localized disease. Longitudinal multicenter studies are warranted to definitely establish the clinical usefulness of TH RT-PCR.

INTRODUCTION

The spread of tumor cells to bone marrow (BM) is a grim prognostic indicator for patients with neuroblastoma (NB), with the exception of infants (1). The search for BM infiltration is therefore of utmost importance for both staging and therapeutic purposes. According to the International Neuroblastoma Staging System (2), BM infiltration is assessed by microscopic examination of both aspirate smears and bone trephine biopsies.

Despite the limited sensitivity of these techniques, routine use of alternative modalities for tumor cell detection has not been recommended (3). However, it has been shown that the presence of >100 NB cells of 105 BM hematopoietic cells at the end of induction therapy, as assessed by immunocytochemical analysis, and the presence of NB cells in the peripheral blood (PB) at diagnosis, as assessed by molecular analysis, are independent unfavorable prognostic factors in NB patients (4–6). Similarly, increasing evidence suggests that NB relapses may be associated with peripheral blood stem cell (PBSC) contamination (7–10), and gene-marked tumor cells present in BM grafts have been detected at relapse (11).

Thus far, two different techniques, namely immunocytochemistry (IC) and reverse transcription-PCRs (RT-PCRs), have been used to identify the small number of neuroblasts in BM or PB samples. Several markers have been tested, including GD2 (12), a disialoganglioside expressed on the surface of NB cells, which is detected by IC, tyrosine hydroxylase (TH; ref. 13), the first enzyme involved in catecholamine synthesis, and pgp9.5 (Pgp), a neuron-specific enolase terminal carboxyl-hydrolase (14), both of which are detected by RT-PCR.

However, no direct comparison of these techniques has yet been done in a large descriptive repeated survey on different types of samples collected at different times during the course of disease. In the present study, we have initially assessed the analytical sensitivity and specificity of IC for GD2 and of RT-PCR for TH and PGP genes in spiking experiments and in
BM samples from leukemia patients. Thus, independent investigators have evaluated blindly the presence of metastatic NB cells in BM, PB, and PBSC samples from NB patients using different techniques. The aims of the study were (a) to evaluate the diagnostic accuracy of different techniques in detecting the presence of metastatic NB cells and (b) to correlate positive results obtained at several disease course time points using different techniques with patient clinical features and outcome.

MATERIALS AND METHODS

Patients. All available samples consecutively centralized at the Hematology-Oncology Division of the Gaslini Institute between February 1997 and May 2000, obtained from 247 Italian children with NB enrolled in the Italian NB-97 (15) or in the Localized European Neuroblastoma Study Group 1, were evaluated.

Diagnosis and staging were done according to the International Neuroblastoma Staging System (2). After relapse or progression, if any, second line regimens were individually tailored, depending on biological and clinical features. For each patient, clinical and demographic data were available from the National Neuroblastoma Study Group Database.

Timing of Sample Collection. BM (aspirates or trephine biopsies), PB, or PBSC samples were collected after informed consent from the patients themselves or their legal guardians, according to the Helsinki declaration. Samples were collected at diagnosis, during therapy, at the time of PBSC harvesting, post-myeloablative therapy (MAT) and off-therapy according to the manufacturer’s recommended procedure. cDNA (10 μL) was separately amplified, in a final volume of 50 μL, with 2.5 units Taq polymerase (Applied Biosystem, Foster City, CA), with primers specific for TH, PGP, or the GAPDH housekeeping gene. Two μL of the first PCR reaction were then amplified with TH and PGP gene-specific primers, internal to those used in the first round. The amplification products were analyzed in a 2% agarose gel stained with ethidium bromide. Primer sequences for TH gene were as follows: outer forward, 5'-TGT-CAGACGGACAAAAGTGT-3'; outer reverse, 5'-GATATTGTCT-TCTCGGATGAC-3'; inner forward, 5'-ACCAAGTTGACCC-TGACCTGGAC-3'; inner reverse, 5'-CCCTTCAGCGTGTTG-TAGACCTCC-3'.

Primer sequences for PGP gene were as follows: outer forward, 5'-AGATCAACCCCGAGATGCTGAAACAGTTG-CTG-3'; outer reverse, 5'-ATTAGGCTCTTGAGACGAGG-CCACGGCAGAGAA-3'; inner forward, 5'-GCTGCTTGTCC-CCTCAC-3'; inner reverse, 5'-CACCGGAAAGGCCATTCG-3'.

Both rounds of amplification were done for 35 cycles with the following profiles: TH first round, 94°C 1 minute, 55°C 1 minute, and 72°C 1 minute; TH nested, 94°C 1 minute and 72°C 2 minutes; PGP first round, 94°C 1 minute, 60°C 1 minute, and 72°C 2 minutes; PGP nested, 94°C 1 minute, 60°C 1 gene minute, and 72°C 1 minute. We did PCR reactions for GAPDH using the primers supplied in the cDNA synthesis kit (Clontech), according to the manufacturer’s instructions.

Negative and positive controls, run with each set of samples, were total RNA extracted from PB mononuclear cells of healthy donors and from the IMR-32 NB cell line, respectively. In the spiking experiments, specificity of the amplified products was assessed by direct sequencing done with the Dye Terminator Cycle Sequencing kit (ABI PRISM; Perkin-Elmer Applied Biosystem, Norwalk, CT). Sequences were resolved and analyzed on the ABI 373A Sequence Apparatus (Perkin-Elmer Applied Biosystem).

RT-PCRs were considered positive when the values of densitometric scanning of the specific amplification product, done with Hewlett-Packard ScanJet 4C (Hewlett-Packard, Palo Alto, CA), was higher than mean + 2SD of the values of negative samples run in the gels. Variability among densitometric values of negative samples did not exceed 3%. The definition of positivity used herein refers to the results obtained after the nested PCRs.
Statistical Analysis. All results obtained with the different techniques were evaluated by four independent researchers (M. V. C. for RT-PCR, L. B. F. for IC, C. R. for morphology, and F. C. for histology) and collected in the National Neuroblastoma Study Group Database.

Descriptive statistics were reported in terms of medians, minimum and maximum value for quantitative variables, and absolute frequencies and percentages for qualitative data. Furthermore, 95% exact binomial confidence intervals [95% confidence interval (CI)] were calculated for frequency data. Comparison of frequency data between categories of patients (i.e., <12 months of age/≥12 months of age) were done by ch² test or Fisher’s exact test, when expected frequencies were <5.

The diagnostic accuracy of each technique was evaluated by calculating the sensitivity, the specificity, and the diagnostic odds ratio (DOR; ref. 16). The DOR of a test is the ratio of the odds of positivity in diseased subjects relative to the odds of positivity in the nondiseased. The value of DOR ranges from 0 to infinity, with higher values indicating better discriminatory test performance. We first assessed the diagnostic accuracy by using “morphology” as reference standard and then by using the results of both “morphology” and “histology” evaluations. The following definition was used: to be considered positive at the morphology and histology test, the sample had to be positive at the morphology or at the histology examination; to be considered negative, the sample had to be negative at both morphology and histology evaluations. For PB samples, the results obtained by morphology and histology in the BM sample simultaneously taken from the same patient were used as a reference standard. For each of these analyses only pairs of samples for which complete evaluation was available (i.e., test and reference standard results) were considered.

Finally, to evaluate the relationship between the test results at diagnosis and at the time of PBSC harvesting with the patient outcome, we drew survival curves separately for different stages of disease (stage 1, 2, or 3 versus stage 4, i.e., localized versus metastatic disease), using (a) life status (dead/alive) as an event of interest; (b) time since diagnosis (in years) as a time variable; and (c) test results (positive/negative), obtained with each technique, as an independent indicator. Survival curves were obtained with the Kaplan-Meier method and compared by log-rank statistics. A P value <0.05 was considered as statistically significant. We did statistical analyses using the Statistica (release 6.0, StatSoft Inc., Tulsa, OK) and the Stata (release 7.0, Stata Corporation, College Station, TX).

RESULTS

Analytical Specificity and Sensitivity of IC and RT-PCR. We assessed the specificity of IC and RT-PCR techniques preliminarily using PB samples from 25 healthy donors and 10 BM aspirates from patients in remission from NB-unrelated tumors (6 acute lymphoblastic leukemias, 1 chronic myeloblastic leukemia, and 3 acute myeloblastic leukemias). None of the samples had detectable GD₂ surface and TH mRNA expression, whereas PGP mRNA expression was detected in 1 of 25 PB samples.

Sensitivity was evaluated in spiking experiments by mixing logarithmic dilutions of the IRM-32 NB cell line with 1 × 10⁷ PB mononuclear cells from healthy donors. A sensitivity of 1 NB cell of 10⁶ total cells was detected for GD₂ IC (Fig. 1A) and PGP RT-PCR (Fig. 1B), whereas the sensitivity for TH RT-PCR was 1 NB cell of 10⁵ total cells (Fig. 1B). The sensitivity of morpholog-

Fig. 1 Representative results obtained in NB-unrelated samples and spiking experiments by IC and RT-PCRs. A, results of IC analysis on a BM aspirate from a patient with unrelated tumor (left) and in a simulated sample containing 10 IMR-32 cells of 10⁷ (middle) or 10⁶ IMR-32 cells of 10⁵ PB mononuclear cells (right). B, nested RT-PCR for TH (top) and PGP (middle) gene expression done on logarithmic dilutions of IMR-32 cells (from 10⁶ to 10⁻¹ cells) in 10⁷ PB mononuclear cells from healthy donor (Lanes 1 to 7), on a BM sample from an acute lymphoblastic leukemia patient (Lane 8), on a PB sample from a healthy donor (Lane 9), and on 10⁷ IMR-32 cells (Lane 10). GAPDH amplification products obtained by RT-PCR on the same samples are shown (bottom). M = ΦX174Hae/digest molecular weigh marker. Sizes of the expected fragments are indicated on (left side) each panel. GAPDH, glyceraldehyde-3-phosphate dehydrogenase.
ical and histologic analyses is approximately 1 NB cell of 10^3 to 10^4 total cells (17).

**Sample Characteristics.** During the study period, at least one sample was centralized from 247 NB patients. Table 1 reports the demographic, biochemical, genetic, and clinical features of the patients. Table 2 reports the type and number of samples collected at diagnosis and at different time periods during the course of disease.

Of the 414 BM morphological evaluations, 286 (69.1%) were negative, 110 (26.6%) were positive, and 18 (4.3%) were not assessable. Histologic analysis was available for 297 evaluations and resulted as negative in 176 (59.3%) cases, positive in 100 (33.7%), and not assessable in 21 (7.1%). Both morphology and histology evaluations were available for 309 examinations: 140 evaluations were negative by both examinations and 169 (54.7%) were positive by either one or both. More precisely, 44 evaluations were positive by both analyses.

With respect to IC and RT-PCR results, the not assessable evaluations accounted for 5.9 and 6.7% in BM samples and for 2 and 14.1% in PB samples, respectively.

**Relationship between Test Results and Sample Types.** The distribution of positive results obtained with the different techniques at the various time points during disease course is shown in Fig. 2. Figure 2A reports data obtained with standard morphology and histology (closed and open bars, respectively). Figure 2B reports data obtained by IC in BM and PB samples. In BM samples, the percentages of IC-positive results (Fig. 2B, open bars) were high at all time points tested, and the pattern of positivity was similar to that obtained by histology (Fig. 2A, open bars).

The percentage of positive results obtained by IC in BM samples (Fig. 2B, open bars) was significantly higher at diagnosis (37.8%; 95% CI, 29.1–47.2%) than during therapy (4.7%; 95% CI, 1–13.1%) and off therapy (3.8%; 95% CI, 0.1–19.6%). The percentage of positive results fell to 0% (95% CI, 0–21.8%) in the post-MAT whereas as expected at relapse, IC positivity was high (45.0%; 95% CI, 23.1–68.5%).

**Table 1** Demographic, biochemical, genetic and clinical features of NB patients (N = 247) and tumors

<table>
<thead>
<tr>
<th></th>
<th>N/total (%)</th>
<th>Median (min-max value)</th>
<th>N patients (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, male</td>
<td>146/247 (59.1)</td>
<td>25.1 (0.13–161.5)</td>
<td>49 (25)</td>
</tr>
<tr>
<td>Infants</td>
<td>86/247 (34.8)</td>
<td>43.6 (0.10–986)</td>
<td>19 (9.7)</td>
</tr>
<tr>
<td>Age (months)</td>
<td>25.1 (0.13–161.5)</td>
<td>68.8 (0.40–1,015)</td>
<td>31 (15.8)</td>
</tr>
<tr>
<td>VMA (mg/24 hrs)</td>
<td>74.2 (7–1,375)</td>
<td>949 (70.7–18,220)</td>
<td>80 (40.8)</td>
</tr>
<tr>
<td>HVA (mg/24 hrs)</td>
<td>121.0 (2–2,537)</td>
<td>67/199 (33.7)</td>
<td>17 (8.7)</td>
</tr>
<tr>
<td>NSE (ng/ml)</td>
<td>67/199 (33.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LDH (IU/L)</td>
<td>121.0 (2–2,537)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ferritin (ng/ml)</td>
<td>67/199 (33.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYC-N amplified †</td>
<td>49 (25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage at diagnosis (N = 196) ‡</td>
<td>19 (9.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 1</td>
<td>49 (25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 2</td>
<td>19 (9.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 3</td>
<td>31 (15.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 4</td>
<td>80 (40.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage 4S</td>
<td>17 (8.7)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| **Abbreviations:** VMA, vanillylmandelic acid; HVA, homovanillic acid; NSE, neuron-specific enolase; LDH, lactate dehydrogenase. * Normal values for healthy children <24 months of age were VMA <3.6 mg/24 hours and HVA <3.9 mg/24 hours; normal values for healthy children ≥24 months of age were VMA and HVA <6.4 mg/24 h; NSE <23 ng/ml; Ferritin 12 to 91 ng/ml; LDH 183–362 IU/L. † Amplification status was determined (according to ref. 18). ‡ Stage at diagnosis is indicated only for children for whom a sample at diagnosis was available.

**Table 2** Distribution of patients and evaluations with respect to time of collection

<table>
<thead>
<tr>
<th></th>
<th>Diagnosis</th>
<th>During therapy</th>
<th>Post-MAT</th>
<th>Off therapy</th>
<th>Relapse</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients</td>
<td>n = 196</td>
<td>n = 84</td>
<td>n = 56</td>
<td>n = 46</td>
<td>n = 51</td>
<td></td>
</tr>
<tr>
<td><strong>Evaluations</strong></td>
<td>688 (35)</td>
<td>819 (66)</td>
<td>211 (10)</td>
<td>343 (17)</td>
<td>186 (7)</td>
<td>2247 (135)</td>
</tr>
<tr>
<td>Morphology</td>
<td>172 (10)</td>
<td>104 (6)</td>
<td>36 (1)</td>
<td>66 (1)</td>
<td>36 (0)</td>
<td>414 (18)</td>
</tr>
<tr>
<td>Histology</td>
<td>75 (8)</td>
<td>89 (4)</td>
<td>36 (3)</td>
<td>65 (3)</td>
<td>32 (3)</td>
<td>297 (21)</td>
</tr>
<tr>
<td>IC BM</td>
<td>189 (9)</td>
<td>182 (14)</td>
<td>52 (3)</td>
<td>87 (7)</td>
<td>48 (0)</td>
<td>558 (33)</td>
</tr>
<tr>
<td>TH BM</td>
<td>44 (0)</td>
<td>79 (8)</td>
<td>27 (0)</td>
<td>32 (1)</td>
<td>20 (2)</td>
<td>202 (11)</td>
</tr>
<tr>
<td>PGP BM</td>
<td>44 (1)</td>
<td>79 (11)</td>
<td>27 (0)</td>
<td>32 (2)</td>
<td>20 (2)</td>
<td>202 (16)</td>
</tr>
<tr>
<td>IC PB</td>
<td>120 (1)</td>
<td>67 (3)</td>
<td>15 (0)</td>
<td>27 (1)</td>
<td>20 (0)</td>
<td>249 (5)</td>
</tr>
<tr>
<td>TH PB</td>
<td>22 (2)</td>
<td>39 (7)</td>
<td>9 (1)</td>
<td>17 (1)</td>
<td>5 (0)</td>
<td>92 (11)</td>
</tr>
<tr>
<td>PGP PB</td>
<td>22 (4)</td>
<td>39 (8)</td>
<td>9 (2)</td>
<td>17 (1)</td>
<td>5 (0)</td>
<td>92 (11)</td>
</tr>
<tr>
<td>IC PBSC</td>
<td>65 (5)</td>
<td>65 (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TH PBSC</td>
<td>38 (0)</td>
<td>38 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PGP PBSC</td>
<td>38 (0)</td>
<td>38 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>688 (35)</td>
<td>819 (66)</td>
<td>211 (10)</td>
<td>343 (17)</td>
<td>186 (7)</td>
<td>2247 (135)</td>
</tr>
</tbody>
</table>

Abbreviations: No, number of performed evaluations; N(%), number of not assessable samples.
Percentages of positivity for TH and PGP RT-PCR evaluations (Fig. 2C and D, respectively) were consistently elevated, both in PB and in BM (closed and open bars, respectively), in the absence of any correlation with morphology and/or histology.

At diagnosis, the percentage of positive results obtained by IC in PB and BM samples was significantly higher in children ≥12 months (51.3 and 63%, respectively) as compared with those <12 months of age (12.2 and 36.1%, P < 0.001). Likewise, in BM samples, positivity by morphology was significantly higher in children ≥12 months (52.9 versus 20.9%; P < 0.001). Moreover, the percentage of positive results by IC in PB samples (64.1 versus 7.3%, P < 0.001) and in BM (88.6 versus 18.5%; P < 0.001), was significantly higher in patients with stage 4 than in patients with stage 1, 2, or 3 disease.

When comparing the two classes of age group or localized versus metastatic disease, no statistically significant difference was observed in terms of percentage of positive results with PGP and TH techniques, either in PB or BM samples.

With respect to PBSC testing, all of the samples from patients <12 months and/or staged 1, 2, or 3 at diagnosis were negative by IC (Table 3). Only 7.6% of PBSC samples tested positive by IC; all were from patients ≥12 months and/or stage 4 at diagnosis (Table 3). On the contrary, TH and/or PGP RT-PCR were positive in ~50% of stage 4 patients at diagnosis and/or ≥12 months of age.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Distribution of positive results obtained in PBSC samples with respect to age and stage at diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>N evaluations/total samples (% NA)</td>
<td>N positive/total evaluations (%)</td>
</tr>
<tr>
<td></td>
<td>Age &lt;12 months</td>
</tr>
<tr>
<td>IC PBSC</td>
<td>60/65 (7.7%)</td>
</tr>
<tr>
<td>TH PBSC</td>
<td>38/38 (0 %)</td>
</tr>
<tr>
<td>PGP PBSC</td>
<td>38/38 (0 %)</td>
</tr>
<tr>
<td>IC PBSC</td>
<td>60</td>
</tr>
<tr>
<td>TH PBSC</td>
<td>38</td>
</tr>
<tr>
<td>PGP PBSC</td>
<td>38</td>
</tr>
</tbody>
</table>
Moreover, several PBSC samples from relapsing patients, which were stage 1, 2, or 3 at diagnosis, tested positive by molecular analyses (Table 3). The difference in the percentage of positive results between IC and RT-PCRs was significant \( (P < 0.001) \), whereas no statistically significant differences were found for any technique between positive and negative results in the two categories of patients \( (i.e., \) by age and tumor stage; Table 3).

**Diagnostic Accuracy of the Different Techniques.** Table 4 summarizes the total number of pairs of evaluations available for the specificity, sensitivity, and diagnostic accuracy analysis of IC, TH, and PGP evaluations, done in BM and PB samples overall and in samples collected at NB diagnosis.

We assessed specificity, sensitivity, and diagnostic accuracy of IC, TH, and PGP assays using as reference standard either the result of BM morphology or the combined results of BM morphology and histology. The highest diagnostic accuracy was observed with the IC technique, in both BM and PB samples. In fact, when morphology was used as the standard, the IC sensitivity in BM specimens was 0.96 and the specificity was 0.58, resulting in a DOR of 33.57 (Table 4). In PB samples IC sensitivity was 0.63 and the specificity was 0.92, respectively, with a DOR of 20.7 (Table 4).

The sensitivity and specificity of TH in BM samples were 0.89 and 0.67, respectively, resulting in a DOR of 16 (Table 4). In PB specimens, the sensitivity of TH was the highest, but the specificity was low, resulting in a DOR of 6 (Table 3). PGP accuracy was quite low, both in BM and PB samples, as a result of low sensitivity and specificity in all types of samples (Table 4).

Taking the results of both morphology and histology evaluations as reference standard, the IC technique was confirmed as the one with the highest DORMH (10.65 in BM and 9.2 in PB samples) and the PGP technique as the one with the lowest (Table 4). A general trend to a decrease in the DORMH values as

![Fig. 3](https://example.com) Survival curves constructed with the Kaplan-Meier method. A, stage 1, 2, or 3 patients grouped by IC results in PB samples at diagnosis. B, stage 4 patients grouped by IC results in PB samples at diagnosis.

<table>
<thead>
<tr>
<th>Reference standard, morphology</th>
<th>Reference standard, morphology + histology</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>Sensitivity *</td>
</tr>
<tr>
<td>All evaluations</td>
<td>358</td>
</tr>
<tr>
<td>TH BM</td>
<td>186</td>
</tr>
<tr>
<td>PGP BM</td>
<td>182</td>
</tr>
<tr>
<td>IC PB</td>
<td>181</td>
</tr>
<tr>
<td>TH PB</td>
<td>70</td>
</tr>
<tr>
<td>PGP PB</td>
<td>66</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>( N )</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>DOR</th>
<th>( N_{MH} )</th>
<th>Sensitivity(_{MH}) \† (95% CI)</th>
<th>Specificity(_{MH}) \† (95% CI)</th>
<th>DORMH</th>
</tr>
</thead>
<tbody>
<tr>
<td>BM</td>
<td>151</td>
<td>0.98</td>
<td>0.78</td>
<td>225.16</td>
<td>98</td>
<td>0.99 (0.92–1)</td>
<td>0.74 (0.55–0.88)</td>
<td>189.75</td>
</tr>
<tr>
<td>TH</td>
<td>44</td>
<td>0.83</td>
<td>0.31</td>
<td>2.2</td>
<td>36</td>
<td>0.83 (0.59–0.96)</td>
<td>0.33 (0.13–0.59)</td>
<td>2.5</td>
</tr>
<tr>
<td>PGP BM</td>
<td>43</td>
<td>0.71</td>
<td>0.42</td>
<td>1.76</td>
<td>35</td>
<td>0.71 (0.44–0.9)</td>
<td>0.44 (0.22–0.69)</td>
<td>1.92</td>
</tr>
<tr>
<td>IC PB</td>
<td>102</td>
<td>0.73</td>
<td>0.91</td>
<td>28.6</td>
<td>67</td>
<td>0.75 (0.60–0.86)</td>
<td>0.84 (0.6–0.97)</td>
<td>16</td>
</tr>
<tr>
<td>TH PB</td>
<td>19</td>
<td>0.80</td>
<td>0.11</td>
<td>0.5</td>
<td>17</td>
<td>0.8 (0.44–0.97)</td>
<td>0.14 (0.004–0.58)</td>
<td>0.7</td>
</tr>
<tr>
<td>PGP PB</td>
<td>17</td>
<td>0.67</td>
<td>0.5</td>
<td>2</td>
<td>15</td>
<td>0.67 (0.30–0.93)</td>
<td>0.5 (0.12–0.88)</td>
<td>2</td>
</tr>
</tbody>
</table>

Abbreviations: \( N \) = total samples evaluated by each method and morphology as reference standard; \( N_{MH} \), total samples evaluated by each method and morphology + histology as reference standard; DORMH, diagnostic odds ratio calculated by using the results of both morphology and histology as reference standard.

\* Sensitivity and specificity, both calculated using morphology as reference standard.

\† Sensitivity\(_{MH}\) and specificity\(_{MH}\) sensitivity and specificity assessed using both morphology and histology: being positive at morphology or at histology = ‘reference result: positive;’ being negative at both morphology and histology = ‘reference result: negative.’ 95% CI, binomial exact 95% confidence interval.
compared with DOR was observed for all techniques (Table 4). This trend was especially evident for TH.

When only the results obtained at diagnosis were considered, the DOR and the DOR$_{BM}$ of IC greatly increased, whereas the TH DOR$_{BM}$ was essentially unchanged compared with DOR$_{BM}$ calculated using the results of all samples (Table 4).

**Survival by International Neuroblastoma Staging System Stage and IC Test Results.** Figure 3 reports the survival analysis of NB patients with localized disease (Fig. 3A) and metastatic disease (Fig. 3B) stratified by results obtained by IC in PB at diagnosis. The survival of the few ($n = 4$) children with localized disease that tested positive at IC in PB samples was significantly different from the overall survival of the 51 patients that were negative by IC (log-rank test, $P < 0.0001$; Fig. 3A). Similar results were obtained when children with localized disease and positive IC in BM at diagnosis were compared with patients with negative IC (log-rank test $P = 0.013$).

On the contrary, among children with metastatic disease at diagnosis, no differences in terms of survival were observed between patients positive ($n = 41$) or negative ($n = 23$) by IC in PB (Fig. 3B, $P = 0.25$) or in BM (log-rank test $P = 0.87$, not shown). In addition, no difference in terms of survival were observed between patients positive or negative by TH or PGP RT-PCR at diagnosis (data not shown).

Survival analysis was done for the stage 4 patients receiving MAT, which were tested by IC ($n = 50$) and RT-PCRs ($n = 33$) at the time of PBSC harvesting. For any techniques, children who tested positive had similar probability of survival of those who were negative (log-rank test; $P = 0.59$ for TH, $P = 0.66$ for IC, and $P = 0.97$ for PGP).

**DISCUSSION**

The primary aim of this large descriptive study was to evaluate the diagnostic accuracy of different techniques for the detection of metastatic NB cells in different types of samples, taken at different times during the disease course.

The currently used techniques for neuroblastoma cell detection are specific but of limited sensitivity; thus, novel and more sensitive assays based on IC or RT-PCR have been developed. Although the results published thus far seem promising, no direct, blind comparison among morphological examination, RT-PCR, and IC has been reported.

In this study GD$_2$ IC, TH, and PGP RT-PCRs displayed similar analytical sensitivity/specifcity in simulated samples and in spiking experiments but often yielded conflicting results in relation to the type of sample, the time of sampling, and the clinical features of the patients. Discrepancies between TH and PGP results, as well as between IC and RT-PCRs, are in accordance with the heterogeneous results reported in previous studies on NB cell detection that used different methods (7, 19) or different molecular markers (19–21). It is noteworthy that the results of BM morphology and histology differed in about 60% of the BM samples tested, in accordance with the report of Aronica et al. (17).

In our study, regardless of the reference standard used and the categories of patients considered, IC showed a better DOR as compared with RT-PCRs, with both BM and PB samples. We thus believe that because of its excellent specificity, IC may become a useful adjunct to morphological and histologic analyses of BM samples.

The lower DOR of RT-PCR assays versus IC was related to the higher percentage of positive results in PB and PBSC samples from all categories of patients, as well as in BM samples from stage 4 patients during therapy. However, whereas the DOR of PGP RT-PCR was very low, indicating that this assay cannot be recommended for additional studies, the DOR of TH RT-PCR was satisfactory. This finding supports the need for further multicentric prospective studies to establish its clinical usefulness, especially because no other factors have been found to predict relapse in stage 4 patients (ref. 22, this report).

False positivity of TH RT-PCR may not be a likely event because (a) the illegitimate transcription of this gene has not been shown in this and previous (5–7, 10, 13, 23–25) studies; (b) in a previous report from some of us (26), several PBSC samples that tested negative by IC turned positive after anti-GD$_2$ immunomagnetic enrichment, indicating the presence of very few tumor cells; and (c) all of the PBSC samples positive by TH RT-PCR were from patients whose BM tested positive 3 to 4 weeks earlier.

A second, but no less important, aim of our study was to evaluate the relationship between the results obtained using different techniques with the patient clinical features and outcomes. The main conclusion was that IC positivity at diagnosis, both in PB and BM samples, significantly correlated with adverse outcome in patients with localized disease, (i.e., stage 1, 2, or 3 according to International Neuroblastoma Staging System criteria; ref. 2). This finding, however, should be taken with caution because it is based on very few patients, and larger prospective studies are needed to confirm this observation.

Our results, however, support the hypothesis that IC may represent a helpful approach to stratify NB patients with localized disease into different risk categories. Thus far only MYC-N amplification has been shown to predict a poor outcome in these patients (27). It is noteworthy that only 1 of 4 and 2 of 16 patients with localized disease, who tested positive by IC in the PB and in BM samples, respectively, showed MYC-N amplification in the primary tumor.

In NB patients with localized disease, TH RT-PCR gave positive results more frequently than IC with both BM and PB samples, but these results did not correlate with patient survival. Thus, the relevance of this observation remains to be addressed in future prospective studies.

Another important finding is that the results obtained with PBSC samples by IC or TH RT-PCR, two techniques with different DOR, did not correlate to patient survival. This lack of correlation has never been reported before for neuroblastoma patients and challenges the tenet that reinfusion of contaminated PBSC increases the likelihood of relapse. In addition, the present results raise doubts about the superiority of selected CD34 cells over unselected PBSC in reducing relapse risk (9, 28) and about the postulated antitumor effect of contaminated PBSC in neuroblastoma patients (29).

The final conclusions of this study are that IC may have a role in prognostic evaluation of patients with localized NB and that TH RT-PCR holds promise for identification of stage 4 patients at risk of relapse. These indications, however, must be
confirmed by large multicenter studies carried out using stan-
dardized procedures and strict quality control.

ACKNOWLEDGMENTS

The authors are indebted to Dr. N. K. Cheung, New York, for the
kind supply of the 5F8 monoclonal antibody, helpful discussions, and
criticism. The collaboration of Dr. A. R. Sementa and A. Cesarini,
Genoa, for photographs and figure composition is greatly appreciated.
The excellent technical support of M. Dagnino, secretarial assistance of
S. Calmanti, and data management of B. Galleni are deeply acknowl-
edged. These three collaborators, as well as M. S. Lo Piccolo and M.
Occhio, were recipients of fellowships from the Fondazione Italiana
per la Lotta al Neuroblastoma. The authors wish to thank T. J. Wiley,
Genoa, for editing supervision.

REFERENCES

(stage IV and IV-S) in the first year of life. Outcome related to age and
stage. Italian Cooperative Group on Neuroblastoma. Cancer (Phil)
1992;70:1625–33.
2. Brodeur GM, Pritchard J, Berthold F, et al. Revisions of the inter-
national criteria for neuroblastoma diagnosis, staging, and response to
3. Brodeur GM. Are bone marrow aspirates or biopsies the preferred
method to detect metastatic neuroblastoma to bone marrow? J Pediatr
4. Seeger RC, Reynolds CP, Gallego R, Stram DO, Gerbing RB,
Matthay KK. Quantitative tumor cell content of bone marrow and blood
as a predictor of outcome in stage IV neuroblastoma: a Children’s Cancer Group Study. J Clin Oncol
cells detected by reverse transcriptase polymerase chain reaction for
tyrosine hydroxylase mRNA are an independent poor prognostic indi-
cator in stage 4 neuroblastoma in children over 1 year. J Clin Oncol
2001;19:1795–1801.
may be predicted by molecular detection of minimal residual disease in
bone marrow in advanced neuroblastoma: a pilot study. J Pediatr Ha-
neuroblastoma cells in CD34+ selected peripheral stem cells using a
combination of tyrosine hydroxylase nested RT-PCR and anti-ganglio-
cells in hematopoietic grafts in neuroblastoma and Ewing’s sarcoma.
9. Tchirkov A, Kanold J, Comaret V, et al. Immunological and mo-
lecular detection of tumor contamination in leukapheresis products from
children with disseminated neuroblastoma: implication for evaluation of
tumor cell depletion after CD34 selection. Bone Marrow Transplant
the time of peripheral blood stem cell harvest in patients with advanced
11. Rill DR, Santana VM, Roberts WM, et al. Direct demonstration that
autologous bone marrow transplantation for solid tumors can return a
12. Cheung NK, Von Hoff DD, Strandjord SE, Coccia PF. Detection of
neuroblastoma cells in bone marrow using GD2 specific monoclonal
hydroxylase mRNA and minimal neuroblastoma cells by the reverse
14. Mattano LA, Moss TJ, Emerson SG. Sensitive detection of rare
cirulating neuroblastoma cells by the reverse transcriptase-polymerase
15. De Bernardi B, Nicolas B, Boni L, et al. Disseminated neuroblas-
toma in children older than one year at diagnosis: comparable results with
three consecutive high-dose protocols adopted by the Italian Co-
16. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PMM. The
17. Aronica PA, Pirotta VT, Yunis EJ, Penchansky L. Detection of
neuroblastoma in the bone marrow: biopsy versus aspiration. J Pediatr
18. Tonini GP, Boni L, Pession A, et al. MYCN oncogene amplification
in neuroblastoma is associated with worse prognosis, except in stage 4s.
19. Cheung IY, Barber D, Cheung NKV. Detection of microscopic
neuroblastoma in marrow by histology, immunocytoLOGY, and reverse
transcription-PCR of multiple molecular markers. Clin Cancer Res
20. Cheung IY, Cheung NKV. Molecular detection of GAGE expres-
sion in peripheral blood and bone marrow: utility as a tumor marker for
merase chain reaction (RT-PCR) analysis of monoamine transporters in
neuroblastoma cell lines: correlations to meta-iodobenzylguanidine
(MIBG) uptake and tyrosine hydroxylase gene expression. Eur J Cancer
22. Simon T, Hero B, Hunneman DH, Berthold F. Tumour markers are
poor predictors for relapse or progression in neuroblastoma. Eur J
23. Burchill SA, Selby PJ. Molecular detection of low-level disease in
24. Burchill SA, Bradbury FM, Smith B, Lewis JJ, Selby P. Neuroblas-
toma cell detection by reverse transcriptase-polymerase chain reaction
(RT-PCR) for tyrosine hydroxylase mRNA. Int J Cancer 1994;57:
671–5.
neuroblastoma cells in bone marrow and peripheral blood at diagnosis
by the reverse transcriptase-polymerase chain reaction for tyrosine hy-
droxylase mRNA. Cancer (Phil) 1995;75:2757–61.
studies and cell sorting-enhanced tumor-cell detection in high-risk neu-
roblastoma patients; implications for leukapheresis strategies. J Clin
is a major prognostic factor in localized neuroblastoma: results of the
French NBL 90 study. Neuroblastoma Study Group of the Societe
of highly purified autologous peripheral CD34(+) progenitor cells:
purging efficacy, hematopoietic reconstitution and long-term outcome
in children with high-risk neuroblastoma. Bone Marrow Transplant
29. Handgretinger R, Leung W, Ihm K, Lang P, Klingebiel T, Nietham-
er D. Tumour cell contamination of autologous stem cell grafts in
children with high-risk neuroblastoma: the good news? Br J Cancer
Detection of Neuroblastoma Cells in Bone Marrow and Peripheral Blood by Different Techniques: Accuracy and Relationship with Clinical Features of Patients

Maria Valeria Corrias, Lawrence B. Faulkner, Angela Pistorio, et al.