

# blood

2012 120: 1274-1281  
Prepublished online June 27, 2012;  
doi:10.1182/blood-2012-01-405555

## Identification of a 3-gene model as a powerful diagnostic tool for the recognition of ALK-negative anaplastic large-cell lymphoma

Luca Agnelli, Elisabetta Mereu, Elisa Pellegrino, Tania Limongi, Ivo Kwee, Elisa Bergaggio, Maurilio Ponzoni, Alberto Zamò, Javeed Iqbal, Pier Paolo Piccaluga, Antonino Neri, Wing C. Chan, Stefano Pileri, Francesco Bertoni, Giorgio Inghirami and Roberto Piva

---

Updated information and services can be found at:  
<http://bloodjournal.hematologylibrary.org/content/120/6/1274.full.html>

Articles on similar topics can be found in the following Blood collections  
[Lymphoid Neoplasia](#) (1192 articles)

---

Information about reproducing this article in parts or in its entirety may be found online at:  
[http://bloodjournal.hematologylibrary.org/site/misc/rights.xhtml#repub\\_requests](http://bloodjournal.hematologylibrary.org/site/misc/rights.xhtml#repub_requests)

Information about ordering reprints may be found online at:  
<http://bloodjournal.hematologylibrary.org/site/misc/rights.xhtml#reprints>

Information about subscriptions and ASH membership may be found online at:  
<http://bloodjournal.hematologylibrary.org/site/subscriptions/index.xhtml>



## Identification of a 3-gene model as a powerful diagnostic tool for the recognition of ALK-negative anaplastic large-cell lymphoma

\*Luca Agnelli,<sup>1</sup> \*Elisabetta Mereu,<sup>2</sup> Elisa Pellegrino,<sup>2</sup> Tania Limongi,<sup>2</sup> Ivo Kwee,<sup>3,4</sup> Elisa Bergaggio,<sup>2</sup> Maurilio Ponzoni,<sup>5</sup> Alberto Zamò,<sup>6</sup> Javeed Iqbal,<sup>7</sup> Pier Paolo Piccaluga,<sup>8</sup> Antonino Neri,<sup>1</sup> Wing C. Chan,<sup>7</sup> Stefano Pileri,<sup>8</sup> Francesco Bertoni,<sup>3,9</sup> Giorgio Inghirami,<sup>2,10</sup> and Roberto Piva,<sup>2,10</sup> on behalf of the European T-Cell Lymphoma Study Group

<sup>1</sup>Department of Clinical Sciences and Community Health, University of Milano; Hematology 1, Fondazione Istituto di Ricovero e Cura a Carattere Scientifico Ca' Granda Ospedale Maggiore Policlinico, Milano, Italy; <sup>2</sup>Department of Molecular Biotechnology and Health Sciences; Center for Experimental Research and Medical Studies, University of Torino, Torino, Italy; <sup>3</sup>Lymphoma and Genomics Research Program, Institute of Oncology Research, Bellinzona, Switzerland; <sup>4</sup>Dalle Molle Institute for Artificial Intelligence, Manno, Switzerland; <sup>5</sup>Pathology and Lymphoid Malignancies Units, San Raffaele Scientific Institute, Milan, Italy; <sup>6</sup>Department of Pathology, University of Verona, Verona, Italy; <sup>7</sup>Department of Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE; <sup>8</sup>Institute of Hematology and Medical Oncology, S. Orsola-Malpighi Hospital, University of Bologna, Bologna, Italy; <sup>9</sup>Lymphoma Unit, Oncology Institute of Southern Switzerland, Bellinzona, Switzerland; and <sup>10</sup>Department of Pathology and New York University Cancer Center, New York University School of Medicine, New York, NY

**Anaplastic large-cell lymphomas (ALCLs) are a group of clinically and biologically heterogeneous diseases including the ALK<sup>+</sup> and ALK<sup>-</sup> systemic forms. Whereas ALK<sup>+</sup> ALCLs are molecularly characterized and can be readily diagnosed, specific immunophenotypic or genetic features to define ALK<sup>-</sup> ALCL are missing, and their distinction from other T-cell non-Hodgkin lymphomas (T-NHLs) remains controversial. In the present study,**

**we undertook a transcriptional profiling meta-analysis of 309 cases, including ALCL and other primary T-NHL samples. Pathway discovery and prediction analyses defined a minimum set of genes capable of recognizing ALK<sup>-</sup> ALCL. Application of quantitative RT-PCR in independent datasets from cryopreserved and formalin-fixed paraffin-embedded samples validated a 3-gene model (TNFRSF8, BATF3, and TMOD1) able to**

**successfully separate ALK<sup>-</sup> ALCL from peripheral T-cell lymphoma not otherwise specified, with overall accuracy near 97%. In conclusion, our data justify the possibility of translating quantitative RT-PCR protocols to routine clinical settings as a new approach to objectively dissect T-NHL and to select more appropriate therapeutic protocols. (*Blood*. 2012; 120(6):1274-1281)**

### Introduction

Systemic anaplastic large-cell lymphomas (ALCLs) are a peripheral T cell–derived malignancy accounting for approximately 12% of all T-cell non-Hodgkin lymphomas (T-NHLs).<sup>1</sup> Originally described by Stein et al in 1985 as ALCL-expressing CD30,<sup>2</sup> its definition and relationship with other T-NHLs has undergone a series of revisions.<sup>1,3,4</sup> Based on genetic and clinical features, 2 different entities are now recognized as systemic forms, the ALK-positive (ALK<sup>+</sup>) and ALK-negative (ALK<sup>-</sup>) ALCL.<sup>1,5</sup> The first entity is characterized by recurrent chromosomal translocations involving the *Anaplastic Lymphoma Kinase (ALK)* gene, which leads to the expression and constitutive activation of anaplastic lymphoma kinase (ALK) fusion proteins.<sup>6,7</sup>

Whereas ALK<sup>+</sup> ALCL are readily diagnosed by anti-ALK Abs, the recognition of ALK<sup>-</sup> ALCL is in some instances subjective. In fact, immunophenotypic or genetic features that define ALK<sup>-</sup> ALCL precisely are missing; accordingly, ALK<sup>-</sup> ALCL has been considered a provisional entity by the World Health Organization (WHO) classification.<sup>1</sup>

Although all ALCLs display strong and diffuse immunoreactivity for CD30, the expression of this marker is not specific for ALCL. Indeed, CD30 is found also in activated nonneoplastic lymphoid cells, in a subset of peripheral T-cell lymphoma not

otherwise specified (PCTL-NOS), in Hodgkin lymphoma, and other neoplasms such as embryonal carcinoma.<sup>8</sup> At present, the diagnosis of ALCL relies on the application of a panel of Abs for B- and T cell–restricted antigens, epithelial membrane antigen (EMA), granzyme B, perforin, and T cell–restricted intracellular antigen-1 (TIA1).<sup>5</sup> ALCLs usually show an aberrant T-cell phenotype with frequent loss of the common T-cell markers such as the pan-T-cell antigens,<sup>9</sup> although in approximately 85%-90% of ALCL clonal TCR gene rearrangement can be detected by PCR. Furthermore, PAX5 negativity is critical for the differentiation of ALCL from common Hodgkin lymphoma and CD30<sup>+</sup> diffuse large B-cell lymphoma.<sup>10</sup>

Recently, chromosomal translocations affecting 6p25.3, which targets *DUSP22* and/or *IRF4* have been described in a subset of ALK<sup>-</sup> ALCLs, with predominant cutaneous involvement<sup>11,12</sup>; however, the effects on the pathogenesis of this lymphoma are still largely unknown.

ALK<sup>-</sup> ALCL distinction is supported by genetic criteria, epidemiologic data, and clinical features. The crude 5-year overall survival of this lymphoma is 49%, a value intermediate between 70% for ALK<sup>+</sup> ALCL and 32% for PTCL-NOS.<sup>1</sup> Nevertheless, when patients are stratified according to the clinical parameters

Submitted January 19, 2012; accepted June 18, 2012. Prepublished online as *Blood* First Edition paper, June 27, 2012; DOI 10.1182/blood-2012-01-405555.

\*L.A. and E.M. contributed equally to this work.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

© 2012 by The American Society of Hematology

(ie, age and/or stage), ALK<sup>+</sup> and ALK<sup>-</sup> ALCL patients display a similar prognosis in terms of failure-free and overall survival.<sup>5</sup> Accordingly, a retrospective study emphasized the primary impact of age and serum  $\beta$ 2-microglobulin levels on the prognosis of ALCL patients.<sup>13</sup> These data suggest that clinical factors other than biologic components might play a prominent role in outcome of ALCL patients and that clearcut distinction of ALK<sup>-</sup> ALCL from PTCL-NOS, in particular for those with strong and uniform CD30 expression, is essential for delivering the most appropriate therapies, thus avoiding unnecessary toxicities or suboptimal treatments.

Gene-expression profiling and comparative genomic hybridization studies have shown that ALK<sup>+</sup> and ALK<sup>-</sup> ALCL share restricted genomic signatures and/or preferential genomic aberrations.<sup>14-19</sup> Nevertheless, these studies have not identified exclusive/specific markers for ALK<sup>-</sup> ALCL. We uncovered several genes similarly expressed in ALK<sup>+</sup> and ALK<sup>-</sup> ALCL samples that are capable of distinguishing ALCL tumors from PTCL-NOS and other NHL samples, suggesting the existence of a common ALCL signature.<sup>20</sup>

Because gene-expression profiling analysis on every patient is currently impractical and not standardized for routine clinical settings, alternative strategies should be considered. Ideally, the stratification of unique entities requires simple, reproducible, and low-cost tests. While this goal has been reached for ALK<sup>+</sup> ALCL, effort is required for the unequivocal distinction of ALK<sup>-</sup> ALCL from CD30<sup>+</sup> PTCL.

In the present study, we undertook a systematic approach to profile the expression signature of a large set of primary T-NHLs and defined a minimum set of genes useful for the stratification of ALK<sup>-</sup> ALCL. Quantitative RT-PCR (RT-qPCR) analysis performed in independent datasets of cryopreserved or formalin-fixed paraffin-embedded (FFPE) samples validated the gene-expression profiling predictions and suggested the possibility of translating RT-qPCR protocols to routine clinical settings as a new avenue to defining T-NHL and to selecting more appropriate therapeutic protocols.

## Methods

### Processing of microarray data

Gene-expression data were obtained from 3 publicly available (GSE6338, GSE14879, and GSE19069) at the National Center for Biotechnology Information Gene Expression Omnibus repository (<http://www.ncbi.nlm.nih.gov/geo/>),<sup>16-18</sup> proprietary,<sup>20</sup> and unpublished datasets including T cells from healthy donors and those from ALCL, PTCL, angioimmunoblastic T-cell lymphoma (AITL), and adult T-cell leukemia/lymphoma patients.

Expression values were extracted from CEL files using Affymetrix chip definition files and manufacturer's annotations for HG-U133Plus arrays (release 31). Normalization was performed with the robust multiarray average procedure using the appropriate functions in the *affy* package for Bioconductor (<http://bioconductor.org>). To strengthen the robustness and the reproducibility of the whole analysis, the expression set has been analyzed using the *arrayQualityMetrics* function in the homonymous package in R software (<http://www.r-project.org>). Samples classified as outliers under standard analysis conditions were removed from computations.

### Microarray data analysis

Differential analysis was carried out using the Comparative Marker Selection suite available as a GenePattern module (<http://www.broad.mit.edu/genepattern>). Genes were ranked based on the value of 2-sided *t* test statistics to assess differential expression. A total of 100 000 permutations were performed to compute the significance (nominal *P* value) of the rank

assigned to each gene. The analysis was adjusted for multiple hypotheses testing using the *q* value as the statistical approach. The selected probe lists were visualized in a heat map format using Heat Map Viewer. The search of classifier genes was executed by Prediction Analysis of Microarrays, as described previously.<sup>21</sup> The optimal value of  $\Delta$  to obtain the minimum cross-validation error was chosen using a leave-one-out cross-validation process.

### Patients and case selection for RT-qPCR analysis

Cryopreserved samples of 20 PTCL-NOS and 40 ALCL (20 ALK<sup>+</sup> and 20 ALK<sup>-</sup>) patients were provided by the Universities of Leuven, Wuerzburg, Torino, Bologna, Verona, Brescia, and Napoli and by the San Raffaele Scientific Institute of Milan. T-NHLs were selected on the basis of stringent criteria: (1) lymph node biopsy site, (2) presence of at least 50% neoplastic cells, (3) RNA preservation, and (4) strong CD30 expression, presence of T cell-associated markers, granzyme B, and TIA-1 positivity and PAX-5 negativity (for ALCL cases).

FFPE tumor samples of 32 PTCL-NOS and 63 ALCL (29 ALK<sup>+</sup> and 34 ALK<sup>-</sup>) patients were from our archive. All samples were obtained at the time of diagnosis before treatment. ALCL cases were submitted to central pathologic review by a panel of 2 expert hematopathologists (S.P. and G.I.). Final diagnoses were assigned according to the criteria of the WHO classification,<sup>1</sup> as described previously.<sup>20</sup> Two unclassifiable cases were excluded from the study. Because CD30<sup>+</sup> PTCL/NOS is not yet characterized in the WHO classification, we defined this subgroup by the following features: (1) CD30 expression in more than 30% of neoplastic elements, (2) pleomorphic morphology with mainly medium-sized cells, and (3) lack of hallmark cells. Informed consent was obtained from all enrolled patients following the procedures approved by the local ethical committees of each participating institution.

### Purification of total RNA and cDNA synthesis

Total RNA from cryopreserved and FFPE samples were extracted using TRIzol reagent (Invitrogen) or the miRNeasy FFPE Kit (QIAGEN), respectively, according to the manufacturer's protocols. cDNA was obtained from 0.5  $\mu$ g of total RNA treated previously with RNase-free DNase (Roche Diagnostic) using reverse transcriptase SuperScript III and random hexamers (Invitrogen) or gene-specific reverse primers, as reported in supplemental Table 1 (available on the *Blood* Web site; see the Supplemental Materials link at the top of the online article).

### RT-qPCR

RT-qPCR was performed with a Thermal iCycler (Bio-Rad) using the Bio-Rad iQ SYBR Green Supermix according to the manufacturer's instructions. The PCR cycling conditions were as follows: 95°C for 5 minutes, followed by 40 cycles at 94°C for 10 seconds and 60°C for 30 seconds.

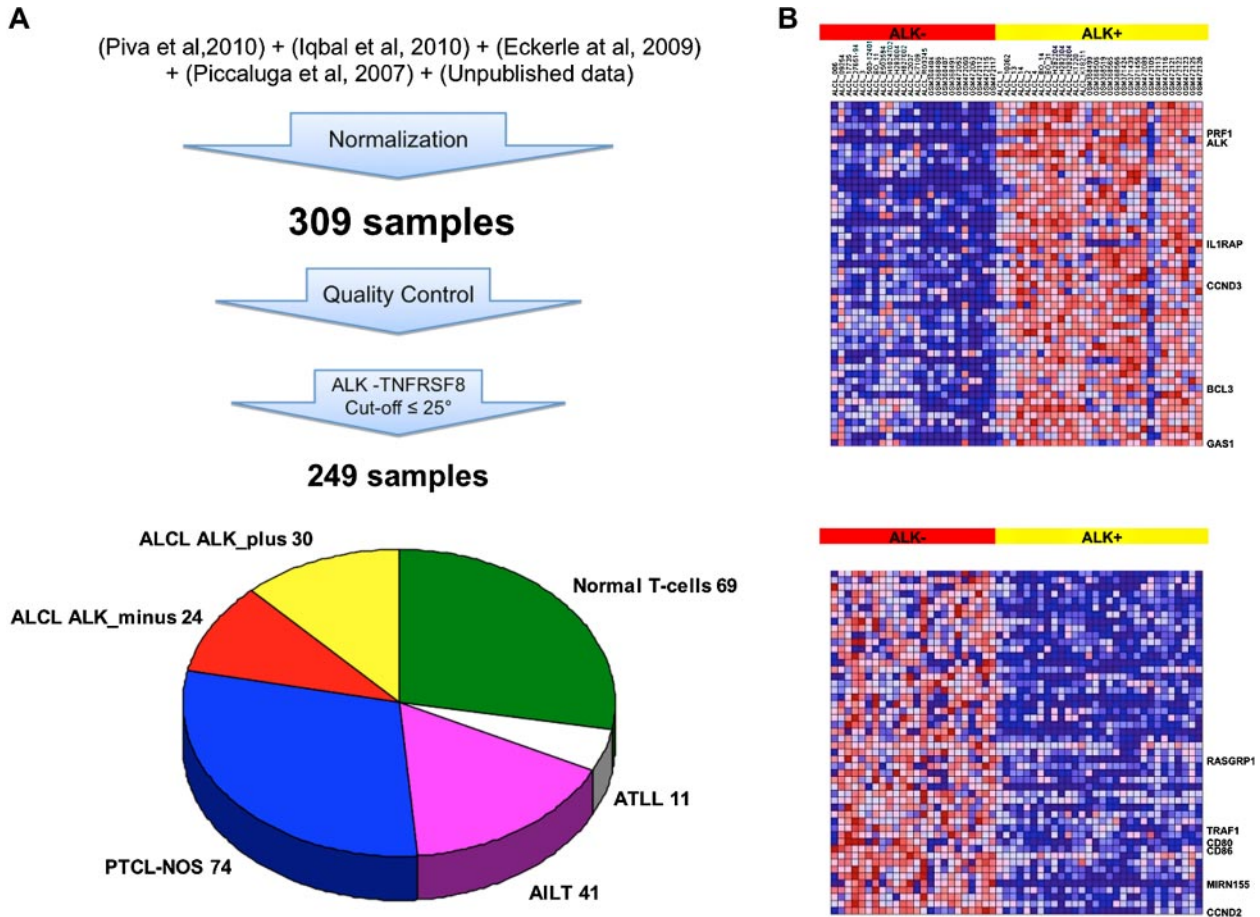
The oligonucleotide primer pairs used for RT-qPCR were designed with the PrimerBLAST (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>) to obtain amplicons of 70-110 bp. Primer sequences are reported in supplemental Table 1.

To confirm the amplification specificity, the PCR products were subjected to the analysis of melting curve, linearity, and slope of standard curve. All PCR assays were performed in triplicate. Gene-expression results were normalized to GAPDH and HUPO expressions and calculated using the  $\Delta$ Ct method according to the manufacturer's instructions.

The quality of cryopreserved and FFPE samples was assessed according to the following qPCR parameters: PTCL-NOS with GAPDH/HUPO Ct value > 30, ALCL with either GAPDH/HUPO Ct value > 30 or TNFRSF8  $\Delta$ Ct value > 5 were excluded from the analysis.

### Statistical analysis

Wilcoxon rank-sum test was used to assess differences in the distribution of 2 sample populations. The predictive power of the investigated genes was tested using the Linear Discriminant Analysis for the classification of multivariate observations, with leave-one-out procedure. ROC analysis was



**Figure 1. Generation of a large T-NHL gene-expression profiling dataset and transcriptional characterization of ALCL patients.** (A) Gene-expression data were extracted from the indicated datasets and normalized with robust multiarray average procedure. Samples classified as outlier on standard quality control analyses were removed from following computations. Expression of TNFRSF8 and/or ALK over the 25th percentile was required to include ALCL cases with higher neoplastic cell content. (B) Heat map of the top 50 genes up-regulated in ALK<sup>+</sup> and ALK<sup>-</sup> ALCL cases (30 ALK<sup>+</sup> and 24 ALK<sup>-</sup>). Each colored cell in the heat map represents the gene-expression value for a probe in a sample. The highest expression values are displayed in red, the lowest values in blue.

performed using DiagnosisMed package. Beeswarm package was used for visualizing stripchart distributions. All calculations were performed in R software.

### Calculation of metagene value

Metagene adjusted  $\Delta C_t$  value corresponds to the weighted mean of  $\Delta C_t$  values of TNFRSF8, BATF3 and TMOD1. Every single gene weight was established as the averaged scaling factors calculated on the training sets for each left-out sample, with the scaling factor defined as the value that transform each  $\Delta C_t$  to discriminant functions in the Linear Discriminant Analysis model. ROC analysis of the metagene value was performed to identify the thresholds leading to the maximum classification accuracy. The unclassifiable samples area (metagene values between 1.64 and 1.66) was interposed between the whiskers of PTCL-NOS and ALK<sup>-</sup> ALCL distributions (delimiting 1.5 times the corresponding interquartile ranges and encompassing > 95% samples). For convenience, online tool is available for smart calculation of the predicted phenotype at <http://cerms.altervista.org>.

## Results

### Generation of a large T-NHL gene-expression profiling dataset

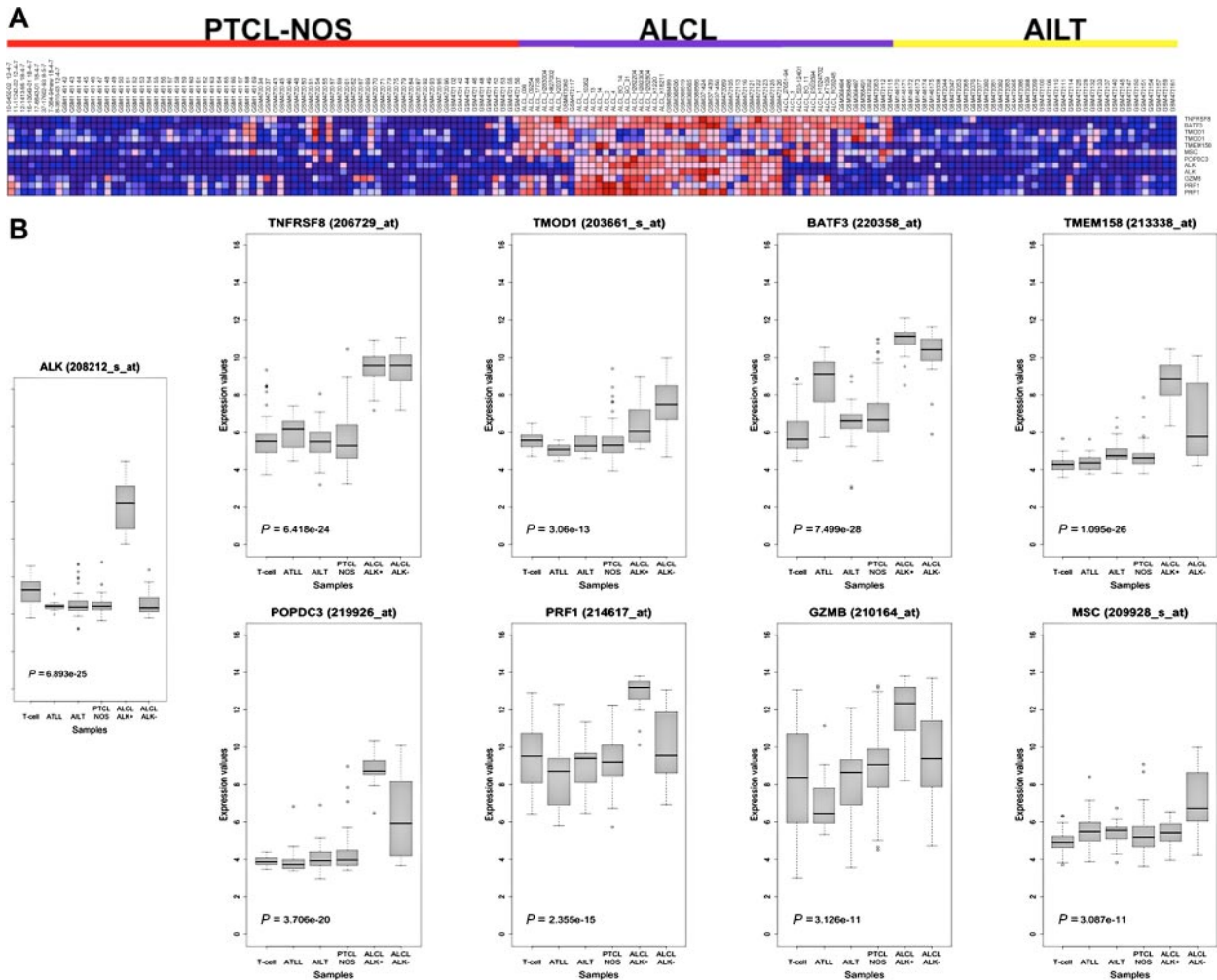
To unravel the regulatory network underlying the ALCL phenotype, and to discover new genomic lesions or biomarkers useful for the recognition of ALK<sup>-</sup> ALCL patients, we undertook a transcriptional meta-analysis assembling 4 available datasets of T-NHL

patients,<sup>16-18,20</sup> integrating them with some proprietary unpublished cases, for a total of 309 samples. Normal T-cell and T-NHL samples were included based on stringent quality controls of the array, followed by a molecular diagnostic verification as described in “Methods.” Substantial expression of CD30/TNFRSF8 and/or ALK over the 25th percentile was further required to select ALCL cases with higher neoplastic cell content, and therefore more likely to display a distinctive ALCL signature. As a matter of fact, exclusion of samples within the lower quartile led to the best false discovery rate values and to the largest number of differentially expressed genes between ALK<sup>+</sup> and ALK<sup>-</sup> samples (supplemental Figure 1). This approach reduced batch effects because of cohort- or laboratory-specific biases, and steered to the selection of 249 samples, among which 69 normal T-cells, 11 T-ALL, 41 AILT, 74 PTCL-NOS, and 54 ALCL (30 ALK<sup>+</sup>, 24 ALK<sup>-</sup>; Figure 1A).

### Identification of a minimum set of genes useful for the stratification of ALK<sup>-</sup> ALCL

In agreement with previous studies,<sup>20</sup> the application of unsupervised analysis to this large dataset was not able to distinguish ALK<sup>-</sup> ALCL from other T-NHL categories (data not shown). First, we investigated the transcriptional pattern distinguishing ALK<sup>+</sup> and ALK<sup>-</sup> ALCL in the large assembled dataset of 54 ALCL. Differential analysis identified 185 genes up-regulated in ALK<sup>+</sup>





**Figure 2. Class-prediction analysis identifies a genomic classifier for ALK<sup>-</sup> ALCL.** (A) Heat map showing the expression of the 6 genes included in the Prediction Analysis of Microarrays classifier, together with ALK, GZMB, and PRF1 in ALCL, PTCL-NOS, and AITL used in the analysis. (B) Box-plot representation of the same genes in the full dataset of 249 samples, including normal donors and adult T-cell leukemia/lymphoma tumors. The Kruskal-Wallis *P* value is indicated for each transcript.

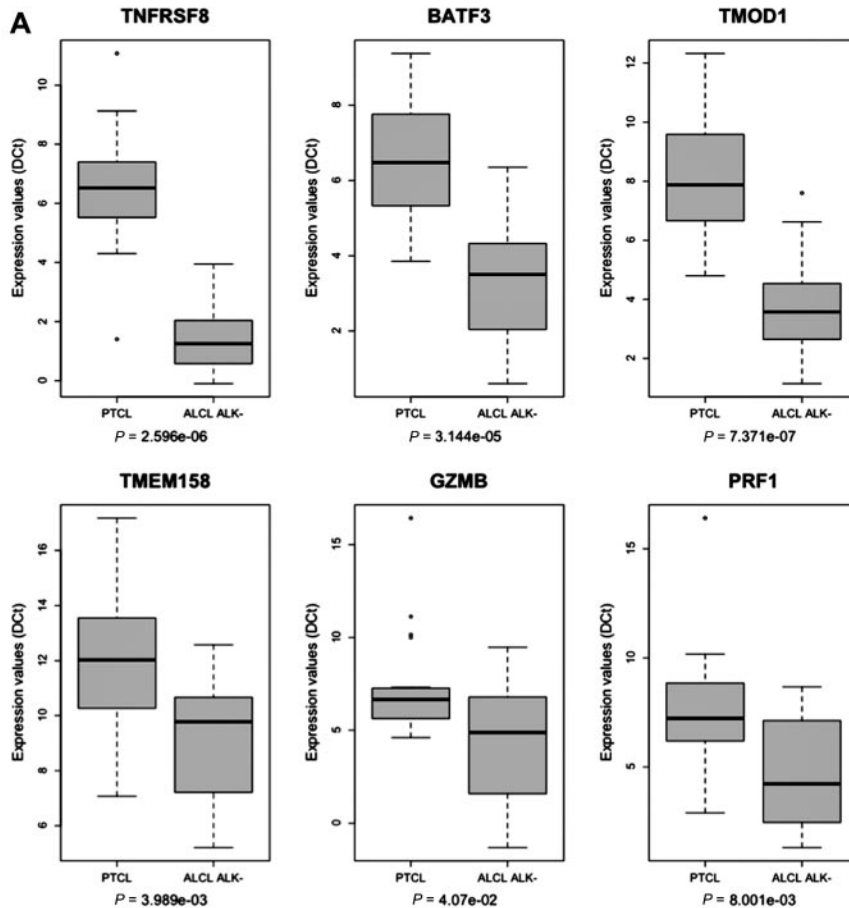
and 95 in ALK<sup>-</sup> ALCL (Figure 1B). PRF1, IL1RAP, CCND3, BCL3, and GAS1 scored within the top up-regulated transcripts of ALK<sup>+</sup> specimens (supplemental Table 2).<sup>20</sup> In ALK<sup>-</sup> samples, CD80, DC86, CCND2 and MIR155HG were among the most significant over-expressed genes (Figure 1B). It has been recently proposed that mir-155 is part of a distinct miRNA expression pattern that characterizes ALK<sup>-</sup> ALCL.<sup>22</sup> Because miR-155 is univocally processed from its noncoding *MIR155HG* host gene,<sup>23</sup> we used the surrogate expression of *MIR155HG* to assess the transcriptional level of the mature miR-155 in the large dataset of T-NHL and normal T-cells. Surprisingly, we found that mir-155 was consistently over-expressed in the majority of samples with the exception of ALK<sup>+</sup> ALCL, suggesting that mir-155 could not be used as a differential marker of ALK<sup>-</sup> ALCL (supplemental Figure 2).

To identify ALK<sup>-</sup> markers, we decided to perform a differential analysis comparing the expression profiles of either whole ALCL or ALK<sup>-</sup> ALCL subset with those of PTCL-NOS and AITL samples (supplemental Figures 3-5). TNFRSF8, BATF3, GGT1, and LGALS1 resulted among the top scorers for both ALK<sup>-</sup> and whole ALCL cases, indicating that a common ALCL signature could be dominant as result of the high heterogeneity of the ALK<sup>-</sup> subgroup.

Provided that immunohistochemical detection of ALK protein unequivocally recognizes ALK<sup>+</sup> ALCL, in the attempt to search for genes characterizing the ALK<sup>-</sup> ALCL fingerprint, we applied the Prediction Analysis of Microarrays software to the assembled dataset. The prediction led to the identification of a 6-gene classifier (TNFRSF8, BATF3, TMOD1, TMEM158, MSC, POPDC3), highly informative in discerning ALK<sup>-</sup> ALCL patients (overall error rate 4.1%). Importantly, the combined expression of the genes included in the classifier outperformed the evaluation of conventional ALCL markers such as granzyme B (GZMB) and perforin (PRF1), which are occasionally expressed also in normal T-cell and other T-NHL (Figure 2).

#### Validation of the classifier by RT-qPCR

The performance of the ALK<sup>-</sup> ALCL fingerprint was challenged using RT-qPCR in an independent set of cryopreserved samples. The expression of the 6 previously identified genes, together with GZMB and PRF1, was evaluated in 19 PTCL-NOS and 28 ALCL (13 ALK<sup>+</sup> and 15 ALK<sup>-</sup>) cases. The Wilcoxon test of RT-qPCR data established that TNFRSF8, BATF3, and TMOD1 genes were expressed the most differentially in ALK<sup>-</sup> ALCL (Figure 3A). Therefore, we verified the predictive power of the 3-gene classifier



**Figure 3. Validation of a 3-gene model classifier by RT-qPCR in cryopreserved samples.** (A) Box plot showing RT-qPCR quantification of TNFRSF8, BATF3, TMOD1, TMEM158, GZMB, and PRF1 expression in PTCL-NOS and ALK<sup>-</sup> ALCL. Statistical significance calculated using the Wilcoxon rank-sum test are shown. Expression levels are given as ΔCt. (B) Classification accuracy measures. positive predictive value, 94.7%; negative predictive value, 93.3%.

**B**

True/Predicted	PTCL-NOS	ALCL ALK <sup>-</sup>	Total
PTCL	18	1	19
ALCL ALK <sup>-</sup>	1	14	15
<b>Total</b>	<b>19</b>	<b>15</b>	

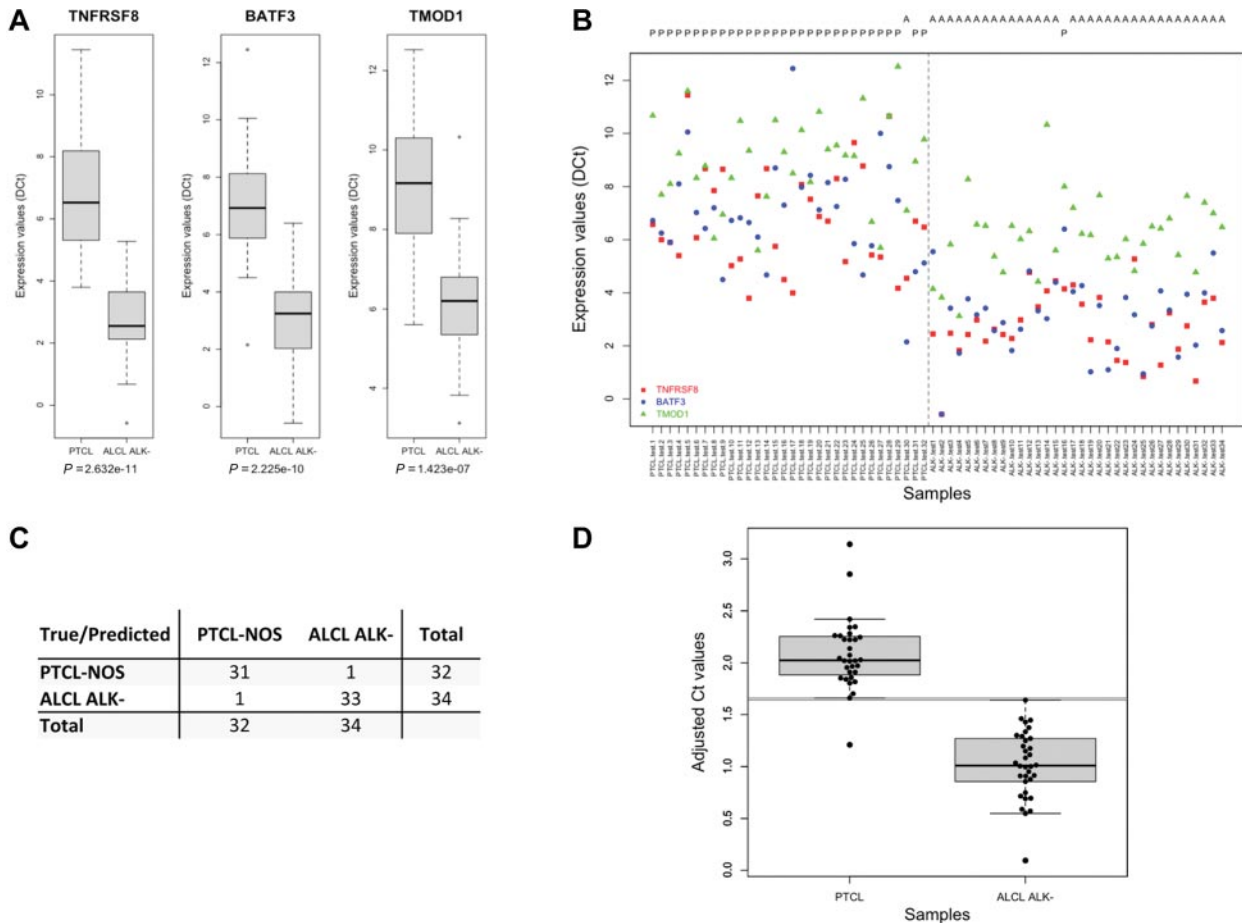
model using leave-one-out and linear discriminant analysis as the cross-validation procedure and prediction method, respectively. This classifier separated robustly either ALK<sup>-</sup> ALCL (accuracy 94.1%) or ALCL overall (accuracy 93.6%) from PTCL-NOS, with higher power than larger sets of genes (Figure 3B and data not shown).

**Translation of the classifier to routine FFPE samples**

The translation of RT-qPCR protocols to the routine is impaired by the poor RNA quality of FFPE tissues, which represent the gold standard for diagnostic procedures. To achieve this goal, we developed a multiple gene-specific retrotranscription protocol using RNA from FFPE sections. This method increased the production of cDNA targets by 250 to 1000 fold, allowing the generation of biologic relevant RT-qPCR data from 80% of FFPE samples (supplemental Figure 6). We verified that up to 10 gene-specific primers could be used in the RT reaction with no impairment of the sensitivity or specificity of the assay (data not shown). The 3-gene classifier model was then evaluated in 32 PTCL-NOS and 63 ALCL (29 ALK<sup>+</sup> and 34 ALK<sup>-</sup>) FFPE samples. Similarly to that data

observed with frozen samples, TNFRSF8, BATF3, and TMOD1 genes exhibited significantly higher expression in ALK<sup>-</sup> ALCL compared with PTCL-NOS (Figure 4A-B). More importantly, either applying leave-one-out cross validation on FFPE samples or using cryopreserved samples and FFPE as training and test set, respectively, the 3-gene model separated either ALK<sup>-</sup> ALCL (accuracy 96.97%) or ALCL overall (accuracy 95.79%) from PTCL-NOS with a very high power (Figure 4C and data not shown). In addition, the molecular classifier was challenged for the efficiency to discriminate CD30<sup>+</sup> PTCL/NOS cases because their distinction from ALK<sup>-</sup> ALCL often represents a major diagnostic challenge. In our panel, we found 7 CD30<sup>+</sup> PTCL-NOS (supplemental Table 3). The classifier correctly identified 6 of 7 cases.

Finally, for convenience in diagnostic procedures, a “metagene” was calculated as the weighted mean of the ΔCt of the 3 genes and expressed by the following equation:  $0.378 \times \Delta Ct_{TNFRSF8} + 0.303 \times \Delta Ct_{BATF3} + 0.177 \times \Delta Ct_{TMOD1}$ , as described in the “Methods” section. Therefore, the overall calculation generated a single value able to discriminate ALK<sup>-</sup> ALCL (< 1.64) from PTCL-NOS (> 1.66) with the maximum classification accuracy (Figure 4D).



**Figure 4. Validation of the 3-gene model classifier by RT-qPCR in FFPE tissues.** (A) Box-plot and (B) point-wise distribution showing RT-qPCR quantification of TNFRSF8, BATF3, and TMOD1 expressions in FFPE samples of PTCL-NOS (n = 32) and ALK<sup>-</sup> ALCL (n = 34), together with the predicted output using LDA and leave one-out procedure. P indicates PTCL; and A, ALCL. (C) Classification accuracy measures. positive predictive value, 96.9%; negative predictive value, 97.1%; (D) Box-plot and strip-chart distributions of the metagene values. Gray band indicates the unclassifiable samples area.

## Discussion

In the present study, we undertook a systematic approach to profiling the expression signatures of a large set of primary T-NHL cases, focusing our attention on systemic ALCL. Specifically, we wondered how related ALK<sup>+</sup> and ALK<sup>-</sup> ALCL are and whether ALK<sup>-</sup> ALCL could be recognized as a distinct subset. To this aim, we identified a robust diagnostic classifier able to distinguish ALK<sup>-</sup> ALCL cases from other common PTCL. Such analysis was first performed using a comprehensive and centrally revised dataset of transcriptional profiles of T-NHL generated on high-density microarrays. The predictions based on gene-expression profiling were subsequently confirmed by RT-qPCR in an independent dataset of cryopreserved samples. These analyses led to the recognition of a 3-gene classifier that successfully separated ALK<sup>-</sup> ALCL from PTCL-NOS with high power. The development of an efficient retrotranscription protocol and the generation of an easy procedure to derive a single metagene value allowed the application of molecular analyses to FFPE tissues, suggesting the possibility of translating the 3-gene model to routine clinical settings as a new approach to precisely defining T-NHL and selecting more appropriate therapeutic strategies.

PTCLs are among the most aggressive NHLs, accounting for 10%-15% of lymphoid neoplasms. Clinically, their response to conventional chemotherapy is discouraging, with 5-year relapse-

free and overall survival rates of 26% and 20%, respectively.<sup>1,24</sup> Although the new WHO classification represents a step forward in the definition of these tumors, several issues remain open. First, PTCL-NOS is considered a basket category embracing an aggressive and heterogeneous group of nodal/systemic PTCL.<sup>25</sup> Second, ALCL is a clinically and biologically heterogeneous disease including the ALK<sup>+</sup> and ALK<sup>-</sup> systemic forms.<sup>5</sup> Whereas ALK<sup>+</sup> ALCLs are molecularly characterized and can be readily diagnosed, immunophenotypic or genetic features to precisely define ALK<sup>-</sup> ALCL are missing.<sup>1</sup> Moreover, no causal molecular events leading to the transformation to ALK<sup>-</sup> ALCL have been demonstrated so far, and it is unclear whether they share common features with other PTCLs, including ALK<sup>+</sup> ALCL,<sup>26</sup> and/or with classic Hodgkin Lymphoma.<sup>27</sup> These uncertainties hamper a successful and reliable diagnostic approach in daily clinical practice.<sup>25</sup>

Several studies focusing on the molecular profiling of PTCL postulated the existence of different subtypes characterized by distinct cellular derivations.<sup>15,16,18,20,28-32</sup> However, most of these studies were largely underpowered to allow definitive diagnostic or prognostic statements. In addition, more convenient methods for the measurement of gene expression need to be developed.

We thought that assembling a larger dataset of T-NHL patients was mandatory to discovering reliable biomarkers for the recognition of ALK<sup>-</sup> ALCL patients. Stringent quality controls of the arrays in public datasets, diagnostic verifications, and detection of high-level expression of known markers, suggestive of high

percentage of neoplastic cells, were the strategies applied to increase the statistical significance of analysis. The application of prediction analysis to this large T-NHL dataset identified a simple and very robust classifier for ALK<sup>-</sup> ALCL (overall error rate, 4.1%), confirming our previous observations.<sup>20</sup> The aim of the present study was to identify a small group of genes the expression of which predicts the diagnosis of ALK<sup>-</sup> ALCL and that can be readily measured in the daily practice. The prediction model generated by gene-expression profiling analysis was validated by RT-qPCR in unrelated groups of patients. The results of this evaluation allowed the design of a model consisting of 3 genes (TNFRSF8, BATF3, and TMOD1) that distinguish ALK<sup>-</sup> ALCL from PTCL-NOS with overall accuracy near 97%. Our analysis clearly indicated that the 3 genes included in the classifier are superior to GZMB and PRF1 in the distinction of the 2 classes.

Because CD30/TNFRSF8 is the conventional (although not exclusive) marker for ALCL, it was not unexpected that TNFRSF8 scored at the top of the 3-gene model. However, its expression does not have an independent predictive power, indicating that BATF3 and TMOD1 contributions are also required. We verified that including more genes did not enhance the performance of the predictive model significantly. On the contrary, a 3-gene classifier resulted in a robust and practical model that was further simplified by the generation of an easy mathematical procedure to convert RT-qPCR values into a single output to discriminate the 2 phenotypes. Moreover, our data suggest that the predictor discriminates ALK<sup>-</sup> ALCL from CD30<sup>+</sup> PTCL-NOS, despite the high levels of CD30/TNFRSF8 expression in these patients, in all likelihood thanks to the pivotal contribution of BATF3 and TMOD1 expression.

The 3-gene model described herein could be easily applied to FFPE sections without the need for shifting or scaling of the expression data to match the values of frozen samples. Although RNA isolated from FFPE is considered a poor material for gene-expression analysis, we have shown that optimization strategies such as limited amplicon size and gene-specific retrotranscription can effectively overcome its limitations. Our experience allowed the generation of biologic relevant data from at least 80% of FFPE samples, supporting the results obtained on frozen samples and suggesting the possibility of applying RT-qPCR protocols to routine clinical settings.

Overall, our present findings support the hypothesis that ALCLs, independent of the presence of ALK fusion proteins, are closely related and may derive from a different set of progenitors compared with PTCL-NOS. In addition, the lack of an exclusive signature for ALK<sup>-</sup> ALCL leads us to speculate that this “provisional entity” could be an heterogeneous group with distinct pathogenetic defects yet to be identified. We foresee that massively parallel DNA and RNA sequencing technologies will be pivotal to a definitive classification and to an effective use of targeted therapeutics in T-NHL.<sup>11,33</sup>

In conclusion, the results of the present study led to the identification of novel diagnostic markers for more objective differential diagnosis within T-NHL. Specifically, the application of RT-qPCR protocols to FFPE tissues will allow the possibility of developing simple and cost-effective molecular diagnostic tools and reducing error and ambiguity in the stratification of T-NHL.

## Appendix

The members of the European T-Cell Lymphoma Study Group are: Barreca A, Cuccuru G, Inghirami G, Medico E, Mereu E, Pellegrino E, Spaccarotella E, Scarfò I, Piva R, Fornari A, Ferreri C, Novero D, Chilosi M, Zamó A, Facchetti F, Lonardi S, De Chiara A, Fulciniti F, Doglioni C, Ponzoni M, Agnelli L, Neri A, Todoterti K, Agostinelli C, Piccaluga PP, Pileri S, Falini B, Tiacci E, Van Loo P, Tousseyn T, De Wolf-Peeters C, Geissinger E, Muller-Hermelink HK, Rosenwald A, Piris MA, Rodriguez ME, Bertoni F, Kwee I, and Boi M.

## Acknowledgments

This work was supported by the Associazione Italiana per la Ricerca sul Cancro (grants IG-8675, 10007, and IG-4569); Regione Piemonte; Compagnia di San Paolo, Torino (Progetto Oncologia); Fondazione Italiana Ricerca sul Cancro; Oncosuisse (grant KLS-02403-02-2009); Fondazione per la Ricerca e la Cura sui Linfomi (Lugano, Switzerland); and the Nelia et Amadeo Barletta Foundation (Lausanne, Switzerland).

## Authorship

Contribution: L.A., F.B., G.I., and R.P., designed the study, interpreted the data, and wrote the manuscript; E.M., E.P., and E.B. performed the experiments; L.A., E.M., E.P., T.L., I.K., A.N., F.B., G.I., and R.P. interpreted the biologic data; M.P., A.Z., J.I., P.P.P., W.C.C., and S.P. provided well-characterized study materials; and P.P.P., S.P., and G.I. reviewed the pathology.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

The current affiliation for T.L. is Istituto Italiano di Tecnologia (IIT), Genova, Italy.

Correspondence: Roberto Piva and Giorgio Inghirami, Department of Molecular Biotechnology and Health Sciences, Center for Experimental Research and Medical Studies, University of Torino, Via Santena 5, Torino 10126 Italy; e-mail: roberto.piva@unito.it or giorgio.inghirami@unito.it.

## References

- Vose J, Armitage J, Weisenburger D. International peripheral T-cell and natural killer/T-cell lymphoma study: pathology findings and clinical outcomes. *J Clin Oncol*. 2008;26(25):4124-4130.
- Stein H, Mason DY, Gerdes J, et al. The expression of the Hodgkin's disease associated antigen Ki-1 in reactive and neoplastic lymphoid tissue: evidence that Reed-Sternberg cells and histiocytic malignancies are derived from activated lymphoid cells. *Blood*. 1985;66(4):848-858.
- Harris NL, Jaffe ES, Stein H, et al. A revised European-American classification of lymphoid neoplasms: a proposal from the International Lymphoma Study Group. *Blood*. 1994;84(5):1361-1392.
- Rizvi MA, Evens AM, Tallman MS, Nelson BP, Rosen ST. T-cell non-Hodgkin lymphoma. *Blood*. 2006;107(4):1255-1264.
- Savage KJ, Harris NL, Vose JM, et al. ALK- anaplastic large-cell lymphoma is clinically and immunophenotypically different from both ALK+ ALCL and peripheral T-cell lymphoma, not otherwise specified: report from the International Peripheral T-Cell Lymphoma Project. *Blood*. 2008;111(12):5496-5504.
- Morris SW, Kirstein MN, Valentine MB, et al. Fusion of a kinase gene, ALK, to a nucleolar protein gene, NPM, in non-Hodgkin's lymphoma. *Science*. 1994;263(5151):1281-1284.
- Chiarle R, Martinengo C, Mastini C, et al. The anaplastic lymphoma kinase is an effective oncoantigen for lymphoma vaccination. *Nat Med*. 2008;14(6):676-680.
- Fornari A, Piva R, Chiarle R, Novero D, Inghirami G. Anaplastic large cell lymphoma: one or more entities among T-cell lymphoma? *Hematol Oncol*. 2009;27(4):161-170.



9. Bonzheim I, Geissinger E, Roth S, et al. Anaplastic large cell lymphomas lack the expression of T-cell receptor molecules or molecules of proximal T-cell receptor signaling. *Blood*. 2004; 104(10):3358-3360.
10. Desouki MM, Post GR, Cherry D, Lazarchick J. PAX-5: a valuable immunohistochemical marker in the differential diagnosis of lymphoid neoplasms. *Clin Med Res*. 2010;8(2):84-88.
11. Feldman AL, Dogan A, Smith DI, et al. Discovery of recurrent t(6;7)(p25.3;q32.3) translocations in ALK-negative anaplastic large cell lymphomas by massively parallel genomic sequencing. *Blood*. 2011;117(3):915-919.
12. Wada DA, Law ME, Hsi ED, et al. Specificity of IRF4 translocations for primary cutaneous anaplastic large cell lymphoma: a multicenter study of 204 skin biopsies. *Mod Pathol*. 2011;24(4):596-605.
13. Sibon D, Fournier M, Briere J, et al. Prognostic factors and long term outcome of 138 adults with systemic anaplastic large-cell lymphoma: a retrospective study by the Groupe d'Etude Des Lymphomes De l'Adulte (GELA) [abstract]. *Blood (ASH Annual Meeting Abstracts)*. 2010;116(21):322.
14. Thompson MA, Stumph J, Henrickson SE, et al. Differential gene expression in anaplastic lymphoma kinase-positive and anaplastic lymphoma kinase-negative anaplastic large cell lymphomas. *Hum Pathol*. 2005;36(5):494-504.
15. Lamant L, de Reynies A, Duplantier MM, et al. Gene-expression profiling of systemic anaplastic large-cell lymphoma reveals differences based on ALK status and two distinct morphologic ALK+ subtypes. *Blood*. 2007;109(5):2156-2164.
16. Piccaluga PP, Agostinelli C, Califano A, et al. Gene expression analysis of peripheral T cell lymphoma, unspecified, reveals distinct profiles and new potential therapeutic targets. *J Clin Invest*. 2007;117(3):823-834.
17. Eckerle S, Brune V, Doring C, et al. Gene expression profiling of isolated tumour cells from anaplastic large cell lymphomas: insights into its cellular origin, pathogenesis and relation to Hodgkin lymphoma. *Leukemia*. 2009;23(11):2129-2138.
18. Iqbal J, Weisenburger DD, Greiner TC, et al. Molecular signatures to improve diagnosis in peripheral T-cell lymphoma and prognostication in angioimmunoblastic T-cell lymphoma. *Blood*. 2010; 115(5):1026-1036.
19. Salaverria I, Bea S, Lopez-Guillermo A, et al. Genomic profiling reveals different genetic aberrations in systemic ALK-positive and ALK-negative anaplastic large cell lymphomas. *Br J Haematol*. 2008;140(5):516-526.
20. Piva R, Agnelli L, Pellegrino E, et al. Gene expression profiling uncovers molecular classifiers for the recognition of anaplastic large-cell lymphoma within peripheral T-cell neoplasms. *J Clin Oncol*. 2010;28(9):1583-1590.
21. Agnelli L, Bicciato S, Mattioli M, et al. Molecular classification of multiple myeloma: a distinct transcriptional profile characterizes patients expressing CCND1 and negative for 14q32 translocations. *J Clin Oncol*. 2005;23(29):7296-7306.
22. Merkel O, Hamacher F, Laimer D, et al. Identification of differential and functionally active miRNAs in both anaplastic lymphoma kinase (ALK)+ and ALK- anaplastic large-cell lymphoma. *Proc Natl Acad Sci U S A*. 2010;107(37):16228-16233.
23. Tili E, Croce CM, Michaille JJ. miR-155: on the crosstalk between inflammation and cancer. *Int Rev Immunol*. 2009;28(5):264-284.
24. Campo E, Swerdlow SH, Harris NL, Pileri S, Stein H, Jaffe ES. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood*. 2011; 117(19):5019-5032.
25. Foss FM, Zinzani PL, Vose JM, Gascoyne RD, Rosen ST, Tobinai K. Peripheral T-cell lymphoma. *Blood*. 2011;117(25):6756-6767.
26. ten Berge RL, de Bruin PC, Oudejans JJ, Ossenkoppele GJ, van der Valk P, Meijer CJ. ALK-negative anaplastic large-cell lymphoma demonstrates similar poor prognosis to peripheral T-cell lymphoma, unspecified. *Histopathology*. 2003;43(5):462-469.
27. Tamaru J, Tokuhira M, Nittsu N, et al. Hodgkin-like anaplastic large cell lymphoma (previously designated in the REAL classification) has same immunophenotypic features to classical Hodgkin lymphoma. *Leuk Lymphoma*. 2007;48(6):1127-1138.
28. de Leval L, Rickman DS, Thielen C, et al. The gene expression profile of nodal peripheral T-cell lymphoma demonstrates a molecular link between angioimmunoblastic T-cell lymphoma (AITL) and follicular helper T (TFH) cells. *Blood*. 2007;109(11):4952-4963.
29. Hartmann S, Gesk S, Scholtysik R, et al. High resolution SNP array genomic profiling of peripheral T cell lymphomas, not otherwise specified, identifies a subgroup with chromosomal aberrations affecting the REL locus. *Br J Haematol*. 2010;148(3):402-412.
30. Martinez-Delgado B, Melendez B, Cuadros M, et al. Expression profiling of T-cell lymphomas differentiates peripheral and lymphoblastic lymphomas and defines survival related genes. *Clin Cancer Res*. 2004;10(15):4971-4982.
31. Piccaluga PP, Agostinelli C, Califano A, et al. Gene expression analysis of angioimmunoblastic lymphoma indicates derivation from T follicular helper cells and vascular endothelial growth factor deregulation. *Cancer Res*. 2007;67(22): 10703-10710.
32. Piccaluga PP, Agostinelli C, Zinzani PL, Baccarani M, Dalla Favera R, Pileri SA. Expression of platelet-derived growth factor receptor alpha in peripheral T-cell lymphoma not otherwise specified. *Lancet Oncol*. 2005;6(6):440.
33. Corless CL. Medicine. Personalized cancer diagnostics. *Science*. 2011;334(6060):1217-1218.