The gene expression profile of nodal peripheral T-cell lymphoma demonstrates a molecular link between angioimmunoblastic T-cell lymphoma (AITL) and follicular helper T (T_{FH}) cells

Laurence de Leval,[1] David S. Rickman,[2] Caroline Thielen,[1] Aurélien de Reynies,[2] Yen-Lin Huang,[3,5] Georges Delsol,[6] Laurence Lamant,[6] Karen Leroy,[3,4,5] Josette Brière,[7] Thierry Molina,[8] Françoise Berger,[9] Christian Gisselbrecht,[10] Luc Xerri,[11] and Philippe Gaulard[3,4,5]

[1]Pathology, Centre Hospitalo-Universitaire (CHU) Sart-Tilman, University of Liège, Belgium; [2]Ligue Contre le Cancer, Paris, France; [3]Institut National de la Santé et de la Recherche Médicale (Inserm), Unité 617, Créteil, France; [4]Assistance Publique–Hôpitaux de Paris (AP-HP), Groupe hospitalier Henri Mondor, Département de Pathologie, Créteil, France; [5]Université Paris 12, Faculté de médecine, Institut Mondor de médecine moléculaire, Créteil, France; [6]Pathology, CHU Purpan, Toulouse, France; [7]Pathology, Hôpital Saint-Louis, Paris, France; [8]Pathology, Hôtel-Dieu, Paris, France; [9]Pathology, Centre Hospitalier Lyon Sud, Pierre Benite, France; [10]Hematology, Hôpital Saint-Louis, Paris, France; [11]Pathology, Institut Paoli Calmettes, Marseille, France

The molecular alterations underlying the pathogenesis of angioimmunoblastic T-cell lymphoma (AITL) and peripheral T-cell lymphoma, unspecified (PTCL-u) are largely unknown. In order to characterize the ontogeny and molecular differences between both entities, a series of AITLs (n = 18) and PTCLs-u (n = 16) was analyzed using gene expression profiling. Unsupervised clustering correlated with the pathological classification and with CD30 expression in PTCL-u. The molecular profile of AITLs was characterized by a strong microenvironment imprint (overexpression of B-cell– and follicular dendritic cell–related genes, chemokines, and genes related to extracellular matrix and vascular biology), and overexpression of several genes characteristic of normal follicular helper T (T_{FH}) cells (CXCL13, BCL6, PDCD1, CD40L, NFATC1). By gene set enrichment analysis, the AITL molecular signature was significantly enriched in published T_{FH}-specific genes. The enrichment was higher for sorted AITL cells than for tissue samples. Overexpression of several T_{FH} genes was validated by immunohistochemistry in AITLs. A few cases with molecular T_{FH}-like features were identified among CD30$^-$ PTCLs-u. Our findings strongly support that T_{FH} cells represent the normal counterpart of AITL, and suggest that the AITL spectrum may be wider than suspected, as a subset of CD30$^-$ PTCLs-u may derive from or be related to AITL. (Blood. 2007;109: 4952-4963)

Introduction

Peripheral T-cell lymphomas (PTCLs) are uncommon malignancies, representing approximately 12% of all lymphomas.[1] In western countries, the most common forms present as nodal tumors and mainly include 3 subtypes: angioimmunoblastic T-cell lymphoma (AITL), anaplastic large cell lymphoma (ALCL), and peripheral T-cell lymphoma, unspecified (PTCL-u).[2] AITL, otherwise known as angioimmunoblastic lymphadenopathy with dysproteinemia, is a systemic disease manifested by B symptoms, polyadenopathy, and various immunologic abnormalities. AITL is characterized by a diffuse polymorphous infiltrate including variable proportions of medium-sized neoplastic cells with abundant clear cytoplasm, prominent arborizing blood vessels, proliferation of follicular dendritic cells (FDCs), and the presence of Epstein-Barr virus (EBV)–positive large B-cell blasts.[3] ALCL shows a broad morphologic spectrum, unified by the recognition of typical "hallmark" cells strongly expressing CD30. They commonly express epithelial membrane antigen (EMA), cytotoxic molecules, and in up to 85% of cases the anaplastic lymphoma kinase (ALK) protein as a consequence of rearrangement of the *ALK* gene.[4] Conversely, PTCL-u, the most common form of nodal T-cell lymphoma, lacks precise defining features and is diagnosed by exclusion of any recognizable ("specified") subtype of T-cell lymphoma. Hence, PTCL-u represents a heterogeneous group of tumors that may include cases with borderline features to ALCL and AITL. Nonanaplastic PTCLs usually carry a dismal prognosis, with 5-year overall survival rates averaging 25%.[1]

The normal cellular derivation of PTCLs is ambiguous.[2] The T-cell system is complex, comprising numerous subsets with different effector, regulatory, or memory functions. The immunophenotypic profiles of PTCLs are heterogeneous and not entirely disease specific. Most cases appear to derive from T cells expressing the αβ form of the T-cell receptor (TCR). PTCLs-u more commonly express CD4 than CD8, but a significant proportion of the cases have an aberrant phenotype (CD4$^+$CD8$^+$ or, less commonly, CD4$^-$CD8$^-$).[5-7] In contrast, it has been suggested that most AITLs derive from mature helper CD4$^+$CD8$^-$ T cells,[8] expressing the Th1-associated chemokine receptors CXCR3 and OX40/CD134.[9,10] However, based on the expression of single markers (namely the transcription factor BCL6,[11] the CXCL13 chemokine,[12-14] the membrane receptor PD-1 [also known as PDCD1[15]], and CXCR5[16]), it has been recently suggested that the neoplastic cells in AITL may derive from a specific subset of T cells normally present in germinal centers with a helper function to follicular B cells (follicular helper T cells, T_{FH} cells).

Submitted October 30, 2006; accepted January 31, 2007. Prepublished online as *Blood* First Edition Paper, February 6, 2007; DOI 10.1182/blood-2006-10-055145.

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.

The genetic alterations and pathogenic mechanisms underlying AITL and PTCL-u are largely unknown. With the exception of ALCL and a small subset of PTCLs with follicular pattern,[17] there are no recurrent translocations reported. Complex karyotypes with numeric and structural chromosomal changes are noted in many cases of PTCL-u with recurrent chromosomal losses and gains as demonstrated by genomic profiling.[18]

In this study, we performed gene expression profiling analysis of a series of AITLs and PTCLs-u in order to delineate their molecular signatures and to gain insight into their ontogeny and molecular mechanisms by an analytic comparison of our data set to available molecular signatures of normal T-cell subsets.[19-22]

Patients, materials, and methods

Patient characteristics and tumor samples

Thirty-four newly diagnosed, previously untreated patients with nodal PTCL-u (n = 16) and AITL (n = 18) were analyzed in this study, according to a protocol approved by the institutional review board of Hôpital Saint-Louis, Paris. Seventeen patients had been enrolled in clinical protocols of the Groupe d'Etude des Lymphomes de l'Adulte (GELA). There were 18 male and 16 female patients, with a median age at diagnosis of 65 years (range, 21 to 83 years). Patients provided informed consent, in accordance with the Declaration of Helsinki.

All cases were reviewed by 2 hematopathologists (L.d.L. and P.G.). The percentage of malignant cells was evaluated by morphology and CD3 immunostaining as exceeding 30% in all AITL cases, and 50% in all PTCLs-u. Lymphomas were classified according to the criteria of the WHO classification.[2] The staining panel included at least CD20 and CD3, and for most AITLs an FDC marker (CD21 and/or CD23 and/or CNA.42) and an EBV marker (latent membrane protein-1 [LMP-1]) and/or EBV-encoded small RNAs (EBERs). FDC expansion and EBV-infected cells were demonstrated in 16 of 17 and 11 of 16 AITLs, respectively. A variable expression of CD10 and BCL6 was demonstrated in 15 of 17 and 10 of 10 AITLs with interpretable staining. CD30 expression in more than 50% of the tumor cells was shown in 6 of 15 PTCLs-u tested. All $CD30^+$ PTCLs-u were EMA-negative large-cell tumors, distinct from ALCL, with a cytotoxic phenotype in 3 of 5 cases tested.

For 17 of 18 AITL patients (S1 to S17) and all PTCL-u patients (S19 to S34), frozen tumor tissue samples were analyzed. For 2 AITL patients with tumor cell suspensions available (S17C and S18C), mononuclear cell suspensions prepared from lymph nodes were enriched in tumor cells by magnetic activated cell sorting. This yielded a 84% $CD3^+CD10^+$ cell population for sample S17C and a 82% $CD4^+CD3^-CD10^+$ cell population for sample S18C.

Microarray procedures

Microarray analyses were performed using 3 μg total RNA as starting material and 10 μg cRNA per hybridization (GeneChip Fluidics Station 400; Affymetrix, Santa Clara, CA). The total RNAs were amplified and labeled following the one-cycle target labeling protocol (http://www.affymetrix.com). The labeled cRNAs were hybridized to HG-U133 plus 2.0 Affymetrix GeneChip arrays (Affymetrix, Santa Clara, CA). The chips were scanned with a Affymetrix GeneChip Scanner 3000 and subsequent images analyzed using GCOS 1.4 (Affymetrix). Except when indicated, all transcriptome analyses were carried out using either an assortment of R system software (v1.9.0; Vienna, Austria) packages including those of Bioconductor[23] (V1.1.1) or original R code. Raw feature data were normalized and $\log_2$ intensity expression summary values for each probe set were calculated using robust multiarray average (RMA[24]; Bioconductor package affy V1.4.32). Probe sets corresponding to control genes or having an "_x_" annotation were masked yielding a total of 50 406 probe sets available for further analyses. Raw data (Affymetrix U133A/B CEL files) from 6 samples corresponding to 3 T-cell subpopulations (T_{FH}, Th1, and Th2) cited by Chtanova et al[22] were normalized, in batch, with another 11 samples of T-cell subpopulations from the same citation using RMA as described above.

Gene expression analyses

Unsupervised classification analyses. Principle component analysis (PCA) was used to classify the tissue samples.[25] The function "prcomp" (R package stats V2.3.0) was used to calculate the 33 principle components for the 33 samples, based on the 25 744 probe sets that yielded a mean intensity level of 3.0 or higher ($\log_2$ value). The first 10 components accounted for an accumulated 70% of the cumulative variance. The first 25 components (95% cumulative variance) were used in the hierarchic clustering (complete linkage) of the samples using the R packages cluster (V1.9.3, Bioconductor[23]) and maptree (V1.3.3, Bioconductor[23]). The resulting dendrogram was cut yielding 3 sample groups, and the enrichment of sample annotations across the different groups was calculated using Fisher exact tests. We also classified the samples using a nonsupervised selection of genes and hierarchic clustering (for more details see Figure S1, available on the *Blood* website; see the Supplemental Materials link at the top of the online article).

Supervised analyses according to pathological features. Univariate *t* tests (BRB ArrayTools v3.4.beta2, http://linus.nci.nih.gov/BRB-ArrayTools.html) were used to define the differentially expressed gene lists with a significance level of each univariate test of $P < .002$ using a random variance model.[26] False discovery rates (FDRs) were calculated based on 1000 sample permutations. Probe sets that yielded a maximal normalized nonlog intensity value of 10 or less were filtered out from further analysis.

Gene ontology (GO) term and KEGG pathways enrichment tests. We used hypergeometric tests (GOstats, V1.1.1; Bioconductor[23]) to measure the enrichment of GO terms in each of the subgroup-specific gene lists. Independent tests were performed for molecular function (MF), biologic process (BP), and cellular compartment (CC) categories. GeneSpringGX (Agilent Technologies, Santa Clara, CA) was used for testing the overrepresentation of genes in a particular KEGG pathway (168 KEGG pathways in total were tested) for a given list of unique gene identifiers (HUGO gene symbols) versus all of the genes on the microarray mapped to that pathway. This test uses a standard Fisher exact test, and the *P* value is adjusted with a Bonferroni multiple testing correction.

Gene set enrichment analysis (GSEA). GSEA was performed as previously described[27] to assess the overrepresentation of T-cell–related gene sets in the different groups of samples (AITL; PTCL-u; $CD30^+$ PTCL-u, and $CD30^-$ PTCL-u). Briefly, this test determines the overrepresentation of a gene set (a list of genes) at the extremes (top or bottom) of the ordered, nonredundant data set (list of all of the genes being used to compare 2 groups of samples). Two different statistics were used to rank the genes for each comparison: "signal-to-noise" ratio (SNR)[27] and a classical *t* test. Fourteen gene sets representative of the Th1 cell, Th2 cell, and T_{FH} subsets derived from literature[19-22] were tested using GSEA for an association with the subgroups of tumors (Table 1). Nonredundant data sets were generated specific to each comparison by first filtering out all probe sets that yielded a maximal intensity value of 10 or less for the 2 groups and then averaging data values for multiple probe sets corresponding to the same gene.

Prediction analysis. Two approaches were used. For the first approach, the sample population was divided into a training group and a validation group. The training group of samples was used for the initial gene selection and building the multigene

Table 1. Gene sets used for gene set enrichment analysis (GSEA)

Gene set	No. of genes	Description and source	Array
GS1_TFH_up	57	Up-regulated in T_{FH}s from Kim et al[20]	cDNA microarray
GS2_TFH_up	100	Top 100 genes up-regulated in T_{FH}s from Chtanova et al[21]	U133A and B Affymetrix
GS3_TFH_up	405	Up-regulated in T_{FH}s from Chtanova et al[21,22]	U133A and B Affymetrix
GS4_TFH_up	443	GS1 plus GS3	cDNA microarray and Affymetrix
GS5_TFH_up	16	Genes common to both GS1 and GS3	cDNA microarray and Affymetrix
GS6_TFH_down	19	Down-regulated in T_{FH}s from Kim et al[20]	cDNA microarray
GS7_TFH_down	45	Down-regulated in T_{FH}s from Chtanova et al[21,22]	U133A and B Affymetrix
GS8_TFH_down	63	GS6 plus GS7	cDNA microarray and Affymetrix
GS9_Th1_up	138	Up-regulated in Th1 from Rogge et al[19]	HuGeneFL array, Affymetrix
GS10_Th1_up	29	Up-regulated in Th1 from Chtanova et al[21]	U133A and B Affymetrix
GS11_Th1_up	159	GS9 plus GS10	cDNA microarray and Affymetrix
GS12_Th2_up	50	Up-regulated in Th2 from Rogge et al[19]	HuGeneFL array, Affymetrix
GS13_Th2_up	28	Up-regulated in Th2 from Chtanova et al[21]	U133A and B Affymetrix
GS14_Th2_up	73	GS12 plus GS13	cDNA microarray and Affymetrix

Gene sets GS1 to GS14 are extracted from publications reporting gene expression signatures associated with purified subsets of $CD4^+$ T cells.[20-22] T_{FH}s are defined as $CD4^+CD57^+CXCR5^+$ cells isolated from tonsils,[20-22] and Th1 and Th2 cells were obtained from cord blood cells cultured in appropriate polarizing conditions.[19,21] The 97 transcripts (GenBank accession numbers) found by Kim et al[20] to be differentially expressed in T_{FH}s ($CD4^+CXCR5^+CD57^+$) versus other $CD4^+$ T-cell subsets (naive $CD4^+CD45RA^+$, early memory $CD4^+CXCR5^+CCR7^+$), and effector memory $CD4^+CCR7^-$) were mapped to 76 unique gene symbols represented on the Affymetrix U133plus2 GeneChip (57 overexpressed genes [GS1] and 19 down-regulated genes [GS6]). We used available gene symbols and/or accession numbers provided by Chtanova et al[21,22] to map the described genes to the U133plus2.0 GeneChip. GS2 represents the top 100 genes overexpressed in T_{FH}s according to Chtanova et al.[21] The ranking of these genes was as follows: summing the individual fold change values between T_{FH}s and the other T-cell samples (the maximum fold change was used for redundant data); ordering the sums from highest to lowest fold change; and selecting the top 100 genes as the first 100 genes in this ordered gene list. GS3 and GS7 correspond to the complete list of overexpressed and underexpressed genes in T_{FH}s compared to other $CD4^+$ T-cell subsets,[21,22] respectively. The union of GS1 (n = 55) and GS3 (n = 405) yielded 443 overexpressed genes (GS4) in T_{FH}s. The intersection of GS1 (n = 55) and GS3 (n = 405) yielded 16 common overexpressed genes in T_{FH}s (GS5). The union of GS6 (n = 19) and GS7 (n = 45) yielded 63 underexpressed genes (GS8) in T_{FH}s. In both studies by Rogge et al[19] and Chtanova et al,[21] the gene expression signature of the Th1 and Th2 subsets was defined by comapping each subset to the other one. Probe set identifiers provided by Rogge et al[19] in the Table S1 were mapped to Entrez Gene identifiers using annotations provided by Affymetrix (NETAFFYX: http://www.affymetrix.com/analysis/index.affx), which were then used to map to U133plus2.0 probe sets yielding 138 genes (GS9) specifically overexpressed in Th1 cells and 50 genes (GS12) specifically overexpressed in Th2 cells. The genes specific to Th1 and Th2 described by Chtanova et al[21] were mapped to the U133plus2.0 GeneChip, yielding, respectively, GS10 (29 genes) and GS13 (28 genes). The union of GS9 (n = 138) and GS10 (n = 29) yielded 159 overexpressed genes (GS11) in Th1 cells. The union of GS12 (n = 50) and GS13 (n = 28) yielded 73 overexpressed genes (GS14) in Th2 cells.

(1-10 genes) predictors. We restricted this analysis to 4 prediction algorithms: nearest shrunken centroids (PAM, pamr, v1.30), k-nearest neighbors (KNN, class, v7.2-27.1), diagonal quadratic/linear discriminant analyses (DLDA/DQDA, sma, v0.5.15). A bottom-up step approach was used to construct the best multigene predictor for each algorithm (Table S5 legend). As a second approach, BRB ArrayTools (v3.5 beta1) was used to apply a leave-one-out cross validation (LOOCV; Table S6 legend). For this, 7 prediction algorithms were used: compound covariate, diagonal linear discriminant analysis, 1-nearest neighbor, 3-nearest neighbors, nearest centroid, support vector machines, or Bayesian compound covariate. The latter algorithm was based on the method previously described by Wright et al.[28]

Validation at the protein level

Tissue microarrays (TMAs) containing three 0.6-mm tissue cylinders of each donor block from 13 AITLs and 15 PTCLs-u were constructed using a manual tissue arrayer (Beecher Instruments, Sun Prairie, WI). Adequate control tissues for specific antibodies were also included. Sections (4-μm each) were stained with the following primary antibodies: anti-CD27 (clone 137B4; Novocastra, Newcastle upon Tyne, United Kingdom), anti-CXCL13 (clone 53610; R&D Systems, Minneapolis, MN), antithrombomodulin (clone 1009; Dako, Glostrup, Denmark), anticlusterin (clone 41D; Upstate Biotechnology, Lake Placid, NY), anti-PDCD1 (kindly provided by Daniel Olive), anti-NFATC1 (clone 7A6; Santa Cruz Biotechnology, Santa Cruz, CA). Immunostainings were performed following heat-induced antigen retrieval, using a standard indirect avidin-biotin peroxidase method. For CXCL13, a tyramide signal amplification system (CSAII kit; Dako) was applied.[13] For thrombomodulin and clusterin, the distribution and extent of immunostaining were recorded. For other markers, cases were regarded as immunoreactive if clearly neoplastic cells exhibited staining. Differences in marker expression were demonstrated by χ^2 analyses.

Results

Unsupervised clustering of PTCLs correlates with pathological features

Similar to previous studies,[29] we classified our tumor samples using PCA. This method typically relies on the first 3 components to spatially represent the samples in a 3-D plot, often, however, at the expense of the total variation being represented (PC1-3 accounted for only 41% of the variation in our data set). Using a larger number of components (PC1-25, 95% of the variation) in combination with hierarchic clustering, we obtained 3 major groups of tumors (designated C1, C2, and C3; Figure 1), showing striking association with the pathologic classification ($P < .001$). Clusters C1 and C2 represented 2 homogeneous groups of tumors consisting of 12 AITLs and 10 PTCLs-u, respectively. Cluster C3 was a heterogeneous group of tumors comprising 5 AITLs and 6 PTCLs-u. Among the PTCLs-u, all but one $CD30^+$ tumor clustered within group C3, whereas all but one PTCL-u sample in C2 were $CD30^-$ ($P = .011$). The observed partition of the lymphoma samples also strongly correlated with the percentage of tumor cells ($P < .001$). When restricting the comparison to the PTCL-u cases distributed across clusters 2 and 3, the percentage of tumor cells (in contrast with the CD30 status) did not correlate with the observed distribution ($P = .28$). These results were reproducible using different proportions of conserved accumulated variation to cluster the samples (as low as 77% [PC1-13], data not shown). In addition, a similar topology was also obtained using a nonsupervised selection of expression profile, based on the 2.5% most varying expression profiles (644 probe sets, Figure S1 and Table S1).

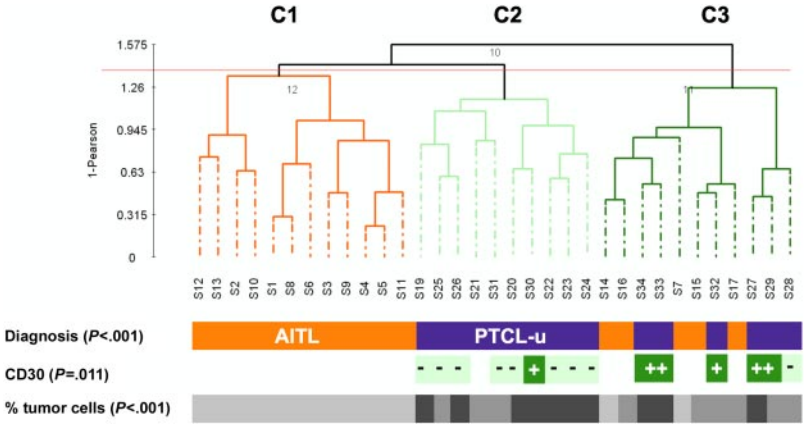

Figure 1. Unsupervised clustering of 16 PTCL-u and 17 AITL tissue samples. Dendrogram of the 33 lymphoma tissue samples based on PCA of the first 25 components (accumulated variance 95%). CD30: + denotes expression in more than 50% of neoplastic cells; - denotes absence of expression or low level of expression of CD30 in neoplastic cells. The percentage of tumor cells is represented by a gray scale (light gray: 30%-50% tumor cells; medium gray: 50%-70% tumor cells; and dark gray: > 70% tumor cells).

AITL and PTCL-u have distinct molecular features

A total of 832 probe sets significantly distinguished AITL tissues from PTCLs-u ($P < .002$; 10.6% maximal FDR). They corresponded to 678 unique gene identifiers, including 442 genes (545 probe sets) overexpressed in AITL (defining the AITL signature) and 236 genes (287 probe sets) overexpressed in PTCL (see Table S2 for a complete gene list and associated statistics). The genes of the AITL signature belonged to several functional categories (Table 2). The most significant overrepresented GO terms in AITL were related to cell-to-cell communication and adhesion ($P < .009$), immune response ($P < .007$), vascular biology ($P < .04$), and extracellular matrix ($P = .005$). Compared to PTCL-u, AITL had higher levels of expression of cell adhesion molecules (cadherins, integrins, CD151) and various membrane receptors (CD10, CD40 ligand, CD200, PDCD1) and proteins involved in membrane signaling.[30] AITL was also characterized by a striking overrepresentation of B-cell– and plasma cell–related genes, and of FDCs as well as complement factors, and higher level of expression of extracellular matrix components (laminin, collagen, ladinin, fibronectin) and of factors and enzymes involved in matrix synthesis and remodeling (TGFβ, fibroblast growth factor, matrix metalloproteinases). Several genes related to vascular biology—including vascular growth factors, endothelium-related genes, coagulation factors—were also overexpressed in AITL. Conversely, the functional categories of the genes overexpressed in PTCL-u were reflective of nonspecific cellular functions, notably protein ubiquination (TRIM proteins[31]), regulation of transcription, and metabolism (Table S2). Few of the genes overexpressed in PTCL-u (Table 3) were related to immune reactions or chemotaxis; they included CCR7, a chemokine receptor mediating homing to lymph nodes,[32] and CD47, an integrin-associated protein involved in interactions between T cells and endothelium.[33]

The microenvironment component has a significant influence on the AITL signature

In 2 AITL cases, the gene expression profile of highly enriched tumor samples enabled us to delineate the components of the AITL molecular signature (ie, tumor cell versus microenvironment-related genes). To this end, we calculated the fold change between the average gene expression level of each of the 545 probe sets that was significantly overexpressed in AITL (compared to PTCL-u), for the 2 AITL cell samples versus the 17 AITL tissues samples. We found 46 genes that were more associated to the "AITL tumor cells" (fold change [FC] > 1, including 8 genes with a FC > 1.5) and 396 genes that were more associated with the "AITL microenvironment" (FC < 1, including 290 genes with a FC < 0.67) (Table 2; see Table S2 for a complete list of the genes).

The AITL transcriptional profile is enriched in genes characteristic of T_{FH} cells

Using immunohistochemistry, we and others have recently demonstrated that neoplastic cells in AITL express CXCL13,[12-14] a chemokine characteristic of T_{FH} cells. Our gene expression profiling data confirmed that AITL expresses higher levels of CXCL13 mRNA than PTCL-u, (Table 2). Moreover, we also identified several other genes overexpressed by AITL (ie, CD200, PDCD1, CD40L, NFATC1, LIF) that had been reported as overexpressed in T_{FH} cells compared to other T-cell subsets.[20-22,34] In order to specifically evaluate the significance of the observed similarity between the AITL and the T_{FH} signatures, we used GSEA.[27] The T_{FH} signature was defined using set of genes independently identified by Kim et al[20] and Chtanova et al[21,22] as differentially expressed in $CD4^+CD57^+$ cells isolated from tonsils compared to other normal purified T-cell subsets with, respectively, cDNA and oligonucleotide microarrays (Table 1). Having ranked the tissue samples according to the AITL versus PTCL-u distinction, we found that AITLs were significantly enriched in genes up-regulated in T_{FH} cells (GS2_ T_{FH} _up and GS5 T_{FH} _up, $P < .014$; Tables 4 and S3). Of the 100 genes comprising the GS2_ T_{FH} _up gene set, the GSEA method identified 42 "core" genes that accounted for enrichment signal (Figure 2A) and were distributed among the top 1722 genes in the ranked data set. Next, we assessed the enrichment of these 42 core genes in the AITL samples by ranking the top 1722 genes according to the type of sample (2 cell suspensions versus 17 tissues) by fold change of gene expression. While a meaningful *P* value cannot be calculated given the limited number of cell suspensions, we observed a higher level of expression in the tumor cell suspensions compared to the tissue samples of these 42 core genes (Figure 2B). The overexpression of T_{FH} genes in AITL tissues and cells is illustrated by a heatmap in Figure 2C.

We found no significant association between any of the different gene sets each overexpressed in Th1 and Th2 T-cell subsets (Table 1, GS9 to GS14), by comparing PTCLs-u and AITLs (33 tissue samples) with either PTCL-u or AITL (Table S3).

$CD30^+$ and $CD30^-$ PTCLs-u have distinct molecular features

By 2 different unsupervised clustering methods, $CD30^+$ PTCLs-u tended to cluster together (Figure 1 and Figure S1). Supervised analysis yielded a list of 241 probe sets (186 genes, 24.8% maximal FDR), comprising 73 genes overexpressed in $CD30^+$ PTCLs-u and

Table 2. Genes overexpressed in AITL compared to PTCL-u

Gene symbol	Gene title	Fold difference of geometric means
Cell-to-cell communication		
CD151	CD151 antigen	1.35
*CD200**	CD200 antigen	2.38
CD24	CD24 antigen (small-cell lung carcinoma cluster 4 antigen)	1.78
CD40LG†*	CD40 ligand (TNF superfamily, member 5, hyper-IgM syndrome)	1.64
CDH11	Cadherin 11, type 2, OB-cadherin (osteoblast)	1.7
CDH3	Cadherin 3, type 1, P-cadherin (placental)	1.47
CLEC4G	C-type lectin superfamily 4, member G	2.02
ITGA5	Integrin, alpha 5 (fibronectin receptor, alpha polypeptide)	1.42
ITGAX	Integrin, alpha X (antigen CD11C [p150], alpha polypeptide)	1.89
ITM2C	Integral membrane protein 2C	2.65
KAL1	CD82 antigen	1.48
L1CAM	L1 cell adhesion molecule	1.56
MAL	Mal, T-cell differentiation protein	4.62
MME†	Membrane metallopeptidase (neutral endopeptidase, CALLA, CD10)	2.26
PCDH17	Protocadherin 17	1.66
PCDHA	Protocadherin alpha subfamily	2.9
PCDHG	Protocadherin gamma subfamily	1.5
PDCD1†*	Programmed cell death 1	1.64
PLEKHC1	Pleckstrin homology domain containing, family C member 1	1.81
Vascular biology		
AMOTL1	Angiomotin-like 1	1.35
ANGPT2	Angiopoietin 2	1.34
EDG2	Endothelial differentiation, lysophosphatidic acid G-protein-coupled receptor, 2	2.03
EDG3	Endothelial differentiation, sphingolipid G-protein-coupled receptor, 3	4.05
EFNB2	Ephrin-B2	2.63
ENG	Endoglin (Osler-Rendu-Weber syndrome 1)	1.56
EPAS1	Endothelial PAS domain protein 1	1.81
EPHA2	EPH receptor A2	1.38
FGF11	Fibroblast growth factor 11	1.65
GAS6	Growth-arrest specific 6	2.04
GJA1	Gap junction protein, alpha 1, 43 kDa (connexin 43)	2.18
GJA4	Gap junction protein, alpha 4, 37 kDa (connexin 37)	2.08
GP1BA	Glycoprotein Ib (platelet), alpha polypeptide	1.98
L1CAM	L1 cell adhesion molecule	1.56
PGF	Placental growth factor, vascular endothelial growth factor-related protein	1.66
SELP	Selectin P (granule membrane protein 140 kDa, antigen CD62)	2.42
SERPINF1	Serpin peptidase inhibitor, clade F (alpha-2 antiplasmin), member 1	1.67
SERPINH1	Serpin peptidase inhibitor, clade H (heat shock protein 47), member 1	1.66
THBD	Thrombomodulin	2.03
THY1	Thy-1 cell surface antigen	2.42
VCAM1	Vascular cell adhesion molecule 1	1.76
VEGF	Vascular endothelial growth factor	1.73
VWA1	Von Willebrand factor A domain containing 1	1.52
Humoral immune response		
C1QTNF1	C1q and tumor necrosis factor-related protein 1	1.8
C1R	Complement component 1, r subcomponent	1.53
C3	Complement component 3	2.67
CLU	Clusterin	3.26
CR2	Complement component (3d/Epstein-Barr virus) receptor 2	6.11
FCAMR	Fc receptor, IgA, IgM, high affinity	4.94
IGH	Immunoglobulin heavy locus	3.09
IGHM	Immunoglobulin heavy constant mu	3.82
IGKC†	Immunoglobulin kappa constant	2.38
IGL	Immunoglobulin lambda locus	2.03
IGLC2	Immunoglobulin lambda constant 1	2.86
IGLC2	Immunoglobulin lambda joining 3	1.8
IGLV3-25	Immunoglobulin lambda variable 3-25	4.05
IGSF3	Immunoglobulin superfamily, member 3	2.21
SDC1	Syndecan 1	3.23
SPIB	Spi-B transcription factor (Spi-1/PU.1 related)	2.94
TCF3	Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)	1.65
TNFRSF17	Tumor necrosis factor receptor superfamily, member 17	4.04

Gene symbol	Gene title	Fold difference of geometric means
Chemokine and cytokine pathways		
CCL19	Chemokine (C-C motif) ligand 19 (EBV-induced receptor ligand chemokine, ELC)	1.99
CCL20	Chemokine (C-C motif) ligand 20 (secondary lymphoid tissue chemokine, SLC)	4.09
CCL22	Chemokine (C-C motif) ligand 22 (macrophage-derived chemokine, MDC)	3.68
CCL24	Chemokine (C-C motif) ligand 24 (eotaxin-2)	3
CCL26	Chemokine (C-C motif) ligand 26 (eotaxin-3)	5.37
*CXCL13**†	Chemokine (C-X-C motif) ligand 13 (B-cell chemoattractant)	2.79
CXCL14	Chemokine (C-X-C motif) ligand 14 (breast and kidney chemokine, BRAK)	2.17
IL4	Interleukin 4	2.41
LIF†	Leukemia inhibitory factor (cholinergic differentiation factor)	3.8
OSMR	Receptor to oncostatin M (an IL-6 type cytokine)	2.22
Extracellular matrix		
ADAM19	ADAM metallopeptidase domain 19 (meltrin beta)	2.23
ADAMTS9	ADAM metallopeptidase with thrombospondin type 1 motif, 9	1.65
CHPF	Chondroitin polymerizing factor	1.47
COL15A1	Collagen, type XV, alpha 1	2.72
COL23A1	Collagen, type XXIII, alpha 1	1.25
COL27A1	Collagen, type XXVII, alpha 1	1.87
COL6A1	Collagen, type VI, alpha 1	2.28
COL6A2	Collagen, type VI, alpha 2	2.27
CTSH	Cathepsin H	1.72
EFEMP2	EGF-containing fibulin-like extracellular matrix protein 2	1.54
FNDC1	Fibronectin type III domain containing 1	2.27
FOXF1	Forkhead box F1	1.74
HS3ST1	Heparan sulfate (glucosamine) 3-O-sulfotransferase 1	1.56
LAD1	Ladinin 1	1.38
LAMB2	Laminin, beta 2 (laminin S)	1.66
LAMC2	Laminin, gamma 2	1.63
MMP1	Matrix metallopeptidase 1 (interstitial collagenase)	3.72
MMP12	Matrix metallopeptidase 12 (macrophage elastase)	5.01
PCOLCE	Procollagen C-endopeptidase enhancer	1.94
TGFB1/1	Transforming growth factor beta 1-induced transcript 1	1.87
TGFB3	Transforming growth factor, beta 3	1.27
TIMP1	TIMP metallopeptidase inhibitor 1	1.57
VTN	Vitronectin (serum spreading factor, somatomedin B)	1.49
Others/miscellaneous		
*NFATC1**†	Nuclear factor of activated T cells cytoplasmic, calcineurin-dependent 1	1.69
PIM2†	Pim-2 oncogene	1.66
VAV2	Vav 2 oncogene	1.82

A complete list of genes is available in Table S2.
*Genes reported as part of the follicular helper T-cell signature.
†Genes of the AITL tumor cell signature.

113 genes overexpressed in CD30$^-$ PTCLs-u (*t* test, $P < .002$; Tables 5-6). The only GO terms overrepresented in CD30$^+$ PTCLs-u were related to molecular function ($P < .009$) and comprised various enzymes with transferase, synthase, and kinase activity. Also overexpressed in CD30$^+$ PTCLs-u were several genes involved in the control of transcription, among which the Jun dimerization protein (acting as an IL2 repressor) showed the highest fold change.[35] GO terms overrepresented in CD30$^-$ PTCLs-u were mostly related to biologic processes (regulation of physiological and metabolic processes, lymphocyte activation, $P < .008$). Particularly, CD30$^-$ PTCLs-u had overexpression of genes associated with T-cell activation and signal transduction, including the costimulatory receptor CD28 (FC > 6.0), the CD69 activation antigen, CD52, and several molecules involved in TCR signal transduction (Itk, Lyn, and Lck). Accordingly, the TCR signaling pathway was the most significant KEGG pathway overrepresented in CD30$^-$ PTCL-u ($P < .001$; Table S4).

Next, the CD30$^+$ and CD30$^-$ subsets of PTCL-u were tested for enrichment in gene sets representative of normal T-cell subsets (Table 1). Although we observed a trend for the signature of CD30$^-$ PTCL-u to be associated with the Th2 signature, no significant correlation was found between either of the 2 PTCL-u subgroups and any of the gene sets representative of the Th1 and Th2 subsets (Table S3). Somewhat surprisingly, the molecular signature of CD30$^-$ PTCL-u was overlapping with that of T_{FH} cells, although the correlation was statistically weaker than for AITL samples (Tables 4 and S3). This observation raised the question as to whether the T_{FH} signature was present in all or most CD30$^-$ PTCLs-u, or if a small subset of PTCL-u cases accounted for this statistical association. All PTCL-u cases were carefully re-examined, and stainings for FDC and EBV were performed for most cases (15/16). Four cases disclosed focal FDC expansion and/or a few EBV-positive large cells (S24, S25, S26, and S28), and, molecularly, these cases (Figure 2C) had some overexpression of the T_{FH} core genes. Leaving out these 4 cases from the GSEA resulted in increased significance of the T_{FH} enrichment of the AITLs (compared to PTCL-u) and reduced significance of the association of the T_{FH} signature to CD30$^-$ PTCL-u (Table S3). Thus, the statistical association between CD30$^-$ PTCL-u and T_{FH} appeared to be mostly contributed by a small subset of cases.

Table 3. Genes overexpressed in PTCL-u compared to AITL

Gene symbol	Gene title	Fold difference of geometric means
Ubiquitination		
MLL3	Myeloid/lymphoid or mixed-lineage leukemia 3	0.58
MYLIP	Myosin regulatory light chain interacting protein	0.57
TRIM25	Tripartite motif-containing 25	0.71
TRIM38	Tripartite motif-containing 38	0.63
TRIM52	Tripartite motif-containing 52	0.71
USP8	Ubiquitin specific peptidase 8	0.78
Regulation of transcription		
ADAR	Adenosine deaminase, RNA specific	0.84
ANKIB1	Ankyrin repeat and IBR domain containing 1	0.6
ELF2	E74-like factor 2 (ets domain transcription factor)	0.74
HOXA7	Homeobox A7	0.54
ID2	Inhibitor of DNA binding 2, dominant-negative helix-loop-helix protein	0.49
POU2F1	POU domain, class 2, transcription factor 1	0.7
PPARBP	PPAR binding protein	0.59
SP100	Nuclear antigen Sp100	0.65
SP110	SP110 nuclear body protein	0.61
SP4	Sp4 transcription factor	0.71
TAF1	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor	0.74
TFAP2A	Transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)	0.52
TRERF1	Transcriptional regulating factor 1	0.42
ZNF12	Zinc finger protein 12	0.57
ZNF148	Zinc finger protein 148 (pHZ-52)	0.68
Immune response: chemotaxis		
CCR7	Chemokine (C-C motif) receptor 7	0.46
CD47	CD47 antigen (Rh-related antigen, integrin-associated signal transducer)	0.61
G1P2	Interferon alpha-inducible protein (clone IFI-15K)	0.58
IFI44	Interferon-induced protein 44	0.36
IFI44L	Interferon-induced protein 44-like	0.17
IFIT2	Interferon-induced protein with tetratricopeptide repeats 2	0.46
IFIT3	Interferon-induced protein with tetratricopeptide repeats 3	0.38

Prediction analysis

Given the occasional difficulty to distinguish AITL from PTCL-u using current criteria, we addressed the question if AITL and PTCL-u could be molecularly classified. As a first approach, we divided the sample population into a training group (S1, 8 AITLs and 8 PTCLs-u [4 $CD30^+$ and 4 $CD30^-$]) and a validation group (S2, 9 AITLs and 8 PTCLs-u). An initial list of 284 probe sets ($P < .001$) was selected using the S1 group and used to build a multigene (1-10 genes) predictor to classify the S2 group (see "Patients, materials, and methods" for details). We obtained 4 top predictors (one for each of the prediction algorithms used) that yielded between 82% and 94% success rate in correctly classifying the S2 group (see Table S5 for the list of genes and classification results). The best of the 4 predictors comprised 8 genes (ITIH5,

Table 4. GSEA results

Comparison	Gene sets	Maximum *P* value	Group enriched
AITL vs PTCL-u	GS2_TFH_up (97); GS5_TFH_up (16)	.014	AITL
AITL cells vs tissues	GS2_TFH_core (42)	.06*	AITL cells
$CD30^+$ vs $CD30^-$ PTCL-u	GS1_TFH_up (53); GS2_TFH_up (94); GS3_TFH_up (374); GS4_TFH_up (411); GS5_TFH_up (15)	.003	$CD30^-$ PTCL-u
AITL vs $CD30^-$ PTCL-u	GS2_TFH_up (98); GS5_TFH_up (16)	< .001	AITL

This table summarizes the relevant GSEA results obtained for the different comparisons (19 AITLs versus 16 PTCLs-u; 2 AITL cell suspensions versus 17 AITL tissue samples; 6 $CD30^+$ versus 9 $CD30^-$ PTCLs-u; 19 AITLs versus 9 $CD30^-$ PTCLs-u). Shown are the enriched gene sets with corresponding size, the maximal *P* value, and the enriched class of samples. A weighted enrichment score (ES) is calculated by walking down the ranked gene list, increasing a running-sum statistic for every gene set encountered and decreasing the statistic for every gene not in the gene set. A *P* value is calculated for an ES by using a sample-based permutation test procedure that preserves the complex correlation structure of the gene expression data. Specifically, we permuted the samples labels (eg, PTCL-u and AITL) and recomputed the ES of the gene set for the permuted data 1000 times, which generated a null distribution for the ES. The *P* value of the observed ES was then calculated based on this null distribution. Other advanced parameters used were as follows: "meandiv" normalization of the data; "timestamp" seed for permutation; and "equalize and balance" randomization mode. The size of a given gene set for a given comparison represents the number of genes out of the total genes in the gene set that were present in the data set following pre-GSEA filtering. After reducing the data set to account for redundant probe sets for a single gene identifier, we obtained the following data sets for each comparison: 13 178 genes for AITL versus PTCL-u; 13 186 genes for AITL versus $CD30^-$PTCL-u; and 13 244 genes for $CD30^+$ PTCL-u versus $CD30^-$ PTCL-u. Since slightly different data sets were used for each comparison (depending on a minimal geometric mean of the 2 groups of samples being compared; see "Patients, materials, and methods"), the number of genes in the table did not necessarily correspond to the number of genes in a gene set that were represented in a data set used in the comparison. For the AITL versus PTCL-u and $CD30^+$ versus $CD30^-$ PTCL-u comparisons, SNR was used to construct the ordered data set. For comparison of the AITL tissues and sorted cells, genes were ordered according to fold change of expression.

*A meaningful *P* value cannot be derived from the permutation tests because the number of samples in one class (cell suspensions) was too small (n = 2).

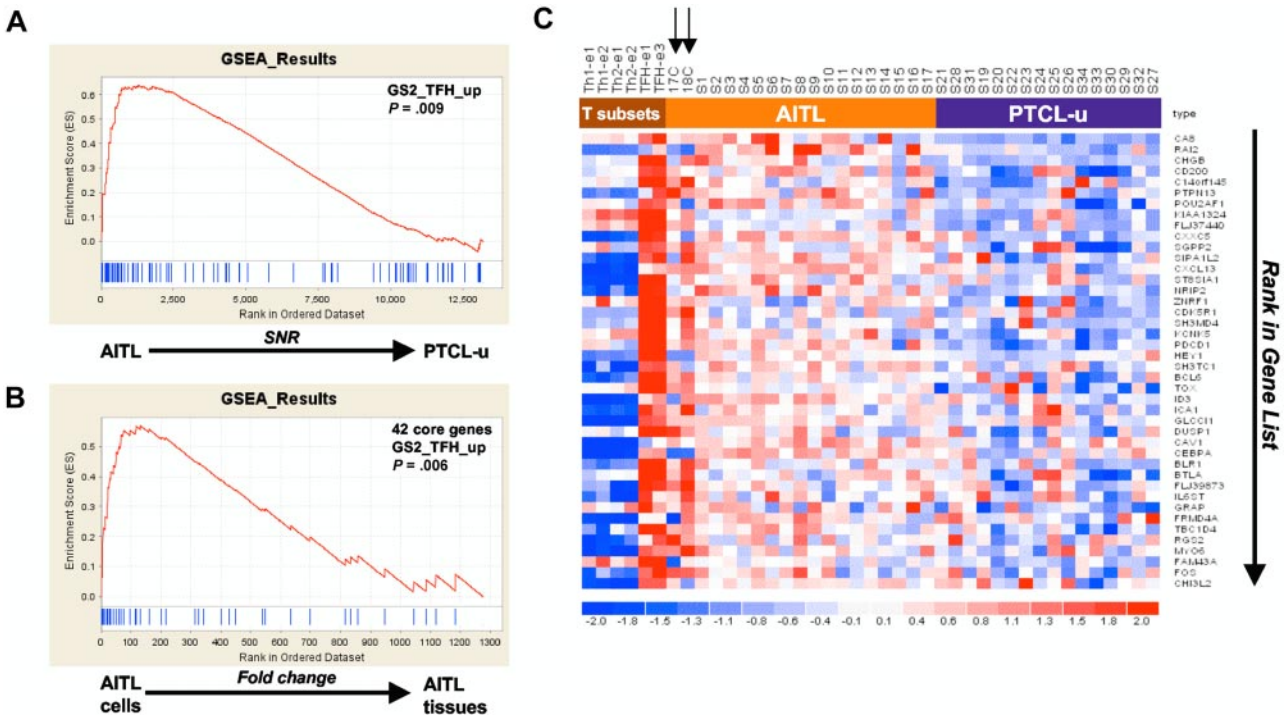

Figure 2. GSEA ES curves and clustering results for the 2 significant gene signatures significantly overexpressed in AITL compared to PTCL-u. (A-B) ES curve (red) or the running sum of the weighted enrichment score obtained from GSEA software. Vertical blue lines indicate the position of each of the 100 genes ("Hits") comprising the GS2_ T_{FH} _up gene set. The graph on the bottom of each panel shows the ranked list metric (gray, SNR for A and fold change for B) for each gene as a function of the rank in the ordered data set. This corresponds to the level of correlation with the groups being tested (the far leftmost gene had the highest correlation with the AITL samples, and the far rightmost gene had the highest correlation with PTCL-u samples). (A) Results obtained for GS2_ T_{FH} _up gene set in the comparison of AITL and PTCL-u using the SNR statistic to rank the genes from left to right (highest [0.9] to the lowest [−0.4] real SNR value, respectively). (B) Results obtained for the 42 core (or leading-edge) genes obtained from the analysis in panel A that were distributed in the top 1722 genes. The core genes in the ranked list represent the genes that are at, or before, the point where the running sum reaches its maximum deviation from zero (see Subramanian et al[27] for more details). Here, the top 1722 genes obtained in panel A are ranked from highest (left) to lowest (right) fold change between the average of the 2 AITL-sorted cell samples (AITL cell) versus the average of the 17 AITL tissue samples and the hits represent the 42 core genes. (C) Clustering and heatmap of these 42 core genes obtained from the analysis shown in panel A. Data from T-cell subsets (T_{FH}, Th1, and Th2; Chtanova et al[21] and this study) were standardized by calculating an independent z-score (mean = 0 and standard deviation = 1) for each data set. Each gene was further standardized by mean-centering its expression profile across all samples. Genes are ordered by their rank calculated in the analysis shown in panel A. DNA-Chip Analyzer (dChip) Version 1.3 (Department of Biostatistics, Harvard School of Public Health, Boston, MA) was used to generate the heatmap and cluster the genes. Standardized expression ranges from −2.0 (blue) to 2.0 (red). Arrows point at the columns representing AITL cell suspensions.

EPIM, THY1, KCNE4, VAV2, LIPT1, C10orf38, MGC16044) and yielded a success rate of 94%. Using a leave-one-out cross-validation (LOOCV) approach, we obtained, at best, an average overall success rate of 88% (see the list of the 220 gene classifier in Table S6).

Immunohistochemical validation

A few genes of the AITL signature were selected for assessment at the protein level by immunohistochemistry on TMAs. Clusterin and thrombomodulin were selected as markers for the microenvironment. CXCL13, PDCD1, and NFATC1 were chosen within the "core" T_{FH} genes, on the basis of available specific antibodies and our experience with these reagents. CD27 was tested because this antigen, routinely assessed by immunohistochemistry, was part of the T_{FH} signature.[21]

An extensive meshwork of clusterin-positive cells with dendritic morphology was evidenced in 8 of 12 AITLs, but in only 1 of 16 PTCLs-u ($P < .001$) (Figure 3A-B). Thrombomodulin antibody produced significant interstitial staining in 6 (46%) of 13 AITLs, but in only 1 (7%) of 15 PTCLs-u ($P = .016$). The neoplastic cells in 12 (92%) of 13 AITL cases stained for CXCL13, whereas only 5 (33%) of 15 PTCLs-u showed some CXCL13 staining in tumor cells ($P < .001$) (Figure 3C-D). Membrane staining for PDCD1 was evidenced in a fraction of the neoplastic cells in 6 (46%) of 13 AITLs, but in only 1 (7%) of 15 PTCLs-u ($P = .016$) (Figure 3E-F). In all AITLs, cytoplasmic staining for NFATC1 was evidenced in atypical medium-sized cells, while only 3 of 9 assessable PTCLs-u were positive ($P < .001$) (Figure 3G-H). Finally, the neoplastic cells were positive for CD27 in 11 (92%) of 12 AITLs and in 5 (66%) of 15 PTCLs-u ($P = .002$) (Figure 3I-J).

Discussion

Over the past few years, genome-wide expression profiling methods have been applied to most types of B-cell lymphomas, leading to significant advances in our understanding of these diseases.[36-39] Yet, few works have been devoted to the molecular profiling of T-cell neoplasms.[40-43] In the 3 largest studies comprising up to 62 patients,[40,41,43] the series of samples were pathologically heterogeneous, including both precursor and various types of PTCLs, and the hybridization platforms consisted of cDNA arrays, containing at the most 8000 spotted clones. Here, we focused our work on the 2 most common forms of nodal PTCLs, and performed gene expression profiling using a highly standardized pangenomic oligonucleotide Affymetrix microarray.

The clusterings obtained by 2 different unsupervised methods correlated to a large extent with the pathological classification, but also reflected some overlap between AITL and PTCL-u. The molecular distinction between AITL and PTCL-u consisted mostly of an overexpression of genes associated with specific functions in

Table 5. Genes overexpressed in CD30$^+$ PTCL-u compared to CD30$^-$ PTCL-u

Gene symbol	Gene title	Fold difference of geometric means
Transferases		
MGST3	Microsomal glutathione S-transferase 3	0.39
NANS	*N*-acetylneuraminic acid synthase (sialic acid synthase)	0.48
NAGA	*N*-acetylgalactosaminidase, alpha	0.54
CDS1	CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 1	0.56
FNTB	Famesyltransferase, CAAX box, beta	0.61
BCKDK	Branched chain ketoacid dehydrogenase kinase	0.64
TGM6	Transglutaminase 6	0.76
PDPK1	3-Phosphoinositide dependent protein kinase-1	0.84
Transcription		
SNFT	Jun dimerization protein p21SNFT	0.13
IMP-3	IGF-II mRNA-binding protein 3	0.2
PAX8	Paired box gene 8	0.72
BCORL1	BCL6 corepressor-like 1	0.76
GATA1	GATA binding protein 1 (globin transcription factor 1)	0.77
Cell adhesion		
CEECAM1	Cerebral endothelial cell adhesion molecule 1	0.68
MADCAM1	Mucosal vascular addressin cell adhesion molecule 1	0.83
Oncogene		
RAB17	RAB17, member RAS oncogene family	0.78
RAB35	RAB35, member RAS oncogene family	0.85
RHOD	Ras homolog gene family, member D	0.74

A complete list of genes is available in Table S4.

AITL. This finding is in fact in accordance with the current definition of PTCL-u as an exclusion diagnosis. The exceptional availability of 2 AITL enriched-tumor cell samples enabled us to show that nearly 90% of the genes of the AITL signature was contributed by nonneoplastic cells, and included genes reflective of the humoral immune response, and genes encoding various chemokines mediating the recruitment of inflammatory cells, and genes involved in the modulation of vasculogenesis and extracellular matrix. These findings are in accordance with the known pathological features of AITL—including an accumulation of FDCs, B

Table 6. Genes overexpressed in CD30$^-$ PTCL-u compared to CD30$^+$ PTCL-u

Gene symbol	Gene title	Fold difference of geometric means
T-cell activation		
CD28	CD28 antigen (Tp44)	6.19
ITK	IL-2-inducible T-cell kinase	3.68
CD69	CD69 antigen (p60, early T-cell activation antigen)	3.31
CD52	CD52 antigen (CAMPATH-1 antigen)	2.97
LCK	Lymphocyte-specific protein tyrosine kinase	2.54
FYN	FYN oncogene related to SRC, FGR, YES	2.53
NFATC3	Nuclear factor of activated T cells, cytoplasmic, calcineurin-dependent 3	1.67
Transcription		
AFF3	AF4/FMR2 family, member 3	7.86
BACH2	BTB and CNC homology 1, basic leucine zipper transcription factor 2	5.48
FOXP1	Forkhead box P1	2.57
Apoptosis		
BTG1	B-cell translocation gene 1, antiproliferative	4.05
FAIM3	Fas apoptotic inhibitory molecule 3	3.91
PDCD4	Programmed cell death 4 (neoplastic transformation inhibitor)	2.45
IL24	Interleukin 24	2.42
MLL5	Myeloid/lymphoid or mixed-lineage leukemia 5 (trithorax homolog, *Drosophila*)	1.59
FLJ39616	Apoptosis-related protein PNAS-1	1.47
Lymphocyte migration		
IL16	Interleukin 16 (lymphocyte chemoattractant factor)	2.71
CCR6	Chemokine (C-C motif) receptor 6	2.64
SELL	Selectin L (lymphocyte adhesion molecule 1)	2.77
Others		
LTB	Lymphotoxin beta (TNF superfamily, member 3)	2.61
CCND3	Cyclin D3	1.83

A complete list of genes is available in Table S4.

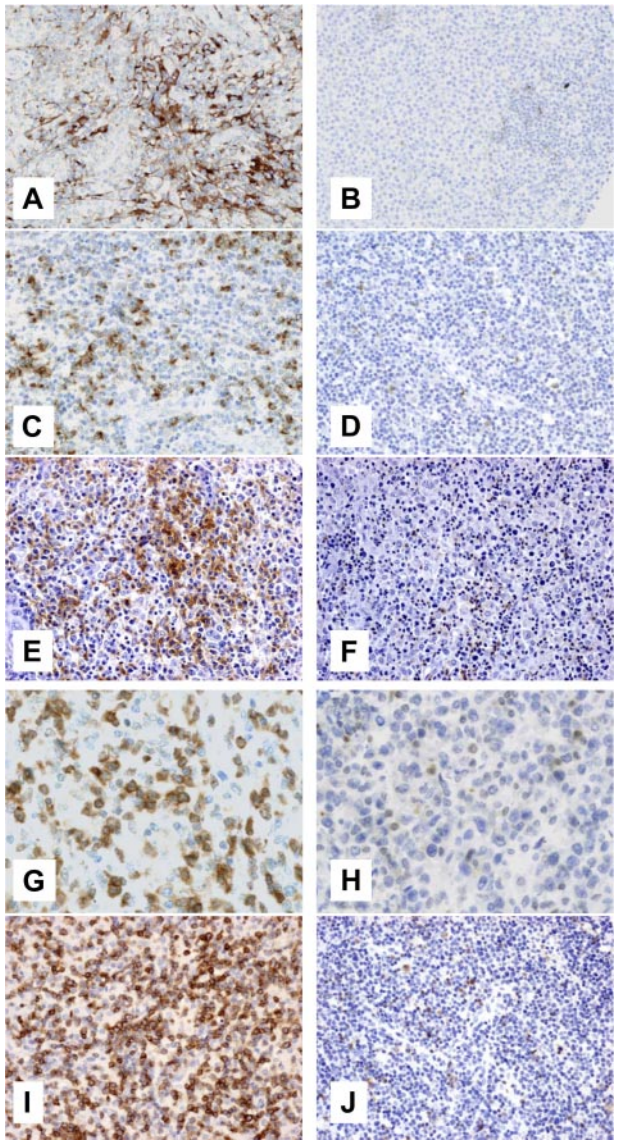

Figure 3. Clusterin, CXCL13, PDCD1, NFTAC1, and CD27 immunohistochemistry in AITL and PTCL-u. Representative AITLs (A, C, E, G, and I) and PTCLs-u (B, D, F, H, and J) were immunostained for clusterin (A-B, original magnification ×200, Zeiss Axioscope; Carl Zeiss, Heidelberg, Germany), CXCL13 (C-D, original magnification ×200, Zeiss Axioscope), PDCD1 (E-F, original magnification ×200, Nikon Eclipse 80i; Nikon, Tokyo, Japan), NFATC1 (G-H, original magnification ×400, Zeiss Axioscope), and CD27 (I-J, original magnification ×200, Nikon Eclipse 80i). For panels A-D and G-H, photographs were taken with a DP70 Olympus camera (Tokyo, Japan); images were acquired using DP Controller 2002 (Olympus) and processed using Adobe Photoshop v7.0 (Adobe Systems, San Jose, CA). For panels E-F and I-J, photographs were taken with a CFW-1310C camera (Scion, Frederick, MD); images were acquired using Histolab 5.131.1 (Alphelys, Plaisir, France) and processed using Adobe Photoshop v7.0. Objectives used were Nikon Plan Fluor 20×/0.50 NA; Zeiss Plan Neufluor 20×/0.40 NA; and Zeiss Plan Neofluor 40×/0.60 NA.

immunoblasts, plasma cells, eosinophils, and vascular proliferation. In addition, we confirmed by immunohistochemistry that clusterin, identified as a marker of AITL microenvironment and otherwise known as a marker for FDCs, was indeed expressed in nonneoplastic stromal cells with FDC morphology, mainly in cases classified as AITL. In agreement with the recent identification of CD10 as a typical marker of AITL,[44] we also confirmed overexpression of the CD10 mRNA in the AITL neoplastic cell component.

In order to gain insight into the ontogeny of AITLs, we have applied the recently described GSEA method[27] to evaluate the strength of the correlation between the AITL signature and that of T_{FH} cells. The demonstration of a significant enrichment of the T_{FH} gene sets not only in the AITL tumor cells but also in the AITL tissue samples, based on wide-genome expression analysis and robust statistical tools, provides definitive molecular evidence that T_{FH} cells are the normal cellular precursors to AITL. Indeed, in keeping with recent reports,[12-15] we showed that the immunohistochemical detection of several T_{FH} markers—especially CXCL13 and PDCD1—help to recognize the (often minimal) neoplastic component of AITL and therefore may provide an important clue for the differential diagnosis with conditions mimicking AITL, including dysimmune reactive conditions.

It may be objected, however, that the profile of AITL tumor cells significantly differs from their postulated normal T_{FH} counterpart. Indeed, whereas T_{FH} cells were initially characterized by their $CD57^+$ phenotype, CD57 expression is not a classical feature of AITL.[8,13] This could be explained by recent data showing phenotypic heterogeneity within T_{FH} cells, which, despite a common transcriptional profile,[21] comprise a subset of $CD57^-$ cells.[34] Of interest, CD57 appears to be dispensable to the follicular B-helper activity, which rather relies upon strong ICOS expression.[45] This is corroborated by our data set, as ICOS transcripts were more abundant in AITLs than in PTCLs-u (FC = 3.06, $P = .08$), and more abundant in AITL cells compared to tissues (FC = 5.09).

AITL encompasses a morphologic spectrum, ranging from atypical lymphoproliferations to overtly malignant lymphomas. Three histologic patterns have been described: with hyperplastic follicles, depleted follicles, or without follicles, and interpreted as reflective of earlier to later disease stages.[44] These observations in fact fit well the proposed ontogenic model, which intuitively would imply that early disease arises in association with germinal centers and that extrafollicular extension parallels disease progression. Our findings also reinforce the hypothesis of the relationship of AITL to a peculiar form of PTCL-u with a follicular growth pattern and suspected to be derived from germinal center T cells,[46] and perhaps to another form of PTCL with a parafollicular pattern of growth[47] or with involvement of the follicular mantle zone.[48]

It is remarkable that a minor subset of normal T cells[34] would be the cell of origin of one of the most frequent type of T-cell lymphoma. In view of the major role of the B-cell follicle in B-cell lymphomagenesis, with more than half of mature B-cell lymphomas deriving from germinal center cells, it is questionable whether mechanisms of genetic alterations similar to those operating in B-cell transformation may also target T_{FH} cells.

Analysis of the molecular features of PTCLs-u considered as a whole has provided deceptively little information. The molecular features of PTCL-u were associated with common biologic functions, and GSEA failed to disclose any correlation with normal T_{FH}, Th1, and Th2 subsets. These results may have been anticipated given the known heterogeneity within the PTCL-u category and the limited number of samples analyzed in this study, and in light of the complexity and variability of normal T-cell subsets, as well as their phenotypic modulation upon activation. The heterogeneity of PTCLs at the molecular level has been highlighted in a previous microarray study.[43] Further larger studies are warranted to better understand the ontogeny of PTCL-u.

The distinct clustering of $CD30^+$ and $CD30^-$ PTCLs-u suggested that these may be associated with distinct molecular features, and, indeed, we found that the $CD30^+$ subset had reduced expression of genes with specific functions, in particular T-cell activation and TCR signal transduction factors such as

Lck, Fyn, and Lyn. Of interest, lack of TCR expression is a feature of both ALK^+ and ALK^- subsets of ALCL, and it has been suggested that this characteristic may provide a unifying feature for ALCL.[49] Expanding on this concept, our findings raise the question as to whether CD30 expression may be an objective criteria to delineate a meaningful PTCL category. The clinical and biologic relevance of this distinction should be investigated in larger series of cases. The reduced transcript levels of CD52 in $CD30^+$ PTCL-u is of particular clinical interest, since pilot studies have confirmed the potential interest of the humanized anti-CD52 monoclonal antibody (alemtuzumab) for the treatment of T-cell lymphoproliferations,[50] and the level of target expression may be an important determinant of response to therapy.[51]

Of importance, the results of our study provide novel pieces of information to substantiate the debate on the pathological classification of PTCLs, and more specifically the borders of the PTCL-u spectrum as opposed to AITL. Indeed, we found that the molecular signature of $CD30^-$ PTCL-u overlaps with that of T_{FH}, although this is less significant than for AITL. Upon review of the cases, we identified a few PTCL-u cases with minimal AITL-like characteristics and molecular T_{FH}-like features, which largely accounted for this statistical association. This suggests that $CD30^-$ PTCLs-u include some lymphomas derived from AITLs. Accordingly, the spectrum of AITL may be wider than suspected. Alternatively, it cannot be formally eliminated that a subgroup of PTCLs-u, distinct from AITL, may also derive from T_{FH} cells but develop along a distinct pathogenic pathway. This raises the questions whether the classic diagnostic criteria used to differentiate both entities may be too stringent,[2] and how to define the border between AITL and PTCL-u. Of interest, we were able to construct molecular predictors for the AITL versus PTCL-u distinction, and an 8-gene predictor yielded a very high success rate. Although these results need to be expanded validated on a larger series of samples, they lay the premise that the molecular classification may supplement other approaches for the classification of PTCLs.

The major finding presented herein is the identification of the putative normal cellular counterpart of AITL. In the future, this might provide a basis for a more precise diagnosis of AITL, and possibly for the elaboration of T_{FH}-specific targeted therapies.

Acknowledgments

This work was supported by the Institut National de la Santé et de la Recherche Médicale (INSERM), by the Ligue Contre le Cancer and the program "Carte d'Identité des Tumeurs," and by the Belgian National Fund for Scientific Research (FNRS). L.d.L. is a senior researcher and C.T. is a research assistant of the FNRS.

The authors wish to thank C. Coppeaux, S. Sousa, N. Martin-Garcia, and J. Marquet for their technical assistance; V. Fattaciolli for tumor banking; and F. Petel for submitting the data to EBI. We also thank our colleagues Drs N. Brousse, F. Charlotte, C. Chassagne-Clément, C. Lavignac, P. Pocachard, and I. Theate for providing samples.

Authorship

Contribution: L.d.L., L.X., and P.G. designed the research; C.T. and Y.-L.H. performed the research; L.d.L., C.T., Y.-L.H., G.D., L.L., K.L., J.B., T.M., F.B., C.G., and P.G. collected the data; L.d.L., D.S.R., C.T., A.R., C.G., L.X., and P.G. analyzed the data; D.S.R. and A.R. contributed vital analytical tools; and L.d.L., D.S.R., L.X., and P.G. wrote the paper.

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

L.d.L. and D.S.R. equally contributed to this work.

Correspondence: Laurence de Leval, Department of Pathology, CHU Sart-Tilman, Tour de Pathologie, +1, 4000 Liège, Belgium; e-mail: l.deleval@ulg.ac.be.

References

1. A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma: The Non-Hodgkin's Lymphoma Classification Project. Blood. 1997;89:3909-3918.
2. Jaffe E, Harris N, Stein H, Vardiman J. Pathology and genetics: tumours of haematopoietic and lymphoid tissues. In: Kleihues P, Sobin L, eds. World Health Organization Classification of Tumours. Lyon, France: IARC Press; 2001:190-235.
3. Dogan A, Attygalle AD, Kyriakou C. Angioimmunoblastic T-cell lymphoma. Br J Haematol. 2003;121:681-691.
4. Stein H, Foss HD, Durkop H, et al. CD30(+) anaplastic large cell lymphoma: a review of its histopathologic, genetic, and clinical features. Blood. 2000;96:3681-3695.
5. Rudiger T, Weisenburger DD, Anderson JR, et al. Peripheral T-cell lymphoma (excluding anaplastic large-cell lymphoma): results from the Non-Hodgkin's Lymphoma Classification Project. Ann Oncol. 2002;13:140-149.
6. Geissinger E, Odenwald T, Lee SS, et al. Nodal peripheral T-cell lymphomas and, in particular, their lymphoepithelioid (Lennert's) variant are often derived from CD8(+) cytotoxic T-cells. Virchows Arch. 2004;445:334-343.
7. Went P, Agostinelli C, Gallamini A, et al. Marker expression in peripheral T-cell lymphoma: a proposed clinical-pathologic prognostic score. J Clin Oncol. 2006;24:2472-2479.
8. Lee SS, Rudiger T, Odenwald T, et al. Angioimmunoblastic T cell lymphoma is derived from mature T-helper cells with varying expression and loss of detectable CD4. Int J Cancer. 2003;103:12-20.
9. Jones D, Fletcher CD, Pulford K, et al. The T-cell activation markers CD30 and OX40/CD134 are expressed in nonoverlapping subsets of peripheral T-cell lymphoma. Blood. 1999;93:3487-3493.
10. Jones D, O'Hara C, Kraus MD, et al. Expression pattern of T-cell-associated chemokine receptors and their chemokines correlates with specific subtypes of T-cell non-Hodgkin lymphoma. Blood. 2000;96:685-690.
11. Ree HJ, Kadin ME, Kikuchi M, et al. Bcl-6 expression in reactive follicular hyperplasia, follicular lymphoma, and angioimmunoblastic T-cell lymphoma with hyperplastic germinal centers: heterogeneity of intrafollicular T-cells and their altered distribution in the pathogenesis of angioimmunoblastic T-cell lymphoma. Hum Pathol. 1999;30:403-411.
12. Grogg KL, Attygalle AD, Macon WR, et al. Angioimmunoblastic T-cell lymphoma: a neoplasm of germinal-center T-helper cells? Blood. 2005;106:1501-1502.
13. Dupuis J, Boye K, Martin N, et al. Expression of CXCL13 by neoplastic cells in angioimmunoblastic T-cell lymphoma (AITL): a new diagnostic marker providing evidence that AITL derives from follicular helper T cells. Am J Surg Pathol. 2006;30:490-494.
14. Grogg KL, Attygale AD, Macon WR, et al. Expression of CXCL13, a chemokine highly upregulated in germinal center T-helper cells, distinguishes angioimmunoblastic T-cell lymphoma from peripheral T-cell lymphoma, unspecified. Mod Pathol. 2006;19:1101-1107.
15. Dorfman DM, Brown JA, Shahsafaei A, Freeman GJ. Programmed death-1 (PD-1) is a marker of germinal center-associated T cells and angioimmunoblastic T-cell lymphoma. Am J Surg Pathol. 2006;30:802-810.
16. Krenacs L, Schaerli P, Kis G, Bagdi E. Phenotype of neoplastic cells in angioimmunoblastic T-cell lymphoma is consistent with activated follicular B helper T cells. Blood. 2006;108:1110-1111.
17. Streubel B, Vinatzer U, Willheim M, et al. Novel t(5;9)(q33;q22) fuses ITK to SYK in unspecified peripheral T-cell lymphoma. Leukemia. 2006;20:313-318.
18. Zettl A, Rudiger T, Konrad MA, et al. Genomic profiling of peripheral T-cell lymphoma, unspecified, and anaplastic large T-cell lymphoma delineates novel recurrent chromosomal alterations. Am J Pathol. 2004;164:1837-1848.
19. Rogge L, Bianchi E, Biffi M, et al. Transcript imaging of the development of human T helper cells using oligonucleotide arrays. Nat Genet. 2000;25:96-101.
20. Kim CH, Lim HW, Kim JR, et al. Unique gene expression program of human germinal center T helper cells. Blood. 2004;104:1952-1960.

21. Chtanova T, Tangye SG, Newton R, et al. T follicular helper cells express a distinctive transcriptional profile, reflecting their role as non-Th1/Th2 effector cells that provide help for B cells. J Immunol. 2004;173:68-78.
22. Chtanova T, Newton R, Liu SM, et al. Identification of T cell-restricted genes, and signatures for different T cell responses, using a comprehensive collection of microarray datasets. J Immunol. 2005;175:7837-7847.
23. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5:R80.
24. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003;4:249-264.
25. Mardia K, Kent J, Bibby J. Multivariate analysis. London, United Kingdom: Academic Press; 1979.
26. Wright GW, Simon RM. A random variance model for detection of differential gene expression in small microarray experiments. Bioinformatics. 2003;19:2448-2455.
27. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545-15550.
28. Wright G, Tan B, Rosenwald A, et al. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci U S A. 2003;100:9991-9996.
29. Pomeroy SL, Tamayo P, Gaasenbeek M, et al. Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature. 2002;415:436-442.
30. Cozier GE, Carlton J, Bouyoucef D, Cullen PJ. Membrane targeting by pleckstrin homology domains. Curr Top Microbiol Immunol. 2004;282:49-88.
31. Meroni G, Diez-Roux G. TRIM/RBCC, a novel class of 'single protein RING finger' E3 ubiquitin ligases. Bioessays. 2005;27:1147-1157.
32. Moser B, Loetscher P. Lymphocyte traffic control by chemokines. Nat Immunol. 2001;2:123-128.
33. Ticchioni M, Raimondi V, Lamy L, et al. Integrin-associated protein (CD47/IAP) contributes to T cell arrest on inflammatory vascular endothelium under flow. Faseb J. 2001;15:341-350.
34. Vinuesa CG, Tangye SG, Moser B, Mackay CR. Follicular B helper T cells in antibody responses and autoimmunity. Nat Rev Immunol. 2005;5:853-865.
35. Bower KE, Zeller RW, Wachsman W, et al. Correlation of transcriptional repression by p21(SNFT) with changes in DNA.NF-AT complex interactions. J Biol Chem. 2002;277:34967-34977.
36. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000;403:503-511.
37. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med. 2002;346:1937-1947.
38. Savage KJ, Monti S, Kutok JL, et al. The molecular signature of mediastinal large B-cell lymphoma differs from that of other diffuse large B-cell lymphomas and shares features with classical Hodgkin lymphoma. Blood. 2003;102:3871-3879.
39. Rosenwald A, Wright G, Leroy K, et al. Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. J Exp Med. 2003;198:851-862.
40. 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;10:4971-4982.
41. Martinez-Delgado B, Cuadros M, Honrado E, et al. Differential expression of NF-kappaB pathway genes among peripheral T-cell lymphomas. Leukemia. 2005;19:2254-2263.
42. Piccaluga PP, Agostinelli C, Zinzani PL, et al. Expression of platelet-derived growth factor receptor alpha in peripheral T-cell lymphoma not otherwise specified. Lancet Oncol. 2005;6:440.
43. Ballester B, Ramuz O, Gisselbrecht C, et al. Gene expression profiling identifies molecular subgroups among nodal peripheral T-cell lymphomas. Oncogene. 2006;25:1560-1570.
44. Attygalle A, Al-Jehani R, Diss TC, et al. Neoplastic T cells in angioimmunoblastic T-cell lymphoma express CD10. Blood. 2002;99:627-633.
45. Rasheed AU, Rahn HP, Sallusto F, et al. Follicular B helper T cell activity is confined to CXCR5(hi)ICOS(hi) CD4 T cells and is independent of CD57 expression. Eur J Immunol. 2006;36:1892-1903.
46. de Leval L, Savilo E, Longtine J, et al. Peripheral T-cell lymphoma with follicular involvement and a CD4+/bcl-6+ phenotype. Am J Surg Pathol. 2001;25:395-400.
47. Rudiger T, Ichinohasama R, Ott MM, et al. Peripheral T-cell lymphoma with distinct perifollicular growth pattern: a distinct subtype of T-cell lymphoma? Am J Surg Pathol. 2000;24:117-122.
48. Ikonomou IM, Tierens A, Troen G, et al. Peripheral T-cell lymphoma with involvement of the expanded mantle zone. Virchows Arch. 2006;449:78-87.
49. 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;104:3358-3360.
50. Dearden CE, Matutes E. Alemtuzumab in T-cell lymphoproliferative disorders. Best Pract Res Clin Haematol. 2006;19:795-810.
51. Ginaldi L, De Martinis M, Matutes E, et al. Levels of expression of CD52 in normal and leukemic B and T cells: correlation with in vivo therapeutic responses to Campath-1H. Leuk Res. 1998;22:185-191.