Defining signatures of peripheral T-cell lymphoma with a targeted 20-marker gene expression profiling assay

Fanny Drieux,[1,2,3*] Philippe Ruminy,[1*] Ahmad Abdel-Sater,[1] François Lemonnier,[3,4] Pierre-Julien Viailly,[1] Virginie Fataccioli,[3] Vinciane Marchand,[1] Bettina Bisig,[5] Audrey Letourneau,[5] Marie Parrens,[6] Céline Bossard,[7] Julie Bruneau,[8] Pamela Dobay,[5] Liana Veresezan,[1,2] Aurélie Dupuy,[3] Anaïs Pujals,[3,9] Cyrielle Robe,[3] Nouhoum Sako,[3] Christiane Copie-Bergman,[3,9] Marie-Hélène Delfau-Larue,[3,10] Jean-Michel Picquenot,[1,2] Hervé Tilly,[1] Richard Delarue,[11] Fabrice Jardin,[1#] Laurence de Leval[5#] and Philippe Gaulard[3,9#]

[1]INSERM U1245, Centre Henri Becquerel, Rouen, France; [2]Service d'Anatomie et Cytologie Pathologiques, Centre Henri Becquerel, Rouen, France; [3]INSERM U955 and Université Paris-Est, Créteil, France; [4]Unité Hémopathies Lymphoïdes, Groupe Hospitalier Henri Mondor, AP-HP, Créteil, France; [5]Institut de Pathologie, Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, Switzerland; [6]Service d'Anatomie et Cytologie Pathologiques, Hôpital Haut-Lévêque, CHU de Bordeaux, France; [7]Service d'Anatomie et Cytologie Pathologiques, CHU de Nantes, France; [8]Service d'Anatomie et Cytologie Pathologiques, Hôpital Universitaire Necker - Enfants Malades, Assistance Publique - Hôpitaux de Paris (APHP), Paris, France; [9]Département de Pathologie, Groupe Hospitalier Henri Mondor, AP-HP, Créteil, France; [10]Département d'Hématologie et Immunologie Biologique, Groupe Hospitalier Henri Mondor, AP-HP, Créteil, France and [11]Service Hématologie Adultes, Hôpital Universitaire Necker - Enfants Malades, Assistance Publique - Hôpitaux de Paris (APHP), Paris, France

**FD and PR contributed equally as co-first authors.*

#FJ, LdL and PG contributed equally as co-senior authors.

 doi:10.3324/haematol.2019.226647

Received: May 9, 2019.

Accepted: September 2, 2019.

Pre-published: September 5, 2019.

Correspondence: *PHILIPPE GAULARD* -philippe.gaulard@aphp.fr

PHILIPPE RUMINY - philippe.ruminy@chb.unicancer.fr

Supplemental Material

Patients and tumor samples

As shown in Figure S1, the cases were divided into two cohorts. The first cohort (classification cohort, n = 230) used to build and train a predictive classification model, was composed of 230 cases, including 30 angioimmunoblastic T-cell lymphoma (AITL), 33 T_{FH}-PTCL, 21 ALK-positive anaplastic large cell lymphomas (ALCL), 34 ALK-negative ALCL, 16 extranodal NK/T-cell lymphomas, nasal-type (NKTCL), 6 hepatosplenic T-cell lymphomas (HSTL), and 13 adult T-cell leukemia/lymphomas (ATLL), and 77 PTCL-NOS cases, according to the WHO 2017 classification. The second cohort (diagnostic cohort) of 40 FFPE PTCL samples (6 ALK-positive ALCL, 4 ALK-negative ALCL, 13 AITL, 9 NKTCL, 6 ATLL, and 2 PTCL-NOS) was used to validate the robustness of the assay and its reproducibility between three independent centers.

Immunohistochemistry and EBV *in situ* hybridization

Deparaffinized tissue sections were stained for a panel of T-cell (CD3, CD2, CD7, CD5, CD8, CD4), cytotoxic (TIA-1, granzyme B, perforin), T_{FH} (PD1, CXCL13, BCL6, ICOS, CD10), follicular dendritic-cell (CD21, CD23), B-cell (CD20, CD79a, PAX5), and other (CD30, CD56, ALK) antigens. Immunostains for GATA3 (clone HG3-31, Santa Cruz Biotechnology, Santa Cruz, CA) and TBX21 (clone 4B10, BD Biosciences, San Jose, CA) were performed on a subset of cases. The scoring system was evaluated as follows: score 0: negative (below threshold of 10% positive tumor cells), score 1: 10-50% positive tumor cells, score 2: >50% positive tumor cells as previously described (Dobay, Haematologica, 2017;102(4): e148-e151). The detection of EBV was performed by *in situ* hybridization using EBER probes and was scored as previously described (de Leval, Haemalogica 2015;100(9):e361-364): score 0: absence of large EBV-positive cells, score 1: up to 5 large EBV-positive cells per high power fields (hpf), score 2 : 5 to 50 per hpf and score 3 : > 50 per hpf , or sheets or aggregates of large EBV-positive cells.

FISH for *DUSP22/IRF4* rearrangement

Laboratory-developed fluorescence *in situ* hybridization FISH-probes using bacterial artificial chromosomes (BACs) were used to explore *DUSP22/IRF4* rearrangements by interphase FISH in FFPE tissue sections of ALK-negative ALCL. Break-apart probes consisted of telomeric

RP3-416J7 labelled with SpectrumOrange™ and centromeric RP11-615C17 and CTD-3139L20 labelled with SpectrumGreen™ (Institute of Pathology, Lausanne).

Allele-specific PCR and targeted deep sequencing

RHOA G17V and *IDH2* R172K/T mutational status were also determined by allele-specific qPCR (AS-qPCR) or next generation sequencing (NGS) on a MiSeq instrument with a mean coverage depth of 1200X. (Dupuy et al. J Mol Diag)

RNA extraction and microarray procedure

Total RNA was extracted from frozen and FFPE tissue samples with Trizol reagent and the Maxwell 16 LEV RNA kit (Promega), respectively. High-throughput gene-expression data (HG-U133 plus 2.0 chips, Affymetrix, Santa Clara, CA) from 72 cases (23 AITL and 49 PTCL-NOS) were previously reported (de Leval et al, Blood. 2007;109(11):4952–4963).

RT-MLPA procedure

All probes consisted of a gene-specific region complementary to the cDNA target linked to a tail. The 5' and 3' tails correspond to the primers U1 (TCCAACCCTTAGGGAACCC) and U2 (GTGCCAGCAAGATCCAATCTAGA) used for the final PCR amplification step. Spacers were added between these tails and the gene-specific regions to allow the separation and identification of the PCR products based on size. The 3' probes were phosphorylated at the 5' ends. For four genes, specific probes without the PCR tails were used as competitors to normalize the amplification signal. The competitors were added to the corresponding probes at a ratio of 1 (MLPA) to 4 (competitor) for *GZB*, *CXCL13*, *TCRα*, and 1 (MLPA) to 3 (competitor) for *PRF*. For the identification of *RHOAG17V* and *IDH2R172K/T* hotspot mutations, 5' probes with the last nucleotide corresponding to the mutated nucleotide were designed.

Bioinformatic analysis

Starting from raw (.fsa) files generated by the genetic analyzer, the web interface provides a graphical representation of the gene expression profile and a table of normalized gene expression calculated as a function of the FAM fluorescence intensity normalized to the

median intensity of the 20 genes of the signature. For each sample, this interface also returns a class prediction deduced from a support vector machine classifier (SVM), established using the e1071 R package with default settings. To minimize the risks of misclassification, this algorithm was coupled with a second nearest centroid classifier to minimize the risks of misclassification. For each class *C*, the coordinates of the centroid μc was calculated so that

$$\mu_C = \frac{1}{N}\sum_{i=1}^{N} S_i$$

where N represents the number of cases belonging to the class *C* defined by the SVM classifier in the training series and *Si*, a sample defined by the expression of the 20 RT-MLPA markers. For each sample *Si*, the distance d(*Si*, *C*) to the centroid μc of the class *C* is given by

$$d(S_i, C) = \sqrt{\sum_{i=1}^{N}(S_i - \mu_C)^2}$$

For each sample, the class prediction returned by the web interface consists of the results of the SVM algorithm completed by the arithmetic distance to the centroids of each class.

A stratified cross validation was performed to assess the accuracy of our assay. A bootstrap resampling process was first used to build 100 independent training series, randomly selecting two-thirds of the samples within the six categories of the unsupervised hierarchical analysis. For each iteration, a SVM predictor was trained and tested against all remaining samples. A definitive SVM predictor was thus trained using the 184 cases classified by hierarchical analysis.

Post-tests were then built to distinguish ALCL from cytotoxic/Th1 cases. A specific threshold was determined using ROC curves, based on the expression of the *CD30* gene values by RT-MLPA (CD30 threshold = 0.8). A second post-test was performed to distinguish ALK-positive from ALK-negative ALCLs, based on the expression of the *ALK* gene (threshold = 0.2).

Another CD30 post-test was designed among the Th2 category to catch misclassified CD30-positive Th2 cases (threshold = 1.4) due to overlapping *FOXP3* and *ICOS* expression in both the CD30-positive and CD30-negative Th2 groups.

An algorithm was finally performed to calculate the distance of the sample to the centroid of the predicted class and compare it to the other intraclass distances. The algorithm constructs a boxplot based on these intraclass distances and calculates the first and third quartile values, deducting the IQR (InterQuartile Range) value, which is equal to Q3-Q1.

Finally, if the distance of the tested sample was higher than Q3 + 1.5 IQR, the sample was considered to be a mild outlier, which defined the NOS category.

Tables

Table S1- Sequences of the oligonucleotides used for the RT-MLPA assay. Genes are listed according to the probes size. Specific sequences are represented in pink (5'probe) and red (3'probe). The blue sequence corresponds to the common 5' and 3' tails. Nucleotides spacers are in green.

Gene	Oligo	probe	Sequence
CD8	CD8E5L	5'probe	GTGCCAGCAAGATCCAATCTAGATCGTGCCGGTCTTCCTGCCAG
	CD8E6R	3'probe	CGAAGCCCACCACGACGCCTCCAACCCTTAGGGAACCC
IDH2R172K	IDH2R172KL	5'probe	GTGCCAGCAAGATCCAATCTAGACCAAGCCCATCACCATTGGCAA
IDH2R172T	IDH2R172TL	5'probe	GTGCCAGCAAGATCCAATCTAGACCAAGCCCATCACCATTGGCAC
	IDH2R172KR	3'probe	GCACGCCCATGGCGACCAGTTCCAACCCTTAGGGAACCC
EBER1	EBER1L	5'probe	GTGCCAGCAAGATCCAATCTAGATACGTAGCCACCCGTCCCGGGTA
	EBER1R	3'probe	CAAGTCCCGGGTGGTGAGGATATCCAACCCTTAGGGAACCC
GATA3	GATA3E3L	5'probe	GTGCCAGCAAGATCCAATCTAGACCTCATTAAGCCCAAGCGAAGGCTG
	GATA3E4R	3'probe	TCTGCAGCCAGGAGAGCAGGGACTCCAACCCTTAGGGAACCC
ALK	ALKE23F	5'probe	GTGCCAGCAAGATCCAATCTAGACCTCCGAGAGACCCGCCCTCGCCCG
	ALKE24R	3'probe	AGCCAGCCCTCCTCCCTGGCCATGCTCCAACCCTTAGGGAACCC
FOXP3	FOXP3E3L	5'probe	GTGCCAGCAAGATCCAATCTAGATAGGACAGGCCACATTTCATGCACCAG
	FOXP3E4R	3'probe	CTCTCAACGGTGGATGCCCACGCTTCCAACCCTTAGGGAACCC
CD4	CD4E4L	5'probe	GTGCCAGCAAGATCCAATCTAGAGAGGAGGTGCAATTGCTAGTGTTCGGAT
	CD4E5R	3'probe	TGACTGCCAACTCTGACACCCACCTTCCAACCCTTAGGGAACCC
CD30	CD30E3L	5'probe	GTGCCAGCAAGATCCAATCTAGATTGTACAGCCTGCGTGACTTGTTCTCGAG
	CD30E4R	3'probe	ACGACCTCGTGGAGAAGACGCCGTACTCCAACCCTTAGGGAACCC
PFR	PFRE2L	5'probe	GTGCCAGCAAGATCCAATCTAGATAACACGGTGGAGTGCCGCTTCTACAG
	PFRE3R	3'probe	TTTCCATGTGGTACACACTCCCCCGTACTACTCCAACCCTTAGGGAACCC
BCL6	BCL6E3Lb	5'probe	GTGCCAGCAAGATCCAATCTAGACATAAAACGGTCCTCATGGCCTGCAG
	BCL6E4Rb	3'probe	TGGCCTGTTCTATAGCATCTTTACAGACCAGTTGTCCAACCCTTAGGGAACCC
RHOA mut	RHOmutL	5'probe	GTGCCAGCAAGATCCAATCTAGAGGTGATTGTTGGTGATGGAGCCTGTGT
RHOA	RHOR	3'probe	AAAGACATGCTTGCTCATAGTCTTCAGCAAGGACCTCCAACCCTTAGGGAACCC
GZMB	GRBE3L	5'probe	GTGCCAGCAAGATCCAATCTAGATACTAACTTCTCCAACGACATCATGCTACTGCAG
	GRBE4R	3'probe	CTGGAGAGAAAGGCCAAGCGGACCAGTACTACTCCAACCCTTAGGGAACCC
TBET	TBETE5L	5'probe	GTGCCAGCAAGATCCAATCTAGATACTACCCAAAGGATTCCGGGAGAACTTTGAGTC
	TBETE6R	3'probe	CATGTACACATCTGTTGACACCAGCATCCCTACTTCCAACCCTTAGGGAACCC
CD56	CD56E11L	5'probe	GTGCCAGCAAGATCCAATCTAGATACTACTACTCACCCCCTCTGCCAGCTATCTGGAG
	CD56E12R	3'probe	GTGACCCCAGACTCTGAGAATGATTTTGGTACTACTCCAACCCTTAGGGAACCC
CXCL13	CXCL13E2L	5'probe	GTGCCAGCAAGATCCAATCTAGATACTACTGGTCAGCAGCCTCTCTCCAGTCCAAG
	CXCL13E3R	3'probe	GTGTTCTGGAGGTCTATTACACAAGCTTGAGGTGTTACTTCCAACCCTTAGGGAACCC
ICOS	ICOSE2L	5'probe	GTGCCAGCAAGATCCAATCTAGAAAGTAACTCTTACAGGAGGATATTTGCATATTTATG
	ICOSE3R	3'probe	AATCACAACTTTGTTGCCAGCTGAAGTTCTGTACTACTTCCAACCCTTAGGGAACCC
TRAC	TRACE3L	5'probe	GTGCCAGCAAGATCCAATCTAGATACTACTACTACTACTACCTGCGGCTGTGGTCCAGCTGAG
	TRACE4R	3'probe	ATCTGCAAGATTGTAAGACAGCCTGTGCTCTACTACTATCCAACCCTTAGGGAACCC
CXCR5	CXCR5E1L	5'probe	GTGCCAGCAAGATCCAATCTAGATACTACTACTACTACTGGACCTCGAGAACCTGGAGGACCTG
	CXCR5E2R	3'probe	TTCTGGGAACTGGACAGATTGGACAACTATAACGTACTACTCCAACCCTTAGGGAACCC
INFg	INFgE3L	5'probe	GTGCCAGCAAGATCCAATCTAGATACTAAACGAGATGACTTCGAAAAGCTGACTAATTATTCG
	IFNgE4R	3'probe	GTAACTGACTTGAATGTCCAACGCAAAGCATACTACTACTACTTCCAACCCTTAGGGAACCC
CCR4	CCR4E1L	5'probe	GTGCCAGCAAGATCCAATCTAGATACTACTACTACTCCTCAGAGCCGCTTTCAGAAAAGCAAG
	CCR4E2R	3'probe	CTGCTTCTGGTTGGGCCCAGACCTTACTACTACTACTACTACTACTCCAACCCTTAGGGAACCC

Table S2. Comparison of *RhoA*G17V (n = 33) and *IDH2*R172K/T (n = 8) mutations detected by RT-MLPA and next generation sequencing (NGS) or allele-specific-qPCR (AS-qPCR) analysis.
Concordant results are represented in green, while discordant result in red.

Id	Pathology	RhoAG17V status by RT-MLPA	RhoA status by AS-qPCR /NGS (allele frequency)	IDH2 R172K/T status by RT-MLPA	IDH2 status with AS-qPCR/ NGS (allele frequency)
UPN001	AITL	+	+/NA	WT	WT/NA
UPN002	AITL	+	+/NA	+	+/NA
UPN003	AITL	+	+/NA	+	+/NA
UPN004	AITL	+	+/NA	WT	WT/NA
UPN005	AITL	+	+/NA	+	+/NA
UPN006	AITL	+	+/+ (6.12%)	WT	NA/NA
UPN007	AITL	+	+/+ (19.27%)	WT	NA/NA
UPN009	AITL	+	+/+ (12.93%)	+	+/+ (5.08%)
UPN010	AITL	+	+/+ (8.53%)	WT	NA/NA
UPN012	AITL	+	+/+ (23.3%)	WT	NA/NA
UPN016	AITL	+	+/+ (22.38%)	+	+/+ (14.9%)
UPN018	AITL	+	+/+ (9.13%)	WT	NA/NA
UPN020	AITL	+	+/+ (12.24%)	+	+/+ (10.23%)
UPN024	AITL	+	+/+ (18.2%)	WT	NA/NA
UPN025	AITL	+	+/+ (7.44%)	WT	WT/+ (2.83%)
UPN026	AITL	+	+/+ (17.52%)	+	+/NA
UPN028	AITL	+	+/NA	WT	NA/NA
UPN029	AITL	+	+/+ (18.31%)	WT	NA/NA
UPN177	AITL	+	+/NA	+	+/NA
UPN117	PTCL TFH	+	+/+ (38.93%)	WT	NA/NA
UPN136	PTCL TFH	+	+/NA	WT	NA/NA
UPN113	PTCL TFH	+	+/NA	WT	NA/NA
UPN120	PTCL TFH	+	+/+ (23.54%)	WT	NA/NA
UPN118	PTCL TFH	+	+/NA	+	+/+ (2.22%)
UPN115	PTCL TFH	+	+/+ (23.15%)	WT	NA/NA
UPN114	PTCL TFH	+	+/NA	WT	NA/NA
UPN138	PTCL TFH	+	+/+ (21.87%)	WT	NA/NA
UPN134	PTCL TFH	+	+/NA	WT	NA/NA
UPN165	PTCL TFH	+	+/NA	WT	NA/NA
UPN116	PTCLnos	+	+/+ (50%)	WT	NA/NA
UPN125	PTCLnos	+	+/NA	WT	NA/NA
UPN137	PTCLnos	+	+/NA	WT	NA/NA
UPN091	ATLL	+	+/NA	WT	NA/NA

Table S3. Clinical parameters of ALK-negative ALCL and comparison of the non-cytotoxic ALCL ALK-negative subgroup according to *DUSP22* status

	ALCL ALK- n=13		Non-cytotoxic ALCL ALK-negative n=24		DUSP22R (n=8)		DUSP22NR (n=8)		p
	median (IQ)	% (n)	median (IQ)	% (n)	median (IQ)	% (n)	median (IQ)	% (n)	
age median (range)	53.8 (40.8-67.2)		58 (52-70)		55 (43-72)		56.45 (53.1-70.3)		0.63
<= 60 years		63.6% (7/11)		54.2% (13/24)		62.5% (5/8)		62.5% (5/8)	1
gender male		53.8% (7/13)		79.2% (19/24)		87.5% (7/8)		75% (6/8)	1
IPI>=3		62.5% (5/8)		33.3% (7/21)		50% (4/8)		16.6% (1/6)	0.3
PIT>=2		85.7% (6/7)		42.8% (9/21)		50% (4/8)		16.6% (1/6)	0.3
extranodal site >=2		85.7% (6/7)		21.7% (5/23)		25% (2/8)		28.6% (2/7)	1
stage >=3		77.8% (7/9)		73.9% (17/23)		75% (6/8)		85.7% (6/7)	1
PS>=2		50% (4/8)		21% (4/19)		28.6% (2/7)		0 (0/5)	0.47
LDH>=1		38.5% (5/13)		50% (11/22)		50% (4/8)		60% (3/5)	1
B symptoms		50% (4/8)		35% (7/20)		57.1% (4/7)		20% (1/5)	0.33

Table S4. Correlation scores of the RT-MLPA values (n = 20) of 40 FFPE samples between three independent centers (very strong correlation (VSC): rho > 0.9, strong correlation (SC) rho > 0.7)

	corr-C2-C3	corr-C1-C3	corr-C1-C2	
CD8	0.9777741	0.9553372	0.9683209	VSC
IDH2	-0.2502684	0.4048631	0.3518407	NA
EBER	0.9951752	0.9954745	0.9877439	VSC
GATA3	0.9771558	0.939488	0.9531494	VSC
ALK	0.9894248	0.9789414	0.9823311	VSC
FOXP3	0.9509514	0.927661	0.9438251	VSC
CD4	0.967501	0.8858032	0.9321464	VSC
CD30	0.9897987	0.9903066	0.9864354	VSC
PRF	0.971159	0.987516	0.9687311	VSC
BCL6	0.9412178	0.8655903	0.8919673	SC
RHOA	0.9324257	0.7869005	0.7019153	SC
GRB	0.9740104	0.9800914	0.9456018	VSC
TBX21	0.9366044	0.9629732	0.9254932	VSC
CD56	0.9362774	0.9749429	0.9381184	VSC
CXCL13	0.9633031	0.9472914	0.9743159	VSC
ICOS	0.9041137	0.8439778	0.8613445	SC
TRAC	0.9728271	0.9564053	0.9828094	VSC
CXCR5	0.9810425	0.9339171	0.9412344	VSC
IFNg	0.8904373	0.8534687	0.8465664	SC
CCR4	0.9921214	0.9822178	0.9819727	VSC

***** There was no correlation for *IDH2 R172m* in the absence of *IDH2* mutation detected in these 40 FFPE samples

Figures

Figure S1. Study design. 270 cases were divided into two cohorts. The classification cohort (n=230 including 30 AITL, 33 PTCL-TFH, 55 ALCL, 13 ATLL, 6 HSTL, 16 NKTCL and 77 PTCL-NOS and consisting mostly of fresh-frozen (FF) samples) was used to train a SVM classifier. The diagnostic cohort (n=40 including 13 AITL, 10 ALCL, 6 ATLL, 9 NKTCL and 2 PTCL-NOS) was used for the independent validation on FFPE samples and for the inter-laboratory reproducibility study.

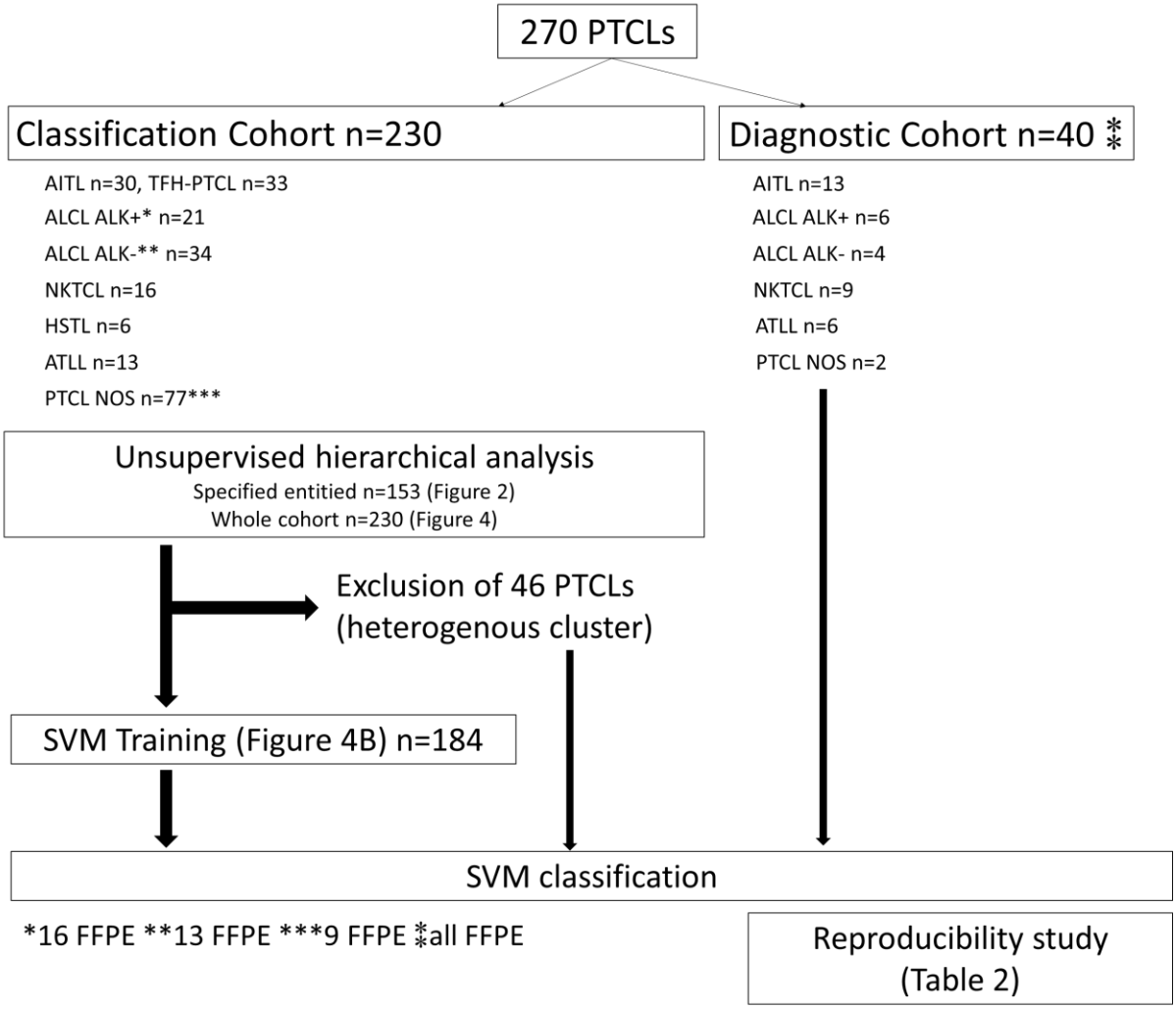

Figure S2. Representation of the RT-MLPA profiles of each PTCL molecular category **A)** AITL/T_{FH} profile, showing the expression of *CXCL13*, *CXCR5*, *ICOS*, and *BCL6*, and in this case the presence of *RHOA* and *IDH2* mutations. **B)** NKTCL signature, characterized by high EBER expression, and that of *CD56* and cytotoxic markers. **C and D)** Cytotoxic ALCL profile, defined by the expression of *CD30* and cytotoxic markers with (C) or without (D) *ALK*. **E)** Non-cytotoxic ALK-negative ALCL signature, characterized by the expression of *CD30* and TH2 markers (*GATA3* and *CCR4*). **F)** Cytotoxic/Th1 signature, defined by the expression of cytotoxic markers with inconsistent expression of Th1 markers. **G)** ATLL/TH2 signature, characterized by the expression of *GATA3*, *CCR4*, and *ICOS*, with inconsistent *FOXP3* expression. **H)** HSTL profile, showing the expression of *CD56*, *TBX21*, *GATA3*, and *BCL6*.

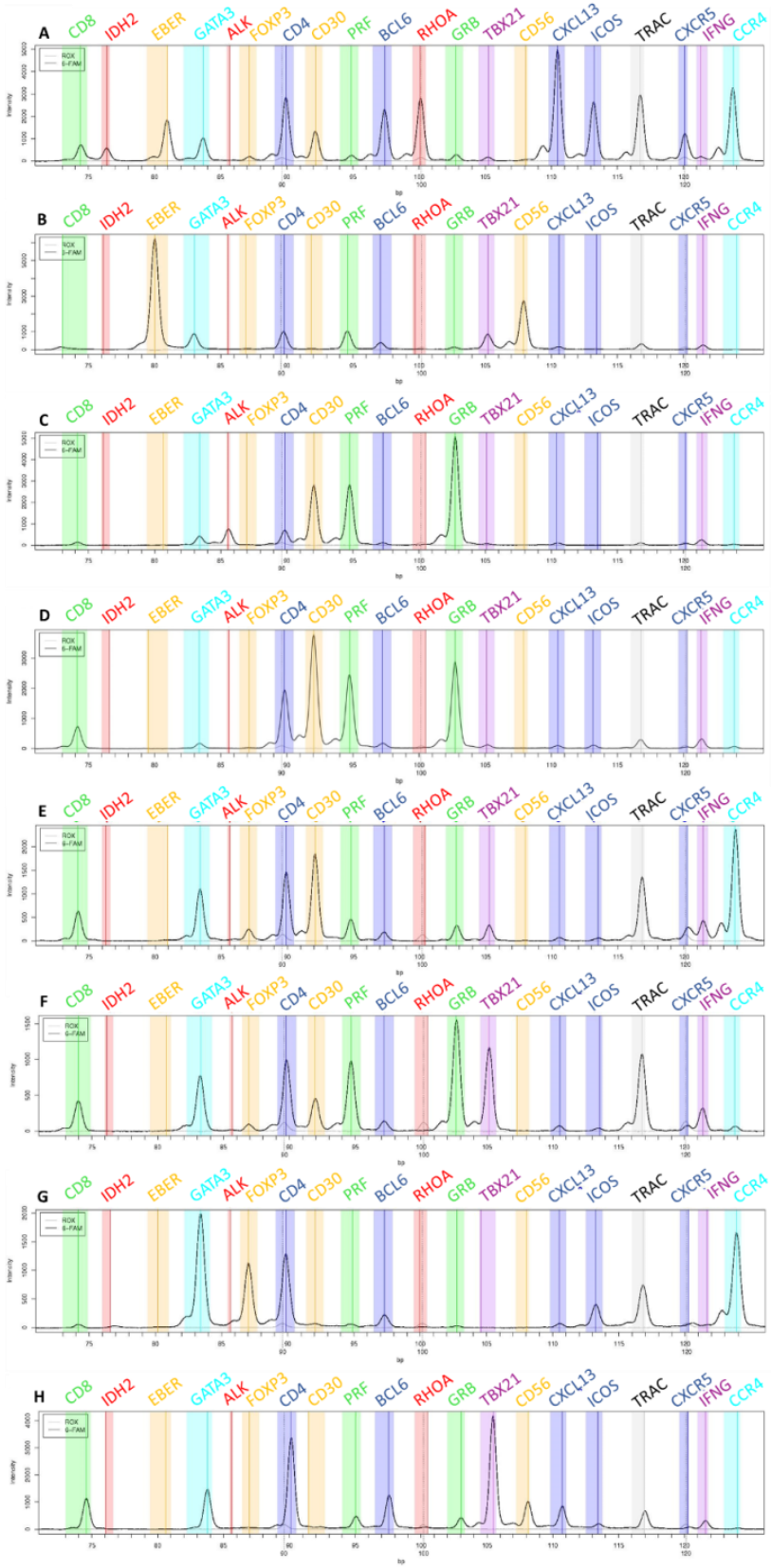
A
B
C
D
E
F
G
H
CD8
IDH2
EBER
GATA3
ALK
FOXP3
CD4
CD30
PRF
BCL6
RHOA
GRB
TBX21
CD56
CXCL13
ICOS
TRAC
CXCR5
IFNG
CCR4
ROX
6-FAM
Intensity
bp

Figure S3. ALCL case reclassified from ALK-negative to ALK-positive based on RT-MLPA assay (UPN051) a) Histopathology of the misdiagnosed ALK- ALCL case, based on the negative immunohistochemistry with ALK1 (performed twice). Strong cytoplasmic staining with the D5F3 clone was obtained, retrospectively. **b)** RT-MLPA profile of the case showing ALK expression (red arrow). **c)** Sequencing of the specific RT-PCR products confirmed an ATIC-ALK fusion transcript. No mutation in the region coding for the epitope of ALK1 was found (data not shown).

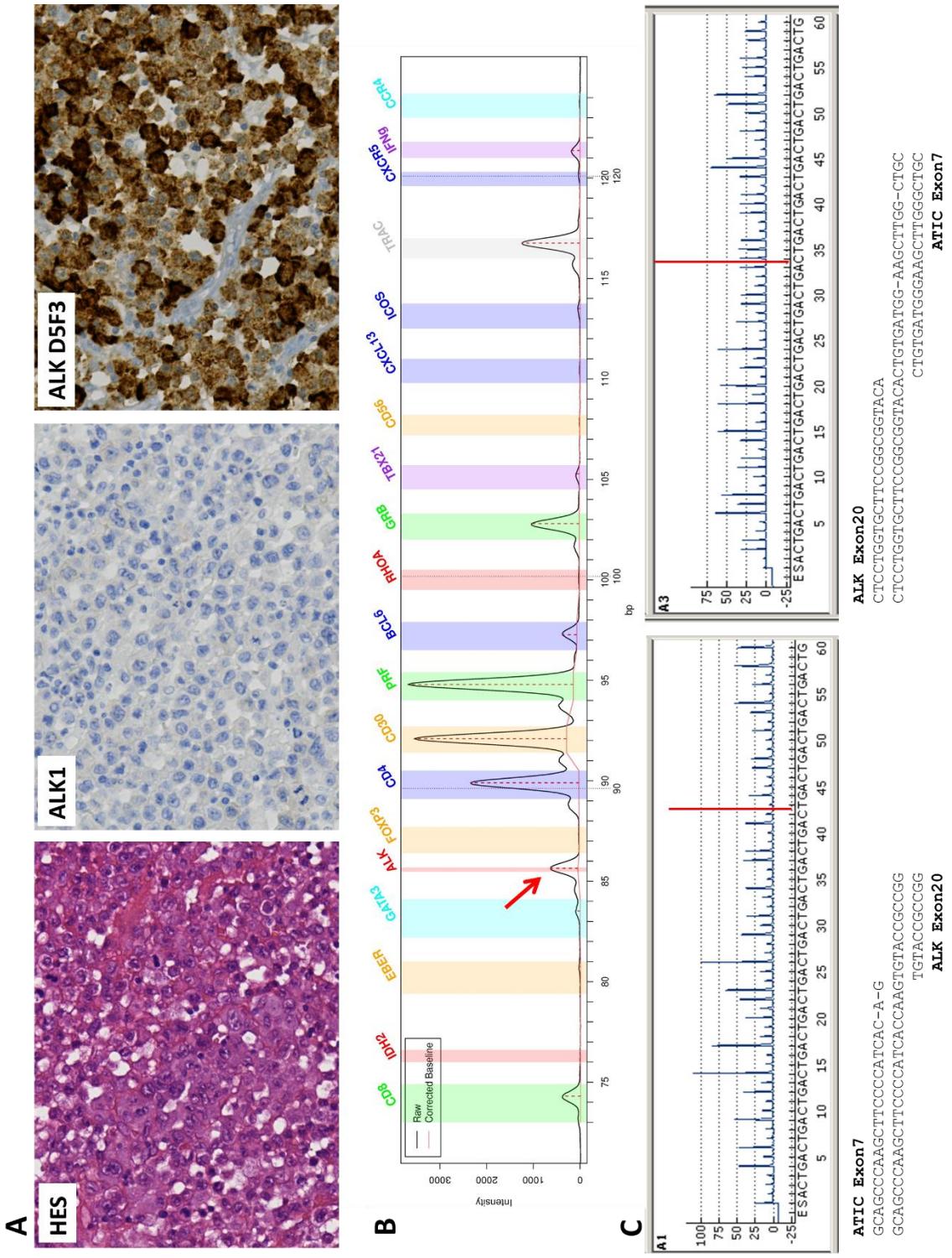

Figure S4. Scatterplot representation of the correlation between RT-MLPA and Affymetrix gene expression values (n = 71 cases, 23 AITL, and 49 PTCL NOS). There were significant correlations (Spearman test, $p < 0.05$) for each gene, with rho > 0.9 for *CXCL13* and *CCR4,* rho > 0.7 for *PRF, GZMB, GATA3, ICOS, CXCR5, BCL6, TNFRSF8/CD30, CD8,* and *TBX21,* and rho > 0.5 for *CD4* and *FOXP3* (x = Affymetrix values log, y = RT-MLPA values). The correlation for *ALK* and *CD56* was not determined because these genes were not evaluated in the Affymetrix series in the absence of ALCL, HSTL, and NKTCL. *RHOA*m, *IDH2*m, TCR, and EBER expression were not studied by the Affymetrix chip.

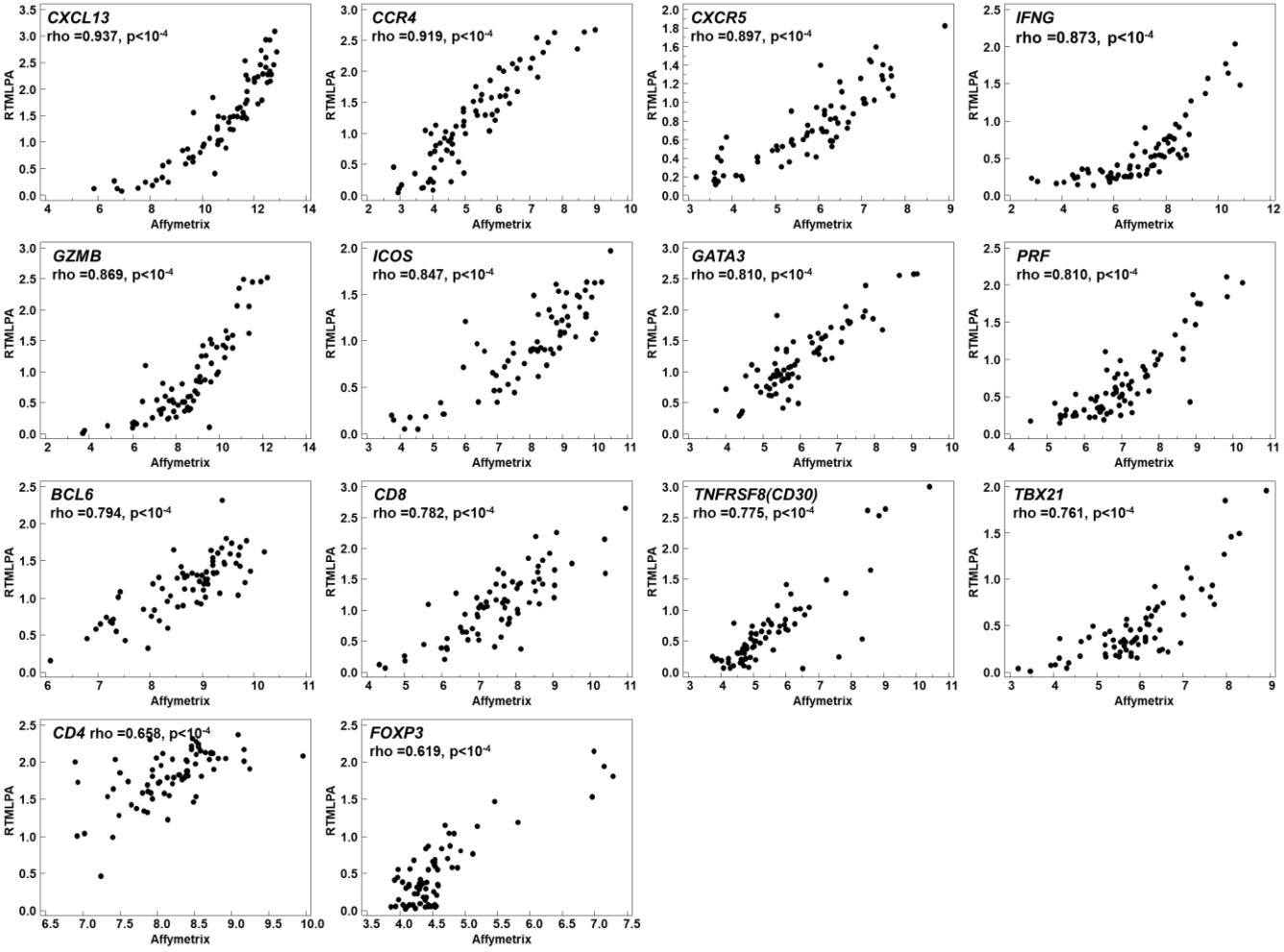

Figure S5. Comparison of individual RT-MLPA gene expression values and immunohistochemical results in 224 PTCLs, including 20 ALK+ ALCLs, 34 ALK- ALCLs, 29 AITLs, 36 PTCL-TFH, 15 NKTCLs, 13 ATLLs, 6 HSTLs, and 70 PTCL-NOS (6 cases with no IHC data were not considered). There was a significant correlation by Wilcoxon's rank- sum test between the RT-MLPA expression level of each gene and negative (-) or positive (+) staining by immunohistochemistry (CD30, TBX21, PRF, GZMB, GATA3, ALK, CXCL13, CD56, ICOS, CD8, CD4, and BCL6) or *in situ* hybridization (EBER).

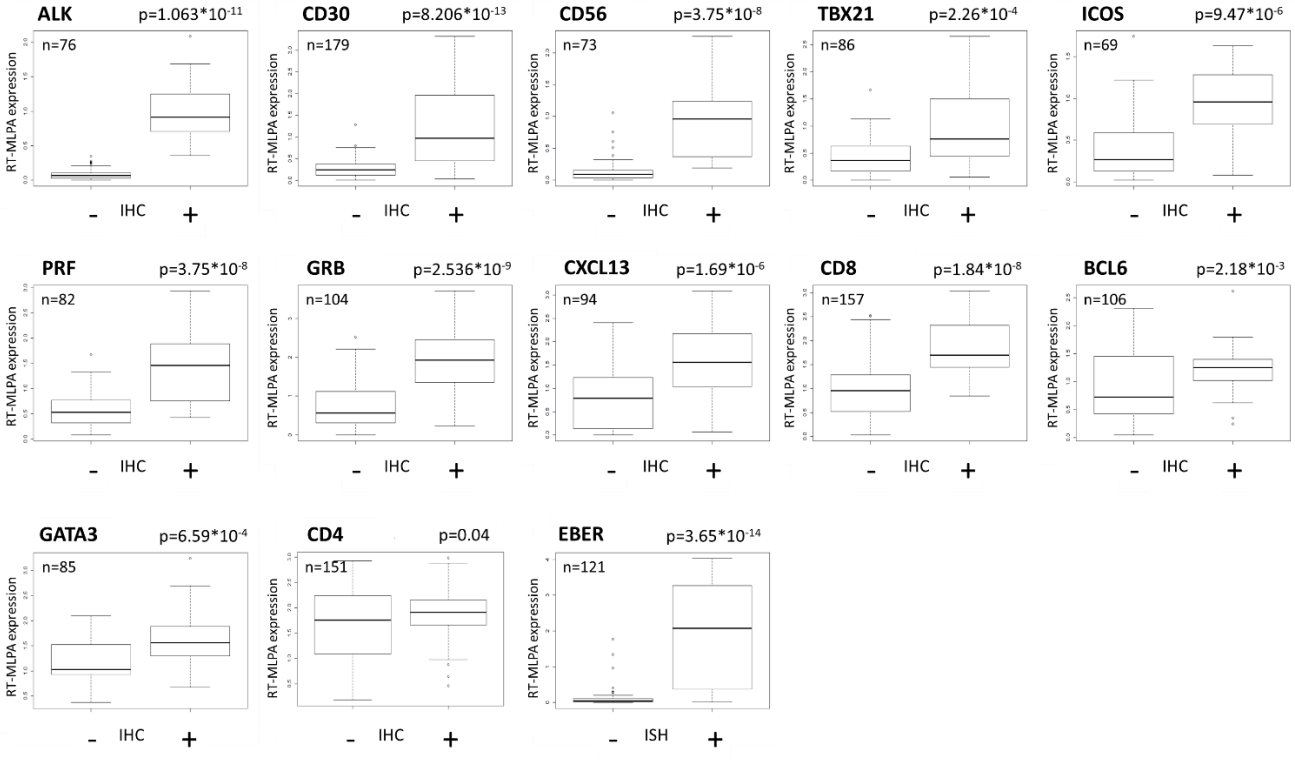

Figure S6. Examples of RT-MLPA profiles for paired FFPE and frozen samples. A) Superimposed profiles showed similar peaks for each 3 paired cases (NKTCL, AITL and non-cytotoxic ALK-negative ALCL). **B**) There was a strong correlation (rho>0.7) of RTMLPA normalized data between frozen (blue) and FFPE (orange) samples.

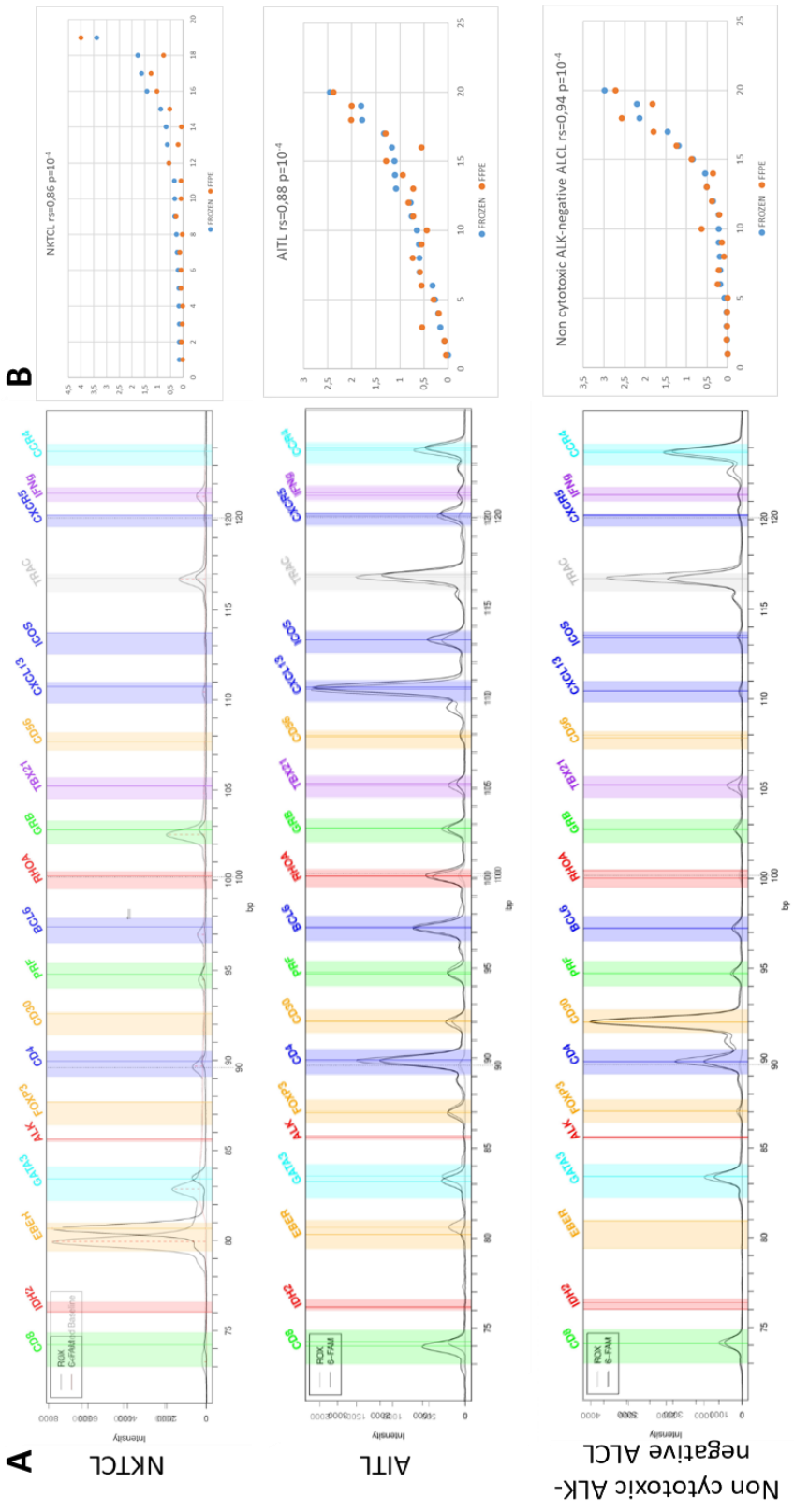

Figure S7. Molecular prediction of PTCL-NOS (n = 77) by the SVM model and correlation with immunohistochemical data. The SVM classification is presented in the top line, the pathological diagnosis in the second line, and immunohistochemical markers in the map). The SVM proposed a molecular class for 92% (69/75) of PTCL-NOS: 17 with a TFH signature, 28 with a cytotoxic/Th1 signature (5 ALK-negative ALCL, 19 cytotoxic/Th1, 5 NKTCL) and 24 with a Th2 signature. Among the 29 cases with a cytotoxic molecular signature, 23 were characterized as cytotoxic by immunochemistry and 6 were undetermined. Among the 24 molecular Th2 PTCL-NOS, 14/18 tested cases had a positive immunostaining for GATA3. Only four discrepancies (5%) were noted: 4 cases with a cytotoxic phenotype were classified in the T_{FH}/ AITL group.