



DNA methylation as a new tool for the differential diagnosis between T-LBL and lymphocyte-rich thymoma

Mehdi Latiri^{1,2†}, Mohamed Belhocine^{3,4†}, Charlotte Smith^{1,2}, Nathalie Gamier⁵, Estelle Balducci^{1,2}, Antoine Pinton^{1,2}, Guillaume P Andrieu^{1,2}, Julie Bruneau⁶, Salvatore Spicuglia³, Stéphane Jamain⁷, Violaine Latapie⁷ , Vincent Thomas de Montpreville⁸, Lara Chalabreysse⁹, Alexander Marx¹⁰, Nicolas Girard^{11,12}, Benjamin Besse^{13,14}, Christoph Plass¹⁵, Laure Gibault^{16,17}, Cécile Badoual^{16,17}, Elizabeth Macintyre^{1,2}, Vahid Asnafi^{1,2}, Thierry Jo Molina⁶ and Aurore Touzart^{1,2*} 

¹ Laboratory of Onco-Hematology, Necker Children's Hospital, Assistance Publique-Hôpitaux de Paris (AP-HP), Paris, France

² Université Paris Cité, CNRS, INSERM UI 151, Institut Necker Enfants Malades (INEM), Paris, France

³ Aix-Marseille University, INSERM, TAGC, UMR1090, Equipe Labélisée Ligue Contre le Cancer, Marseille, France

⁴ Department of Molecular Medicine, Al-Jawhara Centre for Molecular Medicine, Genetics, and Inherited Disorders, Arabian Gulf University, Manama, Bahrain

⁵ Institute of Pediatric Hematology and Oncology, Hospices Civils de Lyon, Claude Bernard Lyon 1 University, Lyon, France

⁶ Department of Pathology, Hôpital Necker-Enfants Malades, Université Paris-Cité, Paris, France

⁷ Univ Paris Est Créteil, INSERM, IMRB, Translational Neuropsychiatry, Créteil, France

⁸ Department of Pathology, Marie-Lannelongue Hospital, Le Plessis-Robinson, France

⁹ Department of Pathology, Groupe Hospitalier Est, Hospices Civils de Lyon, Bron, France

¹⁰ Institute of Pathology, University Medical Center Göttingen, Göttingen, Germany

¹¹ Thorax Institute Curie Montsouris, Paris, France

¹² Université de Versailles Saint Quentin (UVSQ), Paris-Saclay University, Versailles, France

¹³ Department of Cancer Medicine, Gustave Roussy Cancer Campus, Villejuif, France

¹⁴ Paris-Saclay University, Orsay, France

¹⁵ German Cancer Research Center, Division of Cancer Epigenomics, Heidelberg, Germany

¹⁶ Department of Pathology, HEGP Hospital, Paris, France

¹⁷ Université Paris Cité, Paris, France

*Correspondence to: A Touzart, Laboratory of Onco-Hematology, Hôpital Necker-Enfants malades, 149 Rue de Sèvres, 75015 Paris, France.

E-mail: aurore.touzart@aphp.fr

†These authors contributed equally to this work.

Abstract

T-lymphoblastic lymphoma (T-LBL) and thymoma are two rare primary tumors of the thymus deriving either from T-cell precursors or from thymic epithelial cells, respectively. Some thymoma subtypes (AB, B1, and B2) display numerous reactive terminal deoxynucleotidyl transferase-positive (TdT⁺) T-cell precursors masking epithelial tumor cells. Therefore, the differential diagnosis between T-LBL and TdT⁺ T-lymphocyte-rich thymoma could be challenging, especially in the case of needle biopsy. To distinguish between T-LBL and thymoma-associated lymphoid proliferations, we analyzed the global DNA methylation using two different technologies, namely MeDIP array and EPIC array, in independent samples series [17 T-LBLs compared with one TdT⁺ lymphocyte-rich thymoma (B1 subtype) and three normal thymi, and seven lymphocyte-rich thymomas compared with 24 T-LBLs, respectively]. In unsupervised principal component analysis (PCA), T-LBL and thymoma samples clustered separately. We identified differentially methylated regions (DMRs) using MeDIP-array and EPIC-array datasets and nine overlapping genes between the two datasets considering the top 100 DMRs including *ZIC1*, *TSHZ2*, *CDC42BPB*, *RBM24*, *C10orf53*, and *MACROD2*. In order to explore the DNA methylation profiles in larger series, we defined a classifier based on these six differentially methylated gene promoters, developed an MS-MLPA assay, and demonstrated a significant differential methylation between thymomas (hypomethylated; $n = 48$) and T-LBLs (hypermethylated; $n = 54$) (methylation ratio median 0.03 versus 0.66, respectively; $p < 0.0001$), with *MACROD2* methylation status the most discriminating. Using a machine learning strategy, we built a prediction model trained with the EPIC-array dataset and defined a cumulative score taking into account the weight of each feature. A score above or equal to 0.4 was predictive of T-LBL and conversely. Applied to the MS-MLPA dataset, this prediction model accurately predicted diagnoses of T-LBL and thymoma.

© 2024 The Author(s). *The Journal of Pathology* published by John Wiley & Sons Ltd on behalf of The Pathological Society of Great Britain and Ireland.

Keywords: thymoma; T-LBL; diagnosis; biomarker; DNA methylation; epigenetic; prediction model; MS-MLPA

Received 20 October 2023; Revised 31 July 2024; Accepted 6 August 2024

No conflicts of interest were declared.

© 2024 The Author(s). *The Journal of Pathology* published by John Wiley & Sons Ltd on behalf of The Pathological Society of Great Britain and Ireland. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Introduction

Thymic neoplasms are the most common primary tumors of the anterior mediastinum. They mainly include thymomas, thymic carcinomas, lymphomas including T-lymphoblastic lymphomas (T-LBLs), and germ cell tumors. Thymoma and T-LBL, two rare primary tumors, derive either from thymic epithelial cells or from T-cell precursors, respectively. Thymomas are subclassified into the histological subtypes A, AB, B1, B2, and B3 according to the WHO classification [1]. AB, B1, and B2 subtypes display numerous non-tumoral TdT⁺ T-lymphocytes that could mimic T-LBL morphologically and phenotypically, and several articles have reported the difficulty in distinguishing T-LBL from TdT⁺ T-cell-rich thymoma [2,3]. Thus, correct diagnosis can be challenging, especially when assessing a needle biopsy. Usually, thymomas affect adult patients without apparent gender predilection, around the fifth and sixth decades, with a median age of 58 years [4,5] and increasing incidence with age, but younger adults and, in rare cases, children can be affected [1,6]. Thymomas are predominantly characterized by progression in the thymus. However, in a minority of cases, they may invade the pleura, pericardium, lung, and mediastinal lymph nodes, and although rare, distant extrathoracic metastases (lymph nodes, bone marrow, liver, etc.) can be observed [6]. Thymomas are associated with paraneoplastic syndromes such as myasthenia gravis, among others, in about 30% of cases [1,6]. The standard treatment of thymoma is based, depending on stage, on complete surgical resection in combination with post-operative radiotherapy and/or chemotherapy in the case of advanced thymomas [6]. In contrast, T-LBL occurs preferentially in young men, with a highly aggressive clinical presentation combining a severe impairment of well-being, mediastinal mass, pleural/pericardial effusion, and frequent, but not systematic, dissemination (lymphadenopathy, bone marrow infiltration). In T-LBL, neoplastic T-cell progenitors are clonal, as demonstrated by clonal T-cell receptor (TR) rearrangements. However, in at least 10% of T-LBL cases, no clonal rearrangement can be identified [7,8]. Intensive acute lymphoblastic leukemia (ALL)-type poly-chemotherapy is recommended for the treatment of T-LBL. As the prognosis and the treatment differ dramatically between T-LBL and thymoma, making the right diagnosis is essential to avoid wrongly exposing a patient to intensive chemotherapy in the case of an erroneous T-LBL diagnosis. This may eventually require multimodal diagnostic approaches using histological, immunohistochemical (pan-cytokeratin, p63/p40/pax8 expression, and/or other keratin meshworks in thymoma, for instance), and molecular testing. In this study, we propose a new diagnostic tool based on DNA methylation analysis as an additional reliable biomarker for the differential diagnosis between T-LBL and T-cell-rich thymoma.

Materials and methods

Samples

The study was performed in accordance with the Declaration of Helsinki and local laws and was approved by the steering and ethics committee of the RYTHMIC network.

We collected tumoral tissues from 54 T-LBLs and 48 TdT⁺ T-cell-rich thymomas and isolated genomic DNA according to standard procedures at diagnosis, either from fresh/cryopreserved tissues or formalin-fixed, paraffin-embedded (FFPE) samples. For FFPE samples, DNA was extracted from five to ten FFPE tissue slices using a Maxwell[®] RSC DNA FFPE Kit according to the manufacturer's instructions (Promega, Madison, WI, USA).

Pathology review

All thymic epithelial tumors were collected and reviewed by pathologists from the French RYTHMIC network, dedicated to the management of patients with thymic epithelial tumors [9].

Global DNA methylation analysis

DNA methylation using Infinium MethylationEPIC BeadChip (Illumina, San Diego, CA, USA) for a series of seven thymomas and 24 T-LBLs was performed following the manufacturer's protocol.

Identification of differentially methylated regions (DMRs)

We identified DMRs from MeDIP-array and EPIC-array methylation data (supplementary material, Tables S1 and S2). To identify DMRs between distinct conditions, the *lmFit* function from the 'limma' R package was employed to fit the methylation data to a linear model. Subsequently, the moderated *t*-statistics method, incorporating an empirical Bayes approach, was applied to assess differential methylation. The adjusted *p* values were calculated using the empirical Bayes method, and only loci with adjusted *p* values less than 0.05 were retained as significant. Principal component analysis (PCA) was performed using the *prcomp* function in R (R Foundation for Statistical Computing, Vienna, Austria) with data scaling. The resulting PCA scores were extracted, providing a reduced dimensional representation of the methylation profiles. Scatter plots were generated with PCA scores on the first two principal components.

Machine learning model building

From the EPIC-array dataset, methylation data for *ZIC1*, *TSHZ2*, *CDC42BPB*, *C10orf53*, *MACROD2*, and *RBM24* were selected as features for the model building. The PyCaret library in Python (Python Software Foundation, Wilmington, DE, USA) was used for the predictive model development to classify samples as normal or

leukemic. Various machine learning algorithms were trained and compared using the training set (not shown), and the best-performing algorithm, Extra Trees Regressor ($n_jobs = -1$, $random_state = 889$), was selected for further optimization. The Extra Trees Regressor ($n_jobs = -1$, $random_state = 889$) model was fine-tuned using hyperparameter optimization to enhance its predictive performance. The model was optimized based on mean absolute error (MAE). The performance of the tuned Extra Trees Regressor model was evaluated on the testing set. Additionally, interpretation of the model was performed using SHAP (SHapley Additive exPlanations) values to understand the impact of features on model predictions. The effect of the number of features selected on the model performance was evaluated using an optimal prediction with six features (supplementary material, Figure S1). Features' importance (weight) in the prediction was also evaluated (supplementary material, Figure S2).

Validation of the DNA methylation signature

Direct targeted DNA methylation levels were analyzed by methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) using custom probes located in close proximity to DMRs identified in the EPIC-array dataset (<500 bp) (supplementary material, Table S3) and SALSA MLPA P200 Reference-1 Probemix and EK1 Reagent Kits from MRC Holland (Amsterdam, The Netherlands), following the manufacturer's recommendations. Data were analyzed using Coffalyser software (MRC Holland).

TRG and TRD clonality analysis

TRD and TRG clonality was assessed either by one-step next-generation sequencing (NGS) as previously described [10] or/and by multiplex PCR according to EuroClonality protocols [11,12]. NGS clonality data were analyzed and visualized using the Vidjil Platform [13].

Results

Previously, we studied the gene promoters' global DNA methylation in a series of 17 T-LBLs, one lymphocyte-rich thymoma (B1 subtype), and three normal thymi using MeDIP arrays (see supplementary material, Table S4 for clinical characteristics). Unsupervised hierarchical clustering identified two main clusters associated with distinct methylation profiles. All the T-LBL samples ($n = 17$) clustered in one group, whereas the thymoma cases and the three normal thymic tissues clustered in a second group [14]. Based on this observation, we decided to further explore the differential methylation profiles between T-LBL and thymoma in order to epigenetically characterize thymic T-lymphoid populations in the T-LBLs, T-lymphocyte-rich thymoma,

and normal thymic tissues, and to distinguish between T-LBL and non-T-LBL. For this purpose, we performed 'state-of-the-art' global methylation profiling using EPIC arrays in an independent series of 24 T-LBLs and seven thymomas ($n = 2$ AB, $n = 2$ B1, and $n = 3$ B2 subtypes). Clinical characteristics are presented in supplementary material, Table S5. Unsupervised PCA analysis using MeDIP-array or EPIC-array datasets showed clustering of T-LBL samples in a subgroup, and with thymoma and normal thymic samples in a distinct subgroup (Figure 1B). We then identified differentially methylated regions (DMRs) ($p_{adj} < 0.05$) within the two datasets between T-LBL and thymomas and thymus; 54,837 and 13,683 DMRs including 5,214 and 2,705 unique genes were found in the MeDIP-array and EPIC-array datasets, respectively. A significant overlap of 862 differentially methylated genes between the two datasets was observed (Figure 1C,D,E and supplementary material, Tables S1 and S2). Considering the top 100 of the most differentially methylated regions, we identified nine overlapping genes, namely *MACROD2* (mono-ADP ribosylhydrolase 2), *CDC42BPB* (CDC42 binding protein kinase beta), *KLHL34* (kelch-like family member 34), *TSHZ2* (teashirt zinc finger homeobox 2), *RBM24* (RNA binding motif protein 24), *ZIC1* (Zic family member 1), *C10orf53* (chromosome 10 open reading frame 53), *ZIC3* (Zic family member 3), and *CPEB1* (cytoplasmic polyadenylation element binding protein 1) (Figure 1F).

In order to explore the DNA methylation profiles in larger series of T-LBL and thymoma, we then defined a classifier including six among these most differentially methylated gene promoters: *ZIC1*, *TSHZ2*, *CDC42BPB*, *C10orf53*, *RBM24*, and *MACROD2* (Figure 2A). An MS-MLPA assay was designed to explore the methylation status using this six-gene classifier. *ZIC3* and *KLHL34*, located on chromosome X, were excluded, as well as *CPEB1* for technical considerations. Representative MS-MLPA profiles for thymus, thymoma, and T-LBL are shown in Figure 2B. We first tested the MS-MLPA panel in a series of three normal thymi, one TdT⁺ lymphocyte-rich thymoma, and seven T-LBLs from the MeDIP-array series and the seven thymomas and 24 T-LBLs from the EPIC-array series (Figure 2C,D), and this confirmed the hypermethylated profile of T-LBL cases as compared with thymoma cases. We then assessed the MS-MLPA panel in an additional series of 41 lymphocyte-rich thymomas ($n = 21$ AB, 9 B1, and 11 B2 subtypes) and 30 T-LBLs. The epidemiological and biological characteristics of the entire series of thymoma and T-LBL cases analyzed by MS-MLPA are presented in Table 1. As expected, thymoma patients were significantly older than T-LBL patients (median age 61.2 versus 30.8 years, respectively), and T-LBL affected male patients more frequently compared with the patients with thymoma (sex ratio 3.5 versus 0.92, respectively). We also studied the clonal rearrangements of the TRG and TRD loci. As expected, no thymoma case demonstrated clonal TRG, whereas 9/54 T-LBLs

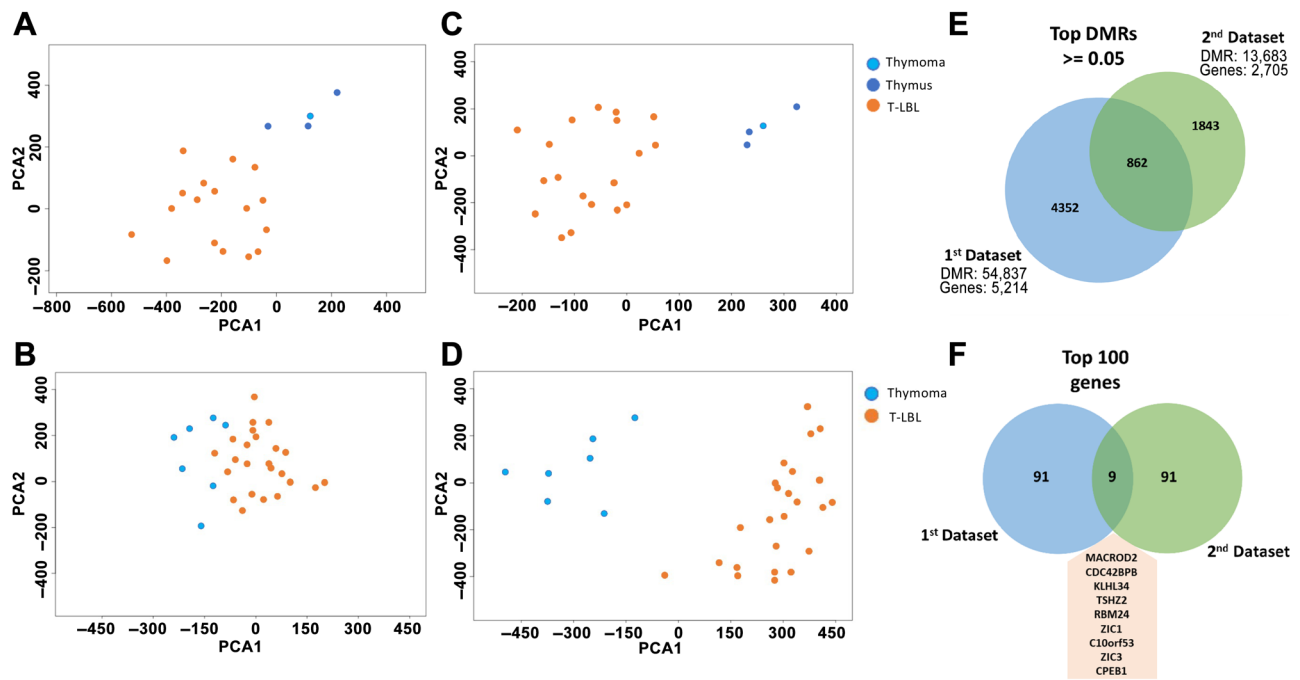


Figure 1. Global methylation analysis in T-LBL, thymoma, and thymus. (A) Unsupervised PCA and (C) supervised PCA using DMRs ($p_{adj} < 0.05$) of MedIP-array data for 17 T-LBLs, one lymphocyte-rich thymoma, and three normal thymi. (B) Supervised PCA and (D) unsupervised PCA using DMRs ($p_{adj} < 0.05$) of EPIC-array data for 24 T-LBL and seven thymoma cases. (E) Venn diagram illustrating overlap between DMRs identified in the MedIP-array dataset (1st dataset) and EPIC-array dataset (2nd dataset). (F) DMR overlap within the top 100 most differentially methylated genes in the two datasets.

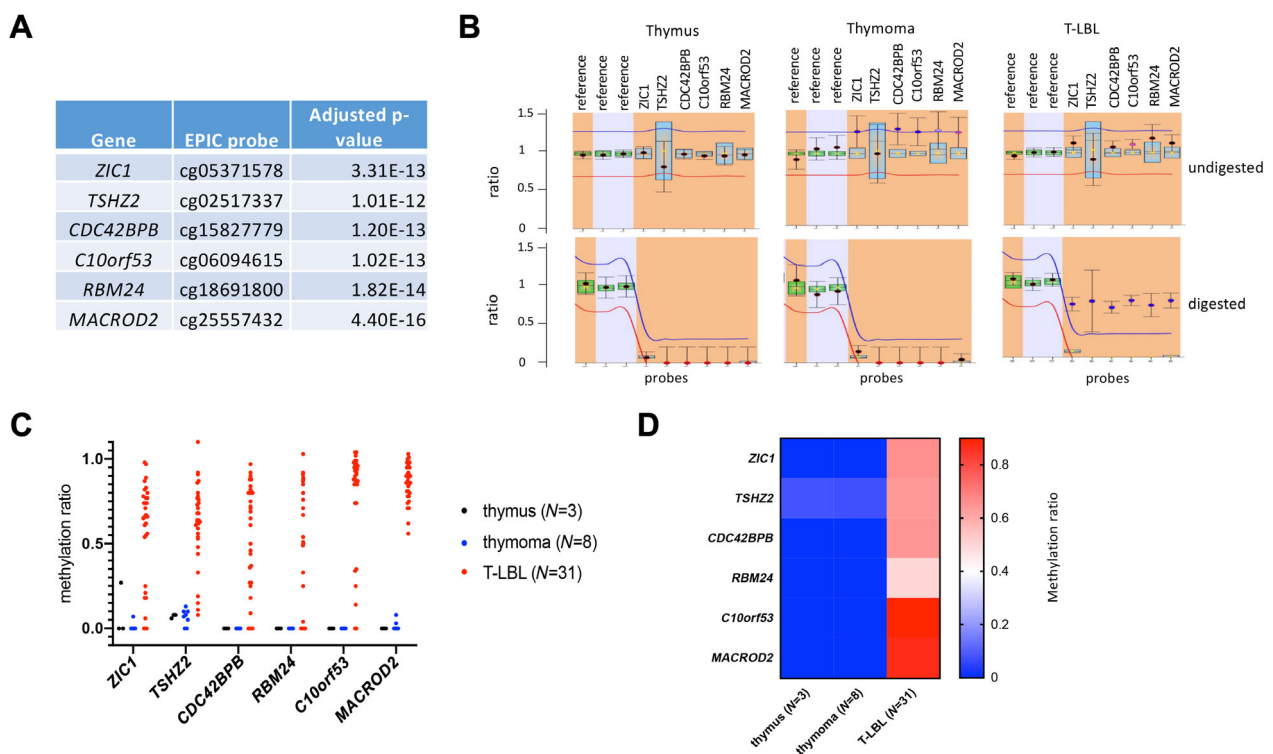


Figure 2. MS-MLPA targeted promoter methylation analysis development to distinguish between T-LBL and T-lymphocyte-rich thymoma. (A) List of the six-gene promoter classifier allowing methylation status prediction with the corresponding most differentially methylated probe in the EPIC-array dataset and significance (adjusted p value). (B) Representative ratio charts of methylation-specific multiplex ligation-dependent probe amplification (MS-MLPA) analysis for one normal thymus, one T-LBL, and one thymoma. The top panels refer to the MLPA (undigested) reference panel and the bottom panel is the MS-MLPA (digested with HhaI restriction enzyme) panel. (C) The methylation ratio of each gene promoter was assessed by MS-MLPA for three thymus, eight thymoma, and 31 T-LBL samples from the MedIP-array and EPIC-array series. (D) Heat map illustrating the methylation ratios by MS-MLPA for the thymus, the thymoma, and the T-LBL subgroups (median).

Table 1. Epidemiological and biological features of the global series of studied T-LBLs and thymomas (EPIC array and/or MS-MLPA).

	T-LBL	Thymoma	p value*
n=	54	48	
Age median (years), (min-max)	30.8 (0.1-71.1)	61.2 (26.5-87.1)	<0.0001
Sex ratio (M/F)	3.5 (42/12)	0.92 (23/25)	0.0021
Clonal TRG	45/54 (83.3%)	0/46 (0%)	<0.0001
Clonal TRD	26/54 (48.1%)	0/46 (0%)	<0.0001
Germline TRG and TRD	8/54 (14.8%)	46/46 (100%)	<0.0001

* χ^2 or Mann-Whitney tests were used where appropriate.

(16.7%) did not display clonal TRG clonal rearrangement. Of note, a vast majority (8/9) of polyclonal TRG T-LBLs also demonstrated non-clonal TRD rearrangement. Overall, 8/54 T-LBLs (14.8%) had germline TRG and TRD rearrangements. Representative clonality profiles for polyclonal, clonal T-LBLs, and polyclonal thymomas are shown in Figure 3. These results highlight that even if TRG clonality could be very helpful in the diagnostic process, it does not constitute a perfect diagnostic biomarker. Regarding DNA methylation status and considering the entire series, methylation ratios were significantly different between thymoma and T-LBL for each gene promoter of the classifier despite some overlap for *ZIC1*, *TSHZ2*, *CDC42BPB*, *C10orf53*, and *RBM24*. Interestingly, no overlap was observed for *MACROD2* (Figure 4A,B). Therefore, *MACROD2* methylation status was the most reliable biomarker for distinguishing between thymoma and T-LBL. Considering the average methylation ratio of the six gene promoters, we confirmed a global significant differential

methylation between thymomas (hypomethylated) and T-LBLs (hypermethylated) (methylation ratio median 0.03 versus 0.66, respectively; $p < 0.0001$) (Figure 4C). We verified the reliability of the assay on DNA isolated from FFPE tissues, as DNA could be of lower quality and hinder molecular analysis. For two thymoma (sample pairs #1 and #2) and two T-LBL (sample pairs #3 and #4) cases with both frozen and FFPE tissues available, similar methylation profiles were obtained in both DNA samples (Figure 4D). In addition, we recently reported methylome data in a large series of adult T-ALLs ($n = 143$) using Illumina Infinium methylation EPIC arrays [15]. Similar results were observed in T-ALL samples displaying significantly hypermethylated promoters for the six-gene classifier compared with normal thymi ($n = 12$), confirming the biological relevance of the aberrant methylation observed in T-LBL (supplementary material, Figure S3).

Finally, in order to build a robust prediction model to distinguish between T-LBL and lymphocyte-rich thymoma, we adopted a machine learning strategy (Figure 5A). The most differentially methylated probe in the EPIC-array dataset for each gene of the six-gene classifier was selected for the model building (training set). Various machine learning algorithms were trained and compared using the training set, and the best-performing algorithm, Extra Trees Regressor, was selected for further optimization (see the Materials and methods section). Each feature weight in the prediction was evaluated and a cumulative score taking into account the relative weight of each feature was calculated; a prediction score below 0.4 was

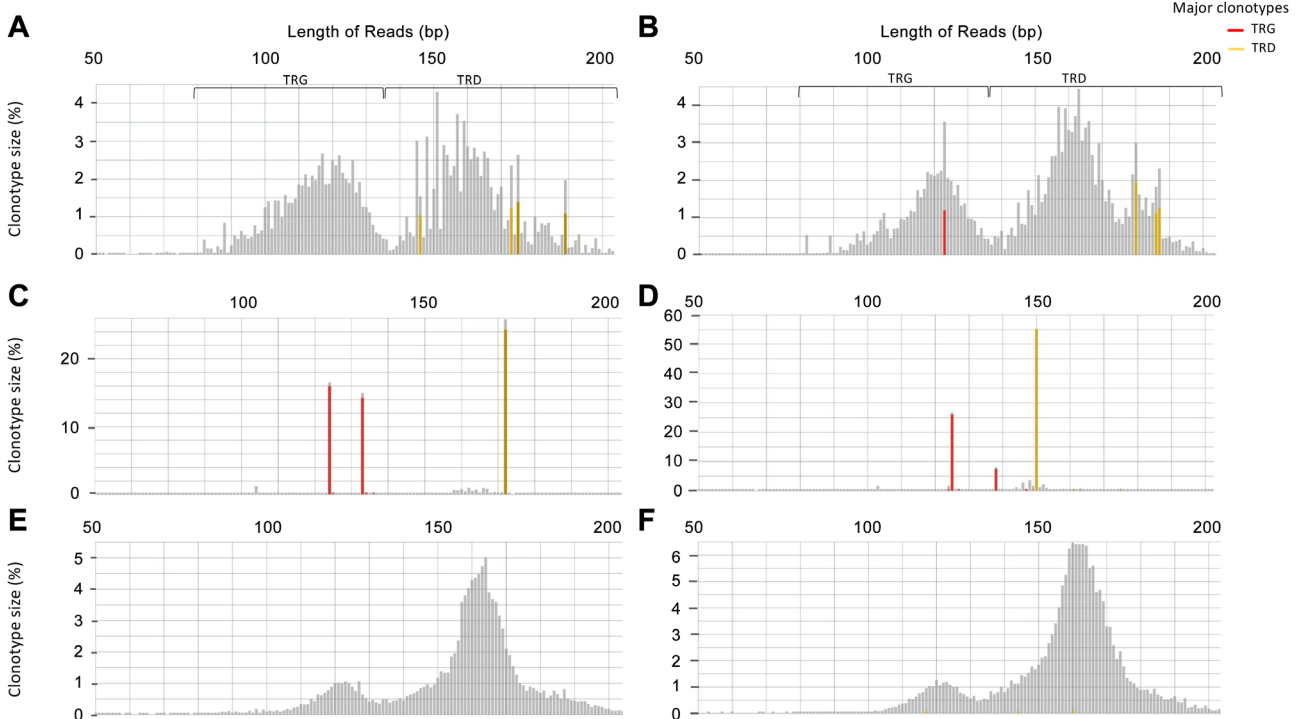


Figure 3. TRG and TRD clonality analysis by NGS. (A-F) Representative TRG and TRD clonotype distributions for two polyclonal T-LBL (A and B), two clonal T-LBL (C and D), and two thymoma samples (E and F). Red and yellow bars represent major TRG and TRD clonotypes, respectively.

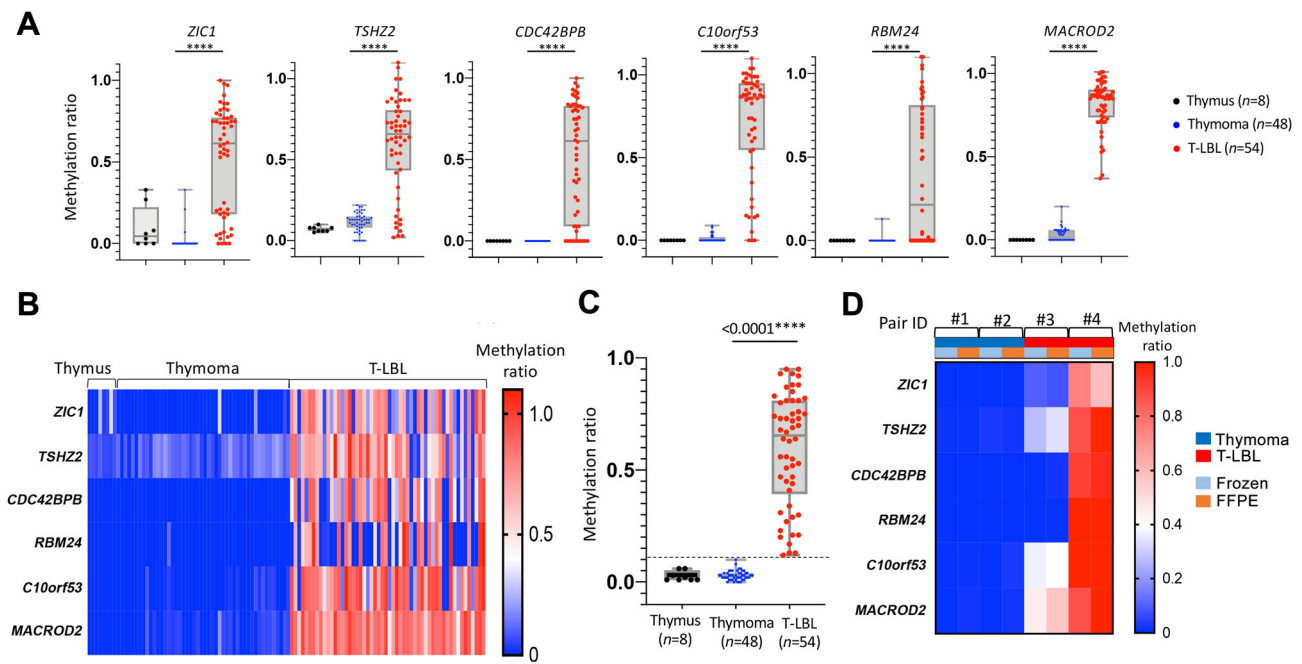


Figure 4. Methylation status by MS-MLPA in lymphocyte-rich thymoma and T-LBL validation series. (A) Methylation ratio for each gene promoter assessed by MS-MLPA in T-LBL ($n = 54$), thymoma ($n = 48$), and thymus samples ($n = 8$). (B) Heat map illustrating the methylation ratios by MS-MLPA for the thymus, thymoma, and T-LBL samples. (C) Average methylation ratio (average methylation of the six-gene panel) for the T-LBL, thymoma, and thymus groups. The dotted line depicts the threshold of 0.1. (D) Heat map illustrating the methylation profiles by MS-MLPA for frozen and FFPE samples for two T-LBL and two thymoma sample pairs. **** $p < 0.0001$ (Mann-Whitney non-parametric test).

considered predictive of a thymoma diagnosis and a score equal to or above 0.4 predictive of a T-LBL diagnosis (Figure 5B). After testing on the MeDIP-array dataset, we applied the model to the MS-MLPA cohort, which clearly and robustly separated thymomas and T-LBLs (Figure 5C).

Discussion

In this study, we confirmed a certain overlap regarding epidemiological and biological features (age, sex, TRG status) between thymoma and T-LBL, underpinning the need for additional biomarkers. Indeed, in this series,

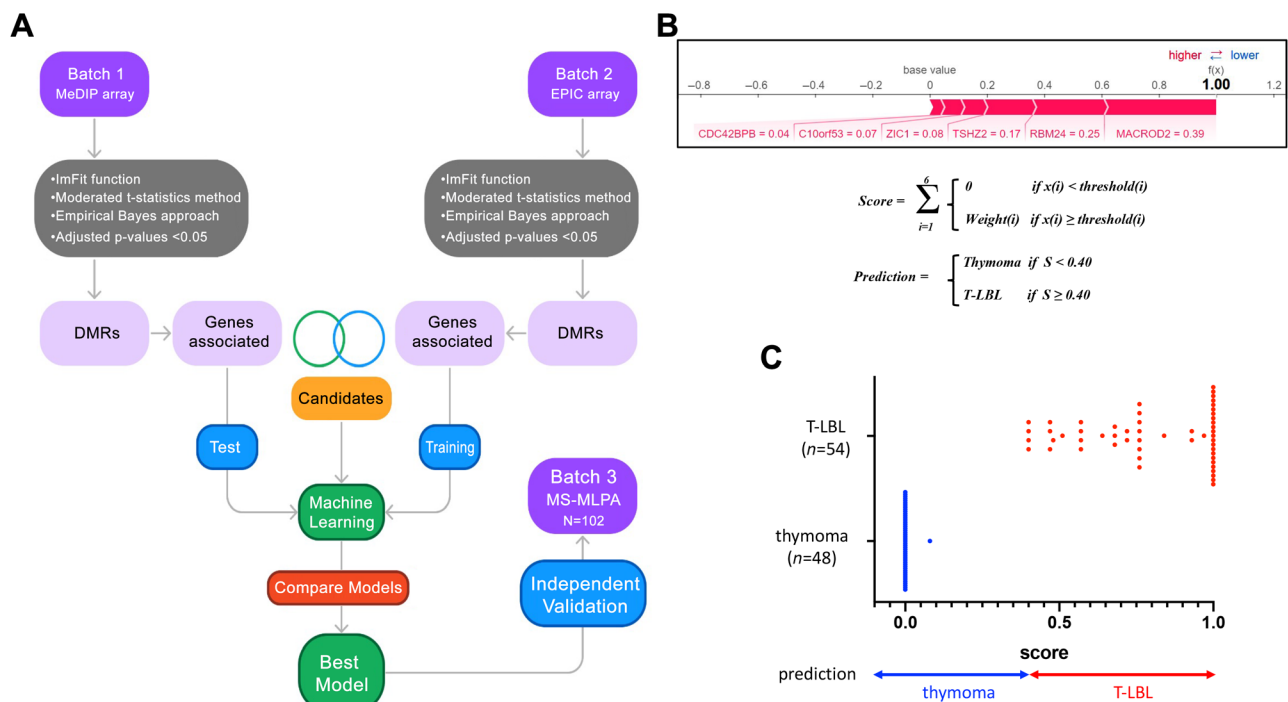


Figure 5. Machine learning prediction model. (A) Flowchart of the machine learning prediction model design. (B) Cumulative score taking into account the relative weight of each feature (red) and mathematical formula to calculate the prediction score. (C) Prediction scores for T-LBL ($n = 54$) and thymoma cases ($n = 48$).

about 14% of T-LBL cases displayed neither TRG nor TRD clonal rearrangement, suggesting a very immature maturation arrest stage. Furthermore, it has been suggested that LMO2 expression could be a specific marker to distinguish between T-LBL/T-ALL and thymomas [16]. Whereas LMO2 expression is frequent in T-ALL and is involved in leukemogenesis, it is not observed in 100% of T-ALLs/T-LBLs but rather in 50–80% of cases [16–18] and never in thymomas. Therefore, even if LMO2 expression could be an additional argument in the differential diagnosis process, it would not be a perfect biomarker because of low sensitivity. In this context, we propose a new prediction model based on DNA methylation analysis as an additional reliable biomarker and diagnostic tool that is both highly specific and highly sensitive for the differential diagnosis between T-LBL and T-cell-rich thymoma. While gene expression is labile and may reflect an alterable cellular state, DNA methylation is robustly stable in patient samples [19]. Moreover, it critically retains cell-of-origin and substantial leukemogenic signatures, reinforcing the evaluation of the methylome as a powerful diagnostic and classifier tool [15]. MS-MLPA offers the advantages of being feasible and adapted for routine activity, since it requires only low amounts of DNA (50 ng), extracted from FFPE or fresh samples and without any specific DNA treatment. Lymphoid populations from T-LBL demonstrate a signature of aberrant DNA hypermethylation within gene promoters in line with the general concept of aberrant methylation as a hallmark of cancer. Conversely, lymphoid populations from TdT⁺ T-cell-rich thymoma show a methylation pattern similar to that of normal thymic tissues, confirming their non-malignant feature in this pathology.

Interestingly, these six most differentially methylated genes are cancer-related. For instance, *ZIC1* was described as a tumor suppressor gene, and its inhibition and epigenetic silencing were associated with chronic myelogenous leukemia [20], anal cancer severity [21], high-risk pediatric rhabdomyosarcoma [22], non-invasive endometrial cancer detection [23], invasive breast cancer [24], cervical cancer detection [25], gastric cancer metastasis [26], glioma cell growth [27], high-risk hepatocellular carcinoma [28], and distinct thyroid cancer subtypes [29]. *TSHZ2* inhibition and epigenetic silencing were associated with mammary tumors [30], a poorer prognosis in lung adenocarcinoma [31], and triple-negative breast cancer [32]. *RBM24* is a splicing regulator. This gene functions as a tumor suppressor gene in liver cancer cells [33] and nasopharyngeal carcinoma [34], and is found to be repressed in colorectal tumors [35]. Somatic variants of C10orf53 have been described in acute myeloid leukemia (AML) and meningioma [36,37]. *MACROD2* seems to be the most discriminating between thymoma and T-LBL. This gene is frequently affected by structural variations (SVs) and is less expressed in hepatocellular carcinoma, promoting cancer cell growth and metastasis [38]. In this latter paper, *MACROD2* protein expression was studied in tumoral tissues using immunohistochemistry (IHC).

In future investigations, it would be interesting to study whether *MACROD2* expression assessed using IHC could also differentiate between thymoma and T-LBL. *MACROD2* deletions were also involved in chromosome instability in colorectal cancer [39] and in gastric cancer [40]. Interestingly, inactivating mutations and structural variation of *MACROD2* were also reported in relapsing pediatric B-cell precursor (BCP)-ALL [41], suggesting a role in leukemogenesis. In conclusion, this ‘six-gene methylation biomarker’ could be a valuable adjunct to the diagnostic toolbox and help to better distinguish between T-LBL and T-cell-rich thymoma.

Acknowledgements

The Rythmic Network was supported by INCA (Institut National du Cancer) (subvention cancers rares).

Author contributions statement

AT wrote the manuscript. AT, ML and SS performed and interpreted molecular analyses. VL and SJ performed EPIC-array studies. SS and MB performed the bioinformatics analysis. LG, LC, VTdM and TJM reviewed thymic epithelial tumors. AT, VA and TJM designed the study. VA and TJM critically reviewed the manuscript. All the authors contributed to the article and approved the submitted version.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Marx A, Chan JKC, Chalabreyse L, *et al.* The 2021 WHO classification of tumors of the thymus and mediastinum: what is new in thymic epithelial, germ cell, and mesenchymal tumors? *J Thorac Oncol* 2022; **17**: 200–213.
- Adam P, Hakroush S, Hofmann I, *et al.* Thymoma with loss of keratin expression (and giant cells): a potential diagnostic pitfall. *Virchows Arch* 2014; **465**: 313–320.
- Boddu P, Thakral B, Alhurajji A, *et al.* Distinguishing thymoma from T-lymphoblastic leukaemia/lymphoma: a case-based evaluation. *J Clin Pathol* 2019; **72**: 251–257.
- de Jong WK, Blaauwgeers JLG, Schaapveld M, *et al.* Thymic epithelial tumours: a population-based study of the incidence, diagnostic procedures and therapy. *Eur J Cancer* 2008; **44**: 123–130.
- Hsu C-H, Chan JK, Yin C-H, *et al.* Trends in the incidence of thymoma, thymic carcinoma, and thymic neuroendocrine tumor in the United States. *PLoS One* 2019; **14**: e0227197.
- Girard N, Ruffini E, Marx A, *et al.* Thymic epithelial tumours: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2015; **26**: v40–v55.

7. Lepretre S, Touzart A, Vermeulin T, et al. Pediatric-like acute lymphoblastic leukemia therapy in adults with lymphoblastic lymphoma: the GRAALL-LYSA LL03 study. *J Clin Oncol* 2016; **34**: 572–580.
8. Callens C, Baleyrier F, Lengline E, et al. Clinical impact of *NOTCH1* and/or *FBXW7* mutations, *FLASH* deletion, and *TCR* status in pediatric T-cell lymphoblastic lymphoma. *J Clin Oncol* 2012; **30**: 1966–1973.
9. Molina TJ, Bluthgen MV, Chalabreysse L, et al. Impact of expert pathologic review of thymic epithelial tumours on diagnosis and management in a real-life setting: a RHYTHMIC study. *Eur J Cancer* 2021; **143**: 158–167.
10. Villares P, Abdo C, Bertrand M, et al. One-step next-generation sequencing of immunoglobulin and T-cell receptor gene recombinations for MRD marker identification in acute lymphoblastic leukemia. *Methods Mol Biol* 2022; **2453**: 43–59.
11. Armand M, Derriex C, Beldjord K, et al. A new and simple TRG multiplex PCR assay for assessment of T-cell clonality: a comparative study from the EuroClonality Consortium. *Hemasphere* 2019; **3**: e255.
12. van Dongen JJM, Langerak AW, Brüggemann M, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* 2003; **17**: 2257–2317.
13. Giraud M, Salson M, Duez M, et al. Fast multiclonal clusterization of V(D)J recombinations from high-throughput sequencing. *BMC Genomics* 2014; **15**: 409.
14. Touzart A, Boissel N, Belhocine M, et al. Low level CpG island promoter methylation predicts a poor outcome in adult T-cell acute lymphoblastic leukemia. *Haematologica* 2020; **105**: 1575–1581.
15. Touzart A, Mayakonda A, Smith C, et al. Epigenetic analysis of patients with T-ALL identifies poor outcomes and a hypomethylating agent-responsive subgroup. *Sci Transl Med* 2021; **13**: eabc4834.
16. Jevremovic D, Roden AC, Ketterling RP, et al. LMO2 is a specific marker of T-lymphoblastic leukemia/lymphoma. *Am J Clin Pathol* 2016; **145**: 180–190.
17. Latchmansingh K-A, Wang X, Verdun RE, et al. LMO2 expression is frequent in T-lymphoblastic leukemia and correlates with survival, regardless of T-cell stage. *Mod Pathol* 2022; **35**: 1220–1226.
18. Morishima T, Krahl A-C, Nasri M, et al. LMO2 activation by deacetylation is indispensable for hematopoiesis and T-ALL leukemogenesis. *Blood* 2019; **134**: 1159–1175.
19. Touzart A, Lengliné E, Latiri M, et al. Epigenetic silencing affects L-asparaginase sensitivity and predicts outcome in T-ALL. *Clin Cancer Res* 2019; **25**: 2483–2493.
20. Choudhury SR, Dutta S, Bhaduri U, et al. LncRNA Hmrhl regulates expression of cancer related genes in chronic myelogenous leukemia through chromatin association. *NAR Cancer* 2021; **3**: zcab042.
21. van der Zee RP, van Noesel CJM, Martin I, et al. DNA methylation markers have universal prognostic value for anal cancer risk in HIV-negative and HIV-positive individuals. *Mol Oncol* 2021; **15**: 3024–3036.
22. Lak NSM, Voormans TL, Zappeij-Kannegieter L, et al. Improving risk stratification for pediatric patients with rhabdomyosarcoma by molecular detection of disseminated disease. *Clin Cancer Res* 2021; **27**: 5576–5585.
23. van den Helder R, Wever BMM, van Trommel NE, et al. Non-invasive detection of endometrial cancer by DNA methylation analysis in urine. *Clin Epigenetics* 2020; **12**: 165.
24. Zhang Y-W, Ma J, Shi C-T, et al. Roles and correlation of FOXA1 and ZIC1 in breast cancer. *Curr Probl Cancer* 2020; **44**: 100559.
25. Dick S, Verhoef L, De Strooper LM, et al. Evaluation of six methylation markers derived from genome-wide screens for detection of cervical precancer and cancer. *Epigenomics* 2020; **12**: 1569–1578.
26. Ge Q, Hu Y, He J, et al. Zic1 suppresses gastric cancer metastasis by regulating Wnt/ β -catenin signaling and epithelial–mesenchymal transition. *FASEB J* 2020; **34**: 2161–2172.
27. Fu J-Q, Chen Z, Hu Y-J, et al. A single factor induces neuronal differentiation to suppress glioma cell growth. *CNS Neurosci Ther* 2019; **25**: 486–495.
28. Wang Y-Y, Jiang J-X, Ma H, et al. Role of ZIC1 methylation in hepatocellular carcinoma and its clinical significance. *Tumour Biol* 2014; **35**: 7429–7433.
29. Rodríguez-Rodero S, Fernández AF, Fernández-Morera JL, et al. DNA methylation signatures identify biologically distinct thyroid cancer subtypes. *J Clin Endocrinol Metab* 2013; **98**: 2811–2821.
30. Uribe ML, Dahlhoff M, Batra RN, et al. TSHZ2 is an EGF-regulated tumor suppressor that binds to the cytokinesis regulator PRC1 and inhibits metastasis. *Sci Signal* 2021; **14**: eabe6156.
31. Zhao S, Guo X, Mizutani K-I, et al. Overexpression of Teashirt Homolog 2 suppresses cell proliferation and predicts the favorable survival of lung adenocarcinoma. *Int J Med Sci* 2021; **18**: 1980–1989.
32. Xie W, Du Z, Chen Y, et al. Identification of metastasis-associated genes in triple-negative breast cancer using weighted gene co-expression network analysis. *Evol Bioinform Online* 2020; **16**: 1176934320954868.
33. Moon SU, Kim JH, Woo HG. Tumor suppressor RBM24 inhibits nuclear translocation of CTNBN1 and TP63 expression in liver cancer cells. *Oncol Lett* 2021; **22**: 674.
34. Hua W-F, Zhong Q, Xia T-L, et al. RBM24 suppresses cancer progression by upregulating miR-25 to target MALAT1 in nasopharyngeal carcinoma. *Cell Death Dis* 2016; **7**: e2352.
35. Xia RM, Liu T, Li WG, et al. RNA-binding protein RBM24 represses colorectal tumorigenesis by stabilising PTEN mRNA. *Clin Transl Med* 2021; **11**: e383.
36. Min J-W, Koh Y, Kim D-Y, et al. Identification of novel functional variants of SIN3A and SRSF1 among somatic variants in acute myeloid leukemia patients. *Mol Cells* 2018; **41**: 465–475.
37. Torres-Martín M, Kusak ME, Isla A, et al. Whole exome sequencing in a case of sporadic multiple meningioma reveals shared *NF2*, *FAM109B*, and *TPRXL* mutations, together with unique *SMARCB1* alterations in a subset of tumor nodules. *Cancer Genet* 2015; **208**: 327–332.
38. Zhou Z-J, Luo C-B, Xin H-Y, et al. MACROD2 deficiency promotes hepatocellular carcinoma growth and metastasis by activating GSK-3 β / β -catenin signaling. *NPJ Genomic Med* 2020; **5**: 15.
39. Sakthianandeswaren A, Parsons MJ, Mouradov D, et al. MACROD2 deletions cause impaired PARP1 activity and chromosome instability in colorectal cancer. *Oncotarget* 2018; **9**: 33056–33058.
40. Hu N, Kadota M, Liu H, et al. Genomic landscape of somatic alterations in esophageal squamous cell carcinoma and gastric cancer. *Cancer Res* 2016; **76**: 1714–1723.
41. Chen C, Bartenhagen C, Gombert M, et al. Next-generation-sequencing of recurrent childhood high hyperdiploid acute lymphoblastic leukemia reveals mutations typically associated with high risk patients. *Leuk Res* 2015; **39**: 990–1001.

SUPPLEMENTARY MATERIAL ONLINE

Figure S1. Effect of the number of features selected on the model performance

Figure S2. Features' importance (weight) in the prediction

Figure S3. Average methylation value analyzed by the EPIC array within transcription start site (TSS) regions for the six genes of the epigenetic biomarker in a series of T-ALL ($n = 143$) and normal thymus ($n = 12$)

Table S1. DMRs in the MeDIP-array dataset (Excel file)

Table S2. DMRs in the EPIC-array dataset (Excel file)

Table S3. MS-MLPA probe sequences

Table S4. Epidemiological features of the T-LBLs and thymoma included in the MeDIP-array series

Table S5. Epidemiological features of the T-LBLs and thymomas included in the EPIC-array series