Biomarker Research

CORRESPONDENCE

Open Access

Tissue- and tumor-type-specific expression of internal-promoter-driven YEATS-domain-devoid isoforms of *MLLT1* and *MLLT3*

Sandra Dailhau[1], Ahmed Zamani[1], Romain Pfeifer[1], Chloé Bessière[1], Thérèse Commes[2], Olivera Casar-Borota[3,4], Christian Touriol[1], Fabienne Meggetto[1], Marina Bousquet[1], Nicolas Gilbert[2*] and Stéphane Pyronnet[1*]

Abstract

As integral parts of the transcription super elongation complex, the ENL and AF9 proteins are two master regulators of gene expression involved in the development and homeostasis of various tissues. They are encoded by *MLLT1* and *MLLT3*, two genes that frequently undergo oncogenic mutations or chromosomal translocations in cancers. In addition, we discovered an internal promoter driving the production of a YEATS (named after the five proteins first shown to contain this domain: Yaf9, ENL, AF9, Taf14, and Sas5)-domain-devoid AF9 isoform massively expressed in healthy hematopoietic stem cells (HSCs), as well as in the blasts of patients with extremely poor-prognosis acute myeloid leukemia (AML). Here, we show that such internal promoter is also active in a subset of digestive tissues, including the small intestine, colon and stomach. We also reveal the existence of a similar internal promoter in the *MLLT1* locus that is inactive in the hematopoietic lineage and AML but functional in other healthy tissues, including the pituitary gland, brain, thyroid, uterus and lung. Furthermore, the YEATS-domain-devoid *MLLT1* isoform may exhibit tumor-type-specific biomarker values, whereas full-length *MLLT1* does not. It is highly expressed in the gonadotroph subtype of pituitary tumors, and its expression in non-small cell lung cancer inversely correlates with the epidermal growth factor mutation and a transcriptomic signature associated with a good prognosis. Thus, both *MLLT1* and *MLLT3* have tissue- and tumor-type-specific internal promoters driving expression of similar YEATS domain-devoid isoforms. This suggests that the functions of both *MLLT1* and *MLLT3* in tissue homeostasis and cancers, as well as their potentials as biomarkers should be re-evaluated taking into account expression of these YEATS-domain-devoid ENL and AF9 transcription factors.

Keywords *MLLT1*, *MLLT3*, ENL, AF9, Hematopoiesis, Acute myeloid leukemia, Pituitary tumor, Non-small cell lung cancer

*Correspondence:
Nicolas Gilbert
nicolas.gilbert@inserm.fr
Stéphane Pyronnet
stephane.pyronnet@inserm.fr
[1]Univ Toulouse, CNRS, INSERM, CRCT (Cancer Research Center of Toulouse), UMR-1037, UMR-5071, Institut Universitaire du Cancer, Laboratoire d'Excellence Toulouse Cancer (Labex TOUCAN), Toulouse 31037, France
[2]IRMB, Université de Montpellier, INSERM, Montpellier, France
[3]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden
[4]Department of Clinical Pathology, Uppsala University Hospital, Uppsala, Sweden

To the Editor,

The homologous transcription factors ENL (*MLLT1*) and AF9 (*MLLT3*) optimize the catalytic rate of transcription by RNApol II at specific genomic loci. They do so *via* an N-terminal chromatin reader domain called YEATS (named after the five proteins first shown to contain this domain: Yaf9, ENL, AF9, Taf14, and Sas5) and a C-terminal transactivation domain called AHD. The YEATS and AHD domains interact with various protein partners of the transcription super-elongation complex or SEC (Fig. 1A) [1]. *MLLT1* and *MLLT3* play essential roles in maintaining cells' homeostasis in the hematopoietic [2] and epithelial [3] systems. They both undergo frequent oncogenic chromosomal translocations or mutations that drive different types of cancer [4, 5]. Furthermore, we recently discovered an internal promoter that produces shorter *MLLT3* transcripts, called *s-MLLT3* in comparison to the full-length *l-MLLT3* [6]. These *s-MLLT3* transcripts encode a YEATS-domain-devoid isoform of AF9 that is massively expressed in healthy hematopoietic

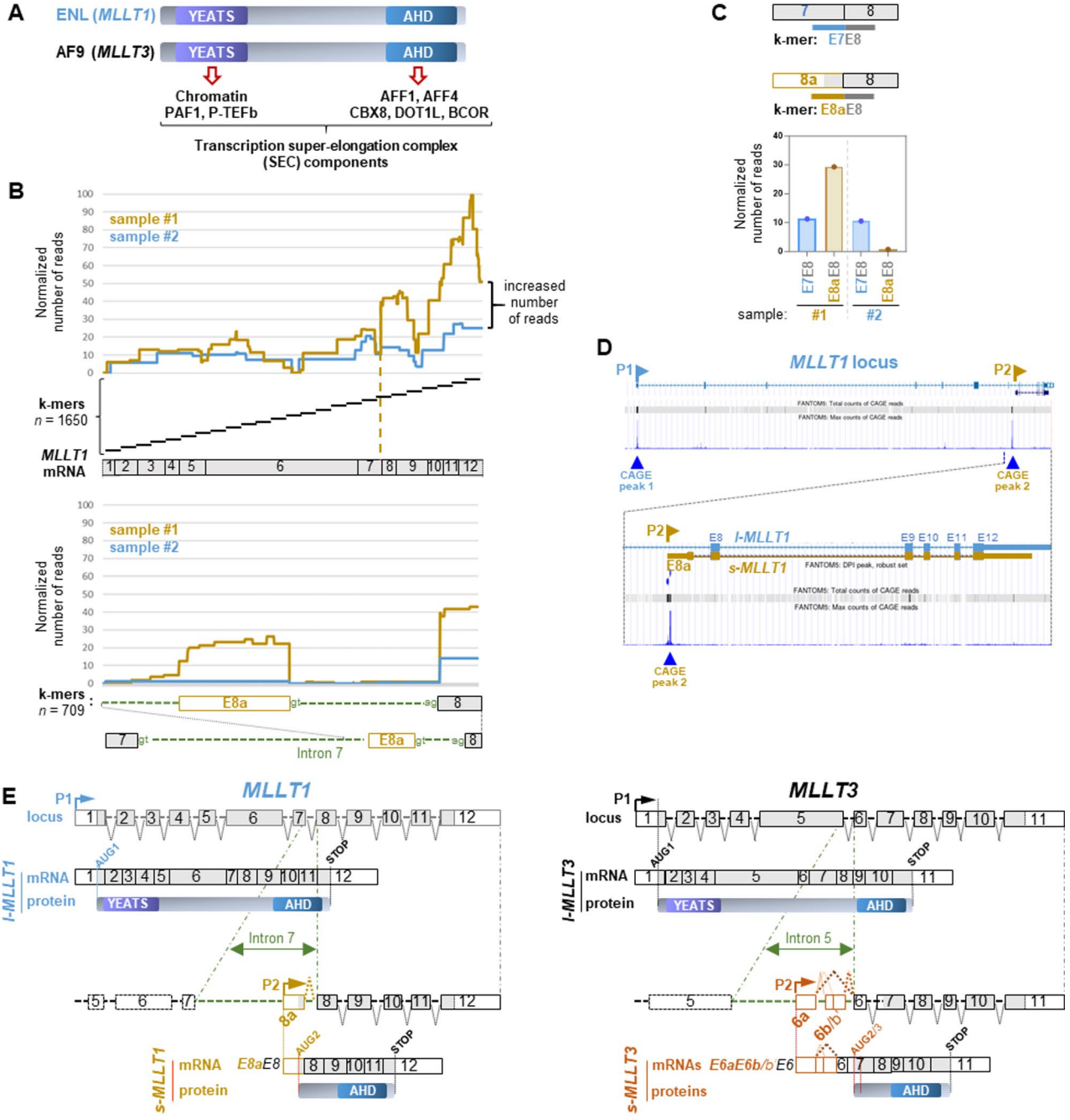

Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Identification of an internal promoter driving expression of *s-MLLT1*, a YEATS-domain-devoid isoform of *MLLT1*. **A** Schematic representation of the ENL and AF9 proteins with the N-terminal YEATS and C-terminal AHD domains and their interacting partners in the super elongation complex (SEC) complex. **B** To see how RNA-Seq reads distribute along the 12 exons (E1-E12) of the *MLLT1* mRNA (X-axis) a k-mer quantification (Y-axis, normalized number of reads) was performed on two healthy pituitary samples (#1, ID: GTEX-13×6J-3126-SM-5Q5EJ; and #2, ID: GTEX-13VXU-3126-SM-5SIA4). The 1680 nucleotides composing the reference *MLLT1* transcript (NM_005934.4, from AUG to stop codon) gave 1650 consecutive k-mers where k = 31 (31-mers). A distinct profile was obtained with sample #1, with a sharp increase in the number of reads occurring at the very first nucleotide of E8 (**top**). To detect potentially novel *MLLT1* sequence segments that could be retained in poly(A)+ RNAs, a k-mer quantification of RNA-Seq reads (Y-axis, normalized number of reads) mapping to *MLLT1* intron 7 (X-axis) was performed. This revealed in sample #1 the existence of one additional segment upstream from E8 and flanked at its 3′ end by a donor (gt) splicing site but lacking a clearly defined 5′ end consistent to a probable transcription start site. This novel exon was named E8a (**bottom**). **C** To verify whether the novel exon E8a could indeed be spliced to exon E8, specific k-mers were designed as follows: the last 16 nucleotides (3′ end) of E7 or 8a joined to the first 15 nucleotides (5′ end) of E8 (i.e. k = 16 + 15 = 31) (Supplementary Information). Then, k-mer quantification of the RNA-Seq reads covering either the putative E8aE8 junction or the reference E7E8 junction confirmed the existence of the alternative E8aE8 junction in pituitary sample #1, but not in sample #2. **D** Since no starting nucleotide could be detected for the alternative E8a, we searched for the presence of potential promoter features in the surrounding area by mining existing data. Visualization using the UCSC genome browser of CAGE peaks detected along the *MLLT1* locus revealed, in addition to a CAGE peak at the canonical P1 promoter, the existence of another peak right upstream of the alternative E8a, consistent with the existence of an internal promoter we named P2. **E** Schematic representation of the canonical P1 promoter-driven full-length transcripts *l-MLLT1* and *l-MLLT3* (**top**), together with their P2 internal promoter-driven, YEATS domain-devoid shorter isoforms *s-MLLT1* and *s-MLLT3* (**bottom**). The complete sequences of *l-MTT1* and *s-MLLT1* transcripts (ENST00000252674.9 and ENST00000585588.1, respectively) and of their corresponding proteins are depicted in Fig. S2

stem cells (HSCs) and in blasts of acute myeloid leukemia (AML) with an extremely poor prognosis [6]. However, there are no data on *s-MLLT3* expression in other healthy or malignant tissues, nor on the possibility that *MLLT1* may possess a similar internal promoter.

An examination of the expression profiles of *MLLT1* and *MLLT3* using the GTEx repository revealed that the *MLLT1* locus exhibited approximately ten times higher global expression than the *MLLT3* locus in almost all tissues. (Fig. S1). Then, we used a k-mer approach we validated recently (Supplementary Methods), and which was designed to visualize how the RNA sequence reads distribute along reference transcripts, quantitatively and at single nucleotide resolution [6]. This uncovered one alternative transcript (called *s-MLLT1*) starting at an internal position in *MLLT1* intron 7, as described here for two typical pituitary poly(A) + RNA-Seq samples (Fig. 1B and C). Exploration of genomic data using the UCSC browser confirmed the existence of an alternative transcription start site, as attested by a CAGE peak in intron 7 (Fig. 1D). The resulting internal transcript starts with an in-frame AUG-containing alternative exon we named exon 8a (E8a) and is predicted to encode an N-terminal-truncated, YEATS-domain-devoid ENL protein, exactly as for AF9 [6] (Figs. 1E and S2).

Quantifications using exon-exon junction-specific k-mers revealed that *s-MLLT1* is expressed in a number of samples of solid tissues, mainly in the pituitary gland and in brain, thyroid, uterus and lung. *s-MLLT3* is hugely expressed in $CD34^+$ HSCs and progenitors, but moderately in a few samples of the small intestine, colon and stomach (Fig. 2A).

Next, we compared the relative expression of *l-MLLT1* and *s-MLLT1* transcripts in a well-defined previously published series of 51 pituitary tumors [7] collected with the support of the Human Protein Atlas (www.proteinatlas.org), and U-CAN (www.u-can.uu.se) [8, 9]. Pituitary tumors can be classified phenotypically as functional tumors that produce hormones (including acromegaly for growth hormone, Cushing's syndrome) or non-functional tumors. They can also be classified molecularly according to their expression of the cell-lineage-specific transcription factors PIT1, SF1 and TPIT. The data revealed that, unlike *l-MLLT1*, which does not fluctuate, *s-MLLT1* expression is high in non-functional tumors (Fig. 2B left) and, more generally, in $SF1^+$ gonadotroph tumors, but low in $TPIT^+$ samples and very low in $PIT1^+$ samples (Fig. 2B right).

We also performed an analysis using a non-small cell lung carcinoma (NSCLC) RNA-Seq dataset, which we made k-mer searchable. This dataset comprises 77 NSCLC samples, each of which is matched with adjacent non-cancerous tissue [10]. Both *l-MLLT1* and *s-MLLT1* appeared globally expressed at lower levels in tumor vs. adjacent tissues (Fig. 2C left). However, in *s-MLLT1*$^+$ tumor samples, a lower expression of *s-MLLT1* (alongside higher expression of *l-MLLT1*) anti-correlated with a previously published [11] NSCLC poor prognosis 3-gene signature (Fig. 2C middle), while the overall expression of *MLLT1* was not discriminating (Fig. S3). In addition, higher *s-MLLT1* expression correlated with *EGFR* heterozygous mutation (the most frequent mutation in NSCLC) (Fig. 2C right), which is generally considered as having a more favorable prognosis value [12].

Thus, both *MLLT1* and *MLLT3* loci produce internal promoter-driven transcripts encoding similar YEATS-domain-devoid ENL and AF9 transcription factors. However, they are tissue-specific and the corresponding alternative transcripts may have biomarker values at least in AML for *MLLT3* and in pituitary and lung tumors for *MLLT1*. The biological and biomarker potential of YEATS-domain-devoid ENL and AF9 offers important new research opportunities, particularly with regard to tissue homeostasis and malignancies of the pituitary

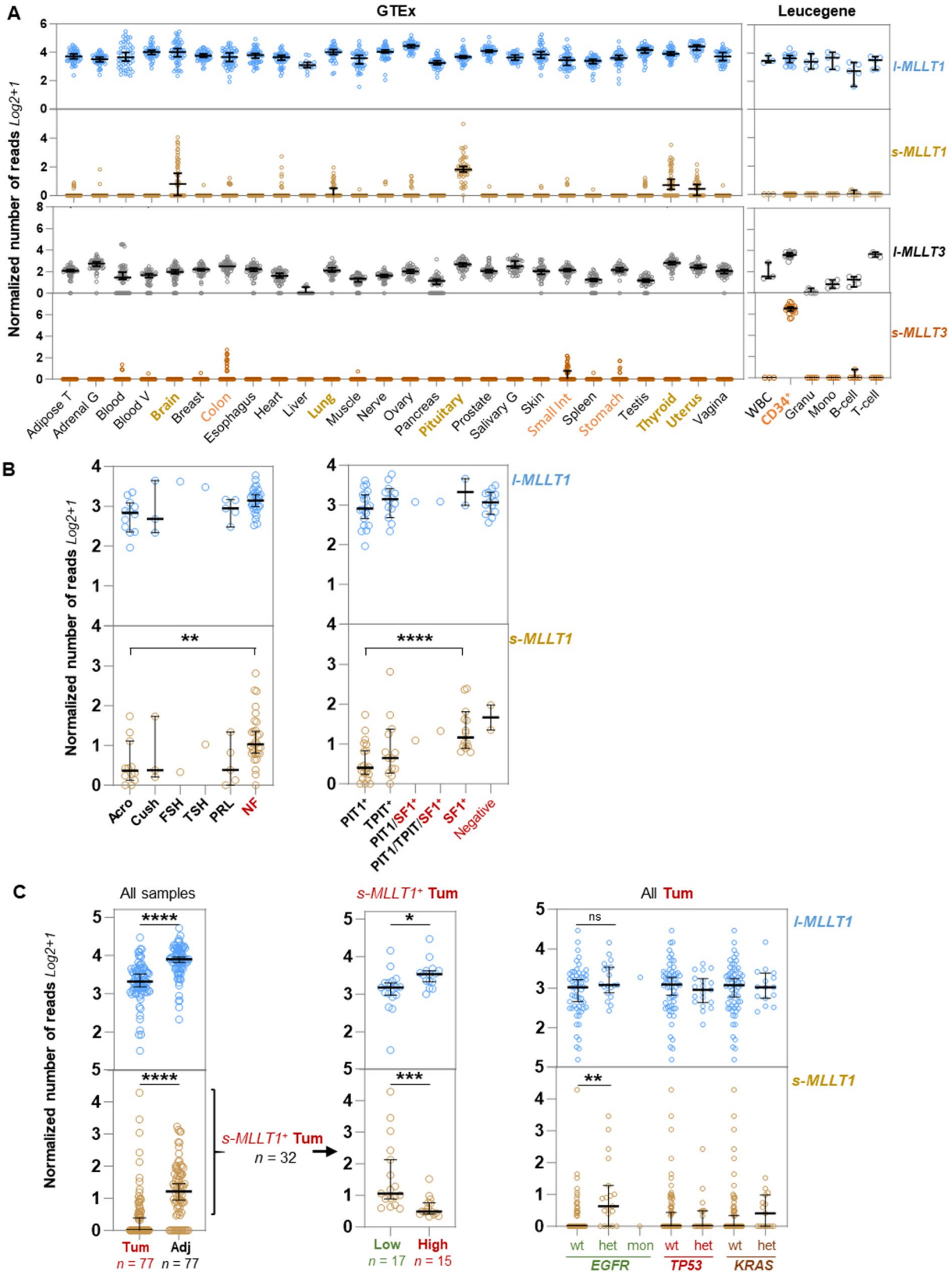

Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Biomarker values of the tissue- and tumor-type-specific expression of internal promoter-driven *s-MLLT1* isoform. **A***l-MLLT1* and *s-MLLT1* were quantified in poly(A) + RNA-Seq samples of the GTEx and Leucegene repositories (Supplementary Information) by using the E7E8 and E8aE8 (for *MLLT1*) and E5E6 and E6aE6 + E6b/b'E6 (for *MLLT3*) k-mers (Supplementary Information). The data revealed a total absence of *s-MLLT1* in the hematopoietic lineage, but expression in a number of samples of other tissues including the pituitary gland, brain, thyroid, uterus and lung. In contrast, the *s-MLLT3* isoform appeared highly expressed in $CD34^+$ cells of the hematopoietic lineage, but moderately expressed in a number of samples of small intestine, colon and stomach. Bars: medians with 95% confidence interval. **B** To search for a potential biomarker value of *l-MLLT1* or *s-MLLT1* in pituitary tumors, a series of $n = 51$ poly(A) + RNA-Seq samples (Supplementary Information) was explored by mapping and counting the *s-MLLT1*- or *l-MLLT1*-specific RNA-sequencing reads using a conventional RNA quantification method. This revealed that the expression of *s-MLLT1*, but not *l-MLLT1*, is higher in non-functional tumors (NF) than in acromegaly (Acro), Cushing's syndrome (Cush) or tumors producing FSH, TSH or PRL (**left**). Consistently, *s-MLLT1* expression is higher in $SF1^+$ samples (**right**). Bars: medians with 95% confidence interval. **C** To verify whether *l-MLLT1* or *s-MLLT1* could have a biomarker potential in lung tumors, a series of $n = 77$ poly(A) + RNA-Seq of non-small cell lung cancer (NSCLC) samples (Tum) together with their matched non-tumor adjacent tissues (Adj) (Supplementary Information) was quantified by the k-mer technique. This revealed that the expression of both *l-MLLT1* and *s-MLLT1* is lower in tumor vs. adjacent tissues (**left**). In *s-MLLT1*$^+$ tumor samples ($n = 32$), a lower expression of *s-MLLT1* but a higher expression of *l-MLLT1* anti-correlated with a 3-gene signature score (defined as Low or High, Supplementary Information) recognized earlier as a marker of bad prognosis (**middle**) [11]. Additionally, significantly higher levels of *s-MLLT1* expression were detected in tumor samples from patients with *EGFR* heterozygous mutations, but not in samples with mutations in *TP53* or *KRAS* (**right**). Bars: medians with 95% confidence interval

gland, lung, brain, thyroid, uterus, small intestine, colon and stomach.

The strengths of this study include a very large sample size of non-pathological human tissues (thousands of samples from the GTEx repository), a well-characterized cohort of lung tumor samples matched with their adjacent non-cancerous tissue, and one of the largest cohorts of well-defined pituitary tumors. The limitations include focusing on a single analytical technique (RNA-seq coupled with k-mer-designed bioinformatics pipelines) and using a unique cohort for each lung and pituitary tumor. Therefore, to validate their robustness and general applicability, it is necessary to validate *MLLT1* and *MLLT3* internal transcripts as biomarkers using another molecular technique and samples collected in different centers.

Abbreviations

AF9	Protein AF9
AFF1/4	ALF (General Transcription Factor IIA 1 Like) Transcription Elongation Factor 1 or 4
AHD	ANC1 (nuclear ANChorage protein 1) Homology Domain
AML	Acute Myeloid Leukemia
BCOR	BCL6 (B Cell Lymphoma 6) Corepressor
CAGE	mRNA 5'Cap Analysis of Gene Expression
CBX8	Chromobox 8
DOT1L	Disruptor Of Telomeric silencing 1 like
EGFR	Epidermal Growth Factor Receptor
ENL	Protein ENL (Eleven-Nineteen Leukemia)
HSC	Hematopoietic Stem Cell
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
NSCLC	Non-Small Cell Lung Cancer
MLLT1/3	Myeloid/Lymphoid or Mixed-Lineage Leukemia (trithorax homolog); Translocated to, 1 or 3
PAF1	RNApol II-Associated Factor 1 homolog
PIT1	Pituitary Transcription Factor 1
P-TEFb	Positive Transcription Elongation Factor b
RNApol II	RNA Polymerase II
SEC	Super Elongation Complex
SF1	Steroidogenic Factor-1
TPIT	T-box transcription factor, Pituitary
TP53	Tumor Protein P53
WBC	White Blood Cells

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40364-026-00928-w.

Supplementary Material 1

Acknowledgements

We thank the Human Protein Atlas and Per Oksvold from KTH, Stockholm University, for extracting the MLLT1 sequencing data from the primary data set.

Author contributions

Stéphane Pyronnet wrote the manuscript. Sandra Dailhau, Ahmed Zamani, Romain Pfeifer and Chloé Bessière, Olivera Casar-Borota, Marina Bousquet and Stéphane Pyronnet contributed to the study design, conception and analysis of the data. Sandra Dailhau, Chloé Bessière, Thérèse Commes and Nicolas Gilbert developed the k-mer-based bioinformatics tools and performed analyses. Olivera Casar-Borota, Christian Touriol and Fabienne Meggetto commented on and edited the manuscript.

Funding

This work was funded by INSERM (MIC/ESCALATE 24CM019-00), The French Ministry of Health and the French National Cancer Institute (INCA CircOma, PRT-K 2022 – 184), Institut Universitaire du Cancer - Toulouse (IUCT), Labex Toucan, Fondation Leucémie Espoir, Ligue Régionale Contre le Cancer (Haute-Garonne), Fondation ARC, Agence Nationale de la Recherche (ANR-22-CE45-0007 Full RNA) and the Cancéropôle GSO (SPLICE-Kmers-2025). Olivera Casar-Borota was supported by the Swedish Cancer Society and by the grant from the Swedish state under the agreement between the Swedish government and the county councils (ALF-agreement). Sandra Dailhau was supported by Ministère de la Santé and Institut National du Cancer (INCA CircOma, PRT-K-2022-184), Ahmed Zamani by Ligue Nationale Contre le Cancer, Romain Pfeifer by Ministère de l'Enseignement Supérieur et de la Recherche and Fondation ARC, Chloé Bessière by Fondation de France.

Data availability

The data on healthy human tissues used for the analyses described in this manuscript were obtained from the GTEx Portal (https://gtexportal.org/home/) on 10/01/25. The publicly available data on healthy human haematopoiesis (GSE51984) were generated by the Leucegene group, which is primarily located at IRIC in Montreal, Canada and supported by Genome Canada and Genome Québec. This dataset was made possible through human AML specimens provided by the BCLQ, Montreal, Canada. RNA-sequencing data for pituitary tumors were generated and published earlier (Tenabi et al. Acta Neuropathol Commun. 2021;9(1):181). The publicly available lung tumors (NSCLC) raw RNA-sequencing data (GSE40419) were published earlier (Seo et al. Genome Res. 2012;22(11):2109-19).

Declarations

Ethics approval and consent to participate
This study was performed in line with the principles of the Declaration of Helsinki. The collection of pituitary tissue material we referred to in this manuscript was approved by the Swedish Ethical Review Authority, Dnr 2018/053, and supported by U-CAN through Uppsala Biobank and the Department of Clinical Pathology at Uppsala University Hospital. All other data used in this study were published earlier with appropriate authorizations.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 24 December 2025 / Accepted: 25 April 2026
Published online: 12 May 2026

References

1. Kabra A, Bushweller J. The Intrinsically Disordered Proteins MLLT3 (AF9) and MLLT1 (ENL) - Multimodal Transcriptional Switches With Roles in Normal Hematopoiesis, MLL Fusion Leukemia, and Kidney Cancer. J Mol Biol. 2022;434:167117.
2. Calvanese V, Nguyen AT, Bolan TJ, Vavilina A, Su T, Lee LK, et al. MLLT3 governs human haematopoietic stem-cell self-renewal and engraftment. Nature. 2019;576:281–6.
3. Wan L, Chong S, Xuan F, Liang A, Cui X, Gates L, et al. Impaired cell fate through gain-of-function mutations in a chromatin reader. Nature. 2020;577:121–6.
4. Meyer C, Larghero P, Almeida Lopes B, Burmeister T, Gröger D, et al. The KMT2A recombinome of acute leukemias in 2023. Leukemia. 2023;37:988–1005.
5. Perlman EJ, Gadd S, Arold ST, Radhakrishnan A, Gerhard DS, Jennings L, et al. MLLT1 YEATS domain mutations in clinically distinctive Favourable Histology Wilms tumours. Nat Commun. 2015;6:10013.
6. Bessière C, Zamani A, Pfeifer R, Dailhau S, Marchet C, Guibert B, et al. A strong internal promoter drives massive expression of YEATS-domain devoid MLLT3 transcripts in HSC and most lethal AML. Cancer Commun (Lond). 2025;45:380–5.
7. Tebani A, Jotanovic J, Hekmati N, Sivertsson Å, Gudjonsson O, Edén Engström B, et al. Annotation of pituitary neuroendocrine tumors with genome-wide expression analysis. Acta Neuropathol Commun. 2021;9:181.
8. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics: tissue-based map of the human proteome. Science. 2015;347:1260419.
9. Glimelius B, Melin B, Enblad G, Alafuzoff I, Beskow A, Ahlström H, et al. Annotation of pituitary neuroendocrine tumors with genome-wide expression analysis. U-CAN: a prospective longitudinal collection of biomaterials and clinical information from adult cancer patients in Sweden. Acta Oncol. 2018;57:187–94.
10. Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. Genome Res. 2012;22:2109–19.
11. Lau SK, Boutros PC, Pintilie M, Blackhall FH, Zhu CQ, Strumpf D, et al. Three-gene prognostic classifier for early-stage non small-cell lung cancer. J Clin Oncol. 2007;25:5562–9.
12. Zhou F, Guo H, Xia Y, Le X, Tan DSW, Ramalingam SS, et al. The changing treatment landscape of EGFR-mutant non-small-cell lung cancer. Nat Rev Clin Oncol. 2025;22:95–116.

Publisher's note