



**HAL**  
open science

# Etude de l'accessibilité chromatinienne des néoplasies lymphoïdes T matures : amélioration de la précision diagnostique et classification par cellule d'origine

Edith Julia

## ► To cite this version:

Edith Julia. Etude de l'accessibilité chromatinienne des néoplasies lymphoïdes T matures : amélioration de la précision diagnostique et classification par cellule d'origine. Cancer. Université Claude Bernard - Lyon I, 2023. Français. NNT : 2023LYO10291 . tel-04844564

**HAL Id: tel-04844564**

**<https://theses.hal.science/tel-04844564v1>**

Submitted on 18 Dec 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THESE de DOCTORAT DE  
L'UNIVERSITE CLAUDE BERNARD LYON 1**

**Ecole Doctorale N° 340  
Biologie Moléculaire Intégrative et Cellulaire**

**Discipline : Biologie**

Soutenue publiquement le 18/12/2023, par :  
**Edith Julia**

---

**Etude de l'accessibilité chromatinienne  
des néoplasies lymphoïdes T matures :  
amélioration de la précision  
diagnostique et classification par cellule  
d'origine**

---

Devant le jury composé de :

Couronné, Lucille - MCU-PH - Université Paris Cité  
Ruminy, Philippe - DR - INSERM  
De Leval, Laurence - PU-PH - Université de Lausanne  
Gaulard, Philippe - PU-PH - Université Paris Est Créteil  
Hermine, Olivier - PU-PH - Université Paris Cité  
Traverse-Glehen, Alexandra - PU-PH - Université Claude Bernard Lyon 1  
Genestier, Laurent - DR - INSERM  
Bachy, Emmanuel - PU-PH - Université Claude Bernard Lyon 1

Rapportrice  
Rapporteur  
Examinatrice  
Examinateur  
Examinateur  
Présidente  
Directeur de thèse  
Co-directeur de thèse

# Remerciements

Merci au Dr Lucille Couronné et à Philippe Ruminy d'avoir accepté d'être rapporteurs de cette thèse.

Merci au Pr Hermine d'avoir accepté d'être examinateur de cette thèse.

Merci aux Pr. De Leval et Gaulard d'avoir accepté de participer à ce jury et surtout pour leur aide dans la mise à disposition des données de TENOMIC, les nombreuses relectures de cas et toutes les discussions intéressantes que nous avons pu avoir sur la complexité des PTCL.

Merci au Pr Traverse-Glehen d'avoir accepté d'être examinatrice cette thèse et pour votre aide dans la mise à disposition et les relectures des nombreux cas de MTCN de Lyon sud.

Merci Laurent et Emmanuel, d'avoir accepté d'encadrer ma thèse, d'avoir toujours été disponibles quand j'en avais besoin malgré vos emplois du temps chargés et d'avoir gardé confiance dans le projet même lorsque je me perdais dans mes énièmes itérations de NMF. Les réunions du vendredi après-midi passées à élaborer des nouvelles théories biologiques toutes plus novatrices les unes que les autres vont rester des moments mémorables de ces trois années de thèses. Merci

Ce travail de thèse présente le projet de toute une équipe que je tiens à remercier chaleureusement.

Merci Amel pour tout le travail de mise en place de l'ATAC-seq, les centaines d'échantillons techniqués et la bienveillance de ton encadrement dans mes premières manipulations de biologies moléculaires. Merci Sarah pour l'ATAC-FFPE, le combat contre les bactéries et ta bonne humeur au labo. Ça a été un plaisir de travailler avec vous.

Merci Sylvain pour les pipelines, les aides au débogage, et pour tout ce que tu m'as appris lors de mon premier stage à la cellule bioinfo et lors de cette thèse. Merci Quentin pour les analyses de CNV. Et merci à toute la cellule bioinformatique, Claire Bardel, Pierre-Antoine Rollat-Farnier, Thomas Simonet pour les aides techniques et les conseils avisés.

Merci Aurélie pour les tris innombrables et Nathalie pour les manipulations et collections d'échantillons. Merci à Camille Lours et à Camille Golfier pour la réalisation de l'ATAC des premiers échantillons de lymphomes et du projet cellule d'origine, ainsi que tous les recueils de données du FAST. Merci Sofian et Manon pour tous les échantillons que vous avez techniqués. Merci David pour ton aide sur les figures finales et bon courage pour la suite du projet.

Merci à Edoardo Missiaglia et Rattina Vimel pour leur aide sur les analyses d'exomes.

Merci Marie Donzel pour ton aide précieuse dans les relectures des cas. J'ai hâte de voir la suite du projet.

Merci Pierre pour toutes les discussions scientifiques et politiques passionnantes.

Merci Manon, d'avoir été là pour les crises de la fin de rédaction, merci pour ton accent italien et tes talents d'artiste plasticienne. Nos duos de chants vont me manquer. Merci Angélique pour les discussions bioinfo et ta réassurance indéfectible lors des pauses cafés, Laetitia pour ton enthousiasme scientifique communicatif et les récits de tes péripéties toujours plus rocambolesques qui ont égayé la période de rédaction, Narimène pour les coups de mains et la solidarité entre thésarde de dernière minutes (et bravo docteur). Et merci Chloé, Emiel, Axel, Erivan, Alexandra, Anne-Marie, Camille, Sarah pour les déjeuners avec les mouches et les soirées apéros. Ces trois années n'auraient pas été les mêmes sans vous.

Merci à tous mes amis pour votre gentillesse et votre soutien surtout dans la dernière ligne droite. Merci aux copain.e.s du CHAL, Juliette, Camille, Lorène, Delphine, Hugo, Marc. Quelques années plus tard je suis si contente de que vous soyez encore là. Merci Noémie et Julien pour votre soutien thérapeutiques et fromager sans faille. On se retrouve bientôt (enfin !) dans la montagne ! Merci à mes colocs, Margot pour le soutien musical, enthousiaste de toutes ces années de cohabitation ; Fabien pour les petits plats et le cinémas, et Julia d'avoir été là dans la dernière ligne droite, sans toi je serais probablement morte de faim avant d'avoir fini de rédiger.

Et enfin un grand merci à mes parents et à mes frère et sœurs d'avoir toujours été là à chaque réussite, à chaque difficulté et ce malgré mon incapacité à répondre aux messages :). Merci tout particulièrement à mon père pour la relecture du manuscrit.

# Résumé

Les néoplasies lymphoïdes matures T (MTCN) sont des maladies hétérogènes au pronostic défavorable. La discrimination entre ces nombreuses entités est difficile et nécessite une expertise anatomopathologique spécialisée. Le séquençage de la chromatine accessible à la transposase (ATAC-seq) est une technique simple qui permet de mesurer l'ouverture des régions chromatinienne et qui s'est déjà avérée très discriminante pour la distinction des tumeurs solides ou des leucémies aigües myéloïdes. Nous avons appliqué l'ATAC-seq aux MTCN afin d'explorer le paysage épigénétique de ces différentes entités et de construire un modèle prédictif pour faciliter le diagnostic. En utilisant des approches de déconvolution non supervisée des profils épigénétiques, nous avons pu discriminer les différents sous-types de MTCN à partir d'échantillons de cellules vivantes triées et de sections tumorales congelées. La plupart des lymphomes périphériques sans spécification (PTCL-NOS) ont été en majorité redistribués dans d'autres sous-types connus, ce qui montre que cette entité ne correspond pas à un groupe unique du point de vue épigénétique. Cependant un nouveau sous-groupe de PTCL-NOS a été découvert qui présente une forte activité du facteur de transcription GATA3, de nombreuses altérations chromosomiques et des mutations de *TP53* et *NCOR1*. Par ailleurs, la comparaison du profil ATAC-seq des MTCN à ceux des sous-types de lymphocytes T et NK de donneurs sains a confirmé la proximité des lymphomes TFH et des lymphocytes TFH normaux et suggère une origine naïve aux T-PLL. Nous avons montré que l'ATAC-seq est réalisable sur des échantillons fixés au formol et inclus en paraffine. Enfin un modèle linéaire généralisé a été entraîné pour prédire le diagnostic de MTCN, montrant une performance de classification élevée sur une cohorte de validation externe.

# Summary

## Genome-wide chromatin accessibility landscape increases diagnostic accuracy and refines cell-of-origin classification of mature T-cell neoplasms

Mature T-cell neoplasms (MTCN) are heterogeneous diseases with dismal prognosis. Differentiating between the many entities requires specialized pathology expertise, and studies show up to 30% of minor or major diagnostic reclassifications following expert review of cases. Assay for transposase-accessible chromatin sequencing (ATAC-seq) is a simple technique to profile open chromatin regions, which has been shown to be highly effective for clustering of solid tumors and acute myeloid leukemias. We applied ATAC-seq to MTCN samples to explore the epigenetic landscape of these different entities and build a predictive model to aid in diagnosis. Using unsupervised deconvolution approaches on epigenetic profiles, we were able to discriminate the different MTCN subtypes from sorted and frozen bulk samples. Most PTCL-NOS clustered with other predefined MTCN subtypes showing that this subgroup is merely virtual, at least from an epigenetic point of view. A novel subset of PTCL-NOS was discovered with high GATA3 transcription factor activity, numerous copy number alterations and *TP53* and *NCOR1* mutations. In parallel, we applied FAST-ATAC to several normal T and NK cell subsets of healthy donors to find MTCN entities normal counterpart. Unsupervised clustering confirmed that AITL are derived from TFH cells. We also found that T-PLL likely derive from the transformation of naïve T cells. Eventually, we showed that ATAC-seq is feasible on formalin-fixed paraffin-embedded samples without batch effect compared to frozen samples. A generalized linear model classifier was trained to predict MTCN diagnosis, showing high classification performance on an external validation cohort.

# Mots clés

## Français

Lymphome T, ATAC-Seq, Classification automatique, Epigénétique

## Anglais

T-cell neoplasia, ATAC-seq, Machine learning, Epigenetic

# Abréviations

5caC	5-carboxylcytosines
5hmC	5-hydroxyméthylcytosines
5mC	5-méthylcytosines
ACI	Analyse Composantes Indépendantes
ACP	Analyse Composante Principale
ADN	Acide Désoxyribonucléique
AITL	Angioimmunoblastic T-cell Lymphoma
ALCL	Anaplastic Large Cell Lymphoma
ALK	Anaplastic Lymphoma Receptor Tyrosine Kinase
ALL	Acute Lymphoblastic Leukemia
ARN	Acide Ribonucléique
ASAP-seq	ATAC with Select Antigen Profiling by sequencing
ATAC-seq	Assay for Transposase-Accessible Chromatin with highthroughput sequencing
ATLL	Adult T-cell Leukemia/Lymphoma
BCR	B Cell Receptor
CAGE	Cap Analysis of Gene Expression
CHIP	Chromatin ImmunoPrecipitation
CMH	Complexe Majeur d'Histocompatibilité
CTCL	Cutaneous T Cell Lymphomas
DMSO	Diméthylsulfoxyde
EATL	Enteropathy-Associated T-cell Lymphoma
EBV	Epstein Bar Virus
EHA	European Hematology Association
ENCODE	Encyclopedia of DNA Elements
ENKTL	Extranodal NK/T cell Lymphoma
eRNA	Enhancer RNA
ETP-ALL	Early T-precursor Acute Lymphoblastic Leukemia
FFPE	Formalin Fixed Paraffin-Embedded
FISH	Fluorescent In Situ Hybridization
GMP	Granulocyte Macrophage Progenitor
H3K27	Lysine 27 de l'histone H3
H3K27ac	Acétylation de la lysine 27 de l'histone H3
H3K4me1	Monométhylation de la lysine 4 de l'histone H3
H3K4me3	Triméthylation de la lysine 4 de l'histone H3
H3K9	Lysine 9 de l'histone H3
HSTL	Hepatosplenic T-cell Lymphoma
HTLV1	Human T-cell Leukemia Virus Type I
ICC	International Consensus Classification
IgHV	Ig Heavy Variable
IHC	Immunohistochemistry
IL	Interleukine



ILCP	Précurseur lymphoïde commun
ILCs	Cellules Lymphoïdes Innées
iTreg	Inducible Regulatory T cells
KIR	Killer Immunoglobulin Like Receptor
LAL-T	Leucémies Aiguës Lymphoblastique à précurseur T
LALB	Leucémies Aiguës Lymphoblastique à précurseur B
LGL	Large Granular Lymphocytic Leukaemia
LLC	Leucémie Lymphoïde Chronique
LMPP	Lymphoid-primed Multipotent Progenitor
MACS2	Model-based Analysis of ChIP-seq 2
ME	Mosaic Ends
MEITL	Monomorphic Epitheliotropic Intestinal T-cell Lymphoma
MF	Mycosis Fungoides
MHC	Major Histocompatibility Complex
ML	Machine Learning
MTCN	Mature T-cell and NK-cell neoplasms
NK	Natural Killer
NK-LGL	NK-Large Granular Lymphocytic Leukemia
NKTCL	Extranodal NK/T-cell Lymphoma
NMF	Factorisation de la Matrice Non Négative
NP40	Nonidet P-40
nTreg	Natural regulatory T cells
NURF	Nucleosome Re-modeling Factors
OMS	Organisation Mondiale de la Santé
pcALCL	Primary Cutaneous Anaplastic Large Cell Lymphoma
PCR	Polymerase Chain Reaction
PRC	Complex répresseur polycomb
PTCL-NOS	Peripheral T-Cell Lymphoma Not Otherwise Specified
RT-MLPA	Reverse Transcriptase Multiplex Ligation-dependant Probe Amplification
SELEX	Systematic Evolution of Ligands by EXponential enrichment
SH	Société d'Hémopathologie
SMILE-seq	Selective Microfluidic-based Ligand Enrichment followed by sequencing
SPTCL	Subcutaneous panniculitis (Panniculite sous-cutanée)
SVM	Support Vector Machine
T-ALL	T-cell Acute Lymphoblastic Leukemia
T-LGL	T-Large Granular Lymphocytic Leukemia
T-PLL	T-cell Prolymphocytic Leukemia
TCM	Lymphocytes T mémoires centraux
TCR	T Cell Receptor
TEMRA	Terminally differentiated effector memory T cells reexpressing 45RA
TF	Facteur de transcription
TFH	Lymphocyte T Folliculaire Helper
TFHL	Lymphomes T Folliculaire
Th	T helper

Treg	Lymphocytes T régulateurs
TRM	Lymphocytes T résidents mémoires
TSS	Transcription Start Site
UV	Ultra Violet
VDJ	Variable, Jonction, Diversité

# Table des matières

Résumé .....	2
Summary .....	5
Mots clés .....	6
Abréviations.....	7
Table des matières.....	10
Introduction.....	14
A    Les néoplasies lymphoïdes T/NK .....	15
1    Introduction générale .....	15
1.1    Classification.....	15
1.2    Epidémiologie.....	17
1.3    Pronostic.....	20
2    Présentation des principales entités .....	20
2.1    Lymphomes et leucémies aiguës lymphoblastiques T (T-cell acute lymphoblastic leukemia ou T-ALL) .....	20
2.2    Leucémie prolymphocytaire T (T-cell prolymphocytic leukemia ou T-PLL) .....	21
2.3    Leucémie à grand lymphocytes granuleux T et NK (NK and T-cell large granular lymphocytic leukemia ou NK-LGL and T-LGL) .....	22
2.4    Leucémie lymphome T de l'adulte lié à l'HTLV1 (Adult T-cell leukemia/lymphoma ou ATLL).....	23
2.5    Lymphomes T et NK EBV positifs (EBV positive NK and T cell lymphomas).....	24
2.6    Lymphomes intestinaux.....	26
2.7    Lymphomes hépatospléniques (Hepatosplenic T-cell lymphoma, HSTL).....	27
2.8    Mycosis fongoïdes et syndromes de Sezary .....	28
2.9    Lymphomes T folliculaires (Follicular helper T-cell lymphoma) .....	29
2.10    Les lymphomes T anaplasiques (Anaplastic large cell lymphoma, ALCL) .....	31

2.11	Lymphome T périphérique sans spécification (Peripheral T-cell lymphoma, NOS, PTCL-NOS).....	32
3	Modalités et difficultés diagnostiques .....	33
3.1	Diagnostic positif de lymphome T.....	33
3.2	Éléments d'orientation clinique .....	33
3.3	Histologie.....	34
3.4	Marqueurs protéiques en immunohistochimie et cytométrie en flux.....	35
3.5	Caryotype et Fluorescence in situ hybridization (FISH).....	39
3.6	Séquençage et variants nucléotidiques.....	40
3.7	Un diagnostic difficile .....	42
4	Différentiation lymphoïde T et cellule d'origine.....	43
4.1	Différentiation normale.....	43
4.2	Cellules d'origine des MTCN .....	49
B	Epigénétique et mesures de l'accessibilité chromatinienne .....	53
1	Structure de la chromatine et impact sur la régulation transcriptionnelle.....	53
1.1	Organisation du génome .....	53
1.2	Modifications épigénétiques et facteurs de transcription .....	55
1.3	Éléments régulateurs de la transcription .....	59
1.4	Facteurs de transcription, coactivateurs et corépresseurs. ....	63
2	Mesure de l'accessibilité chromatinienne .....	66
2.1	Généralités .....	66
2.2	Historique .....	66
2.3	L'ATAC-seq .....	67
2.4	Traitement bioinformatique des données d'ATAC-seq .....	70
2.5	Exemples d'application de l'ATAC-seq.....	74
C	Omique et classification en cancérologie.....	78

1	Apprentissage non supervisé.....	78
1.1	Clustering.....	78
1.2	Réduction de dimension et visualisation .....	80
2	Classification supervisée.....	83
2.1	Etapes de l'entraînement d'un modèle .....	83
2.2	Spécificité de l'application du machine learning aux mesures omiques.....	84
2.3	Algorithmes principaux .....	87
2.4	Classification multiclasse et probabilité de classe .....	88
3	Exemples de publications .....	89
3.1	Classification des lymphomes T .....	89
3.2	Performance des profils épigénomiques pour la classification de cancers .....	92
	Objectifs.....	94
	Article .....	95
	Introduction.....	99
	Results .....	101
	Discussion .....	110
	Methods .....	113
	Acknowledgements.....	123
	References.....	124
	Figures .....	129
	Supplementary data .....	140
	Discussion .....	149
1	Eclairage apporté par les points de divergence entre classification OMS et clusters épigénétiques.....	150
1.1	MEITL.....	150
1.2	ALCL.....	151

2	Caractérisation épigénétique des PTCL-NOS en ATAC-seq .....	152
2.1	Les PTCL NCOR, une nouvelle entité ? .....	152
3	Découverte d'un nouveau marqueur des lymphomes TFH.....	154
4	ATAC-seq, microenvironnement et pronostic des lymphomes TFH .....	155
5	Identification de la cellule d'origine des lymphomes T par ATAC-seq.....	156
6	Vers une utilisation de l'ATAC-seq en routine ? .....	159
6.1	Faisabilité.....	159
6.2	Validité.....	160
	Conclusion .....	162
	Sources .....	163
	Liste des tableaux .....	192
	Liste des figures .....	193
	Liste des genes et proteines .....	194

# INTRODUCTION

---

# A LES NEOPLASIES LYMPHOÏDES T/NK

## 1 Introduction générale

### 1.1 *Classification*

Les hémopathies lymphoïdes T/NK sont un groupe hétérogène de néoplasies dérivées de lymphocytes T et NK. Ils sont premièrement sous divisés en deux groupes en fonction du stade de maturation de la cellule atteinte :

- Les hémopathies lymphoïdes T immatures dérivées de précurseurs thymiques des lymphocytes T, appelées leucémies aiguës lymphoblastiques T
- Les hémopathies lymphoïdes T/NK matures

Elles peuvent atteindre la moelle osseuse, le sang périphérique (leucémies), les organes lymphoïdes (ganglions et rate) et les tissus extra lymphoïdes (foie, intestin, peau, cavité nasale). Elles étaient initialement classifiées selon leur site d'atteinte : leucémique, ganglionnaire ou extraganglionnaire. Mais les apports récents de la biologie moléculaire tendent à regrouper certaines entités d'après leur mécanisme physiopathologique et ce malgré leurs présentations cliniques variées. Il existe actuellement deux classifications principales des hémopathies lymphoïdes T :

- la classification de l'organisation mondiale de la santé (OMS - Tableau 1) (Alaggio et al. 2022)
- la classification consensus internationale (ICC) constitué de l'association européenne d'hématologie (EHA) et la société d'hémathologie (SH) (Tableau 2) (Campo et al. 2022).



Precursor T-cell neoplasms
<b>T-lymphoblastic leukaemia/lymphoma (T-ALL)</b>
T-lymphoblastic leukaemia / lymphoma, NOS (T-ALL-NOS)
Early T-precursor lymphoblastic leukaemia / lymphoma (ET-PLL)
Mature T-cell and NK-cell neoplasms (MTCN)
<b>Mature T-cell and NK-cell leukaemias</b>
T-prolymphocytic leukaemia (T-PLL)
T-large granular lymphocytic leukaemia (T-LGL)
NK-large granular lymphocytic leukaemia (NK-LGL)
Adult T-cell leukaemia/lymphoma (ATLL)
Sezary syndrome (SS)
Aggressive NK-cell leukaemia (NKL)
<b>Primary cutaneous T-cell lymphomas</b>
Primary cutaneous CD4-positive small or medium T-cell lymphoproliferative disorder
Primary cutaneous acral CD8-positive lymphoproliferative disorder
Mycosis fungoides (MF)
Primary cutaneous CD30-positive T-cell lymphoproliferative disorder: Lymphomatoid papulosis
Primary cutaneous CD30-positive T-cell lymphoproliferative disorder: Primary cutaneous anaplastic large cell lymphoma
Subcutaneous panniculitis-like T-cell lymphoma (SPTCL)
Primary cutaneous gamma/delta T-cell lymphoma
Primary cutaneous CD8-positive aggressive epidermotropic cytotoxic T-cell lymphoma
Primary cutaneous peripheral T-cell lymphoma, NOS
<b>Intestinal T-cell and NK-cell lymphoid proliferations and lymphomas</b>
Indolent T-cell lymphoma of the gastrointestinal tract
Indolent NK-cell lymphoproliferative disorder of the gastrointestinal tract
Enteropathy-associated T-cell lymphoma (EATL)
Monomorphic epitheliotropic intestinal T-cell lymphoma (MEITL)
Intestinal T-cell lymphoma, NOS
<b>Hepatosplenic T-cell lymphoma</b>
Hepatosplenic T-cell lymphoma (HSTL)
<b>Anaplastic large cell lymphoma</b>
ALK-positive anaplastic large cell lymphoma (ALCL-ALK+)
ALK-negative anaplastic large cell lymphoma (ALCL-ALK-)
Breast implant-associated anaplastic large cell lymphoma (BIA-ALCL)
<b>Nodal T-follicular helper (TFH) cell lymphoma (TFHL)</b>
Nodal TFH cell lymphoma, angioimmunoblastic-type (AITL)
Nodal TFH cell lymphoma, follicular-type
Nodal TFH cell lymphoma, NOS (PTCL-TFH)
<b>Other peripheral T-cell lymphomas</b>
Peripheral T-cell lymphoma, not otherwise specified (PTCL-NOS)
<b>EBV-positive NK/T-cell lymphomas</b>
EBV-positive nodal T- and NK-cell lymphoma
Extranodal NK/T-cell lymphoma (NKTCL)
<b>EBV-positive T- and NK-cell lymphoid proliferations and lymphomas of childhood</b>
Severe mosquito bite allergy
Hydroa vacciniforme lymphoproliferative disorder
Systemic chronic active EBV disease
Systemic EBV-positive T-cell lymphoma of childhood

**Tableau 1 - Classification OMS 2022 des néoplasies T**

Mature T-cell and NK-cell neoplasms
T-cell prolymphocytic leukemia
T-cell large granular lymphocytic leukemia
Chronic lymphoproliferative disorder of NK cells
Adult T-cell leukemia/lymphoma
Epstein-Barr virus–positive T-cell/NK-cell lymphoproliferative disorders of childhood
Hydroa vacciniforme lymphoproliferative disorder
Classic
Systemic
Severe mosquito bite allergy
Chronic active Epstein-Barr virus disease, systemic (T-cell and NK-cell phenotype)
Systemic Epstein-Barr virus–positive T-cell lymphoma of childhood
Extranodal NK/T-cell lymphoma, nasal type
Aggressive NK-cell leukemia
Primary nodal Epstein-Barr virus–positive T-cell/NK-cell lymphoma
Enteropathy-associated T-cell lymphoma
Type II refractory celiac disease
Monomorphic epitheliotropic intestinal T-cell lymphoma
Intestinal T-cell lymphoma, NOS
Indolent clonal T-cell lymphoproliferative disorder of the gastrointestinal tract
Indolent NK-cell lymphoproliferative disorder of the gastrointestinal tract
Hepatosplenic T-cell lymphoma
Mycosis fungoides
Sézary syndrome
Primary cutaneous CD30+ T-cell lymphoproliferative disorders
Lymphomatoid papulosis
Primary cutaneous anaplastic large cell lymphoma
Primary cutaneous small/medium CD4+ T-cell lymphoproliferative disorder
Subcutaneous panniculitis-like T-cell lymphoma
Primary cutaneous gamma-delta T-cell lymphoma
Primary cutaneous acral CD8+ T-cell lymphoproliferative disorder
Primary cutaneous CD8+ aggressive epidermotropic cytotoxic T-cell lymphoma
Peripheral T-cell lymphoma, NOS
Follicular helper T-cell lymphoma
Follicular helper T-cell lymphoma, angioimmunoblastic type (angioimmunoblastic T-cell lymphoma)
Follicular helper T-cell lymphoma, follicular type
Follicular helper T-cell lymphoma, NOS
Anaplastic large cell lymphoma, ALK positive
Anaplastic large cell lymphoma, ALK negative
Breast implant–associated anaplastic large cell lymphoma

**Tableau 2 - Classification ICC 2022 des néoplasies matures T**

## 1.2 *Epidémiologie*

Les néoplasies matures T sont beaucoup moins fréquentes que les lymphomes B (Figure 1) et représentent environ 10% des nouveaux cas de lymphomes en France. En 2018, d'après des estimations issues des registres du réseau cancer Francim, il y a eu 1945 nouveaux cas de lymphome T/NK matures dont un peu moins de la moitié (809 cas) sont des formes cutanées.

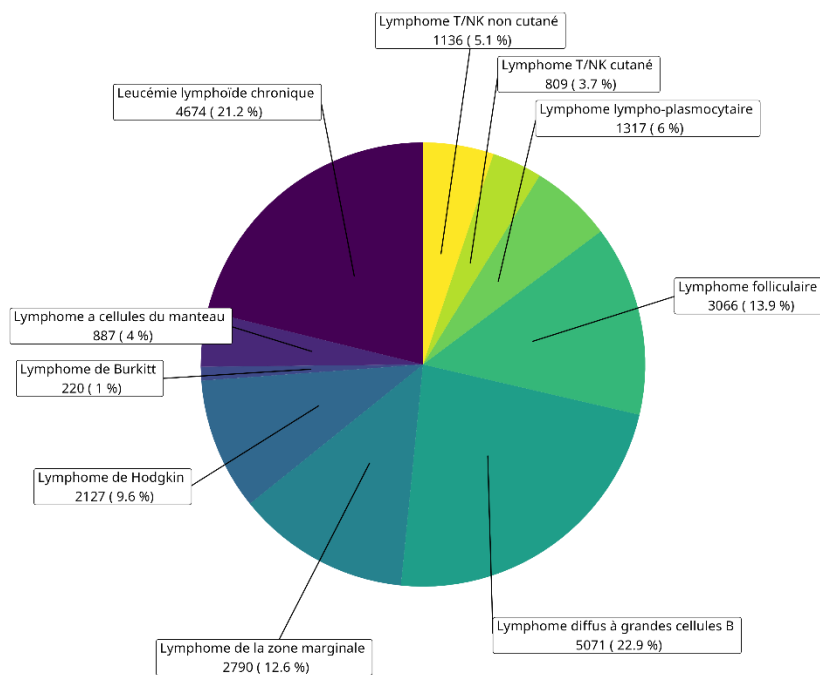


Figure 1 - Répartition des nouveaux cas d'hémopathies matures B et T en France en 2018

Comme pour d'autres hémopathies lymphoïdes, on observe actuellement une nette augmentation du taux d'incidence des lymphomes T non cutanés, respectivement de 1,8% par an chez l'homme et 4,3% par an chez la femme (Figure 2) qui ne peut être expliqué entièrement par le vieillissement de la population.

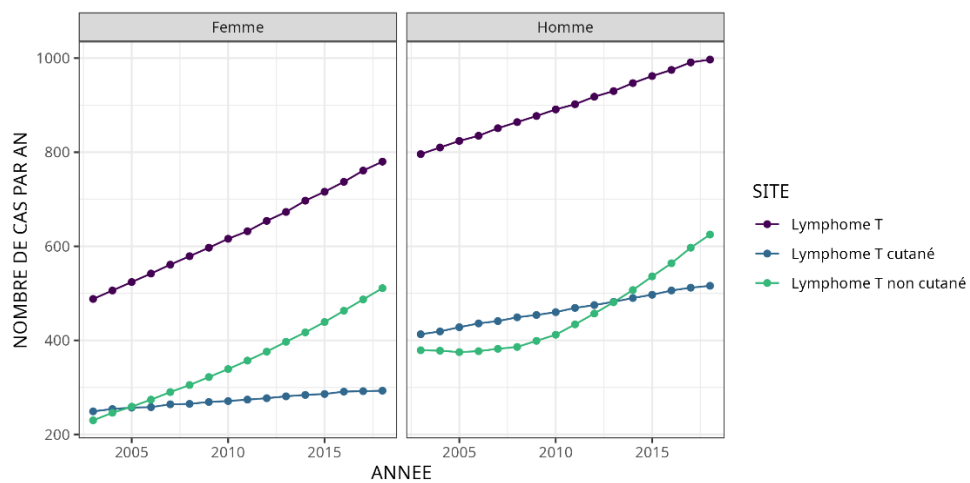


Figure 2 - Evolution du nombre de cas incidents de lymphomes T en France

Comme on peut le voir sur la Figure 2, il existe en France une prédominance de lymphomes T chez les hommes avec un sex-ratio à 1,8 pour les lymphomes T cutanés et 1,4 pour les non cutanés. La médiane d'âge des patients atteints de lymphome T est située entre 65 et 70 ans, et leur incidence augmente constamment avec l'âge. Certaines sous-entités touchent néanmoins des patients plus jeunes tels les lymphomes hépatospléniques (âge médian 29-38 ans) et les NKTCL (âge médian 50 ans) (J. Lee et al. 2005; Pro, Allen, and Behdad 2020; Yabe, Miranda, and Medeiros 2018).

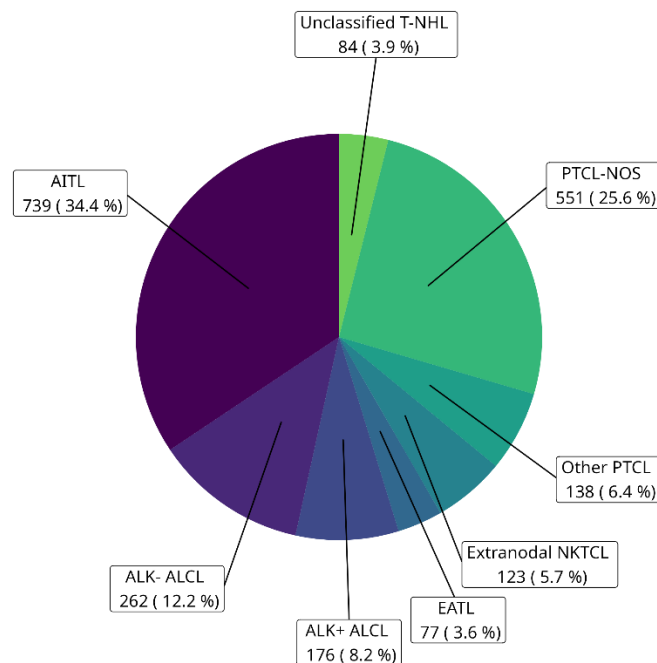


Figure 3 - Distribution des sous-entités de lymphomes T périphériques diagnostiqués dans le réseau Lymphopath entre 2010 et 2013 (Laurent et al. JCO 2017)

En France, les lymphomes T angioimmunoblastiques représentent la catégorie la plus fréquente. Cette prédominance est nettement moindre aux États-Unis et en Asie (16% et 17.9% respectivement) (Vose et al. 2008) possiblement en partie en raison de disparités diagnostiques. De même, les NKTCL et les ATLL sont nettement plus fréquents dans les pays d'Asie où ils représentent respectivement 22 et 25% des cas (Vose et al. 2008). À l'heure actuelle, plus d'un quart des lymphomes T ne peuvent être rattachés à une entité particulière avec les méthodes diagnostiques disponibles et sont classés comme lymphome T sans spécification (Peripheral T-cell Lymphome Not Otherwise Specified ou PTCL-NOS). Par ailleurs, la catégorie des lymphomes T « autres » de la figure 3 représente seulement 6.4%

mais regroupe plus de 5 sous-entités qui sont individuellement extrêmement rares en France (ATLL, MEITL, HSTL...). Les formes leucémiques d'hémopathies T matures (LGL, T-PLL ...) ne sont pas représentées dans ce diagramme, mais elles sont tout aussi rares.

### 1.3 *Pronostic*

Le pronostic des hémopathies T matures est défavorable avec une survie globale médiane à 5 ans inférieure à 50% (Vose et al. 2008; Sibon et al. 2012; Advani et al. 2021). Certaines entités ont un pronostic très sombre malgré l'âge relativement jeune des patients atteints : les lymphomes hépatospléniques, les ATLL et lymphomes associés aux entéropathies (survie médiane à 5 ans 7%, 14% et 20% respectivement). Il existe cependant quelques entités au pronostic favorable : les ALCL positifs pour ALK (survie globale à 5 ans de 81% (Sibon et al. 2019)), les ALCL primitifs cutanés (pcALCL 90%), et les panniculite sous-cutanées (SPTCL). Les lymphomes cutanés ont des évolutions différentes en fonction de leur stade d'avancement, en effet la survie médiane des patients atteints de mycosis fongoïdes est de plus de 10 ans (Y. H. Kim et al. 2003) alors que les syndromes de Sézary et les stades avancés de mycosis fongoïdes ont une survie à 5 ans de 77% (Scarlsbrick et al. 2015). Les LGL font aussi exception et se présentent sous la forme d'une maladie indolente avec des survies médianes supérieures à 10 ans.

## 2 Présentation des principales entités

Nous allons présenter brièvement les entités de MTCN les plus fréquentes qui seront évoquées dans la suite de ce travail.

### 2.1 *Lymphomes et leucémies aiguës lymphoblastiques T (T-cell acute lymphoblastic leukemia ou T-ALL)*

Les lymphomes et leucémies aiguës lymphoblastiques T sont des proliférations de cellules précurseurs des lymphocytes T. Ils ne font pas partie des hémopathies T/NK matures mais ont été inclus dans notre étude en raison des rares cas de confusion diagnostique avec les PTCL-NOS. Ils touchent principalement les enfants et jeunes adultes. Ils sont appelés leucémies dans leur présentation majoritairement sanguine et médullaire (> 25% de cellules immatures dans la moelle) ou lymphomes dans les présentations essentiellement tumorales

(la localisation majoritaire étant thymique). En cytologie, les cellules sont de taille moyenne avec un cytoplasme basophile peu abondant au noyau rond et convolué. Les T-ALL sont caractérisées en cytométrie de flux et immunohistochimie par la présence des marqueurs d'immaturation T (CD1a, CD34, CD99, CD117 et TdT) et peuvent être doublement négatives, doublement positives, ou simplement positives pour l'un ou l'autre des marqueurs CD4 et CD8 (Tembhare et al. 2021). Un nouveau sous-groupe de leucémies aiguës a été récemment identifié comme dérivant des précurseurs thymiques précoces (early T-precursor type ALL ou ETP-ALL). Elles sont définies par un engagement dans le lignage T et l'expression concomitante de marqueurs myéloïdes (Coustan-Smith et al. 2009), et sont reconnues par l'absence de CD1a et CD8 et l'expression faible ou absente de CD5.

Un des mécanismes principaux d'oncogenèse des T-ALL est l'activation de la voie Notch via des mutations activatrices de *NOTCH1* ou inactivatrices de son régulateur *FBXW7* (Weng et al. 2004). Il existe par ailleurs de très fréquents réarrangements du locus TCR avec celui de facteurs de transcription (*TAL1*, *TAL2*, *TLX1*, *TLX3* et les gènes à homeobox *HOXA*) ou les gènes *MYC* et *MYB* (Liu et al. 2017). Des délétions des gènes suppresseurs de tumeurs *CDKN2A* et *CDKN2B* sont aussi fréquentes, ainsi que de nombreuses altérations génétiques dans les voies PI3K, /AKT, JAK/STAT, RAS et MAPK.

On notera qu'il existe à ce jour une incertitude sur l'existence de leucémies aiguës NK. Car il n'existe pas de marqueurs spécifiques des précurseurs NK et les caractéristiques NK (absence de réarrangement T, expression du CD56) ne sont pas rares dans les leucémies les plus immatures du nouveau sous-groupe ETP-ALL.

## 2.2 *Leucémie prolymphocytaire T (T-cell prolymphocytic leukemia ou T-PLL)*

La leucémie pro lymphocytaire T est une prolifération clonale de lymphocytes T  $\alpha\beta$  de localisation essentiellement sanguine et médullaire. Dénommée prolymphocytaire en raison de son aspect cytologique particulier (Catovsky et al. 1973; Galton et al. 1974), elle est constituée de cellules de taille moyenne à grande, à cytoplasme basophile et irrégulier caractérisé par des protrusions et un noyau vésiculé à chromatine plutôt condensée. Les patients se présentent souvent avec une lymphocytose très élevée, supérieure à 100 G/L. L'infiltration médullaire induit fréquemment une anémie et une thrombocytopénie et il existe dans 15% des cas une atteinte de la peau associée (Matutes et al. 1991). Les cellules

sont majoritairement CD4+ mais peuvent être double positifs CD4+CD8+ dans (25-50% des cas) ou CD8+ seul dans 15% des cas. Elles expriment toujours un TCR $\alpha\beta$ . Cette maladie est associée dans la plupart des cas à une anomalie caryotypique t(14;14)(q11.2;q32) ou inv(14)(q11.2q32) juxtaposant le gène *TCL1A* au locus du TCR. L'expression constitutive de *TCL1A* active la voie du TCR en l'absence de ligand et favorise la prolifération cellulaire. Dans 5 % des cas c'est le gène *MTCP1* sur le chromosome X qui est associé au locus du TCR (t(X;14)(q28;q11.2)) et joue un rôle semblable à *TCL1A*. S'y associent fréquemment des anomalies de la réparation de l'ADN (mutation ou délétion du gène *ATM*, de *TP53* et plus rarement de *CHEK2*). L'ataxie télangiectasie, maladie génétique liée à une mutation inactivatrice d'*ATM*, est d'ailleurs une condition prédisposante à la T-PLL (Taylor et al. 1996). La troisième voie atteinte par des mutations activatrices est la voie IL2R – JAK 1/3 - STAT5B favorisant la survie des cellules de T-PLL (Bellanger et al. 2014; Kiel et al. 2014). Cette maladie extrêmement agressive, parfois précédée d'une phase indolente de quelques mois voire années, est de pronostic très défavorable malgré le traitement par anticorps anti-CD52 qui permet une rémission initiale chez une majorité des patients.

### 2.3 *Leucémie à grand lymphocytes granuleux T et NK (NK and T-cell large granular lymphocytic leukemia ou NK-LGL and T-LGL)*

Les leucémies à grands lymphocytes granuleux sont une prolifération essentiellement sanguine et médullaire de cellules de grande taille dont le cytoplasme est rempli de granules cytotoxiques. Les LGL sont divisés en deux classes selon leur origine T ou NK, mais ces deux entités partagent de nombreux aspects cliniques, cytologiques et génétiques.

Les T-LGL sont issus de lymphocytes T le plus fréquemment TCR $\alpha\beta$ , rarement TCR $\gamma\delta$  (10%) et majoritairement CD8+. Les NK-LGL sont issus de lymphocytes NK et n'ont pas de réarrangement du TCR et sont CD16+ sCD3- CD56+/. Qu'elles soient T- ou NK, les-LGL, sont caractérisées par une forte expression du CD94.

Plusieurs mécanismes induisent une survie et une cytotoxicité anormale des cellules LGL. Ces cellules ont la particularité d'être résistantes à l'apoptose induite par Fas (T. Lamy et al. 1998) même si les mutations inhibitrices de *FAS* sont rares (Cheon et al. 2022). Elles portent des mutations activatrices de *STAT3*, plus rarement *STAT5B* qui promeuvent leur survie (Koskela

et al. 2012) et d'autres mutations de gènes contrôlant l'épigénétique (*KMT2D* et *TET2*) (Cheon et al. 2022).

Dans cette pathologie, ce n'est pas la taille du clone tumoral qui induit les symptômes mais les manifestations immunologiques associées. En effet, plusieurs mécanismes supposés (augmentation du relargage de FasL soluble, TNF $\alpha$  et de l'IFN $\gamma$  par les cellules leucémiques, défaut de granulopoïèse et cytotoxicité directe des LGL sur les neutrophiles (Thierry Lamy, Moignet, and Loughran 2017)) induisent une neutropénie et/ou une anémie. Hormis les cytopénies, on observe d'autres manifestations immunologiques associées : polyarthrite rhumatoïde, lupus ou syndrome de Gougerot Sjögren. Les patients sont asymptomatiques dans 15-30% des cas. Les traitements principalement utilisés sont essentiellement à visée immunosuppressive et non antitumorale (cyclophosphamide et ciclosporine notamment). Contrairement aux deux leucémies décrites précédemment, cette pathologie est indolente et les rares décès sont principalement dus à des infections en lien avec la neutropénie.

#### *2.4 Leucémie lymphome T de l'adulte lié à l'HTLV1 (Adult T-cell leukemia/lymphoma ou ATLL)*

Cette entité a la particularité d'être causée par le virus HTLV1 (human T-cell leukemia virus type 1), un virus endémique dans certaines zones du monde (Japon, Caraïbes, Afrique subsaharienne, Moyen-Orient, Amérique du Sud et Nouvelle-Guinée) (Gessain and Cassar 2012; Catovsky et al. 1982; Uchiyama et al. 1977). La distribution géographique des ATLL suit cette répartition géographique. Les patients atteints d'ATLL ont le plus souvent été infectés par le virus par contamination du lait maternel (Ando et al. 1987; Bartholomew et al. 1998). Les lymphomes se développent chez 3-5% des personnes infectées par le virus plusieurs dizaines d'années après l'infection initiale. Les présentations cliniques sont variables et se répartissent en quatre sous types différents : aiguës, lymphomes, 'smouldering' et chroniques (Uchiyama et al. 1977; Kawano et al. 1984; Yamaguchi et al. 1983; 1986). La présentation clinique est essentiellement leucémique pour les formes aiguës, et ganglionnaire pour les formes lymphomateuses. Il existe des atteintes cutanées fréquentes similaires à celles retrouvées dans les syndromes de Sézary et les mycosis fongoïdes. Les cellules sont typiquement décrites comme ayant une morphologie pléomorphe et des noyaux lobulés prenant parfois la forme typique d'un trèfle (Jaffe et al. 1984). Certaines



cellules peuvent tendre vers une morphologie anaplasique tandis que d'autres restent cytologiquement proches de cellules normales. Phénotypiquement, ces cellules sont CD2+CD3+CD5+, ont perdu l'expression du CD7 et sont CD4+. Elles expriment fréquemment le marqueur CD25 (Lando et al. 1983).

La tumorigénèse a été initialement attribuée à l'activation de gène par la protéine virale Tax médiée par NFκB (M. Yoshida et al. 1989; Wano et al. 1988; Ballard et al. 1988; Ruben et al. 1988). Cependant des études plus récentes soutiennent l'hypothèse d'un rôle prédominant de l'ARN et de la protéine virale HBZ (Ma, Yasunaga, and Matsuoka 2016). S'associent à ces deux facteurs de transcription, des mutations activatrices de la voie du TCR et de la voie NFκB, des fusions des molécules de costimulation (CTLA4 ::CD28 ou ICOS ::CD28) ainsi que des mutations et variants structuraux favorisant l'évasion immunitaire (HLA-A, HLA-B et PD-L1) (Kataoka et al. 2015; Kogure et al. 2022; Marçais et al. 2021).

## 2.5 *Lymphomes T et NK EBV positifs (EBV positive NK and T cell lymphomas)*

### a. Lymphome T/NK extranodal (Extranodal NK/T-cell lymphoma, ENKTL)

Les lymphomes T/NK extranodaux sont retrouvés chez des patients jeunes, d'origine est-asiatique, ou d'Amérique latine (Laurini et al. 2012; Au et al. 2005). Ils sont liés au virus Epstein-Barr (EBV). Ils peuvent être divisés en deux sous-types selon leur localisation dans la cavité nasale ou non. Hors cavité nasale, ils touchent fréquemment la peau, l'estomac, l'intestin, ou les testicules.

Ce type de lymphome envahit fréquemment les sites muqueux et cutanés, conduisant souvent à des ulcérations. L'histologie est caractérisée par un microenvironnement hautement inflammatoire, de la nécrose et une angio-invasion importante. En immunohistochimie, en fonction de leur phénotype NK (60%) ou T (40% des cas), on retrouve un marquage positif pour le CD56 ou le CD3 de surface et dans les deux cas un marquage cytotoxique (TIA1, granzyme B, perforine). Le CD3ε intracytoplasmique peut être retrouvé dans les cellules NK. Les autres marqueurs T et NK sont généralement négatifs (CD4, CD16, CD57) (Pongpruttipan et al. 2012). Le CD8 peut être positif dans les cas d'origine T qui présentent un réarrangement du TCR.

Des sous-groupes ont récemment été décrits en fonction des profils d'expression génique, mutationnel et du profil de latence d'EBV (Xiong et al. 2020)

- Le sous-groupe TSIM présente une prédominance de mutations activatrices et/ou d'amplification des gènes de la voie JAK-STAT (STAT3, STAT5B, et JAK3) et des régulateurs épigénétiques. L'origine prédominante est NK et l'EBV a un profil de latence de type 2.
- Le sous-groupe MB présente des mutations de *MGA*, des pertes d'hétérozygotie de *BRDT* et une surexpression de *MYC*. L'origine peut être T ou NK et le profil de latence EBV est de type 1.
- Le sous-groupe HEA présente essentiellement des mutations des histones acétylases et déacétylases (mutations de *HDAC9*, *EP300* et *ARID1A*) et une plus forte expression des marqueurs T.

Malgré des traitements spécifiques et notamment l'adjonction d'asparaginase à la chimiothérapie, leur pronostic reste sombre avec une survie à 5 ans de 52% (Oishi et al. 2023). L'intérêt des checkpoints inhibiteurs (anti-PD1 notamment) reste à confirmer.

- b. Lymphomes T/NK nodaux liés à l'EBV (Primary nodal Epstein-Barr virus-positive T-cell/NK-cell lymphoma)

Le groupe des PTCL EBV nodaux est une nouvelle entité depuis la dernière classification OMS 2022. Comme ces derniers, ils sont liés au virus EBV et sont constitués de cellules tumorales au phénotype cytotoxique. En revanche, ils sont de localisation essentiellement ganglionnaire, et présentent une invasion limitée des sites extranodaux. Ils sont plus fréquemment d'origine T (85% des cas (Kato et al. 2015)) et positifs pour le CD8. Leur profil mutationnel est légèrement différent de celui des ENKTL avec une prédominance de mutations de *TET2* (64%), *PIK3CD* (33%), mais un taux similaire de mutations de *DDX3X*, *STAT3* et *JAK3* (Wai et al. 2022), beaucoup plus rarement des mutations ou des délétions de *TP53* comparativement aux ENKTL. La rareté et la définition récente de cette entité limitent nos connaissances, mais son pronostic semble très défavorable (Wai et al. 2022).

## 2.6 *Lymphomes intestinaux*

- a. Lymphome T associé à une entéropathie (Enteropathy-associated T-cell lymphoma, EATL)

Cette pathologie survient dans un contexte de maladie cœliaque même si celle-ci peut ne pas avoir été diagnostiquée antérieurement dans 15% des cas (Moffitt et al. 2017). Les EATL sont fréquemment précédés d'une phase de lymphome de bas grade appelée « sprue réfractaire » qui est caractérisée par la prolifération de lymphocytes intraépithéliaux clonaux de phénotype aberrant NKP46+ ne présentant pas d'atypie cytologique (Cheminant et al. 2019).

Les lésions lymphomateuses atteignent principalement le jéjunum ou d'autres régions de l'intestin grêle et plus rarement l'estomac et le colon. Il existe des lésions extra-intestinales dans 1/3 des cas (Malamut et al. 2013). L'atteinte anatomopathologique est caractérisée par l'accumulation de lymphocytes atypiques de grande taille ou de morphologie anaplasique dans la *lamina propria* et la sous-muqueuse, d'une importante angio-invasion et angio-destruction. Les manifestations de maladie cœliaque habituelles sont visibles en périphérie de la tumeur (atrophie villositaire, hyperplasie des cryptes et augmentation de la lymphocytose intraépithéliale). Les cellules sont généralement CD3+ mais CD4-CD8- et expriment les protéines cytotoxiques et le CD103. La plupart sont TCR négatives. Le CD30 est souvent exprimé sur les cellules de morphologie anaplasique.

Les mutations 'drivers' principales atteignent l'axe JAK1-STAT3, les régulateurs épigénétiques (*TET2*, *KMT2D*), l'ARN hélicase *DDX3X*, la voie NF-KB (*TNFAIP3/A20*, *TNIP3*), les voies de dommages/réparation de l'ADN (*POT1*, *TP53*) ou de l'échappement immunitaire (*CD58*, *FAS*, *B2M*) (Cording et al. 2022; Roberti et al. 2016; Moffitt et al. 2017). On retrouve un réarrangement monoclonal du TCR dans la plupart des cas, qui n'est pas toujours productif (Ettersperger et al. 2016).

En raison de l'état de dénutrition des patients et de l'incidence importante des perforations avant ou sous chimiothérapie, cette entité présente un mauvais pronostic qui pourrait s'améliorer grâce à l'adjonction de brentuximab-vedotin, un anticorps anti-CD30 couplé à de la chimiothérapie (Sibon et al. 2021).

- b. Lymphome T intestinal épithéliotrope monomorphe (Monomorphic epitheliotropic intestinal T cell lymphoma, MEITL)

Cette entité a récemment été distinguée des EATL précédemment décrits. Antérieurement appelée EATL de type 2, l'absence d'association avec la maladie cœliaque et son phénotype cellulaire particulier ont conduit à la renommer lymphome T intestinal épithéliotrope monomorphe. En effet, elle est constituée d'une prolifération de cellules monomorphes, de taille petite à moyenne au noyau rond et à la chromatine dispersée associées à un cytoplasme modérément abondant. Contrairement aux EATL, la nécrose et l'infiltrat inflammatoire sont peu communs. Comme les EATL, les lymphocytes tumoraux résident initialement dans l'épithélium intestinal et expriment fréquemment le CD103, mais *a contrario* la majorité expriment un TCR $\gamma\delta$  (50%) ou un TCR $\alpha\beta$  (32%) (Veloza et al. 2023). Les lymphocytes sont la plupart du temps CD2+, CD3+, CD4-, CD5-, CD7+, CD8+, CD56+ (Veloza et al. 2023) et n'expriment pas toujours le granzyme B ou la perforine. Ils sont caractérisés, contrairement aux EATL, par des mutations très fréquentes de *SETD2* et une plus grande fréquence de mutation de *STAT5B* et *JAK3* par rapport à *STAT3* et *JAK1* (Moffitt et al. 2017).

### 2.7 Lymphomes hépatospléniques (*Hepatosplenic T-cell lymphoma, HSTL*)

Les lymphomes hépatospléniques atteignent des patients jeunes (Yabe, Miranda, and Medeiros 2018). Il existe une prédominance masculine et une fraction des cas (20%) survient dans un contexte d'immunosuppression (azathioprine, infliximab) et/ou de maladie auto-immune (maladie de Crohn, rectocolite hémorragique ou polyarthrite rhumatoïde) suggérant une implication de la stimulation antigénique chronique ou d'un défaut d'immunosurveillance dans la lymphomagénèse (Yabe, Medeiros, Daneshbod, et al. 2017). Les lymphocytes T, essentiellement TCR  $\gamma\delta$  (80%) et plus  $\alpha\beta$  (20%), infiltrent la rate, le foie et la moelle osseuse provoquant un tableau clinique bruyant marqué par une hépatosplénomégalie, des symptômes B et des cytopénies. Les cellules tumorales sont retrouvées principalement au niveau des sinusoides du foie et de la rate. La morphologie de ces cellules est particulière puisqu'elles présentent des contours irréguliers, une chromatine dense et un cytoplasme pâle avec parfois une morphologie blastique. Les formes TCR $\gamma\delta$  expriment principalement une chaîne V $\delta$ 1 (Przybylski et al. 2000) conformément au sous-type majoritaire observé dans une rate normale. Le phénotype des cellules est la plupart du

temps CD2+, CD3+, CD4-, CD5-, CD7+/-, CD8-, CD16+/-, CD38+ et CD56+. Les expressions du CD8, du CD5, et du CD57 ont plus rarement été rapportées. Le profil des marqueurs cytotoxiques est immature : TiA1+ mais granzyme B- et perforine- (Vega, Medeiros, and Gaulard 2007) et on retrouve souvent une positivité de granzyme M. Il existe aussi dans la moitié des cas des anomalies du chromosome 7q (isochromosome 7q, chromosome 7q en anneau, amplifications isolées du chromosome 7q) à la suite d'une erreur de réarrangement du locus du TCRG (7p14) ou TCRB (7q32). Ces anomalies pourraient être à l'origine de la surexpression des gènes *PTPN12*, *CHN2*, *ABCB1* et *PPP1R9A* impliqués dans la tumorigenèse et la résistance aux traitements (Ferreiro et al. 2014; Travert et al. 2012). Les trisomies 8q ou amplifications du chromosome 8 sont la deuxième altération la plus fréquente (McKinney et al. 2017). Les principales mutations atteignent les gènes de la voie JAK-/STAT, tels que *STAT5B* (30-40%) et *STAT3* (9%), et des modificateurs épigénétiques, *SETD2* (25%), *INO80* (21%), *ARID1B* (21%) (McKinney et al. 2017). Les HSTL présentant un récepteur V $\delta$ 1 sont capable de reconnaître la molécule CD1d chargée avec du sulfatide, qui est la cible habituelle des lymphocytes NKT, ce qui amène l'hypothèse du rôle d'une stimulation chronique du TCR dans la lymphomagénèse de cette entité (Bachy et al. 2016).

Malgré le jeune âge des patients atteint d'HSTL, la survie médiane est de 11 mois et le seul traitement qui semble conférer un meilleur pronostic est l'allogreffe de moelle osseuse (McKinney et al. 2017; Yabe et al. 2016).

## 2.8 *Mycosis fongoïdes et syndromes de Sezary*

Le lymphome T cutané (CTCL) le plus fréquent est le mycosis fongoïde (50% des cas). Il est caractérisé par l'apparition de plaques et de tumeurs cutanées constituées de cellules tumorales infiltrant l'épithélium. Les cellules sont initialement alignées le long de l'épithélium basal puis forment des amas, colonisent les couches supérieures du derme formant des lésions macroscopiquement plus épaisses.

Le syndrome de Sézary est une deuxième entité de lymphome T cutané, nettement moins fréquente (2-3% des lymphomes T cutanés). Il se présente sous la forme d'une érythrodermie et d'une circulation sanguine de cellules tumorales. Dans certains stades avancés, il existe des atteintes ganglionnaires ou d'autres organes.

Les mycosis fongoïdes peuvent aussi évoluer vers une dissémination sanguine et remplir alors les critères diagnostics de syndrome de Sézary ou vers une transformation à grandes cellules, appelée alors mycosis fongoïde transformé.

L'immunophénotype de ces deux pathologies est similaire: CD2+, CD3+, TCR beta+, CD5+, CD4+, CD8-, CD7-/dim et CCR4+. Il existe souvent une expression de PD1 et une perte de CD26, caractéristique essentiellement retrouvée dans les syndromes de Sézary. De rares formes CD8+ ou TCR  $\gamma\delta$ + ont été décrites.

Ces deux pathologies dérivent de lymphocytes T CD4+ mémoires CCR7+ expriment l'antigène lymphoïde cutané (CLA), un récepteur de « homing » à la peau. Il est noté une importante signature mutationnelle liée aux UV suggérant un rôle de l'exposition solaire dans cette pathologie (Jones et al. 2021). Les anomalies génétiques retrouvées dans ces deux entités sont similaires avec la présence notamment de mutations activatrices de la voie du TCR et de NF-KB (*PLCG1*, *CARD11*, *CD28*, *TNFRSF1B*, *RLTPR*) et de modificateurs épigénétique (*ARID1A*, *DNMT3A*, *TET2*, *TRRAP*). Les syndromes de Sézary présentent plus fréquemment des délétions du chromosome 17 et des mutations de *TP53* et de *FAS* (Chang et al. 2018; Choi et al. 2015; da Silva Almeida et al. 2015; Park et al. 2017a).

## 2.9 Lymphomes T helper folliculaires (*Follicular helper T-cell lymphoma*)

Les lymphomes T folliculaires (TFHL) ont été regroupés depuis la classification OMS 2016 grâce à plusieurs travaux montrant un même profil mutationnel similaire et un profil d'expression génique suggérant une origine commune à partir d'un lymphocyte T folliculaire helper (Tfh) (de Leval et al. 2007; Dobay et al. 2017; Lemonnier et al. 2012; Palomero et al. 2014; Rodríguez-Pinilla et al. 2008; Zhan et al. 2011).

Les cellules sont des lymphocytes T TCR $\alpha\beta$ + et CD4+. Le diagnostic positif requiert en outre la positivité d'au moins 2 marqueurs Tfh parmi PD1, ICOS, CXCL13, CD10, BCL6, CXCR5, une forte expression de PD1 étant requise.

La pathologie est supposée liée à des mutations des régulateurs épigénétiques TET2 et DNMT3A au niveau d'un progéniteur commun et une activation de la voie du TCR (Lemonnier et al. 2012; Odejide et al. 2014; Quivoron et al. 2011). La mutation la plus fréquente est la mutation RhoA G17V (70%) spécifique des lymphomes TFH. La forme mutée

de la protéine RhoA induit, via son interaction avec Vav1, une activation de la voie du TCR (Palomero et al. 2014; Sakata-Yanagimoto et al. 2014). On observe souvent d'autres mutations de cette voie (*VAV1*, *PLCG1*, *FYN*) et des mutations ou fusions des molécules de costimulation *ICOS*, *CTLA4*, *CD28* (Palomero et al. 2014; Vallois et al. 2016). Les mutations de *TET2* et *DNMT3A*, survenant souvent au niveau des précurseurs hématopoïétiques sont aussi retrouvées dans les cellules du microenvironnement (lymphocytes B, macrophages) et favorisent probablement la croissance tumorale (Fujisawa et al. 2022).

La présentation clinique de ces lymphomes est essentiellement nodale avec une atteinte fréquente de la rate, du foie et de la peau, et plus rarement des autres organes périphériques. Il existe des manifestations immunologiques cliniques et/ou biologiques associées fréquentes (anémie hémolytique autoimmune, anticorps anti-muscle lisse, facteur rhumatoïde ou cryoglobuline), supposées en lien avec une stimulation non contrôlée des lymphocytes B par les cellules Tfh tumorales.

L'analyse histologique est le principal critère qui détermine les trois sous entités de lymphomes TFHL (TFHL) :

- Les lymphomes T angioimmunoblastiques (Follicular helper T-cell lymphoma, angioimmunoblastic type, AITL) présentent une structure typique avec un effacement de l'architecture normale, remplacée par un infiltrat inflammatoire polymorphe, une prolifération des veinules à endothélium épais, une expansion du réseau de cellules folliculaires dendritiques, la présence d'immunoblastes B positifs ou non pour l'EBV, et parfois des histiocytes épithélioïdes (Dogan, Attygalle, and Kyriakou 2003).
- Les lymphomes TFH folliculaires (Follicular helper T-cell lymphoma, follicular type, TFHL-F) prennent la forme d'une prolifération nodulaire de cellules lymphoïdes de taille moyenne, monomorphes, et ne présentent pas de prolifération endothéliale, ni d'expansion des cellules folliculaires dendritiques.
- Les lymphomes TFH non spécifiés (Follicular helper T-cell lymphoma, NOS, TFH-NOS) se caractérisent par l'expression de CD4 et d'au moins deux marqueurs TFH, mais n'ont pas les critères permettant la catégorisation en tant que lymphome T angioimmunoblastique ou lymphome TFH folliculaire.

Il existe des variabilités inter-observateurs sur la catégorisation de ces trois entités et il est probable qu'elles représentent un *continuum* anatomopathologique plutôt que trois entités radicalement distinctes.

### 2.10 Les lymphomes T anaplasiques (*Anaplastic large cell lymphoma, ALCL*)

Les lymphomes T anaplasiques sont définis historiquement par une morphologie commune : des cellules de grande taille, à cytoplasme abondant, pléomorphes, avec des noyaux en forme de fer à cheval ou réniformes. Il a par ailleurs été mis en évidence une forte positivité du marqueur CD30 en immunohistochimie. En revanche, ce groupe rassemble des entités assez hétérogènes par leurs pronostics et leurs caractéristiques moléculaires.

- a. Lymphomes anaplasiques ALK+ (*Anaplastic large cell lymphoma, ALK positive, ALK+ ALCL*)

Les ALCL ALK+ sont caractérisés par la présence d'une translocation créant une protéine de fusion entre le gène *ALK* (*Anaplastic Lymphoma Receptor Tyrosine Kinase*) d'une part et de l'autre *NPM1* (t(2;5)(p23;q35), 84% des cas), *TPM3* (t(1;2)(q25;p23), 13% des cas) ou encore d'autres partenaires plus rares (*ATIC, TFG, CLTC, MSN, TPM4, MYH9, TRAF1, EEF1G, RNF213*) (Lamant et al. 1999; Morris et al. 1994). Cette fusion mène à l'activation constitutive de l'activité kinase de ALK qui induit de nombreuses voies pro-oncogéniques. Notamment l'activation de STAT3 qui favorise l'activité du facteur de transcription C/EBPβ indispensable à la survie de la cellule (Bonzheim et al. 2013).

Ce lymphome est diagnostiqué par une immunohistochimie positive pour la protéine ALK. Le phénotype est par ailleurs caractérisé par la perte de marqueurs T (CD3 et TCR), malgré la présence majoritaire d'un réarrangement alpha beta. On retrouve l'expression de CD2 et CD4 et un phénotype cytotoxique (TiA1+ GZMB+ et PRF+).

- b. Lymphomes anaplasiques ALK- (*Anaplastic large cell lymphoma, ALK negative – ALK- ALCL*)

Ce groupe est nettement plus hétérogène que le précédent (Parrilla Castellar et al. 2014). Deux grandes catégories d'anomalies moléculaires ont été décrites dans les lymphomes anaplasiques :ALK-:



- Les réarrangements *DUSP22::IRF4* t(6;7)(p25.3;23.3) (Feldman et al. 2011),
- Une activation constitutive de la voie STAT3 via des mutations de *JAK1* et *STAT3*, ou des fusions (*NFKB2::ROS1* et *NCOR2::ROS1*, *PABPC4::TYK2*, *NFKB2::TYK2*, *NPM1::TYK2*) (Crescenzo et al. 2015; Velusamy et al. 2014).

Ces deux catégories d'anomalies semblent mutuellement exclusives (Luchtel et al. 2018). Les lymphomes avec un réarrangement *DUSP22* sont une entité provisoire dans la nouvelle classification ICC en raison d'un phénotype légèrement différent, une morphologie moins anaplasique et plus souvent en 'doughnut', une faible cytotoxicité et une expression plus fréquente du CD3 (Sibon et al. 2023; Feldman et al. 2011). Leur profil transcriptomique est aussi très différent des autres lymphomes anaplasiques ALK- (Luchtel et al. 2018). Les ALCL *DUSP22* réarrangés présentent dans la majorité des études une meilleure survie globale que les autres ALCL ALK- (Parrilla Castellar et al. 2014; Pedersen et al. 2017; Onaindia et al. 2019; Hapgood et al. 2019; Sibon et al. 2023). Des translocations du gène *TP63* ont aussi été décrites dans les ALCL mais celles-ci ne semblent pas définir un groupe à part et sont retrouvées dans d'autres entités de PTCL.

### 2.11 Lymphome T périphérique sans spécification (*Peripheral T-cell lymphoma, NOS, PTCL-NOS*)

Le diagnostic de lymphome T périphérique NOS est un diagnostic par défaut et représente par conséquent le groupe le plus hétérogène de cette classification. Le phénotype majoritaire est CD4, ou plus rarement CD8, voire double positif, essentiellement TCR $\alpha\beta$  mais parfois TCR $\gamma\delta$  ou TCR négatif. Le diagnostic repose essentiellement sur l'élimination de toutes les autres entités.

Les classifications successives identifient de nouvelles entités qui sont exclues progressivement de ce groupe comme les PTCL-TFH rapprochés des AITL en 2016 et les PTCL EBV+ des NKTCL en 2022. Plusieurs études transcriptomiques récentes ont proposé la sous division des PTCL-NOS en deux groupes selon l'expression du facteur de transcription TBX21 ou GATA3 (Iqbal et al. 2014). Cette classification peut être reproduite en immunohistochimie par les marqueurs GATA3, CCR4 vs TBX21 et CXCR3 avec une concordance de 85% (Amador et al. 2019). Les PTCL-GATA3 ont un plus grand nombre d'anomalies caryotypiques et de fréquentes pertes de copie ou mutations des gènes *CDKN2A/B* et *TP53* et de la voie PTEN-

PI3K. Les PTCL-TBX21 sont enrichis en mutations touchant les gènes modificateurs épigénétiques (*TET2*, *TET3*, *TET1*, *DNMT3A*) (Heavican et al. 2019). Le pronostic des PTCL-GATA3 est défavorable par rapport à celui des PTCL-TBX21. Il existe cependant une sous partie du groupe TBX21 avec des marqueurs de cytotoxicité qui serait également de mauvais pronostic (Iqbal et al. 2014).

### 3 Modalités et difficultés diagnostiques

#### 3.1 *Diagnostic positif de lymphome T*

Le diagnostic positif de lymphome T repose sur un faisceau d'arguments et notamment la présence de marqueurs du lignage T : CD3, CD2, CD5, CD7 et les marqueurs CD8 et/ou CD4 sur les cellules tumorales. Chacun de ces marqueurs peut manquer de manière occasionnelle ou parfois systématique dans certaines entités comme la perte du CD7 dans les lymphomes T cutanés et les ATLL ou, et la perte du CD3 de surface dans les lymphomes anaplasiques. La recherche d'un réarrangement clonal du TCR et l'absence de réarrangement du BCR (B cell receptor) est une information importante permettant d'orienter vers une pathologie lymphomateuse T. Cependant, il est à noter qu'il existe une fraction d'adénopathies purement réactionnelles (inflammatoire ou autoimmune) qui présentent un certain degré d'oligoclonalité voire une monoclonalité. Par ailleurs, les lymphomes T angioimmunoblastiques présentent souvent une expansion d'un clone B pouvant orienter à tort vers une pathologie lymphoïde B et fausser le diagnostic dans un nombre non négligeable de cas (Laurent et al. 2017).

Une fois le diagnostic positif de lymphome T posé, préciser la sous-entité d'hémopathie T requiert l'intégration d'informations cliniques, histologiques, de marqueurs immunohistochimiques et viraux, d'anomalies chromosomiques et mutationnelles. Nous allons tenter de décrire quelques éléments d'orientation diagnostique et leurs limites.

#### 3.2 *Éléments d'orientation clinique*

Le contexte de survenue du lymphome peut être un élément d'orientation important. La présence d'une infection HTLV1 oriente en premier lieu vers un ATLL, une maladie cœliaque vers un EATL.

En revanche certains éléments d'orientation sont nettement moins formels. La prise d'un traitement immunosuppresseur est fréquente (20%) dans les lymphomes hépatospléniques mais peut aussi être retrouvé dans d'autres pathologies notamment les PTCL-EBV. Les pathologies auto-immunes sont fréquentes dans les AITL mais aussi dans les LGL.

La localisation tumorale est un élément essentiel du diagnostic :

- Les lymphomes cutanés, sont par définition de localisation cutanée. Cependant les T-PLL ont une phase circulante mais présentent des lésions cutanées dans 15% des cas. Les lymphomes T angioimmunoblastiques sont aussi très fréquemment associés avec une atteinte cutanée et peuvent être difficiles à différencier d'un lymphome cutané primitif de type MF ou syndrome de Sézary avec atteinte ganglionnaire.
- Les atteintes intestinales caractéristiques des MEITL et des EATL peuvent aussi être retrouvées dans les ENKTL. La différence principale sera la positivité du marquage EBV.
- Une localisation hépatosplénique et médullaire est en faveur d'un lymphome hépatosplénique surtout si les cellules expriment un TCR $\gamma\delta$ , mais il existe des leucémies à grand lymphocytes granuleux avec infiltration hépatique dans les stades avancés, certaines étant possiblement TCR $\gamma\delta$  (Yabe, Medeiros, Wang, et al. 2017). La plupart des autres lymphomes dans leurs stades avancés peuvent aussi se présenter avec une infiltration hépatique et/ou splénique.

### 3.3 *Histologie*

L'histologie est aussi un des éléments prédominant du diagnostic, mais elle est rarement spécifique.

Certaines morphologies histologiques sont décrites comme typiques d'une entité :

- La morphologie de cellules en « doughnut » pour les lymphomes anaplasiques DUPS22 réarrangés.
- La cytologie monomorphe et régulière des MEITL par rapport EATL qui ont une morphologie pléomorphe.
- Les cellules en trèfle des ATLL.
- Les noyaux cérébriformes des syndromes de Sézary.

- Les grandes cellules pléomorphes des lymphomes anaplasiques.

Ces formes typiques orientent le diagnostic mais elles sont souvent prises en défaut. Par exemple, certains lymphomes anaplasiques ALK+ sont constitués de petites cellules portant pourtant la translocation typique t(2;5)(p23;q35)(Kinney et al. 1993). De même, il existe 10% d'ALCL ALK+ à la morphologie lymphohistiocytaire (Falini et al. 1998). Lors des rechutes successives ces patients présentent souvent une forme anaplasique à grandes cellules, ce qui suggère que le phénotype cellulaire n'est pas présent dès le début de la maladie.

La composition du microenvironnement tumoral est aussi déterminante, notamment dans le diagnostic des lymphomes T angioimmunoblastiques. Depuis la mise en évidence de leur point commun avec les lymphomes TFH, il semble que des aspects histologiques extrêmement différents recourent des proliférations tumorales issues de mécanismes moléculaires similaires (de Leval et al. 2007). Par ailleurs, il n'est pas rare qu'un lymphome TFH ne rechute pas sous la même forme alors même que le même clone tumoral est identifié (Alaggio et al. 2022; Huang et al. 2009).

Il est donc très difficile d'attribuer sur la morphologie seule un diagnostic précis de lymphome T et des techniques additionnelles, au premier rang desquelles les marqueurs immunohistochimiques, sont indispensables.

### 3.4 *Marqueurs protéiques en immunohistochimie et cytométrie en flux*

La positivité des marqueurs de lignage T (CD3, CD2, CD5, CD7, CD4, CD8) et ou NK (CD56) est essentielle au diagnostic de lymphome NK ou T. On remarque cependant qu'il est très fréquent qu'un ou plusieurs de ces marqueurs soient perdus. Ces trous phénotypiques aident à affirmer le caractère anormal et donc probablement tumoral de la prolifération T par rapport à un contexte réactionnel. Ces trous phénotypiques peuvent aussi être des éléments d'orientation diagnostique quand ils sont caractéristiques de certaines sous-entités (absence d'expression du CD7 dans les syndromes de Sézary, ATLL et MF ; perte de multiples marqueurs T CD2, CD3, CD5, CD7 dans les lymphomes anaplasiques).

C'est la combinaison de plusieurs marqueurs immunohistochimiques qui est centrale dans la définition des différentes entités de lymphomes T. Les phénotypes observés les plus fréquemment sont résumés dans le tableau 3.

Il est notable qu'à l'exception du marquage ALK dans les lymphomes anaplasiques ALK+ (Pittaluga et al. 1997), aucun marqueur immunohistochimique n'est spécifique d'une entité. L'expression du CD30, systématiquement retrouvée dans les lymphomes anaplasiques, est par exemple retrouvée également dans 43% des NKTCL (Li et al. 2013), à la surface de certaines cellules de la majorité des mycosis fongoides surtout dans leurs formes transformées (Diamandidou et al. 1998; Edinger et al. 2009), dans 70-80% des lymphomes associés à une entéropathie (Bossard et al. 2014; Malamut et al. 2013), et enfin près de 50% des AITL (Bossard et al. 2014). Il est alors important de considérer l'intensité du marquage qui peut orienter vers le diagnostic de lymphome anaplasique, et le nombre de cellules positives. En effet dans les ALCL, l'anticorps anti-CD30 marque la très grande majorité des cellules, mais on observe aussi un marquage > 75% dans 13% des PTCL NOS, 15% des ENKTL ou 43% des EATL (Bossard et al. 2014).

Aucun marqueur n'étant spécifique à lui seul et certains pouvant manquer dans chaque cas, une combinaison de marquages est donc essentielle au diagnostic précis. Ainsi, les lymphomes TFH sont caractérisés par la présence d'au moins 2 marqueurs parmi les 5 marqueurs suivants : PD1, CXCL13, ICOS, BCL6 et CD10. Un marquage fort de PD1 est requis dans les lymphomes TFH-NOS (Alaggio et al. 2022). Cette combinaison est nécessaire car chaque marqueur est individuellement très peu spécifique. En effet, les NKTCL expriment très fréquemment CXCL13 (59% des formes d'origine NK) et PD1 (40% des formes d'origine T) mais rarement les deux (Pongpruttipan et al. 2012). BCL6 a été décrit initialement comme un marqueur restreint aux lymphomes anaplasiques CD30+ car il est positif dans 45% des cas (Carbone et al. 1997) juste avant sa description dans les lymphomes angioimmunoblastiques (Ree et al. 1999). Le CD10 est positif dans les leucémies aiguës lymphoblastiques T. Les lymphomes/leucémies de l'adulte liés à l'HTLV1 (ATLL) sont aussi fréquemment positifs pour les marqueurs PD1 (34%), et ICOS (17.1%), mais plus rarement BCL6 (5%). Il est rapporté que 10% des cas sont positifs pour deux marqueurs TFH parmi ICOS, BCL6 et PD1 (Tamaki et al. 2023) et un des cas décrits dans cette étude était triple positif avec la positivité de CXCL13. Hormis la sérologie HTLV1, il n'y a alors aucun marqueur permettant de différencier formellement ces cas d'un PTCL-TFH nodal.

Néoplasie T	Phénotype predominant	Exceptions / particularité	référence
T-cell large granular lymphocytic Leukemia	CD3+, CD2+, CD5-, CD7+, CD8+, TCRαβ++, CD16+, CD57+, CD56-, CD45RA+, CD62L-, TiA-1+, PRF+, GZMB+, GZMM+	Certains cas TCR gamma (<10%), ou CD4. Expression occasionnelle de CD56+ ou CD26+.	(Bourgault-Rouxel et al. 2008; Lima et al. 2003; Yabe, Medeiros, Wang, et al. 2017)
NK-cell large granular lymphocytic Leukemia	sCD3-, cCD3e±, CD2-, CD7-, CD8±, CD16+, CD56-, CD57+ , TiA1+, GZMB+	Expression occasionnelle de CD7, CD2, CD56dim	(F. Cao et al. 2015)
Extranodal NK/T-cell lymphoma, nasal type	- NK cell origin : sCD3-, cCD3e+, CD2+, CD5-, CD56+, CD16-, CD57-, TiA1+, GZMB+, PRF±, EBER+ - T cell origine : sCD3±, cCD3e+, CD2+, CD5±, CD8±, TCRαβ+ ou TCRγδ+, CD56±, CD16-, CD57-, TiA1+, GZMB+, PRF±, EBER+,	Négativité du CD56 dans 67% des cas au TCR réarrangé, et occasionnellement CD7+ (40%), CD30+(43%), CD8+(33%), CD4+(29%).	(Chan, Tsang, and Ng 1996; Jaffe et al. 1996; Li et al. 2013; Pongpruttipan et al. 2012)
T-cell prolymphocytic leukemia	CD3+, CD2+, CD5+, CD4+ ou CD4+/CD8+, CD7+, TCRαβ++, TCL1+, CD34-, CD1a-, T dT-, CD52+	Trous phénotypiques possibles CD2, CD3, CD5, CD7. Certains cas sont double positifs CD4/CD8 (25-40%) ou plus rarement CD8+ (15%)	(Herling et al. 2008; Matutes et al. 1991)
Adult T-cell leukemia/lymphoma	CD3+, CD2+, CD5+, CD7-, CD4+, CD26-, CD25+, FOXP3±, CD30±, Cytotox-, CCR4+	Possible trous phénotypiques CD2, CD3, CD5. Quelques cas CD7+. Marqueurs d'agressivité CCR7+	(Kagdi et al. 2017; Karube et al. 2004; Takeshita et al. 1995)
Hepatosplenic T-cell lymphoma	CD3+, TCRγδ+, CD4-, CD5-, CD8±, CD56± (70%), CD16± (60%), TIA-1±, GZMM+, GZMB± (40%), PRF-	Rares cas TCRαβ. Possible perte de CD2 et 7.	(Yabe, Miranda, and Medeiros 2018)
Enteropathy-associated T-cell Lymphoma	CD3+, CD7+, CD8-, CD4-, CD5-, CD103+, TiA1+, GZMB+, TCR-, CD30±, NKP46+	Une proportion des cellules exprime CD30 dans la plupart des cas. CD103 peut être perdu. 25% des cas sont CD8+. Certains cas TCR+	(Malamut et al. 2013)
Monomorphic epitheliotropic intestinal T-cell lymphoma	CD3+, CD4-, CD5-, CD7+, CD8+, CD56+, CD103+, TCRγδ+, TiA1+, GZMB±, NKp46+, NKG2D+, CD30-	TCRαβ dans 25% des cas. Occasionnellement CD20+	(S.-Y. Tan et al. 2013; Vellozo et al. 2023)
Mycosis fungoides	CD2+, CD3+, CD4+, CD5+, CD7- , TCRαβ+, PD1+	Dans les stades avancés, perte de CD2, CD5, TCRβ, acquisition du CD30 et perte de CCR4. Rares cas CD8+ ou TCRγδ+ cytotoxiques	(Yaohua Zhang et al. 2012)
<b>Néoplasie T</b>	<b>Phénotype predominant</b>	<b>Exceptions / particularité</b>	<b>Reference</b>

Sézary syndrome	CD2+, CD3+, CD4+, CD5+, CD7-, CD26-, TCRαβ, CCR4+, CCR7+	Pertes fréquentes de CD2, CD5 et CD7	(Cetinözman et al. 2012; Klemke et al. 2015)
Nodal TFH lymphoma	CD3+, CD4+, CD2+, CD5+, CD7±, CD8-, au moins 2/5 parmi [strong PD1, ICOS, BCL6, CXCL13 and CD10]	Le marquage PD1 n'est spécifique que lorsqu'il est fortement positif.	(Basha et al. 2019)
Anaplastic large cell lymphoma, ALK-positive	CD30+, ALK+, EMA+, CD3±(25%), CD2±, CD4±, CD5±, CD43+, TiA1+, GZMB+, PRF+, CD56-, EBV-		(Foss et al. 1996; Pittaluga et al. 1997)
Anaplastic large cell lymphoma, ALK-negative	CD30+, ALK-, CD43+, CD3±, CD2±, CD4± (70%), CD5±, CD43+, TCR-, GZMB±, PRF±, Clusterin+, MUM1+, EMA±, CD56-, EBV-	Rare cas CD8+. 20% TCR αβ+	(Amador and Feldman 2021)
Peripheral T-cell lymphoma, NOS	CD3+, CD4+, CD5±, CD7±, CD8-, TCRαβ++	Ce phénotype est le plus fréquent mais le diagnostic est un diagnostic d'exclusion des autres entités.	

**Tableau 3 - marqueurs immunophénotypiques des principales néoplasies T**

Le dernier cas qui reste le plus problématique dans l'utilisation des marqueurs TFH est celui des lymphomes cutanés. En effet, PD1 est positif dans plus de 50% des cellules chez 13% des mycosis fongoïdes, et 89% des syndromes de Sézary (Cetinözman et al. 2012). Dans ces cas BCL6 et CXCL13 marquent 25 à 50% des cellules PD1+. Ces deux pathologies peuvent présenter des similarités cliniques puisque les AITL sont fréquemment accompagnés de lésions cutanées et il pourra alors être difficile de discriminer un lymphome TFH avec localisation secondaire cutanée ou un lymphome cutané avec localisation ganglionnaire.

Ces marqueurs protéiques sont donc insuffisants et des techniques supplémentaires telles que le caryotype et la biologie moléculaire sont utiles pour compléter le diagnostic.

### 3.5 Anomalies structurales

Hormis les LGL, les néoplasies lymphoïdes T portent très fréquemment une ou plusieurs anomalies chromosomiques. Il existe peu d'anomalies caryotypiques qui soient spécifiques de sous entités de néoplasie T. Les principales sont décrites dans le Tableau 4.

Entité	Translocation	Fréquence	Effet
T-PLL	- t(14 ;14)(q11 ;q32) - inv(14)(q11.2 ;q32) - t(X;14)(q28;q11.2)		Expression constitutive de TCL1A (chr14) ou de MTCP1 (chrX) sous le contrôle du promoteur du TCR, induisant une activation de la voie du TCR
ALCL ALK+	- t(2;5)(p23;q35) - t(1;2)(q25;p23) - Inv(2)(p23q53) - Autres fusions impliquant 2p23	84% 13% 1% 2%	Protéine de fusion impliquant ALK et NPM1 (chr5), TPM3 (chr1), ATIC (chr2) ... induisant une activation constitutive de ALK
ALCL ALK- DUSP22R	- t(6;7)(p25.3;q32.3)	28% of ALCL ALK-	Adjonction de <i>DUSP22</i> ou <i>IRF4</i> dans 45% des cas au locus FRA7H. Diminution de l'expression de <i>DUSP22</i> .
HSTL	- Isochromosome 7q - i(7)(q10) ou ring chromosome 7q r(7) - 7q amplification	47-60%	Altération de l'expression des gènes du chromosome 7q

Tableau 4 - Principales anomalies chromosomiques ayant une valeur diagnostique dans les néoplasies matures T

Les trois premiers groupes d'anomalies caryotypiques sont très spécifiques de leur entité. En revanche, des isochromosomes 7q ont déjà été retrouvés dans des cas de lymphomes NK/T



et des lymphomes anaplasiques ALK- (Feldman et al. 2008). La plupart des autres anomalies chromosomiques retrouvées dans les lymphomes T sont aspécifiques. Cependant, certaines peuvent orienter le diagnostic : par exemple, les gains du chromosome 9q31 sont récurrents dans les EATL et absents des MEITL qui ont plus fréquemment des gains de *MYC*, de 1q et 5q (Moffitt et al. 2017).

### 3.6 Séquençage et variants nucléotidiques

Les méthodes de biologie moléculaire et de recherche de variants nucléotidiques sont devenues ces dernières années des aides incontournables au diagnostic en hématologie et les lymphomes T ne font pas exception à cette règle.

Les voies de signalisation et processus cellulaires les plus fréquemment atteints dans les lymphomes T sont :

- La voie du TCR et de NFκB (*RHOA, CARD11, PLCG1, CD28, VAV1, PRKCB, TNFRSF1B, IRF4*),
- La régulation épigénétique (*TET2, DNMT3A, EP300, SETD1B, SETD2, SMARCA4, SMARCB2, CREBBP, ARID1A, BCOR, KMT2D, KMT2C, IDH2, MLL2, INO80*),
- Les voies JAK/STAT (*STAT3, STAT5B, JAK3, JAK1*),
- Les altérations de la réparation de l'ADN (*TP53, POT1, ATM, CHEK2*),
- La migration lymphocytaire T (*CCR4*),
- Les hélicases ARN (*DDX3X*).

Il n'existe pas d'étude génomique comparant toutes les entités, mais la grande majorité de ces mutations se retrouvent dans de nombreuses entités différentes. Le spectre mutationnel est souvent partagé entre entités de phénotypes similaires. L'activation de la voie du TCR et de NF-κB dans les lymphomes de phénotypes T CD4 helper (TFHL, ATLL et CTCL), l'activation de la voie JAK/STAT dans les lymphomes au profil plutôt cytotoxique (ALCL, NKTCL, HSTL, EATL, MEITL, LGL) mais aussi dans les T-PLL et de rares CTCL.

Le type de mutation de la voie JAK/STAT peut être un élément d'orientation bien que peu spécifique (Tableau 5) :

		Taux de mutation			
		STAT5B	JAK3	JAK1	STAT3
<b>Prédominance de mutations STAT5B</b>	<b>T-PLL</b> (Kiel et al. 2014)	36%	18%	10%	0%
	<b>HSTL</b> (McKinney et al. 2017)	30%			10%
	<b>MEITL</b> (Moffitt et al. 2017; Roberti et al. 2016)	60%	40-46%	13%-30%	0-13%
	<b>LGL CD4</b> (Andersson et al. 2016)	50%			
<b>Prédominance de mutations STAT3</b>	<b>EATL</b> (Moffitt et al. 2017)	12%	12%	16%	16%
	<b>LGL CD8 et NK</b> (Jerez et al. 2012; Koskela et al. 2012)	3%			30-40%
	<b>NKTCL</b> (Jiang et al. 2015; Xiong et al. 2020)	2%			15-17%
	<b>ALK-ALCL</b> (Crescenzo et al. 2015)			15%	11%
	<b>ATLL</b> (Kataoka et al. 2015)		2.5%		21%
<b>Faible taux de mutation JAK/STAT</b>	<b>CTCL</b> (Park et al. 2017a)	4%	3%	1%	1%
	<b>TFHL</b> (Vallois et al. 2016)		<2%	<2%	2.7%

Tableau 5 - Fréquence des mutations des protéines JAK et STAT retrouvées dans les néoplasies matures T

La découverte d'une mutation de STAT5B peut ainsi favoriser le diagnostic d'HSTL par rapport à celui d'une LGL, ou celui d'un MEITL plutôt qu'un EATL.

Il existe cependant quelques mutations qui sont très spécifiques d'entités, au premier rang desquels RHOA G17V qui est présente dans plus de 60% des lymphomes d'origine TFH et n'est pas retrouvée dans d'autres entités qui peuvent présenter d'autres mutations de RHOA (Palomero et al. 2014; Sakata-Yanagimoto et al. 2014; Odejide et al. 2014; Vallois et al. 2016).

Les mutations d'*ATM* sont aussi nettement plus fréquentes dans les T-PLL (70%) que dans les autres entités à l'exception des PTCL-NOS qui portent des mutations dans 25% des cas (Heavican et al. 2019; Kiel et al. 2014). Enfin les mutations de *DDX3X* retrouvées dans 25% des lymphomes NKTCL et 0-50% des EATL (Cording et al. 2022; Jiang et al. 2015) ne sont pas rapportées dans d'autres entités. Au total l'ajout du séquençage à haut débit est d'une grande aide dans le diagnostic de néoplasies T, mais il n'apporte qu'occasionnellement un diagnostic spécifique d'entité.

### 3.7 *Un diagnostic difficile*

Le diagnostic précis des sous entités de lymphome T nécessite donc l'intégration d'un grand nombre d'informations cliniques, histologiques, immunophénotypiques, caryotypiques et moléculaires et reste extrêmement difficile. La relecture par un anatomopathologiste spécialiste de ces pathologies est démontrée comme étant essentielle au diagnostic exact (Laurent et al. 2017). En effet, dans une étude du réseau français d'anatomopathologie des lymphomes (Lymphopath), la relecture menait à une modification de diagnostic dans 30% des cas de lymphomes T. Un nombre important de modifications avaient un impact thérapeutique direct. On peut noter notamment qu'il n'est pas rare d'avoir reclassé des néoplasies matures T comme leucémies aigües lymphoblastiques T. Les plus faibles taux de concordance (47 et 47,2% respectivement) étaient retrouvés pour le diagnostic des lymphomes anaplasiques ALK- et ATLL. La concordance, inter-experts cette fois-ci, a été évaluée dans une autre étude du projet international du lymphome T. Les plus faibles taux de concordance variaient entre 66% pour les lymphomes anaplasiques primitifs cutanés, 72% pour les lymphomes hépatospléniques, 74% pour les lymphomes anaplasiques ALK- et 75% pour les PTCL-NOS (Vose et al. 2008). Il semble donc intéressant de développer de nouveaux outils d'aide au diagnostic des lymphomes T.

## 4 Différentiation lymphoïde T et cellule d'origine.

### 4.1 *Différentiation normale*

#### a. Différentiation thymique initiale

Les lymphocytes T sont issus de la différenciation des cellules souches hématopoïétiques, à partir d'un précurseur lymphoïde commun aux lymphocytes B et T. Les précurseurs T qui ne sont pas encore engagés dans le lignage T migrent de la moelle vers le thymus où ils reçoivent des signaux des cellules épithéliales thymiques activant la voie Notch. Le signal Notch induit l'expression de deux facteurs de transcription essentiels à l'identité T : T-cell factor 1 (TCF1) et GATA3 qui vont activer entre autres l'expression de Rag. Grâce à la protéine Rag, la recombinaison des TCR $\gamma$ ,  $\beta$  et  $\delta$  est initiée. C'est à ce stade que se décide la différenciation entre les lymphocytes  $\alpha\beta$  et  $\gamma\delta$  (Ciofani et al. 2006).

#### b. Différentiation des lymphocytes T $\alpha\beta$

Si la chaîne  $\beta$  du TCR est correctement réarrangée, la signalisation du préTCR (chaîne  $\beta$  + chaîne pré- $\alpha$ ) permet la prolifération cellulaire et la progression vers le stade double positif CD4+CD8+. La chaîne alpha va alors être réarrangée et les thymocytes doubles positifs vont subir la sélection négative et positive permettant la calibration du niveau d'affinité des lymphocytes T pour les complexes majeurs d'histocompatibilité (CMH) présentant des peptides du soi. Certains lymphocytes T régulateur ayant une forte affinité pour ces complexes vont être dirigés vers la lignée des lymphocytes T régulateurs décrits plus loin. Les cellules ayant un TCR reconnaissant le CMH de classe I vont se différencier vers le stade simple positif CD8 et perdre leur CD4 grâce à l'expression du facteur de transcription Runx3. A l'inverse les thymocytes reconnaissant le CMH de classe II vont se diriger vers la lignée CD4 via l'expression de ThPOK. Les thymocytes simples positifs vont subir une étape de maturation finale puis exprimer les récepteurs S2PR1 et CD62L (L-selectine) pour émigrer du thymus vers la circulation périphérique.

Un type particulier de lymphocyte  $\alpha\beta$  issus du thymus est formé par les lymphocytes NKT qui ne reconnaissent pas de complexe CMH-peptide, et n'ont donc pas suivi les étapes de sélection positive et négative. En revanche ils reconnaissent des molécules homologues CMH, les molécules CD1 qui ne présentent pas de peptides mais des glycolipides aux lymphocytes T (Stax et al. 2017). Les lymphocytes NKT sont sélectionnés dans le thymus via

l'expression des molécules CD1 par les thymocytes et ont déjà un phénotype mémoire à la sortie du thymus (Pellicci, Koay, and Berzins 2020).

c. Différentiation extra thymique des lymphocytes T $\alpha\beta$

Les lymphocytes T simples positifs sortent du thymus grâce à l'expression de CD62L (L-selectine), puis circulent dans le sang et dans les organes lymphoïdes secondaires. Ils sont appelés lymphocytes T naïfs et sont des cellules de petite taille à la chromatine condensée. Leur première activation (priming) est induite par la combinaison de trois signaux : premièrement le contact d'un TCR avec le complexe CMH/peptide correspondant, deuxièmement un signal fourni par des voies de costimulation (via notamment CD28/B7) et troisièmement la présence de cytokines environnantes et notamment l'IL-2. Suite à leur activation, ils entrent en prolifération et se différencient en lymphocytes T effecteurs. Le priming des lymphocytes CD8 génère des lymphocytes effecteurs cytotoxiques capables de tuer directement les cellules infectées par des pathogènes. Alors que les lymphocytes CD4 se diversifient en une variété des lymphocytes effecteurs « helper » en fonction de leur environnement cytokinique et via l'expression de facteurs transcription drivers. Les lymphocytes T helper CD4 sont ainsi sous divisés en plusieurs catégories Th1, Th2, Th17, TFH et iTreg (induced regulatory T cell). On notera que les lymphocytes Th1 expriment le facteur de transcription T-bet (ou TBX21) et les lymphocytes Th2 expriment le facteur de transcription GATA3 (Figure 4).

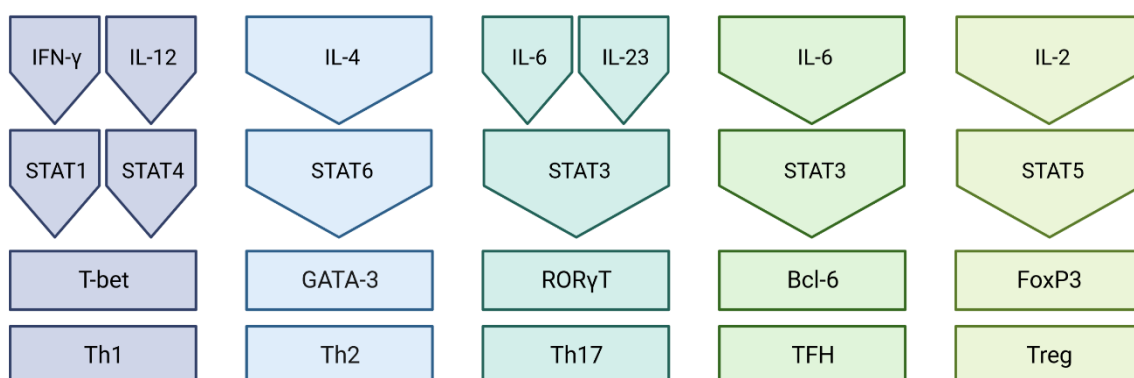


Figure 4 - Cytokines et facteurs de transcription impliqués dans la différenciation T helper.

*Adapté de Janeway's Immunobiology 2017*

L'activation des lymphocytes T naïfs est associée à la modification de plusieurs molécules de surface et notamment :

- L'expression de l'isoforme CD45RO, un variant d'épissage de la tyrosine phosphatase CD45,
- La perte d'expression du récepteur à l'IL7 (IL7R $\alpha$  ou CD127),
- L'acquisition de CD44, un récepteur à l'acide hyaluronique et d'autres ligands,
- La perte de CD62L, ce qui empêche une nouvelle recirculation à travers les ganglions ,
- L'acquisition de la P-selectin glycoprotein ligand (PSGL1) qui permet l'adhésion à l'endothélium,
- L'augmentation du niveau d'expression de LFA-1 et CD2 ainsi que des intégrines VLA-4 qui se lient aux molécules d'adhésion VCAM1 permettant la migration vers les tissus non lymphoïdes.

Différentes intégrines peuvent être exprimées en fonction du site du ganglion ou les lymphocytes T naïfs sont primés ce qui permet un ré-adressage au site initial de l'activation :

- L'intégrine  $\alpha 4:\beta 7$  permet une migration dans les muqueuses.
- L'antigène lymphocytaire cutané (cutaneous lymphocyte antigen CLA) en association à CCR4 et CCR10 permet une migration vers l'épiderme.

Les lymphocytes TFH, décrits en détail plus loin, vont eux acquérir des molécules permettant d'entrer la zone B des tissus lymphoïdes.

A la fin d'une infection, la plupart des lymphocytes effecteurs entrent en apoptose en raison de la disparition de la stimulation antigénique, la diminution de la quantité d'IL2, la perte de l'expression de CD25 et la perte d'expression du récepteur à l'IL7. Certains lymphocytes effecteurs survivent et se différencient en lymphocytes T mémoires qui expriment toujours CD44/PGP1, mais perdent l'expression des marqueurs d'activation tel CD69. La survie de ces lymphocytes est permise par la réacquisition du récepteur à l'IL7 : IL7-R $\alpha$ .

On distingue trois types de cellules à mémoires :

- les lymphocytes T mémoires centraux (TCM) qui expriment CCR7, ce qui leur permet une recirculation similaire aux lymphocytes T naifs dans les tissus lymphoïdes périphériques (Sallusto, Geginat, and Lanzavecchia 2004).
- Les lymphocytes T effecteurs mémoires qui n'expriment pas CCR7, mais de hauts niveaux d'intégrine  $\beta 1$  et  $\beta 2$ , leur permettant une entrée rapide dans les tissus inflammatoires. En cas d'activation, ils sont capables d'acquérir des fonctions effectrices nettement plus rapidement que les TCM et de sécréter en quantité diverses cytokines.
- Les lymphocytes T résidents mémoires (TRM), sont des cellules qui résident à long terme dans des sites épithéliaux. Ils expriment des récepteurs de cytokines et des intégrines spécifiques des tissus dans lesquels ils sont localisés comme l'intégrine  $\alpha 4:\beta 7$  (qui se lie à MadCAM-1) et le récepteur au chimiokine CCR9 dans la *lamina propria* de l'intestin, ou l'intégrine  $\alpha E:\beta 7$  (CD103 qui se lie à l'E-cadhérine) dans l'épithélium intestinal. Les TRM continuent aussi à exprimer CD69, habituellement exprimés dans les cellules activées, ce qui réduit l'expression de S2PR et permet leur rétention dans les tissus périphériques (Sathaliyawala et al. 2013)

Le lien entre les différents états Th1, Th2, Th17 des lymphocytes TCD4 helper et cette trajectoire de différenciation (naïve, central mémoire, effecteur mémoire) n'est pas complètement élucidé. Il semble possible que les TCM CD4 ne soient pas engagés dans une lignée effectrice particulière (Th1, Th2Th1, TH2 ou TH17). De plus, il est démontré que les lymphocytes T effecteurs CD4 peuvent modifier leur phénotype cytokinique en fonction du contexte local (Messi et al. 2003). En effet, la plupart des lymphocytes Th1 peuvent sécréter de l'IL4 en plus de l'IFN $\gamma$  et la plupart des lymphocytes Th2 de l'IFN $\gamma$  en plus de l'IL4, en fonction du contexte cytokinique dans lequel ils sont stimulés. Cette plasticité étant particulièrement prédominante entre les différenciations Th17 et Th1. En revanche, un groupe de cellules CRTH2+ semble engagé irréversiblement dans la lignée Th2. Ces études ont amené l'hypothèse que les lymphocytes T CD4 mémoires seraient générés séparément des lymphocytes effecteur Th1, Th2 ou Th17 et seraient initialement non polarisés.

Certains lymphocytes CD8 T effecteurs mémoires réacquièrent l'expression du CD45RA et sont appelés CD8 TEMRA. Cette population CD45RA+CCR7- est beaucoup moins présente

parmi les lymphocytes CD4, mais s'expand dans certains contextes infectieux notamment les expositions au cytomégalo virus et à la dengue. Ce sont des cellules qui expriment très fortement les marqueurs cytotoxiques perforine et granzyme ainsi que les facteurs de transcription RUNX3 et TBX21 (Tian et al. 2017).

d. Lymphocytes T folliculaires helper (TFH)

Suite à leur activation dans les ganglions, certains lymphocytes T CD4 naïfs vont commencer à exprimer la molécule CXCR5 qui permet leur entrée dans les zones B des ganglions via l'interaction avec la chimiokine CXCL13. L'expression de CXCR5 nécessite le facteur de transcription Bcl6. Dans les follicules B, ces lymphocytes T vont rencontrer des lymphocytes B exprimant la molécule ICOSL qui, après interaction avec ICOS va activer des signaux permettant la différenciation finale en lymphocytes TFH. Le contact maintenu entre lymphocytes TFH et B va permettre la prolifération des lymphocytes B et leur différenciation en plasmocytes via notamment la sécrétion d'une grande quantité d'interleukine 21, et l'interaction CD40-CD40L.

De manière surprenante, les lymphocytes TFH sont aussi décrits comme pouvant être polarisés en lymphocytes Th1, Th2 ou Th17 et ce en fonction du contexte cytokinique ganglionnaire. Cette polarisation permet d'induire des biais dans le choix des isotypes des lymphocytes B (Reinhardt, Liang, and Locksley 2009).

e. Lymphocytes T régulateurs (Treg)

En comparaison avec les lymphocytes T naïfs, les lymphocytes T régulateurs expriment constitutivement la chaîne CD25 du récepteur à l'IL2 ce qui leur permet de capter de faible dose d'interleukine 2 et de maintenir leur survie. Il existe deux types de lymphocytes T régulateurs : ceux qui sont induits dans le thymus, aussi appelés natural regulatory T cells (nTreg) et ceux qui sont induits en périphérie (induced regulatory T cells, iTreg) à partir de lymphocytes T CD4 naïfs dans certains contextes cytokiniques (présence de TGF $\beta$  et absence d'interleukine 6). Le facteur de transcription spécifique des lymphocytes T régulateur est FoxP3. Similairement aux TFH, il a été décrit que l'existence d'une coexpression des facteurs de transcription Foxp3 et Tbet dans les lymphocytes T régulateurs était nécessaire à leur action régulatrice lors d'une réaction Th1 (Koch et al. 2009), de même que FoxP3 et Irf4 dans un contexte Th2 (Y. Zheng et al. 2009).



f. Différentiation gamma delta

Les lymphocytes T gamma delta sont issus d'un processus de différenciation spécifique. Ils sont générés en plusieurs vagues lors de la vie fœtale et néonatale à partir d'organes hématopoïétiques différents (foie, thymus) dont certains dans des processus thymo-indépendants. Il était décrit que les lymphocytes T $\gamma\delta$  présentent un phénotype mémoire à la sortie du thymus. En revanche des études récentes remettent cette hypothèse en question avec la démonstration de l'existence de populations matures et immatures de lymphocytes T $\gamma\delta$  V $\gamma$ 9V $\delta$ 2 similairement à ce qui est décrit pour les lymphocytes NKT.

Les lymphocytes T  $\gamma\delta$  ont la particularité de porter fréquemment des réarrangements invariants du TCR et ne sont pas dépendants du CMH pour reconnaître des antigènes. Certains reconnaissent directement des antigènes à la manière d'un anticorps, d'autres reconnaissent des antigènes présentés par des molécules de CMH non conventionnels (Uldrich et al. 2013; Luoma et al. 2013).

g. Différentiation des lymphocytes NK et des lymphocytes innés

Les lymphocytes NK dérivent du même progéniteur que les lymphocytes B et T. Plusieurs hypothèses existent quant aux signaux orientant les NK vers leur différenciation. Un modèle suggère qu'il existe un précurseur lymphoïde engagé dans la différenciation NK marqué par la perte d'expression de CD34 et l'acquisition de LFA-1. Ce précurseur acquerrait la molécule CD94 puis CD65 donnant lieu aux NK CD56bright. Puis les lymphocytes NK diminueraient leur expression de CD94, acquerrait CD16 et les killer immunoglobulin like receptor (KIR) (Cichocki, Grzywacz, and Miller 2019).

Les cellules lymphoïdes innées (ILCs) se développent aussi à partir du même progéniteur lymphoïde commun aux lymphocytes B et T. L'expression du facteur de transcription Id2 réprime le devenir T ou B et permet la formation d'un précurseur lymphoïde commun (ILCP). Ces cellules n'expriment pas de récepteur T ou B, mais expriment l'IL7R et parfois le CD3 intracytoplasmique. Trois groupes sont définis parmi les ILC matures qui semblent dirigés par les mêmes facteurs de transcription que les CD4 helper :

- Les ILC1 qui sécrètent de l'IFN $\gamma$ , en réponse à l'activation par l'IL12 et l'IL18. Les lymphocytes NK sont proches des ILC1 mais auraient un développement différent. Les ILC1 ressembleraient aux lymphocytes CD4 Th1 alors que les NK sont plus

proches des lymphocytes CD8. Les lymphocytes NK nécessitent en plus du facteur Id2, la présence d'IL-15 et du facteur de transcription Nfil3 et eomesodermin (Eomes) alors que les ILC1 nécessitent seulement l'IL7 et le facteur de transcription Tbet.

- Les ILC2 sont similaires aux Th2 et produisent de l'IL4, de l'IL5 et de l'IL13.
- Les ILC3 produisent de l'IL17 et de l'IL22.

Il existe donc de nombreux stades de différenciation et sous types de lymphocytes T, NK et ILC orchestrés par différents facteurs de transcription. Il est intéressant d'essayer de rattacher chaque entité d'hémopathie T à ces différents stades pour permettre une meilleure compréhension des processus inhérents au développement lymphocytaire normal qui sont communs à la lymphomagénèse.

#### 4.2 Cellules d'origine des MTCN

La classification des hémopathies T est basée en partie sur des marqueurs qui font référence à différents stades de maturation des lymphocytes T et NK. Ainsi le diagnostic de lymphomes TFH est basé sur la présence de marqueurs de cette cellule (CD4, CXCR5, PD1, BCL6). Mais elle utilise aussi certains marqueurs qui font plutôt référence au statut d'activation des lymphocytes T (CD25, CD30), ou à des mécanismes de lymphomagénèse précis qui ne sont pas retrouvés dans des lymphocytes T ou NK normaux (ALK). Pour identifier les contreparties normales des MTCN, de nombreuses études se sont basées sur l'expression protéique de différents marqueurs de différenciation T (Tableau 6). Ainsi l'utilisation des marquages CCR7, CD62L, CD45RA et CD45RO ont été utilisés pour définir le stade de différenciation des cellules issus de lymphocytes conventionnels. Les syndromes de Sézary, sont principalement CCR7+CD62L+ dans le sang mais aussi dans la peau et sont donc supposés dériver de lymphocytes T centraux mémoire avec capacité de « homing » à la peau grâce à l'expression d'intégrines comme l'antigène lymphoïde cutané (CLA). Les mycosis fongoïdes en revanche sont négatifs pour CCR7 et CD62L et sont décrits comme résidents mémoires de la peau (Campbell et al. 2010). Il est intéressant de noter que les autres marqueurs de différenciation T CD45RO et CD45RA étaient très hétérogènement exprimés dans ces deux pathologies et cette étude a été remise en question par plusieurs travaux de cytométrie de flux qui montrent une forte hétérogénéité intra patient de l'expression des marqueurs de naïveté. Ainsi 88% des patients présentaient des cellules tumorales de phénotype central mémoire et

63% des cellules de phénotypes naives, une grande partie présentant donc les deux phénotypes cellulaires simultanément (Najidh et al. 2021). Une autre étude montre une fluctuation entre les phénotypes naïfs et mémoire des syndrome de Sézary sur différents prélèvement successifs chez 40% des patient et émet l’hypothèse que les cellules souches du syndrome de Sézary seraient des lymphocytes T « stem-cell » mémoire (CD158+) (Roelens et al. 2017). Utiliser seulement l’expression des marqueurs de surface de la différenciation normale T semble donc insuffisant pour déterminer la cellule d’origine des lymphomes T.

D’autres arguments ont été utilisés pour rattacher une entité à sa cellule d’origine. Notamment des tests fonctionnels de réponse à la stimulation du TCR. En effet, les cellules centrales mémoire produisent plus rapidement des cytokines en réponse à un stimulation CD3/CD28. Ainsi Oberbeck démontre que les leucémies prolymphocytaires T, en plus d’avoir un phénotype prédominant central mémoire (CD45RO+CD45RA-CCR7+CD62L+), ont une réponse rapide et forte à la stimulation CD3/CD28 (Oberbeck et al. 2020).

Entité	Contrepartie cellulaire normale	Marqueurs immnuophénotypiques majoritaires	Référence
T-PLL	Lymphocyte T central mémoire	CD45RO+ CD45RA- CCR7+ CD62L+ CD45RO+/RA+ (21%)	(Oberbeck et al. 2020)
AITL/PTCL-TFH	Lymphocyte T folliculaire helper	CD4+ CXCR5+ PD1+ CXCL13+ BCL6+	(de Leval et al. 2007; Dorfman et al. 2006; Grogg et al. 2005; Krenacs et al. 2006; Ree et al. 1999)
MF	Lymphocyte T CD4 résident mémoire	CCR7- CD62L- CCR4+ CLA+	(Campbell et al. 2010)
SS	Lymphocyte T CD4 central mémoire	CCR7+ CD62L+ CCR4+ CLA+/-	(Campbell et al. 2010; Najidh et al. 2021; Roelens et al. 2017)
ATLL	Lymphocyte T CD4 helper	CD45RO CD25+, FOXP3+ /-(65%) CCR7- IL7R-	(Abe et al. 2008; Karube et al. 2004; Nagai et al. 2015; B. J. Tan et al. 2021)
ALCL ALK+	Divergence par rapport au développement T normal	CD3- CD4+ CD8- CD30+ GZMB+	
RCD2 / EATL	Cellules lymphoïdes innées de type 3 CD103+	sCD3-cCD3ε+TCR-CD8-CD122+CD127- NKP46+GZMB+	(Ettersperger et al. 2016)
MEITL	Lymphocyte intraépithélial CD8αα Tγδ (ou αβ)	CD8αα+CD56+CD57-TiA-1+TCRγδ+ (αβ+)	(S.-Y. Tan et al. 2013)
HSTL	Lymphocyte T γδ splénique	CD2+, CD3+, CD4-, CD5-, CD7+, CD16+, CD57-, CD25-, TCRγδ ou plus rarement αβ	(Cooke et al. 1996)
NKTCL	Lymphocyte T αβ cytotoxique ou lymphocyte NK	CD3+, CD2+, GZMB+,CD56+,TCRαβ+/-	(Pongpruttipan et al. 2012)
LGL	Lymphocyte T effecteur mémoire exprimant le 45RA	CD8+CD57+CD45RA+	(Brenchley et al. 2003; Melenhorst et al. 2001)

Tableau 6 - Principales origines cellulaires suspectées des différentes hémopathies matures T

L'existence d'une cellule d'origine unique pour chaque lymphome T est directement remise en cause par l'hétérogénéité observée dans l'expression de marqueurs pourtant caractéristiques de la différenciation T et NK. Ainsi les leucémies prolymphocytaires T peuvent être CD4+ ou double positif CD4 CD8 ou plus rarement simple positif CD8. Les LGL et NKTCL peuvent être avoir un TCR $\alpha\beta$  ou plus rarement  $\gamma\delta$  ou aucun TCR.

Les lymphomes anaplasiques représentent une des classes pour laquelle l'authentification de la cellule d'origine est la plus difficile. En effet, l'absence d'expression du CD3 et du TCR de surface avait initialement fait douter de l'origine T qui a ensuite été démontrée par la découverte de réarrangements TCR. Le phénotype de surface avec expression de CD30 et MUC1 évoque un phénotype T activé (ten Berge et al. 2001; Agrawal et al. 1998). L'expression de granzyme B évoque plutôt un profil de lymphocytes T cytotoxiques. Et enfin les analyses transcriptomiques favorisent un profil Th17 (expression de *IL17A*, *IL17F*, *IL26*, *IL22*, et ROR $\gamma$ t) (Iqbal et al. 2014). Cependant une étude extensive des réarrangements du TCR a démontré que les lymphomes anaplasiques avaient un réarrangement du TCR $\alpha$  et  $\beta$  mais que la majorité des réarrangement du TCR $\beta$  n'étaient pas productifs (Malcolm et al. 2016). Dans un modèle murin déficient pour Rag1, la translocation NPM1-ALK permettait le développement de lymphocytes T en absence de recombinaison TCR $\beta$ , ce qui conduit les auteurs à postuler l'origine thymique des ALCL-ALK+. Enfin, une autre équipe décrit le processus inverse de dédifférenciation à l'origine de la lymphomagénèse des lymphomes anaplasiques, en démontrant que la translocation NPM-ALK1 permet la dédifférenciation de lymphocytes T CD4 matures vers un phénotype thymique précoce d'après des profils de méthylation de l'ADN et ARN (Congras et al. 2020).

Par ailleurs, il est décrit que certains lymphomes T sont issus d'un processus de transdifférenciation vers un phénotype innate-like. En effet, de nombreux lymphomes T expriment des récepteurs des lymphocytes NK. Dans un modèle de lymphomagénèse dépendante de P53, il a été démontré par notre équipe que la stimulation chronique du TCR induit une transdifférenciation des lymphocytes T vers un phénotype innate-like avec l'expression des molécules SYK et TYROBP et de récepteurs NK et une réduction de l'expression des molécules habituellement T (CD3,CD28) (Carras et al. 2021).

L'identification des cellules d'origine des MTCN est donc une tâche difficile et selon la méthode utilisée (transcriptomique, marqueurs protéiques, réarrangement TCR), les hypothèses formulées peuvent être différentes. Les mesures épigénomiques qui présentent une grande spécificité de type cellulaire pourraient apporter un éclairage nouveau à cette question.

## B EPIGENETIQUE ET MESURES DE L'ACCESSIBILITE CHROMATINIENNE

### 1 Structure de la chromatine et impact sur la régulation transcriptionnelle

A l'origine, l'épigénétique était définie comme l'étude des changements héréditaires de phénotypes non liés à des modifications de la séquence de l'ADN. Depuis, le sens de ce mot a dérivé et il est communément utilisé pour décrire l'ensemble des informations portées par le génome qui ne sont pas codées dans l'ADN. En effet, l'expression du génome est en grande partie déterminée par son organisation au sein du noyau.

#### 1.1 Organisation du génome

L'ADN de tous les chromosomes d'une cellule mis bout à bout mesurerait environ 2m de long. Comment est-il donc possible qu'ils tiennent dans un noyau de 10 $\mu$ m ? Il existe en fait de nombreux niveaux de repliements de l'ADN dont les mécanismes sont précisément orchestrés (Misteli 2020) :

1. Le premier niveau est celui du nucléosome qui est constitué d'une unité de 146 paires de bases d'ADN enroulé autour d'un octamère formé de 4 histones principaux (H2A, H2B, H3 et H4). La position des nucléosomes est variable au cours des processus cellulaires. Ils sont moins présents dans les zones actives de l'ADN (enhancers, insulators et zones transcrites des gènes). L'ADN internucléosomal est souvent occupé par des facteurs de transcription, l'ARN polymérase ou par des histones « linker ». L'histone linker H1 est nettement plus mobile que les octamères d'histones (Phair et al. 2004).
2. Cette fibre chromatinienne forme ensuite des boucles de tailles variables dans le noyau, qui rapprochent des domaines éloignés d'une même molécule d'ADN. Les boucles de petite taille (10-100 pb) se localisent souvent au niveau de zones permettant la régulation de la transcription appelées enhancers (décrits page 61). Des boucles plus grandes (jusqu'à 1 Mb) contribuent au repliement du génome ou à

la régulation coordonnée de clusters de gènes lors du développement des organismes.

3. Les boucles sont ensuite regroupées en domaines chromatiniens appelés TAD pour « Topologically associated domains », généralement de plusieurs centaines de kilobases.
4. Les domaines chromatiniens s'associent entre eux pour former des compartiments chromatiniens qui peuvent être classés selon l'intensité de leur compaction en compartiments A et B et qui correspondent approximativement à la distinction hétérochromatine (silencieuse) / euchromatine (active) décrite depuis de nombreuses années (Lieberman-Aiden et al. 2009; Shopland et al. 2006).
5. Enfin chaque chromosome est formé de plusieurs compartiments chromatiniens, et va prendre une place spécifique, dans le noyau interphasique. Un territoire chromosomique est relativement compact (2-3 $\mu$ m), positionné de manière non aléatoire dans le noyau et bien séparé des autres chromosomes (Cremer and Cremer 2010).

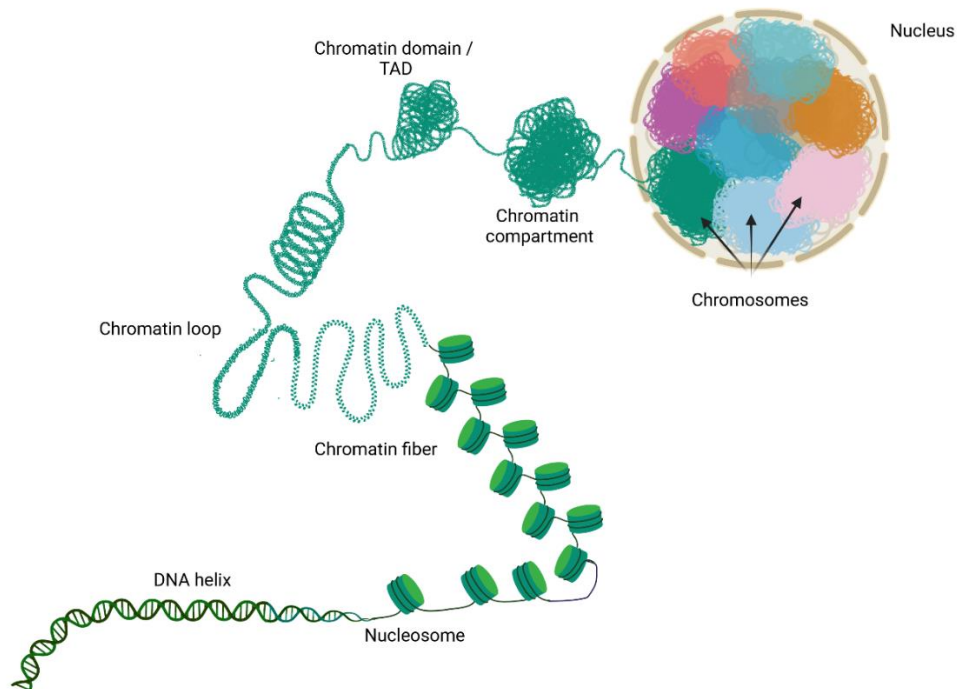


Figure 5 - Organisation du génome d'après Misteli (Misteli 2020)

Cette structure n'est pas fixe et des modifications organisationnelles sont nécessaires à chaque niveau pour permettre la réalisation des différents processus cellulaires qui touchent l'ADN (transcription, réparation de l'ADN, réplication de l'ADN, différenciation).

## 1.2 *Modifications épigénétiques et facteurs de transcription*

Il existe trois grands types de processus qui modifient la chromatine et vont jouer sur cette conformation : la méthylation de l'ADN, les modifications des histones et le remodelage chromatinien.

### a. Méthylation de l'ADN

Le cinquième carbone des cytosines de l'ADN peut être méthylé. Cette réaction a lieu habituellement au sein de séquences répétées de dinucléotides CG ou de trinucleotides CNG appelés îlots CpG. Elle est catalysée par différentes ADN méthyltransférases. L'ADN méthyltransférase 1 (DNMT1), est impliquée dans la maintenance de la méthylation de l'ADN lors de sa réplication agissant sur un brin d'ADN hémiméthylé. Les DNMT3 a et b sont responsables de la méthylation de novo. Ainsi DNMT3a est notamment capable de méthylé de novo les promoteurs des gènes, processus indispensables à la différenciation cellulaire lors du développement (Luo, Hajkova, and Ecker 2018). La déméthylation de l'ADN est un processus plus complexe impliquant l'oxydation des 5-méthylcytosines (5mC) en 5-hydroxyméthylcytosines (5hmC) et puis en 5-carboxylcytosines (5caC) par les enzymes Tet méthylcytosine dioxygénase (TET), suivie de l'excision de la base par la thymidine DNA glycosylase et son remplacement par une cytosine non méthylée par la machinerie de « base excision repair ».

Cette marque de l'ADN est en général associée à une forme compactée de la chromatine et réprime la transcription des gènes adjacents. Les régions transposées et répétées sont souvent riches en GC et sujettes à une forte méthylation.

Des anomalies de la méthylation sont fréquemment retrouvées dans les cancers. Elles induisent souvent une réexpression d'oncogènes. Dans les pathologies lymphoïdes T, l'ADN méthyltransférase 3A (DNMT3A) et la Tet méthylcytosine dioxygénase (TET2) sont très fréquemment la cible de mutations pertes de fonction (Couronné, Bastard, and Bernard 2012).



b. Modification des histones

Les histones canoniques (H2A, H3B, H3, H4) sont des protéines à la conformation globulaire excepté au niveau de leur queue N-terminale dont les acides aminés peuvent être soumis à de nombreuses modifications (acétylation, méthylation, ubiquitinylation, phosphorylation, déamination...). Ces modifications, en altérant les contacts entre nucléosomes (charges électrostatiques différentes ou encombrement allostérique), ou via le recrutement d'autres protéines, vont avoir un effet sur l'organisation globale de la chromatine et sur plusieurs processus liés à l'ADN ( Kouzarides 2007).

Modification	Résidu modifiés	Fonction régulée
Acétylation	K-ac	Transcription, réparation, condensation, réplication
Méthylation (lysines)	K-me1 K-me2 K-me3	Transcription, réparation
Méthylation (arginines)	R-me1 R-me2a R-me2s	Transcription
Phosphorylation	S-ph T-ph	Transcription, réparation, condensation
Ubiquitinylation	K-ub	Transcription, réparation
Sumoylation	K-su	Transcription
ADP ribosylation	E-ar	Transcription
Désimination	R > Cit	Transcription
Proline isomérisation	P-cis > P-trans	Transcription

Tableau 7 - Fonctionnalité des différentes classes de modification d'histones  
d'après Kouzaride 2007

Ainsi, l'acétylation de la lysine 27 de l'histone H3 (H3K27ac) se retrouve principalement au niveau des promoteurs et enhancers des gènes et favorise la transcription (Heintzman et al. 2009). La triméthylation de la lysine 4 de l'histone H3 (H3K4me3) est classiquement localisée autour des sites d'initiations de la transcription des gènes actifs. A l'inverse, les enhancers actifs sont enrichis en H3K4me1 plutôt qu'en H3K4me3 (Creyghton et al. 2010; Rada-Iglesias et al. 2011).

De nombreuses enzymes sont responsables de l'ajout (writers) ou de l'enlèvement (erasers) de ces marques épigénétiques. L'action des enzymes est plus ou moins spécifique de résidus. Ainsi les lysines méthyltransférases sont extrêmement spécifiques, comme EZH2 qui triméthyle H3K27 ou SETDB1 qui triméthyle H3K9, alors que les acétyltransférases peuvent souvent acétyler plusieurs résidus. Les acétyltransférases p300 (aussi appelée EP300 ou E1A

binding protein p300) et CBP (alias CREB-bind protein ou CREBBP) sont ainsi capables d'acétyler H3 (K14, K18), H4 (K5, K8), H2A (K5), H2B (K12, K15) (Kouzarides 2007).

Les protéines modificatrices d'histone sont très fréquemment atteintes dans les hémopathies T et plusieurs traitements les ciblent. On peut citer les inhibiteurs d'histones désacétylases comme la Romidepsine et le Vorinostat déjà utilisés depuis de nombreuses années dans le traitement des lymphomes T périphériques et cutanés (Coiffier et al. 2012; Olsen et al. 2007) ou les inhibiteurs d'EZH2 en cours de développement dans les lymphomes T notamment liés à l'HTLV1 (Izutsu et al. 2022; Song et al. 2022).

#### c. Positionnement des nucléosomes et remodeleurs de la chromatine

Le positionnement des nucléosomes n'est pas aléatoire ni uniforme le long de l'ADN. Ceux-ci sont en effet densément réarrangés dans l'hétérochromatine mais plus espacés au niveau des locus régulateurs (enhancers, insulators) et des gènes en cours de transcription. Ce positionnement est précisément régulé par les remodeleurs de la chromatine, des protéines capables de transloquer les nucléosomes et d'assurer trois fonctions principales :

- L'assemblage et l'organisation des nucléosomes : Lors de la division cellulaire, les histones sont acheminées au niveau de l'ADN par des chaperonnes d'histones. Puis, se forme un prénucléosome entre un octamère d'histone et l'ADN (Fei et al. 2015). L'intervention des remodeleurs de la chromatine est alors indispensable pour permettre le superenroulement de l'ADN autour de l'octamère d'histones. Ces remodeleurs vont ensuite assurer l'espacement correct des nucléosomes le long de l'ADN.
- La mobilisation des nucléosomes pour permettre l'accès à l'ADN : Une autre famille d'histones est capable de coopérer avec les facteurs de transcription et les modificateurs d'histone pour faire glisser ou éjecter les histones pour que des protéines (facteurs de transcription, ARN polymérase puissent avoir accès à l'ADN).
- L'édition des nucléosomes. Une troisième famille de remodeleurs chromatiniens permet lors de la transcription l'échange de certaines histones canoniques par des variants d'histones aux propriétés spécifiques.

On distingue ainsi 4 grandes familles de remodeleurs de la chromatine à activité ATP-ase et aux fonctions diverses (Clapier et al. 2017) :

- Les ISWI (imitation switch) sont responsables principalement de l'assemblage et de l'espacement des nucléosomes pour limiter l'accessibilité chromatinienne et l'expression des gènes. La sous famille NURF a des sous-unités accessoires qui peuvent, à l'opposé, promouvoir la transcription (Tableau 8).
- Les CHD (chromodomaine hélicase DNA binding) sont aussi responsables de l'assemblage des nucléosomes et de leur espacement. Certains complexes peuvent modifier leur disposition pour permettre l'accessibilité des promoteurs, et peuvent incorporer le variant d'histone H3.3.
- Les SWI/SNF (switch/sucrose non fermentable) facilitent l'accès à la chromatine. Ils font glisser et éjectent les protéines nucléosomales de l'ADN lors de la répression ou l'activation des gènes.
- Les INO80 : cette famille est composée de deux complexes SWR1 et INO80. Le complexe SWR1 (TRRAP chez l'humain) remplace les histones H2A par un variant H2A.Z qui est enrichi au niveau des promoteurs et enhancers. INO80 catalyse la réaction inverse et est aussi impliqué dans la réparation de l'ADN car il est capable d'enlever le variant H2A.X qui marque les cassures double brin (van Attikum et al. 2004).

La composition de ces complexes est très variable et peut modifier totalement leur activité. Ainsi le complexe NuRD (membre de la famille CHD) peut incorporer différentes protéines MTA. Les complexes contenant MTA1 répriment la transcription induite par les récepteurs aux œstrogènes et contribuent ainsi à la croissance invasive des tissus cancéreux. Par opposition les complexes contenant MTA-3 interagissent avec BCL6 et répriment l'expression des gènes inducteurs de métastases (Snail) et maintiennent un phénotype non invasif du tissu tumoral (Toh and Nicolson 2009).

Les complexes remodeleurs de la chromatine sont fréquemment mutés dans les cancers (Lai and Wade 2011; Mittal and Roberts 2020; Wilson and Roberts 2011). On observe notamment des mutations du complexe SWI/SNF (ARID1A, ARID1B, SMARCB1, SMARCA2 et

SMARCA4) dans toutes les hémopathies T et, dans certaines, des mutations de TRRAP (CTCL, PTCL-NOS) et d'INO80 (notamment dans les HSTL).

Famille	ISWI		SWI/SNF	INO80		MI-2/CHD	
Sous-famille	ACF/CHRAC	NURF	SWI/SNF	INO80	SWR1	CHD1	NuRD
Complexe	ACF	NURF	BAF	INO80	TRRAP/Tip60	CHD1	NuRD
Sous-unités	SNF2H ACF1 CHARC15 CHAR17	SNF2L BPTF RbAP46, RbAP48	<b>BRG1 (SMRACA4)</b> ou <b>BRM (SMARCA2)</b> <b>BAF250A (ARID1A)</b> ou <b>ARID1B</b> BAD155, BAF170 BAF60a,b,c BAF53a,b <b>SNF5 (INI1 / SMARCB1)</b> BAF57 B-actin  pBAF : +BAF180 <b>+BAF200 (ARID2)</b> +BRD7	<b>Ino80</b> Tip49a, Tip49b BAF53a, Arp5,8 Ies2 Ies6 Amida NFRKB MCRS1 FLJ90652 FLJ20309	P400 TIP49a, Tip49b BAF53a, Actin GAS41 DMAP1 YL-1 Brd8/TRCp120 <b>TRRAP</b> Tip60 MRG15, MRGX FLJ11370 MRGBP EPC1, EPC-like ING3	CHD1	NuRD Mi2 $\alpha$ /CHD3 Mi-2 $\beta$ /CHD4 MBD3 MTA1,2,3 HDAC1 HDAC2 RbAP46,48 P66 $\alpha,\beta$ DOC-1

Tableau 8 - Composition des principaux complexes remodeleurs de la chromatine

*d'après (Bao and Shen 2007; Clapier et al. 2017), en bleu les protéines mutés dans les MTCN*

Les enzymes responsables de la méthylation de l'ADN, les modificateurs d'histones et les remodeleurs de la chromatine agissent de manière interconnectée. Par exemple, les ADN méthyltransférases sont recrutées par l'histone méthylase G9A qui triméthyle H3K9 au niveau des zones répétées de l'ADN (Dong et al. 2008). A l'opposé, SWI/SNF se lie aux histones H3 acétylées des gènes activés pour éjecter les nucléosomes et permettre la transcription (Chatterjee et al. 2011). L'interconnexion de tous ces éléments orchestre ainsi la régulation transcriptionnelle.

### 1.3 Eléments régulateurs de la transcription

Le génome humain porte deux types d'informations fondamentales : les séquences des ARN et les éléments régulateurs qui permettent la transcription de celles-ci. Les éléments cis-régulateurs sont des séquences qui jouent le rôle de plateforme pour la fixation des

protéines nécessaires à la transcription. Ils peuvent être proximaux (« core » promoteur et éléments proximaux du promoteur) ou distaux (enhancers, silencers, insulators).

a. Site d'initiation de la transcription et promoteurs :

Les promoteurs sont des loci génomiques qui chevauchent les sites d'initiation de la transcription des gènes (transcription start site ou TSS) où se fixe l'ARN polymérase. On distingue la partie centrale du promoteur (core promoter) qui se situe environ 35 paires de bases en amont et en aval du site d'initiation de la transcription et les autres éléments situés dans les quelques centaines de paires de bases autour du TSS (promoter-proximal elements).

Le complexe de préinitiation de la transcription, composé de l'ARN polymérase et de facteurs de transcription (TFIIA, TFIIB, TFIID, TFIIIE, TFIIIF et TFIIH), est recruté au niveau du promoteur grâce à la présence de différents éléments de séquence. Le premier élément de « core promoter » à avoir été découvert est la boîte TATA en 1979 : la séquence TATAWAW est en amont de la plupart des gènes codants dans les organismes modèles (levure, drosophile). Il a été démontré dans ce modèle et *in vitro* que la transcription démarrait systématiquement une trentaine de base en aval de ces boîtes TATA. Depuis, on sait que les TATA box sont moins fréquentes chez l'humain et que de nombreuses autres séquences participent au recrutement du complexe de préinitiation.

Les études de séquençage des ARN en 5' (Cap analysis of gene expression, CAGE) ont montré que la transcription n'est pas initiée à une seule base fixe au niveau du promoteur, mais peut débiter au niveau de plusieurs nucléotides adjacents. Les TSS ont ainsi été classés en deux catégories en fonction de la longueur sur laquelle l'initiation de la transcription a été observée (Carninci et al. 2006). Des TSS larges sont corrélés avec la présence d'ilots CpG et un caractère ubiquitaire des gènes concernés, alors que les TSS étroits sont liés à des gènes tissus spécifiques et ont souvent des boîtes TATA. Pour ajouter à cette complexité, la plupart des gènes ont plusieurs promoteurs forts qui vont ainsi donner lieu à plusieurs transcrits à partir d'un même gène, complétant l'épissage alternatif pour générer de la diversité dans les ARN.

Enfin on peut classifier les promoteurs en fonction des marques épigénétiques qu'ils portent (Lenhard, Sandelin, and Carninci 2012):

- Les gènes qui sont exprimés spécifiquement dans les cellules à des stades de différenciation terminale ont des promoteurs de type 1 et portent des marques H3K4me3 seulement en aval du TSS. Ce sont les gènes aux TSS étroits.
- Les gènes du second type, exprimés ubiquitairement, portent la marque H3K4me3 sur tout leur promoteur au niveau d'ilôts CpG courts et ont souvent plusieurs sites d'initiation de la transcription.
- Enfin, les gènes impliqués dans le développement des organismes ont la particularité d'être associés à une répression par les complexes Polycomb. Ils font partie de larges ilôts CpG et sont recouverts de marques répressives H3K27me3 et/ou activatrices H3K4me3.

#### b. Enhancers

Chez les métazoaires (animaux), la régulation transcriptionnelle des gènes est fréquemment découplée de la seule région promotrice proximale et distribuée entre divers éléments distaux appelés enhancers. Ils se trouvent souvent dans les introns des gènes qu'ils régulent ou ceux des gènes adjacents et dans des cas extrêmes jusqu'à plus d'1Mb du promoteur (Lettice et al. 2003). Malgré leur découverte il y a plus de 40 ans (Banerji, Rusconi, and Schaffner 1981), la connaissance des enhancers a été longtemps limitée par l'absence d'outil exact de mesure de leur localisation. Ils sont identifiés par une forte accessibilité chromatinienne (nous y reviendrons) et une signature spécifique des marques chromatiniennes : la monométhylation de la lysine 4 (H3K4me1) en l'absence de triméthylation (H3K4me3) (Heintzman et al. 2009). Les enhancers actifs étant caractérisés de plus par l'ajout d'H3K27 acétylé et d'une faible quantité d'H3K4me3 (Creyghton et al. 2010; Rada-Iglesias et al. 2011). Ces modifications sont variables d'un type cellulaire à l'autre et ce nettement plus que les différents profils de modification d'histone visibles au niveau des promoteurs. Il est donc probable que ce soient les enhancers qui soient les plus importants dans la régulation tissu spécifique des gènes (Heintzman et al. 2009).

Les enhancers sont de courts segments d'ADN (une à trois cent paires de bases) qui servent de plateforme de recrutement des facteurs de transcription via la présence de motifs spécifiques de l'ADN (Bulger and Groudine 2011). Ces facteurs de transcription interagissent avec le complexe coactivateur Mediator ou le facteur de transcription IID (TFIID) du complexe initiateur de la transcription qui permet ensuite le recrutement de l'ARN polymérase (Malik

and Roeder 2010). Par ailleurs, l'effet activateur des enhancers passe aussi indirectement par le recrutement de modificateurs de la chromatine (histone acetyltransférases, complexes remodeleurs), au niveau du promoteur du gène cible, le rendant plus accessible.

Les super-enhancers sont des regroupement de plusieurs enhancers rapprochés par des boucles de la chromatine et qui ont un fort pouvoir d'activation de la transcription (Hnisz et al. 2013).

#### c. Silencers

Par opposition aux enhancers, les silencers sont définis comme des régions reconnues par des facteurs de transcription mais capables de réprimer des gènes cibles (Pang et al. 2023). Les facteurs de transcription qui s'y fixent recrutent des facteurs corépresseurs ou diminuent l'accessibilité des promoteurs. Les silencers sont les moins étudiés de tous les éléments régulateurs distaux et ne sont pas identifiés à l'échelle du génome car ils n'ont pas de marques d'histone spécifiques. Ils peuvent cependant être reconnus par une forte accessibilité chromatinienne associée à la présence de facteurs de transcriptions répresseurs. Des études confirmatives de corrélation négative avec l'expression d'ARN sont ensuite nécessaires.

#### d. Insulators

Les « insulators » sont des séquences d'ADN qui sont capables de protéger un gène des signaux environnements (West, Gaszner, and Felsenfeld 2002). Il existe deux mécanismes principaux :

- Le premier mécanisme consiste à bloquer l'action d'un enhancer distal. Il nécessite que l'insulator soit situé entre l'enhancer et le promoteur et recrute la protéine CTCF (Bell, West, and Felsenfeld 1999). Les mécanismes de ce blocage ne sont pas complètement élucidés et pourraient impliquer un blocage physique de la formation de boucle, ou une inhibition de la vague de modification d'histone induite par les enhancers.
- Le deuxième rôle des insulators est de bloquer les contacts d'un gène avec la chromatine condensée environnante qui pourraient induire une compaction chromatinienne et une répression de longue durée.

Enfin les insulators jouent un rôle dans l'attachement à la matrice nucléaire

e. Transcription non conventionnelle : ARN enhancers et promoteurs rétrotransposons  
Les études par CAGE montrent un contrôle de la transcription nettement moins strict que ce qui était décrit antérieurement. Ainsi des études ont montré que l'expression des gènes ont un certain degré de stochasticité, que la plupart des promoteurs sont le lieu d'une transcription bidirectionnelle, et que l'on retrouve de la transcription au niveau de zones qui ne sont pas classées comme des promoteurs.

La transcription se retrouve entre autres au niveau de plus de 200 000 rétrotransposons du génome qui sont capables de jouer le rôle de promoteurs alternatifs d'ARN non codants ou même d'ARN codants (Frith et al. 2008). Ces transcrits sont retrouvés à un niveau d'expression faible ou moyen par rapport aux ARN messagers et sont très spécifiques de tissus (Cohen, Lock, and Mager 2009; Faulkner et al. 2009).

Par ailleurs, il est démontré qu'une fraction importante de l'activité transcriptionnelle débute en dehors de tout promoteur au niveau des exons des gènes codants les protéines et de leur région 3'UTR (Carninci et al. 2006). Cette transcription n'est pas toujours corrélée à celle du transcrit conventionnel et semble spécifique des tissus. Les transcrits antisens initiés en 3'UTR ont été décrits comme pouvant réguler l'expression du transcrit sens par un mécanisme antisens.

Enfin les enhancers sont le lieu de transcription bidirectionnelle à partir de leur séquence régulatrice et donnent lieu à des ARN non codants, les enhancers RNA (eRNA) (T.-K. Kim et al. 2010; Kristjánsson et al. 2020). Ces transcrits se distinguent de ceux issus du promoteur par leur caractère instable (Core et al. 2014). Pour la plupart, aucune fonction n'a été démontrée. Cependant des études montrent la possibilité qu'ont ces ARN d'établir directement des interactions longues distances avec le promoteur et de réguler la transcription (Ørom et al. 2010). Ce phénomène pourrait participer au mécanisme d'action des enhancers.

#### **1.4 Facteurs de transcription, coactivateurs et corépresseurs.**

Aussi importants que soient les complexes modificateurs de la chromatine et les méthylases d'ADN dans les processus de régulation transcriptionnelle, ce ne sont pas eux qui vont permettre de choisir les sites spécifiques d'initiation de la transcription. Les protéines qui



sont responsables de cette tâche sont les facteurs de transcription. Ces protéines possèdent deux domaines caractéristiques : un domaine de liaison à l'ADN et un deuxième domaine qui est responsable de la régulation transcriptionnelle (Mitchell and Tjian 1989). Les motifs d'ADN reconnus par les facteurs de transcription ont été caractérisés à grande échelle grâce à deux types d'expériences, l'immunoprécipitation de la chromatine (CHIP-seq) et les mesures d'interaction *in vitro* couplées à des algorithmes bio-informatiques de sélection (SELEX (Slattery et al. 2011), SMILE-seq (Isakova et al. 2017)). Les facteurs de transcription d'une même famille ont souvent des motifs de fixation extrêmement similaires *in vitro* (Figure 6). Cependant *in vivo* on observe une fixation plus différenciée, ce qui laisse supposer que le motif de fixation à l'ADN des facteurs de transcription dépend aussi de l'interaction avec d'autres protéines régulatrices.

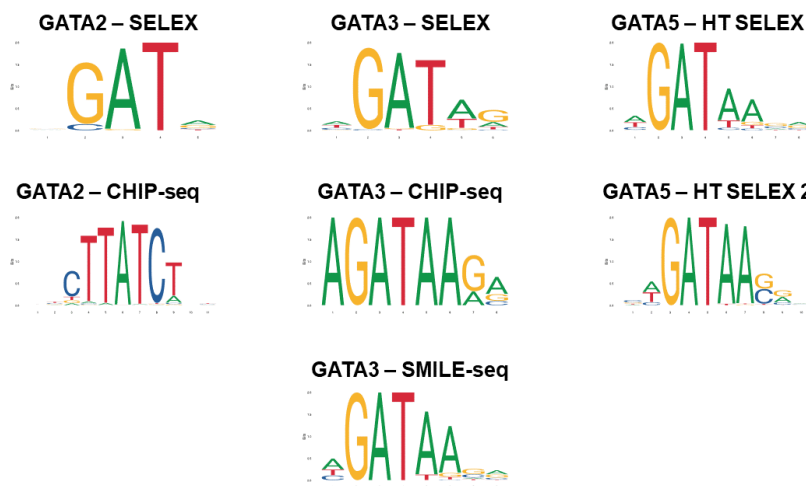


Figure 6 - Exemples de motifs reconnus par les facteurs de transcription GATA3 d'après différentes méthodes (issu de la base JASPAR 2022)

Les facteurs de transcription peuvent être constitutivement exprimés (Sp1, CCAAT, NF1) dans toutes les cellules ou leur action peut être différenciellement régulée en fonction de l'état de différenciation de la cellule (GATA, HNF, HOX...) ou de certains signaux sécrétés (famille des récepteurs nucléaires aux stéroïdes GR, œstrogènes ER, progestérone PR, hormones thyroïdiennes (TR) ou à l'acide rétinoïque RARs, RXRs), des signaux internes (p53, SREBP), ou encore l'activation de voies de signalisation via des récepteurs de surface (STATs, SMADs, NFκB, Notch, NFAT) (Brivanlou and Darnell 2002). Les signaux d'activation peuvent modifier leur capacité d'activation de la transcription en favorisant leur translocation dans le noyau, en modifiant leur conformation (acide rétinoïque et RXR), et/ou en apportant des

modifications post-traductionnelles (phosphorylation des protéines STAT). Dans leur forme active, les facteurs de transcription se lient à l'ADN et recrutent des coactivateurs

Un des principaux complexes recruté est le complexe Mediator qui permet le lien avec les facteurs généraux d'initiation de la transcription (TFIID, TFIIB, TFIIF, TFIIIE et TFIIH) qui eux-mêmes interagissent avec l'ARN polymérase II. Les autres protéines coactivatrices nécessaires sont des modificateurs d'histones, en particulier des histones acétyltransférases, et des remodeleurs de la chromatine notamment SWI-SNF (Cosma 2002; Torchia, Glass, and Rosenfeld 1998).

Les mécanismes d'activation de la transcription ont été extensivement décrits, mais la répression de l'expression des gènes a été moins étudiée. Elle peut être liée au recrutement d'histones déacétylases retrouvées notamment dans les complexes répresseurs NuRD, Sin3 et coREST, d'histones méthylases EZH1 et EZH2 membres du complexe répresseur polycomb2 (PRC2) ou des histones ubiquitines ligases RING1A et RING1B membres du complexe PRC1 (Lai and Wade 2011; Margueron and Reinberg 2011). Certains de ces complexes contiennent aussi des remodeleurs de la chromatine comme le complexe NuRD qui comprend les unités de remodelage de la chromatine Mi-2a et Mi-2b.

Certains facteurs de transcription en fonction de leur état d'activation vont recruter tantôt des facteurs activateurs tantôt des facteurs répresseurs. On peut citer l'exemple des récepteurs aux hormones nucléaires dont fait partie le récepteur à l'acide rétinoïque. En l'absence d'acide rétinoïque, il est associé à des complexes corépresseurs composés de protéine de liaison N-COR ou SMRT et à des molécules dont l'histone désacétylase HDAC3, TBL1X et TBL1XR1 qui inhibent la transcription (Yoon et al. 2003). La liaison de l'acide rétinoïque favorise un changement conformationnel qui désengage le complexe corépresseur, et le récepteur RAR se lie alors à un complexe coactivateur composé des protéines NcoA ou SRC (Nuclear receptor coactivator) et d'activateurs de la transcription (CREBBP/EP300, cARM1).

Enfin certains facteurs de transcription ont un effet direct sur la formation de boucles de l'ADN. Ainsi, il a été démontré que les facteurs de transcription de la famille GATA sont capables de lier deux molécules d'ADN à deux endroits éloignés du génome (Y. Chen et al. 2012) ce qui pourrait permettre la formation de boucles entre les promoteurs de gènes

régulés conjointement. On retrouve notamment ce phénomène au niveau de la région génomique contenant les gènes IL4, IL5 et IL13 dont l'expression est sous le contrôle de GATA3 (G. R. Lee et al. 2003).

## 2 Mesure de l'accessibilité chromatinienne

### 2.1 Généralités

L'accessibilité chromatinienne correspond au degré avec lequel l'ADN chromatinisé peut être atteint physiquement par des macromolécules. Elle est ainsi déterminée par le positionnement des nucléosomes et des facteurs liant la chromatine qui limitent l'accès à l'ADN, mais aussi par l'état de compaction global de la chromatine. Le génome accessible correspond seulement à 2-3% du génome total (Thurman et al. 2012). A l'inverse, plus de 90% des endroits du génome où sont liés des facteurs de transcription sont caractérisés comme accessibles. La mesure de l'accessibilité chromatinienne est ainsi un moyen direct de cibler les régions régulatrices du génome.

### 2.2 Historique

Les premières mesures de l'accessibilité chromatinienne ont été réalisées dans les années 1970 grâce à l'utilisation d'endonucléases pour fragmenter l'ADN suivie de Southern Blot. Elles ont permis de démontrer pour la première fois l'existence de protéines régulièrement positionnées le long du génome et gênant l'accès de l'endonucléase (Hewish and Burgoyne 1973). Ces protéines étaient en fait les octamères d'histone, et cette expérience est la première mise en évidence de la périodicité du positionnement des nucléosomes. Après l'endonucléase c'est la DNase I qui a été utilisée pour fragmenter la chromatine et découvrir les régions régulatrices (appelés alors sites d'hypersensibilité à la DNase) toujours en association au Southern Blot (C. Wu, Wong, and Elgin 1979). Il a fallu attendre 2006 pour avoir la première méthode de mesure de l'accessibilité chromatinienne à grande échelle grâce au développement du DNase-CHIP, quantifiant sur une micropuce les fragments obtenus par digestion à la DNase et amplifiés par PCR (Crawford et al. 2006; Sabo et al. 2006). Ce protocole a ensuite été adapté au séquençage à haut débit (Boyle et al. 2008; Hesselberth et al. 2009) permettant de démontrer que la majorité (80%) des sites

d'hyperméthylation à la DNase ne se trouvent pas dans les promoteurs mais au niveau d'enhancers distaux.

En parallèle une deuxième approche a été développée pour mesurer l'accessibilité chromatinienne en profitant de phénomène de « crosslinking » entre protéines et ADN créé par le formaldéhyde pour séparer les fragments d'ADN libres des fragments liés aux histones (Giresi et al. 2007; Nagy et al. 2003).

Ces deux méthodes étaient relativement longues et nécessitaient un très grand nombre de cellules et ont donc progressivement été remplacées par l'ATAC-seq. L'ATAC-seq (assay for transposase-accessible chromatin using sequencing ou quantification de l'accessibilité chromatinienne à la transposase couplée au séquençage) a été décrite pour la première fois en 2013 (Buenrostro et al. 2013). Utilisant la Tn5, une version mutée hyperactive de la transposase bactérienne Tn5 (Goryshin and Reznikoff 1998), cette méthode a permis la mesure de l'accessibilité chromatinienne en partant d'une quantité minime de cellules (500-2000). Deux protocoles ont ensuite été proposés, le premier pour augmenter la rapidité de la méthode sur un faible nombre de cellules sanguines, le FAST-ATAC-seq (Corces et al. 2016) et le deuxième pour augmenter la qualité du signal en partant de sections de tissus congelés, l'OMNI-ATAC-seq (Corces et al. 2017; Cusanovich et al. 2015). Enfin le protocole a rapidement été adapté pour une utilisation à l'échelle unicellulaire (Buenrostro et al. 2015). Depuis le single-cell ATAC-seq peut être couplé à l'étude des protéines, ASAP-seq (Mimitou et al. 2020) ou du transcriptome (J. Cao et al. 2018; S. Chen, Lake, and Zhang 2019).

### 2.3 L'ATAC-seq

Le principe de l'ATAC-seq est donc d'exposer des cellules ou noyaux perméabilisés à la transposase Tn5 qui, par diffusion, entre dans les noyaux et sectionne l'ADN accessible en insérant des amorces qui seront par la suite utilisées pour l'amplification.

- a. Plusieurs protocoles de préparation des noyaux et cellules

Il existe plusieurs protocoles d'ATAC-seq qui diffèrent essentiellement par la première phase expérimentale, celle de la préparation des noyaux ou cellules (Figure 7).

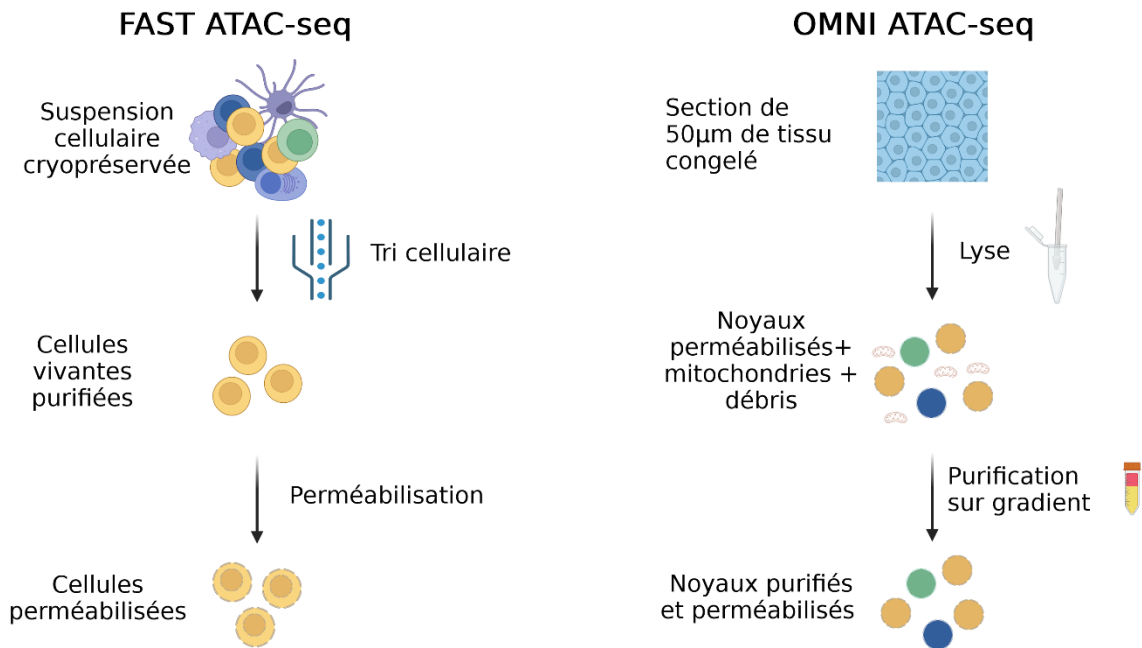


Figure 7 - Préparation des noyaux et cellules pour l'ATAC-seq

Le FAST ATAC-seq est un protocole qui nécessite d'utiliser des cellules fraîches ou qui ont été décongelées à partir d'échantillons cryopréservés en DMSO (diméthylsulfoxyde). Les suspensions cellulaires sont triées par cytométrie en flux sur leur viabilité et éventuellement d'autres marqueurs antigéniques. Ceci permet de sélectionner des populations cellulaires d'intérêt. Dix-mille cellules triées sont ensuite perméabilisées et utilisées pour la suite du protocole

L'OMNI-ATAC seq permet d'exploiter des tissus congelés en azote liquide qui sont plus fréquemment disponibles en cancérologie notamment dans les pathologies lymphomateuses. Des sections de 50µm d'épaisseur sont lysées dans un tampon contenant du NP40 qui permet de perméabiliser les noyaux résiduels dans le même temps. La destruction du tissu est aidée par l'utilisation d'un mortier homogénéisateur (dounce). Le lysat est purifié sur gradient d'iodixanol pour éliminer les mitochondries et les débris et ne garder que les noyaux.

#### b. Transposition et amplification

L'enzyme Tn5 a été initialement retrouvée dans les bactéries *Escherichia* et *Shewanella*. Elle fait partie de la famille des transposases qui comprend aussi entre autres la transposase

Sleeping beauty. Ces protéines sont capables de cliver une molécule donneuse d'ADN au niveau de deux séquences consensus, emporter le fragment d'ADN médian (transposon) et le réinsérer à un autre endroit du génome (sans restriction de séquence particulière). Il existe tout de même un biais, la Tn5 insère plus facilement son transposon dans des régions riches en GC. Étonnement le fragment d'ADN entier n'est pas nécessaire à l'insertion dans le nouveau génome ; seules les deux séquences consensus appelées mosaic ends (ME) le sont. L'utilisation d'une Tn5 hyperactive chargée avec des molécules d'ADN contenant les séquences ME et des adaptateurs de séquençage a ainsi permis de regrouper en une seule étape la fragmentation et la ligation des adaptateurs des bibliothèques de séquençage (Adey et al. 2010).

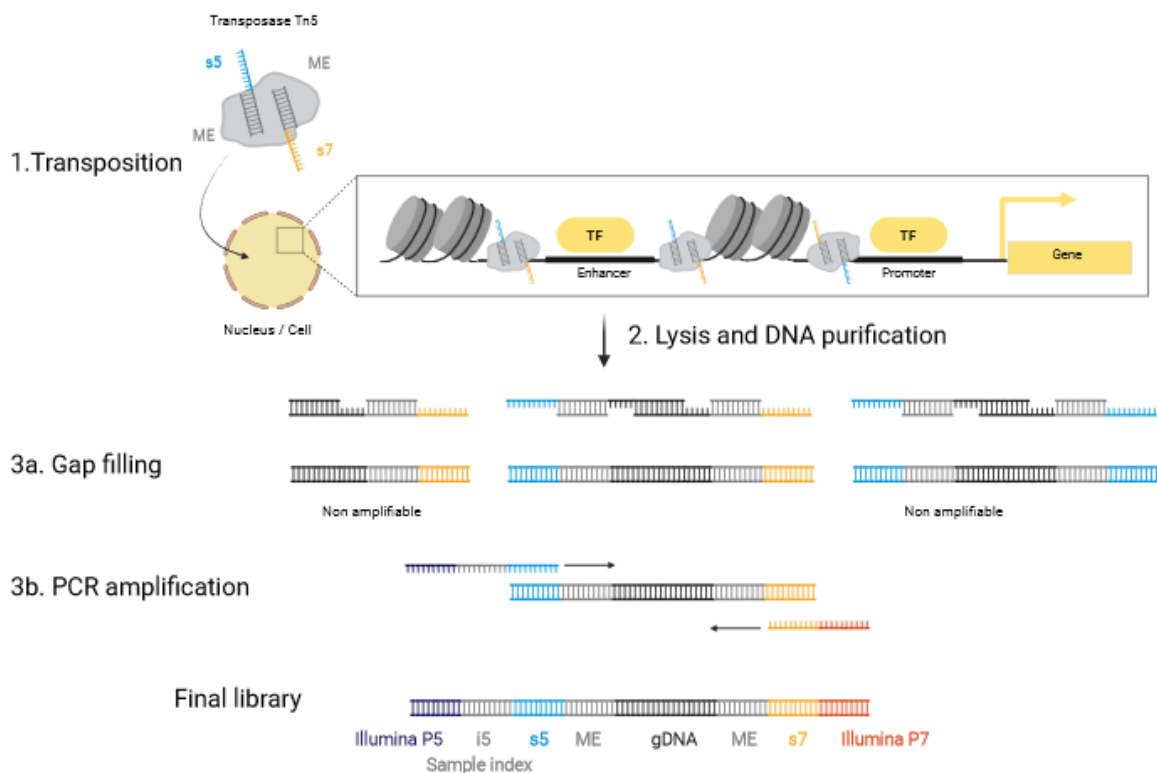


Figure 8 - Transposition et amplification en ATAC-seq

Pour en revenir à l'ATAC-seq, une fois les cellules/noyaux perméabilisés, ils sont incubés avec la Tn5 qui fragmente l'ADN s'il est accessible ajoutant dans un même temps des séquences s5 et s7 (Figure 8, Etape 1). Les noyaux sont ensuite lysés et l'ADN purifié (Figure 8, Etape 2). Après une première phase d'élongation permettant de remplir les brèches créées par la Tn5 (3.a), une seule réaction de PCR va amplifier les fragments qui ont été générés (3.b). Cette

réaction peut varier en nombre de cycles en fonction des protocoles et de la quantité d'ADN obtenu, mais elle utilise toujours un temps d'élongation court (< 1 minute) qui permet d'enrichir en fragments courts (zones très accessibles). Enfin la PCR est aussi l'occasion d'ajouter les index d'échantillons qui serviront au démultiplexage et les séquences P5 et P7 qui permettent la clusterisation sur les « flow cell » Illumina. Le produit de PCR est ensuite quantifié et les bibliothèques de plusieurs échantillons sont mélangées et séquencées en « paired-end » sur 75 ou 150 cycles (2 reads de 37 ou 75 bases).

## 2.4 *Traitement bioinformatique des données d'ATAC-seq*

### a. Principales étapes préanalytiques

Une fois le séquençage terminé, les données de chaque échantillon sont démultiplexées grâce à l'index i5. Le pipeline d'analyse contient ensuite les étapes habituelles de traitement des données de séquençage :

- « Trimming » des séquences adaptateurs (ME,s5,s7) restant sur les reads (cutadapt, trimmomatic)
- Alignement des reads au génome de référence en utilisant des logiciels d'alignement adaptés aux reads courts (bowtie, bwa-mem2)
- Marquage des duplicats (samblaster, markduplicate)
- Elimination des reads non ou mal alignés, des alignements secondaires, et des reads couvrant l'ADN mitochondrial (samtools).

Il existe ensuite des particularités d'analyse. Contrairement au RNA-seq, la couverture ne peut être calculée sur toute la longueur des reads car l'information d'accessibilité est donnée par la position des extrémités du read. Pour prendre un exemple, la transposase peut couper l'ADN de part et d'autre d'un nucléosome, le read séquencé couvrant alors les parties accessibles aux extrémités mais aussi toute la région couverte par le nucléosome qui, elle, n'est pas accessible. On ne s'intéresse donc qu'aux sites de transposition à l'extrémité des reads pour la suite des analyses (Figure 9). Ensuite ces positions doivent être décalées de + 4pb pour le brin positif de l'ADN et -5bp pour le brin négatif pour compenser l'ajout de nucléotides de l'étape de « gap filling » (Figure 8) (Buenrostro et al. 2013).

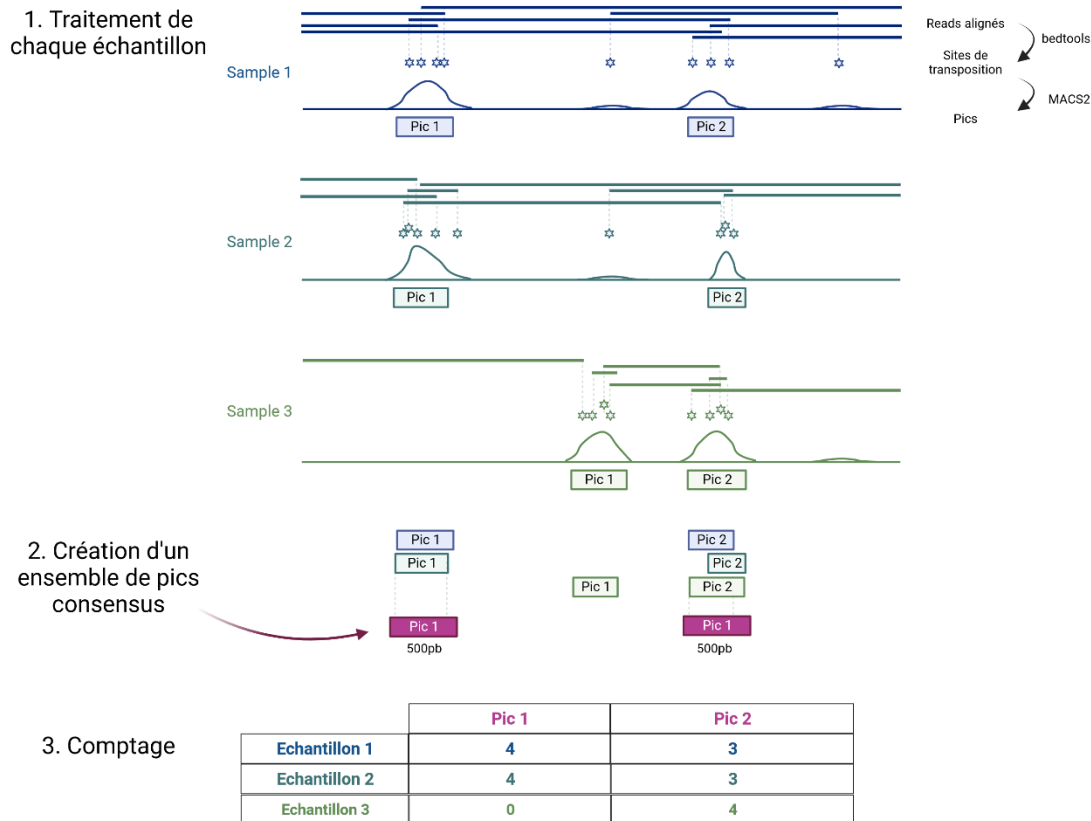


Figure 9 - Traitement bioinformatique des données d'ATAC-seq

L'étape suivante consiste à trouver quelles sont les régions ouvertes. Contrairement au RNA-seq, pour lequel on utilise des bases de données de transcrits bien identifiées, il n'existe pas de liste universelle des régions ouvertes de la chromatine. On doit les identifier à chaque échantillon en recherchant des zones où il existe une accumulation de sites d'insertion qui indiquent une grande accessibilité, grâce à des logiciels de « peak calling ». L'outil le plus fréquemment utilisé, notamment par le consortium d'étude épigénétique ENCODE, est le logiciel MACS2 (Yong Zhang et al. 2008). Ce logiciel cherche les endroits où le nombre des sites d'insertion dévie par rapport au bruit de fond en se basant sur une distribution de Poisson pour définir leur significativité. Contrairement au CHIP-seq, il n'existe pas de contrôle négatif et le bruit de fond doit donc être évalué par la couverture moyenne locale sur une fenêtre de quelques kilobases autour du pic ou globale le long de tout le génome. Il existe d'autres outils d'identification des régions ouvertes, basés sur les comptes comme Homer et Zinba, sur la forme des pics (Polyapeak) ou bien créés spécifiquement pour l'ATAC-seq à partir de modèles de Markov (Tarbell and Liu 2019).



Une fois les pics définis individuellement pour chaque échantillon (Figure 9), il est nécessaire de créer un set consensus de régions ouvertes en comparant la position des pics entre tous les échantillons. L'approche la plus communément décrite consiste à sélectionner un pic s'il est présent dans 2 échantillons différents et de choisir les coordonnées du plus significatif (iterative overlap filtering) (Corces et al. 2016). Pour faciliter les analyses suivantes, les pics sont ensuite réduits à une taille de 500 paires de bases ce qui correspond à la taille moyenne des régions ouvertes observées au niveau des enhanceurs et des promoteurs. Les pics situés dans des régions répétées ou caractérisées comme problématiques en ATAC-seq sont éliminés.

Enfin, le nombre de sites d'insertion dans chaque pic est compté pour chaque échantillon pour obtenir la matrice de compte pics\*échantillons. Elle donne le degré d'accessibilité de chaque échantillon au niveau des régions ouvertes de la chromatine et est utilisée comme principale base des analyses d'ATAC-seq, similairement aux matrices de comptes du RNA-seq.

#### b. Analyse de facteurs de transcription

Une autre donnée inférée à partir des données d'ATAC-seq est l'activité des facteurs de transcription. Comme expliqué précédemment, les facteurs de transcription sont capables de reconnaître des séquences spécifiques de l'ADN ou motifs. En recherchant la présence de ces motifs au sein des régions ouvertes de la chromatine découvertes en ATAC-seq, on peut estimer indirectement l'« activité » de ces facteurs de transcription. En effet, qu'un facteur de transcription soit répresseur ou activateur, on observe une ouverture chromatinienne au niveau de son motif (Buenrostro et al. 2015). Mais les facteurs répresseurs induisent autour de leur motif une empreinte sur l'accessibilité chromatinienne différente de celle des activateurs. Ainsi l'accessibilité chromatinienne diminue très rapidement en dessous de la moyenne du génome dans les 10 bases autour du motif de facteurs de transcription répresseurs (Berest et al. 2019). A l'opposé, les facteurs activateurs ont une augmentation de l'accessibilité chromatinienne autour de leur motif et ce jusqu'à 100 bases de part et d'autre.

L'empreinte d'un facteur de transcription est classiquement obtenue en sommant l'accessibilité chromatinienne au niveau de toutes les positions du génome contenant son motif (Bentsen et al. 2020; 2020; Berest et al. 2019). Ces méthodes nécessitent une couverture de séquençage importante et souvent d'assembler les résultats de plusieurs réplicats. D'autres méthodes, adaptées notamment au single-cell, assimilent l'activité d'un

facteur de transcription à la moyenne de l'ouverture des pics qui contiennent son motif (Schep et al. 2017).

Quelle que soit la méthode, le biais d'action de la transposase Tn5 qui fragmente préférentiellement les zones riches en GC doit être compensé.

Toutes ces méthodes sont basées sur des motifs qui peuvent être similaires entre facteurs de transcription de la même famille et ne permettent donc pas d'affirmer formellement le statut d'activation d'un facteur de transcription (exemple de la famille GATA, Figure 6). L'intégration avec des données de RNA-seq permet en partie de compenser ces limites (Berest et al. 2019; Corces et al. 2016).

### c. Contrôles qualité

Lors de l'analyse des données d'ATAC-seq, plusieurs paramètres de contrôle qualité sont mesurés (qualité d'alignement, quantité d'ADN mitochondrial, taux de duplicats, nombre de site d'insertion et enrichissement au TSS). Le paramètre déterminant est l'enrichissement aux sites d'initiation de la transcription (TSS). Celui-ci est calculé en moyennant le signal d'accessibilité au niveau de tous les sites d'initiation de la transcription du génome rapporté au bruit de fond adjacent (Figure 10).

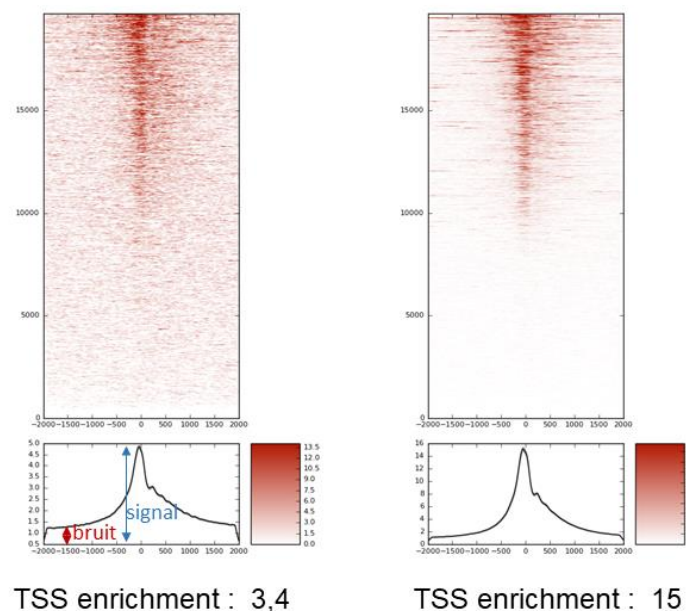


Figure 10 - Exemples de profils d'enrichissement au TSS pour un échantillon de mauvaise (à gauche) et bonne qualité (à droite)

Ce paramètre peut être influencé par une mauvaise conservation des échantillons avant congélation, la quantité de nécrose des prélèvements mais aussi le taux de neutrophiles de l'échantillon.

Le nombre de sites d'insertion corrélé à la taille de la librairie semble moins déterminant pour identifier des types cellulaires. Ainsi, des librairies de différents sous-types de lymphocytes T sont différenciables à partir des profils ATAC-seq où on a échantillonné aléatoirement seulement 1000 - 10000 fragments (Satpathy et al. 2018). En revanche, il faut une certaine profondeur pour permettre les analyses des empreintes de facteurs de transcription.

## 2.5 Exemples d'application de l'ATAC-seq

En réduisant le nombre de cellules nécessaire, la mise au point de l'ATAC-seq a permis d'étendre l'étude de l'accessibilité chromatinienne à des échantillons précieux, notamment issus de prélèvements de patients ou de populations cellulaires rares.

### a. Accessibilité chromatinienne et lymphocytes T

Plusieurs études se sont attachées à étudier la différenciation lymphocytaire T grâce à l'ATAC-seq. Toutes ces études montrent qu'il existe une grande différence épigénétique entre des cellules naïves et des cellules mémoires. Cette différence est nettement supérieure à celle observée entre lymphocytes CD4+ et CD8+, entre plusieurs phénotypes helper ou encore entre lymphocytes effecteurs et centraux mémoires. En effet, on retrouve systématiquement un grand nombre de régions différentiellement ouvertes entre populations mémoires qu'elles soient effectrices ou centrales et les lymphocytes naïfs (Calderon et al. 2019; Giles et al. 2022; Moskowitz et al. 2017). Dans les lymphocytes CD4+, ce changement d'état épigénétique au cours de la transition naïve-mémoire, correspond à la fermeture d'un grand nombre de régions chromatiniennes (Satpathy et al. 2018). Les pics les plus ouverts dans les lymphocytes T naïfs sont associés à la prolifération T, au signal du TCR et à la recombinaison VDJ (Moskowitz et al. 2017). A l'inverse, les pics les plus ouverts dans les lymphocytes CD8+ TCM comparés aux CD8+ naïfs se situent dans des gènes impliqués dans la réponse immunitaire adaptative et en particulier dans l'adhésion cellulaire.

Un changement d'activité de facteurs de transcription accompagne cette transition. Ainsi les motifs des facteurs de transcription de la famille bZIP ( AP-1, FOS/JUN..) ont une forte accessibilité dans toutes les populations non-naïves (Giles et al. 2022). Les facteurs de transcription qui semblent les plus actifs dans les lymphocytes naïfs sont SOX4, KLF3, KLF12, HSF2. Par ailleurs d'autres facteurs de transcription comme NF-KB sont plus actifs dans les lymphocytes centraux mémoires.

Les populations centrales mémoires et effectrices mémoires des lymphocytes T CD8+ sont aussi bien différenciées par l'activité des facteurs de transcription. Les TEM montrant une plus forte activité de TBX21. Les TEMRA à l'extrémité de cette trajectoire de différenciation portent une forte ouverture des pics associés aux gènes de la famille KLR (KLRC2, KLRF1, KIR3DL1, etc.) en plus d'une très forte accessibilité aux motifs de TBX21.

La différenciation des lymphocytes T CD4 helper est plus difficile à mettre en évidence en ATAC-seq. En effet, les différences entre Th17 et Th1 ne sortent que sur les composantes 3 et 4 d'une analyse en composantes principales. Mais on peut de manière supervisée observer tout de même les différences attendues : une ouverture de pics dans le gène IL2RA dans les Treg, IFNG dans les TH1 et IL26 et IL22 dans les TH17 (Satpathy et al. 2018).

Enfin, il a été démontré que lors d'une stimulation des lymphocytes T avec des billes CD3/CD28 et l'IL-2, on observe une forte transformation du paysage épigénétique liée à l'ouverture d'un grand nombre de régions chromatinienne. Une grande proportion de ces régions ouvertes inductibles par stimulation est partagée entre les lymphocytes T CD4+, CD8+ et les lymphocytes T  $\gamma\delta$  (Calderon et al. 2019). Ce résultat n'était pas attendu puisqu'une étude avait antérieurement montré que le profil d'accessibilité chromatinienne de différentes cellules lymphoïdes innées était très stable après stimulation par des interleukines contrairement aux importantes différences transcriptomiques observées (Shih et al. 2016).

b. Accessibilité chromatinienne, tissu spécificité et classification en cancérologie

Les premières études publiées sur l'ATAC-seq ont montré un fort potentiel de cette technique pour discriminer des sous-types cellulaires normaux. La pureté des clusters obtenus en utilisant des profils d'accessibilité chromatinienne de précurseurs hématopoïétiques était même supérieure à celle obtenue à partir de RNA-seq (Corces et al. 2016). Ce même constat

a été fait pour la différenciation des sous-types de lymphocytes T dans une étude (Calderon et al. 2019) mais n'est pas rapporté dans une autre (Giles et al. 2022). Il semble que l'information sur les sous-types cellulaires se situe essentiellement dans des régions éloignées des TSS qui correspondent en partie à des enhanceurs et dont l'accessibilité est corrélée à l'expression des gènes adjacents (Corces et al. 2017; Giles et al. 2022; H. Yoshida et al. 2019). En effet, le clustering des cellules en utilisant seulement ces régions distales est plus pure que lorsque l'on inclut les régions proches des TSS. Ces dernières donnant lieu à un clustering proche de celui obtenu en RNA-seq.

Cette information de spécificité de tissu semble très intéressante pour aider à classifier les différents types de cancers. L'hypothèse a été testée en mesurant les profils d'accessibilités de 23 types de tumeurs solides de l'ATLAS TCGA (Corces et al. 2018). Le clustering résultant des pics distaux montrait une forte concordance avec les types de cancers démontrant la capacité discriminante de l'ATAC-seq. De plus, cette étude a démontré le potentiel de l'ATAC-seq à découvrir de nouveaux sous-types de cancers tels que des carcinomes rénaux à cellules papillaires différenciées, des carcinomes invasifs mammaires ou des adénocarcinomes de prostate. En hématologie, il a été démontré que les leucémies lymphoïdes chroniques (LLC), présentent des profils d'accessibilité différents en fonction du statut mutationnel de leur immunoglobuline. D'autre part, l'ATAC-seq a permis d'identifier un sous-type intermédiaire de LLC précédemment décrit dans des études de méthylation (Rendeiro et al. 2016).

Il est intéressant de noter que dans les leucémies lymphoïdes chroniques, comme dans les leucémies aiguës myéloïdes, la majorité de la variabilité entre profils d'accessibilité d'échantillons cancéreux est expliquée par les trajectoires de différenciation hématopoïétique normale (Corces et al. 2016; Rendeiro et al. 2016). L'ATAC-seq a ainsi été utilisé pour identifier les contreparties normales de ces pathologies. Cependant, en utilisant le single-cell ATAC-seq on observe entre les cellules tumorales une hétérogénéité des stades de maturation. Ainsi les cellules souches de leucémies aiguës myéloïdes présentent un profil proche des précurseurs myéloïdes (GMP, LMPP) alors que les blastes présentent une distribution plus étalée variant des GMP aux monocytes (Corces et al. 2016). Le profil d'accessibilité chromatinienne, lorsqu'il est réalisé en bulk, représente une moyenne de l'accessibilité sur une population cellulaire et peut ainsi cacher des informations sur les cellules d'origine.

Au total, la mesure de l'accessibilité chromatinienne par ATAC-seq semble être efficace pour discriminer différents types de cancers et les comparer à leurs contreparties normales.

# C OMIQUE ET CLASSIFICATION EN CANCEROLOGIE

Les approches dites omiques (génomique, transcriptomique, protéomique, métabolomique et épigénomique) visent à caractériser et/ou quantifier à large échelle différents groupes de molécules biologiques (ADN, ARN, protéines...). Les améliorations technologiques majeures des dernières décennies ont permis d'obtenir une large quantité de données et ont nécessité le développement de méthodes spécifiques pour leur analyse. Des méthodes d'apprentissage non supervisé ont joué un rôle important pour affiner la classification de nombreux cancers tandis que des méthodes supervisées offrent la possibilité d'améliorer l'évaluation diagnostique, pronostique ou théranostique des patients.

## 1 Apprentissage non supervisé

### 1.1 Clustering

L'analyse par clustering des échantillons est une approche extrêmement répandue pour classer des échantillons et découvrir des sous-types cancéreux en omique. Elle consiste à essayer de regrouper des échantillons par clusters basés sur leur similarité. Pour évaluer la proximité de deux individus, il faut déterminer la distance qui les sépare grâce à différentes métriques. Nous pouvons citer ici trois des distances les plus utilisées en omique :

- La distance euclidienne :  $d_{ab} = \sqrt{\sum_i (e_{ai} - e_{bi})^2}$
- La distance de Manhattan :  $d_{ab} = \sum_i |e_{ai} - e_{bi}|$
- La distance de Pearson :  $d_{ab} = 1 - r_{ag}$ , avec  $r_{ab} = \frac{\sum_i (e_{ai} - \bar{e}_a)(e_{bi} - \bar{e}_b)}{\sqrt{\sum_i (e_{ai} - \bar{e}_a)^2} \sqrt{\sum_i (e_{bi} - \bar{e}_b)^2}}$

où  $d_{ab}$  est la distance entre un échantillon a et un échantillon b, et  $e_{ai}$  le niveau d'expression (de transcription, d'accessibilité, de méthylation...) d'un gène i pour l'échantillon a. Le choix de l'une ou l'autre des métriques de distance dépend de ce que l'on veut observer dans les données. La distance euclidienne et la distance de Manhattan sont sensibles aux différences de niveau moyen d'expression qui peuvent être dues à des difficultés de normalisation. La distance de Pearson en revanche ne l'est pas.

Les méthodes de clustering peuvent être divisées en quatre classes : les clustering hiérarchiques, les méthodes de partitionnements (partitioning), les méthodes basées sur des modèles et les méthodes basées sur la densité. Dans le clustering hiérarchique, les individus

sont reliés entre eux grâce à une fonction de « linkage » qui dépend de la distance. Le clustering hiérarchique agglomératif commence avec des clusters constitués d'un seul échantillon, relie entre eux les deux clusters les plus proches puis réitère jusqu'à ce que tous les clusters soient connectés en un arbre. La proximité entre deux clusters est définie différemment selon les algorithmes (single, complete, average, centroid...). Ainsi le clustering par « complete linkage » considère la distance la plus grande entre les membres les plus éloignés des deux clusters. Le single linkage considère la distance la plus courte et le 'average linkage' considère la distance moyenne entre paires d'individus. La méthode de Ward, fréquemment utilisée, consiste à regrouper à chaque étape les clusters qui vont permettre de minimiser l'augmentation de la variance intracluster à l'étape suivante.

La deuxième classe de méthode est celle par partitionnement. Elle consiste à définir au préalable un nombre de clusters et ne suppose pas de lien hiérarchique entre eux. La question du nombre de cluster à définir est donc difficile. La méthode habituelle consiste à réaliser plusieurs partitionnements avec différents nombres de clusters et de choisir le meilleur selon des critères qui ne sont pas consensuels (stabilité, score de silhouette...). Parmi ces algorithmes on compte les méthodes k-means et le partitionning around medoids (Kaufman 1987; Macqueen 1967) qui sont très efficaces dans certaines situations, mais souvent limitées par le fait qu'elles supposent que tous les clusters aient la même taille.

Les méthodes de clustering basées sur des modèles estiment la probabilité d'être dans un cluster à partir d'un modèle qui peut varier entre chaque cluster. La distribution la plus fréquemment utilisée est une distribution gaussienne dont la variance peut être différente d'un cluster à l'autre ce qui permet d'obtenir des clusters de tailles différentes. Une limite de cette approche est qu'elle considère que les clusters auront toujours la même forme.

Une autre méthode de clustering basée sur la densité a été inventée pour permettre la découverte de clusters de formes arbitraires (Ester et al. 1996). Elle part de l'hypothèse que les centres des clusters présentent une forte densité et sont éloignés les uns des autres, mais que la densité de tous les clusters est similaire.

Si les densités sont différentes entre clusters, la méthode précédente peut être mise en défaut. Des méthodes basées sur la recherche de voisins proches partagés (shared nearest



neighbors) ont ainsi été décrites par Ertöz et al (Ertöz, Steinbach, and Kumar 2003). Les étapes de cette méthode sont :

- Rechercher les plus proches voisins de chaque individu.
- Construire le graphe des plus proches voisins : les individus sont reliés entre eux s'ils ont un minimum de k plus proches voisins en commun.
- Trouver la densité de chaque point : en fonction d'un paramètre Eps (à optimiser), trouver pour chaque point, le nombre d'autres points qui ont une similarité à ce point supérieure à Eps.
- Définir les points centraux ('core points') qui ont une densité supérieure à un MinPts (un troisième paramètre à définir).
- Former les clusters : si deux points centraux sont dans un rayon inférieur à Eps l'un de l'autre, ils sont considérés dans le même cluster. Tous les autres points sont assignés au cluster du point central le plus proche. Les points qui sont éloignés de plus de Eps de tous les points centraux ne sont pas assignés à un cluster.

Cette méthode se libère donc des définitions habituelles de distance et définit une nouvelle forme de similarité entre points, éliminant ainsi les problèmes de densités variables et permettant de ne pas inclure les points trop éloignés qui sont considérés comme du bruit. C'est cette méthode que nous avons utilisée dans notre étude.

## 1.2 Réduction de dimension et visualisation

En génomique, le nombre de variables est très élevé, de l'ordre d'une ou de plusieurs dizaines de milliers de gènes en transcriptomique et de centaines de milliers de pics en ATAC-seq. Les mesures habituelles de distance (euclidienne notamment) vont alors souffrir de la « malédiction » de la dimensionnalité (curse of dimensionality). En effet, les distances notamment euclidiennes peuvent ne plus refléter correctement la similarité entre les points lorsque l'on augmente le nombre de variables considérées (Aggarwal, Hinneburg, and Keim 2001).

Il est donc nécessaire de réduire le nombre de variables. Pour cela deux approches sont possibles : la sélection de variables et la réduction de dimensions. En transcriptomique et épigénomique, on utilise souvent une méthode de sélection de 500-10 000 variables suivie

ou non d'une étape de réduction de dimension qui permet de passer à un nombre de variables de l'ordre de 2-100 selon les analyses. La sélection de variables consiste à choisir celles qui vont contenir le plus d'informations intéressantes. En général, les variables constantes sont enlevées et on sélectionne les variables qui ont la plus grande variabilité (estimées par la variance, le coefficient de variation (écart-type rapporté à la moyenne), l'espace interquartile, la déviation par rapport à la médiane...).

Plusieurs méthodes de réduction de dimension sont utilisées en épigénomique dont les principales sont l'analyse en composantes principales, l'analyse en composantes indépendantes, et la factorisation de matrices non négatives (NMF). Ces trois méthodes ont pour but de définir des nouvelles variables qui sont des combinaisons linéaires des variables initiales (niveau d'expression génique par exemple). Ces nouvelles variables sont nommées composantes.

L'analyse en composantes principales (ACP) tend à maximiser la variance le long d'un premier axe, puis un second et ainsi de suite. Chaque nouvel axe étant orthogonal au premier. Cette méthode concentre donc le maximum de variabilité à chaque étape ce qui est moins efficace lorsque l'on veut séparer des signaux. En single-cell ATAC-seq par exemple, la première composante reflète souvent à la fois des biais de qualité mais peut aussi contenir un signal important de l'identité cellulaire (Granja et al. 2021).

L'analyse en composantes indépendantes (ACI) est une méthode qui vise à estimer en même temps les différentes combinaisons linéaires de variables. Ici aussi les composantes sont orthogonales. Plusieurs algorithmes existent et notamment l'algorithme Fast-ICA qui cherche à trouver de nouvelles composantes avec une distribution qui s'éloigne le plus possible d'une gaussienne. Cette méthode permet efficacement de séparer différents signaux à partir de données transcriptomiques de patients atteint de cancers (sexe, type de cancer, composition cellulaire...) (Biton et al. 2014; Nazarov et al. 2019).

Enfin, introduite dans le domaine de la transcriptomique par Brunet (Brunet et al. 2004), la factorisation de matrice non négative (NMF) est la méthode que nous avons utilisée dans notre étude. Elle a la particularité de chercher des combinaisons linéaires et positives de variables appelées métagènes, composantes ou patterns. On obtient ainsi une matrice H représentant le score de chaque échantillon sur une composante et une matrice W qui

montre le poids de chaque gène dans la construction des composantes. La contrainte de non-négativité facilite nettement l'interprétation des composantes, par rapport à l'ACP ou l'ACI. Différents algorithmes permettent cette décomposition et ont pour but de minimiser la différence entre la matrice initiale (A) et le produit de la matrice des échantillons (H) par la matrice des variables (W) (Figure 11).



Figure 11 - Exemple de factorisation non négative de matrice de comptes à partir de profils transcriptomiques de lymphocytes B, T CD4 et T CD8 pour un rang k=2 et k=3.

Dans son article, Brunet démontre que les échantillons peuvent ainsi être clusterisés en fonction de leur niveau d'expression de chaque métagène. Le choix du paramètre k, nombre de composantes, est déterminant. Ainsi en utilisant deux composantes, cet auteur a montré la clusterisation d'une part des leucémies aiguës myéloïdes et d'autre part des leucémies aiguës lymphoïdes, alors qu'à trois composantes un nouveau cluster apparaît et on obtient la séparation entre leucémies aiguës lymphoblastiques T et B. Un exemple similaire Figure 11 montre la ségrégation B vs T puis B vs T CD4 et T CD8. Il n'existe pas de méthode consensuelle pour déterminer le paramètre k. On peut s'aider de différentes mesures (résiduelles, silhouette, coefficient de corrélation cophénétique entre différentes itérations). Différents rangs k donnent des informations différentes, qui peuvent être biologiquement pertinentes à différents niveaux. Ainsi les signaux forts comme le sous type tumoral sont mieux capturés grâce à des modèles entraînés avec un faible nombre de

composantes alors que les signaux plus subtils comme ceux du microenvironnement peuvent nécessiter un grand nombre de composantes pour être découverts (Way et al. 2020).

Plusieurs algorithmes de NMF ont été ainsi appliqués à la découverte de nouvelles classes de cancers (Brunet et al. 2004; Gao and Church 2005; P. M. Kim and Tidor 2003). Des formulations bayésiennes de la NMF ont ensuite été développées qui semblent plus à même d'identifier des patterns associés aux phénotypes et processus biologiques (Fertig et al. 2010; Kossenkov and Ochs 2009).

Hormis ses propriétés de réduction de dimension et de clustering, la NMF a aussi démontré une grande capacité à déconvoluer de manière non supervisée le profil transcriptomique de différents types cellulaires composant un échantillon tumoral (Houseman et al. 2016; Peng et al. 2019).

Ainsi, les méthodes de réduction de clustering non supervisées permettent l'identification de nouveaux sous-groupes de cancers. Pour pouvoir ensuite reconnaître ces sous-groupes de cancers sur de nouveaux échantillons, il faut en revanche utiliser des techniques d'apprentissage supervisé.

## 2 Classification supervisée

### 2.1 *Etapes de l'entraînement d'un modèle*

La classification supervisée nécessite une phase d'entraînement qui va permettre de construire un modèle capable de reconnaître dans les données, des motifs qui sont associés à chaque étiquette (ici à chaque sous type cancéreux). L'entraînement du modèle consistera à montrer les données, tester des paramètres du modèle et optimiser itérativement ces paramètres pour minimiser une fonction de coût qui prend différentes formes selon les modèles.

Les étapes d'entraînement d'un modèle de machine learning (ML) sont les suivantes :

- Séparation du jeu de données en deux parties, un jeu d'entraînement et un jeu de test.
- Choix des variables (sélection de variables et/ou ou réduction de dimensions)

- Normalisation des variables
- Entraînement de différents modèles sur le jeu d'entraînement avec optimisation des paramètres de chaque modèle par crossvalidation (k-fold ou leave-one-out)
- Comparaison des performances et choix du modèle
- Application du modèle choisi à la cohorte test.

Ces étapes sont similaires quel que soit le jeu de données sur lequel on travaille (analyse d'images, de sons, de données numériques, textuelles). Mais, dans le cas de la classification de cancers par méthode omique, il existe de nombreuses spécificités.

## 2.2 *Spécificité de l'application du machine learning aux mesures omiques.*

### a. Dimensionnalité

Tout comme en clustering non supervisé, le nombre de variables mesurées va être un des problèmes clefs. En effet une augmentation linéaire du nombre de variables conduit à une augmentation exponentielle du nombre de combinaisons possibles de paramètres d'un modèle. Ce nombre de combinaisons dépasse alors dramatiquement le nombre d'échantillons observés et les risques de surapprentissage (overfitting) sont majeurs. Les solutions possibles consistent à diminuer le nombre de variables ou augmenter le nombre d'échantillons.

#### i. *Diminution du nombre de variables*

Une première solution consiste à sélectionner les variables intéressantes pour la prédiction en utilisant les connaissances du domaine étudié. On peut même choisir de ne mesurer qu'un nombre réduit de variables en travaillant notamment sur des panels de séquençage. Cette méthode nécessite une très bonne connaissance a priori des maladies étudiées.

Les solutions alternatives se déclinent selon une combinaison de variables non supervisées ou supervisées. Contrairement au clustering non supervisé, on peut en effet en machine learning utiliser une sélection supervisée qui permet d'enrichir en variables utiles à la distinction entre les classes d'intérêt. En génomique il existe notamment deux méthodes préférentielles :

- L'analyse univariée qui utilise des tests statistiques simples (t-test) ou plus complexes (modèles bayésiens d'analyses différentielles) (Aref-Eshghi et al. 2018; Saeys, Inza, and Larrañaga 2007)
- L'entraînement de plusieurs modèles de prédiction (random forest, support vector machine...) sur des sous-ensembles du jeu de variables, suivi de la sélection pour chaque partition des variables les plus importantes à la prédiction (Díaz-Uriarte and Alvarez de Andrés 2006)

Dans le cas d'un choix supervisé de variables, le risque est important de surestimer les performances du modèle sur le jeu d'entraînement. Pour éviter ceci, il est nécessaire de sélectionner les variables à chaque niveau de partitionnement du test (Ambroise and McLachlan 2002).

Le deuxième outil de réduction du nombre de variables est la réduction de dimension. Plusieurs méthodes non supervisées de réduction de dimensions ont été rapportées comme efficaces notamment l'ACP et l'ACI décrites plus haut (C.-H. Zheng et al. 2008). D'autres méthodes, comme celle des moindres carrés partiels (PLS), utilisent l'information de label pour réduire l'espace dimensionnel. Pour éviter l'overfitting ces méthodes ont souvent été dérivées sous des formes parcimonieuses (ou « sparse »), qui limitent le nombre de variables contribuant à chaque composante : sparse-PCA, sparse-ICA, sparse-PLS (Chung and Keles 2010; Hotelling 1933) ce qui permet d'éviter les problèmes de convergence liés à la nette supériorité du nombre de variables par rapport au nombre d'échantillons.

#### *ii. Augmentation du nombre d'échantillons*

Le machine learning a surtout été développé dans des domaines où les sources de données sont importantes (reconnaissance d'image, de texte...). Les approches sur un faible nombre d'échantillons (few sample learning) sont moins développées. Une règle commune des modèles de classification conventionnels suppose que dix observations doivent être disponibles pour chaque paramètre du modèle à optimiser afin d'éviter les risques d'overfitting. Ce nombre d'échantillons est difficilement atteignable en classification en cancérologie, ce qui favorise l'utilisation de modèles extrêmement simples (avec peu de paramètres).

Des méthodes de création de données synthétiques ont été développées pour essayer de palier à ce problème en modifiant ou mélangeant des données de différents échantillons pour augmenter artificiellement le nombre d'observations. On peut citer notamment l'approche de synthetic minority oversampling utilisée pour augmenter la taille d'une classe sous-représentée (Chawla et al. 2002).

#### b. Effet batch et normalisation

Les problèmes d'effet batch entre deux séries d'échantillons qui ont été étudiées avec des méthodes, à des endroits ou des temps différents est un problème majeur en machine learning supervisé tout comme il l'est en analyse différentielle omique.

Un des biais communs en expérience de génomique est la profondeur de séquençage. En effet la profondeur de séquençage n'a pas un effet linéaire sur le niveau d'expression de gène (en RNA-seq) ou la hauteur de pics (en CHIP-seq et ATAC-seq) et peut être difficile à corriger malgré la normalisation.

Idéalement les échantillons de diverses classes devraient être randomisés et distribués uniformément au sein de chaque batch d'entraînement, et les jeux de tests et de validation devraient faire partie de batch différents. Si nécessaire, des méthodes de correction de batch peuvent être appliquées comme la normalisation par quantile, l'ajustement bayésien sur des variables mesurées par Combat, ou l'analyse de variables alternatives (surrogate variable analysis) (Johnson, Li, and Rabinovic 2007; Leek et al. 2012). Il est intéressant de noter que si cette correction est réalisée dans un même temps sur le jeu de training et de test, elle va induire une fuite d'information (data leakage) et une surestimation des performances du modèle par rapport à son application sur une cohorte externe où le preprocessing aurait été fait séparément. Une autre solution consiste à inclure les variables confondantes dans le modèle de machine learning (Dinga et al. 2020; Listgarten et al. 2010).

#### c. Variables confondantes et biais liés à la composition cellulaire

Les variables confondantes sont des variables non observées qui altèrent la structure de dépendance entre les variables observées. Ces variables confondantes peuvent être liées à la variabilité technique lors du preprocessing (cf. paragraphe précédent). Mais elles peuvent aussi être liées à des problèmes intrinsèques aux échantillons. Ainsi le contenu tumoral est

un facteur confondant bien connu pour fausser les analyses différentielles et le clustering en transcriptomique et épigénomique (Aran, Sirota, and Butte 2015). De nombreuses méthodes existent pour estimer la composition cellulaire d'un échantillon (Gaujoux and Seoighe 2013; 2012; Newman et al. 2015) et corriger les analyses différentielles en la prenant en compte (Shen-Orr et al. 2010), voire essayer de séparer le profil d'expression de chaque cellule au sein de chaque échantillon (Chu et al. 2022; Luca et al. 2021; Rahmani et al. 2019; Zaitsev et al. 2019). Cependant elles ne sont pas communément utilisées dans les méthodes de classification automatique.

Une des raisons est que les cohortes utilisées pour entraîner des algorithmes de classification cancéreuse requièrent souvent une haute pureté tumorale. Ainsi le consortium pancancer TCGA avait initialement fixé un seuil minimal de 80% de noyaux tumoraux pour l'inclusion dans la cohorte avant de le réduire à 60%. Des seuils similaires sont retenus dans la plupart des études de classification de tumeurs solides que nous avons étudiées (Capper et al. 2018; Koelsche et al. 2021). Ceci exclut directement l'application des modèles à des échantillons ne remplissant pas ce critère, notamment dans certaines pathologies hématologiques où l'infiltration tumorale est parfois faible. Ainsi le « classifier » des tumeurs du système nerveux central basé sur la méthylation dont nous reparlerons plus bas est nettement mis en défaut lorsque l'infiltration tumorale est faible (Capper et al. 2018; Z. Wu et al. 2022). Ce problème se posait déjà dans l'analyse non supervisée, puisque certains glioblastomes de faible infiltration étaient mélangés avec des gliomes de bas grades dans les analyses en T-sne. La correction de l'infiltration tumorale par l'usage d'algorithmes de déconvolution permet de corriger ce biais (Z. Wu et al. 2022) et semble intéressante pour les applications en hématologie.

Il est à noter que la composition cellulaire non tumorale peut aussi être informative du sous-type tumoral ou revêtir une valeur pronostique notamment lorsqu'elle reflète l'infiltration immunitaire (Fridman et al. 2012) et n'est donc pas systématiquement à éliminer des algorithmes de classification en cancérologie.

### 2.3 *Algorithmes principaux*

Les algorithmes utilisés couramment pour la classification en cancérologie sont les suivants :



- Elastic net
- Partial least squares differential analysis ou GLM
- Naive Bayes estimator
- Support vector machine (SVM)
- Random forest
- Boosted trees (XGboost)

Les modèles de deep-learning pourtant extrêmement efficaces pour le traitement d'autres données de santé comme l'imagerie médicale, ne sont presque pas utilisées pour les données omics car ils requièrent un nombre très conséquent d'échantillons pour être entraînés. Par ailleurs, ces modèles ont comme principal intérêt de pouvoir être pré-entraînés à reconnaître des structures de données communes à plusieurs problématiques (reconnaissance de forme dans une image). Cependant ces structures ne sont pas évidentes au sein des données transcriptomiques et épigénomiques.

De nombreuses études ont comparé les performances de différents modèles notamment en transcriptomique et épigénomique (Boulesteix et al. 2008; Kruppa et al. 2014; Maros et al. 2020; Statnikov, Wang, and Aliferis 2008). Le meilleur modèle varie d'un jeu de données à l'autre et il est donc recommandé de tester plusieurs modèles et de choisir le meilleur par crossvalidation.

#### 2.4 *Classification multiclasse et probabilité de classe*

La plupart des études de classification en machine learning ont concerné des tâches binaires. Dans le cas des diagnostics de sous-types de cancers, on cherche à identifier plusieurs classes. L'approche de classification multiclasse apporte plusieurs problématiques particulières. Premièrement tous les algorithmes ne peuvent pas être formulés de manière multiclasse. Ainsi la régression logistique multinomiale et le random forest permettent directement de prédire plusieurs classes, alors que les algorithmes de support vector machine ne le permettent pas. Dans ces derniers il faut alors entraîner plusieurs modèles qui prédisent chacun l'appartenance à une classe par rapport à toutes les autres (one versus all) ou une classe par rapport à chacune des autres (one versus one). Chaque algorithme donne une probabilité qui doit ensuite être combinée aux autres pour donner un score final reflétant la probabilité d'appartenir à chaque classe.

De plus se pose le problème de l'interprétation de ce score final. En effet il est essentiel en médecine que les algorithmes utilisés donnent un niveau de confiance associé à leur prédiction pour permettre d'aider les cliniciens dans leurs choix (Dankowski and Ziegler 2016; Kruppa et al. 2014; Simon 2014). En effet la plupart des algorithmes de machine learning donnent des scores d'appartenance à une classe qui varient entre 0 et 1. Mais ces scores ne peuvent pas toujours être interprétés comme des probabilités. Les algorithmes de régression logistique ou d'analyses linéaires discriminantes sont construits sur des estimations de probabilités et sont donc utilisables dans ce cas. En revanche, les méthodes telles que le random forest ou le support vector machine ne donnent pas directement des probabilités, surtout dans le cas des analyses multiclassées. Une deuxième étape de calibration avec un algorithme probabiliste comme la régression logistique est alors nécessaire pour obtenir ces probabilités (Maros et al. 2020).

Par ailleurs l'estimation de ces probabilités dépend de la distribution des classes au sein du jeu d'entraînement et il peut être nécessaire de recalibrer un algorithme sur un nouveau jeu de données plus représentatif de la réalité pour pouvoir en améliorer la validité externe.

Enfin les métriques utilisées en classification multiclassée doivent être adaptées surtout s'il existe un déséquilibre entre les classes. En effet l'« accuracy » ou proportion de prédiction correcte qui mesure usuellement la performance d'un classifieur peut être biaisée si une classe est sous-représentée et peut être remplacée par des mesures d'accuracy moyenne pour chaque classe (Mortaz 2020).

### 3 Exemples de publications

De nombreuses études ont été publiées pour diagnostiquer les cancers à partir de méthodes omiques. Nous donnerons quelques exemples d'études appliquées aux lymphomes T ou utilisant des mesures épigénomiques.

#### 3.1 *Classification des lymphomes T*

Trois études se sont attaquées à la problématique de la classification automatique des lymphomes T. Leurs principales caractéristiques sont résumées dans le Tableau 9. Les trois méthodes se basent sur la quantification de transcrits ARN par différentes méthodes qui

peuvent être utilisées sur des échantillons fixés en formol et inclus en paraffine (FFPE) : la micropuces, la « reverse transcriptase multiplex ligation-dependent probe amplification » ou RT-MLPA, et le « digital gene expression profiling » de Nanostring.

	Piccaluga	Drieux	Amador
Technique	Gene expression microarray	Reverse transcriptase multiplex ligation-dependent probe amplification (RTMPLA) (gene expression + mutation)	Digital gene expression profiling (nCounter, NanoString)
Classes finales	<ul style="list-style-type: none"> <li>- AITL</li> <li>- ALCL-ALK<sup>-</sup></li> <li>- PTCL-NOS</li> </ul>	<ul style="list-style-type: none"> <li>- AITL/TFHL</li> <li>- ALCL ALK<sup>-</sup> non cytotoxiques</li> <li>- Th1 /cytotoxic PTCL</li> <li>- ALCL ALK<sup>-</sup> cytotoxiques</li> <li>- ALCL ALK<sup>+</sup></li> <li>- ENKTCL</li> <li>- PTCL Th2 /ATLL</li> <li>- CD30 Th2</li> </ul>	<ul style="list-style-type: none"> <li>- AITL</li> <li>- ALCL ALK<sup>-</sup></li> <li>- ALCL ALK<sup>+</sup></li> <li>- ENKTCL</li> <li>- ATLL</li> <li>- PTCL-NOS TBX21</li> <li>- PTCL-NOS GATA3</li> </ul>
Algorithme	Support vector machine	Support vector machine + distance au centroïdes	Prédicteurs linéaires binaires $S_{12j} = \sum_i x_{ij} \frac{\mu_{1i} - \mu_{2i}}{\sigma_i^2}$
Approche multiclasse	ALCL ALK <sup>-</sup> vs PTCL-NOS et AITL vs PTCL-NOS	SVM sur 6 classes + Seuil CD30 pour ALCL cytotoxiques vs PTCL-Th1/cytotoxiques + Seuil ALK pour ALCL ALK <sup>-</sup> vs ALK <sup>+</sup> au sein des cytotoxiques Seuil CD30 pour Th2 vs CD30	One vs one : Entité de PTCL vs PTCL-NOS puis Si PTCL-NOS : PTCL-TBX21 vs PTCL-GATA3 Si 2 sous-types prédits : sous-type 1 vs sous type 2
Type d'échantillons	Entraînement sur FFPE, validation sur congelé	Entraînement congelé + FFPE Validation FFPE	Entraînement FFPE (signatures établies sur congelé lors d'un précédent travail) Validation FFPE
Performance sur validation	AITL : Se 72%, Sp 80%, Ac 77% ALCL ALK <sup>-</sup> : Se 73% Sp 96%, Ac 93%	4/40 erreurs de classification 2/40 discordances intercentre 1/40 échec 1/40 patient non classifié	Se 75-100% (par classe) Sp 93-100% (par classe) Ac 93-99%
Limites principales	Nécessité d'hypothèse préalable sur la catégorie à tester Précède la définition des PTCL-TFH	Technique additionnelle nécessaire pour distinguer PTCL Th2 et ATLL	Exclusion des PTCL-TFH

**Tableau 9 - Comparaison des principales études d'algorithmes diagnostics des PTCL basés sur la transcriptomique**

*$S_{12j}$  est le score d'appartenance à la classe 1 vs la classe 2 de l'échantillon  $j$ ,  $x_{ij}$  le niveau d'expression du gène  $i$  pour l'échantillon  $j$ ,  $\mu_{1i}$  et  $\mu_{2i}$  les niveaux moyens d'expression des gènes  $i$  dans les classes 1 et 2 et  $\sigma_i$  la variance dans les deux classes assemblées.*

La première étude, publiée en 2012, utilise deux modèles basés sur les Support Vector Machines pour distinguer d'un côté les ALCL ALK- des PTCL-NOS et de l'autre les AITL des PTCL-NOS. Elle n'est donc utilisable que pour une fraction des PTCL nodaux, lorsqu'il y a une hésitation entre ces trois sous-types. Ces résultats intéressants ont pu être validés sur une cohorte importante. Mais cette méthode est peu sensible (72-73%) notamment car elle a été réalisée avant la description des PTCL-TFH du groupe PTCL-NOS. De plus, elle est basée sur une technologie qui est rarement utilisée depuis l'avènement du séquençage à haut débit.

La deuxième étude a inclus un nombre nettement plus important d'entités différentes de lymphomes T : AITL, ENKTCL, ALCL ALK- et ALK+, HSTL, ATLL et PTCL-NOS (Drieux et al. 2020). Elle utilise la RT-MLPA qui permet en plus de la quantification de transcrits, la détection des mutations RHOA G17V et IDH2 R172K/T pour améliorer la détection des lymphomes TFH. Les performances sur la cohorte de validation sont très bonnes hormis sur quelques AITL/PTCL-TFH qui sont reclassés en PTCL Th2 en raison de l'expression de CCR4 et GATA3 et d'un AITL avec une forte infiltration CD8 cytotoxique reclassé en PTCL TH1/cytotoxique. Il est aussi intéressant de noter que cet algorithme n'utilise pas de classe unique PTCL-NOS mais les sépare en deux groupes PTCL cytotoxiques/Th1 et PTCL-Th2. De plus, les PTCL-Th2 et ATLL sont regroupés au sein d'une même classe en raison d'un profil d'expression similaire et doivent être différenciés à posteriori par d'autres méthodes telles que la recherche du virus HTLV1. Par ailleurs, établir une signature claire des ALCL semble difficile car les ALCL n'ont pu être classifiés en une seule étape. Les ALCL ALK- non cytotoxiques constituent une classe du modèle SVM. Mais il existe une deuxième classe comprenant les PTCL-Th1 cytotoxiques, les ALCL ALK- cytotoxiques et les ALCL ALK+ qui doivent être secondairement différenciés grâce à des seuils de leur expression du CD30 et de ALK. Le nombre de gènes limités évaluables par RT-MLPA pourrait être un des facteurs qui empêche de mieux discriminer ces entités.

Enfin la dernière étude est basée sur la technologie propriétaire Nanostring (Amador et al. 2022). Le classifieur est réalisé en deux étapes. La première étape du classifieur consiste à utiliser des signatures basées sur la différence de profils d'expression entre chaque entité et les PTCL-NOS. Les échantillons obtiennent ainsi un score dans 5 comparaisons binaires (entité considérée vs PTCL-NOS). Ceux ayant une expression suffisante d'une signature spécifique d'entité sont prédits comme cette entité. Si les échantillons peuvent appartenir à deux catégories, un nouveau test binaire est appliqué pour comparer le profil de ces deux sous-entités. Enfin si les échantillons sont classés PTCL-NOS dans toutes les comparaisons, un deuxième test sépare les PTCL-TBX21 des PTCL-GATA3. Les sensibilités et spécificités de cet algorithme sont aussi extrêmement élevées au-dessus de 80% et 93% respectivement, sauf pour les ALCL ALK- pour lesquels la sensibilité est de 75% seulement. Une des principales limites de ce classifieur est que les PTCL-TFH ont été exclus de la cohorte d'entraînement. Lors d'une évaluation à posteriori, seulement 17% des PTCL-TFH avaient une signature d'AITL. Cela pourrait être lié au fait que la signature utilisée pour marquer les AITL comporte de nombreux gènes exprimés par les cellules du microenvironnement et non les cellules TFH elles-mêmes. Mais la liste des gènes utilisés n'est pas communiquée par les auteurs ce qui représente une autre limite importante de cette étude.

Au total ces études démontrent à la fois le potentiel des analyses omiques en classification diagnostique mais aussi la complexité de leurs applications dans les lymphomes T qui représentent une entité extrêmement hétérogène. Elles démontrent aussi que les cas de misclassification peuvent correspondre à de nouvelles entités non encore décrites, par exemple les PTCL-TFH définis Th2 en RTMLPA.

Compte tenu de la capacité des données épigénomiques à discriminer les différents types cellulaires évoqués dans la partie B de cette introduction, de nombreuses équipes se sont légitimement intéressées à l'utilisation de ces méthodes pour des problèmes de classification en cancérologie.

### **3.2 Performance des profils épigénomiques pour la classification de cancers**

En effet, en dehors du champ des néoplasies T, de nombreuses études ont démontré le potentiel de l'application diagnostique du machine learning sur des données épigénomiques. Un des cas les plus emblématiques de cette approche est celui de la classification des tumeurs

du système nerveux central basée sur la méthylation de l'ADN (Capper et al. 2018). En effet, ces tumeurs sont extrêmement difficiles à classer morphologiquement et par immunohistochimie, reflétant un besoin médical très important. Cette étude a montré la capacité à reconnaître 14 grandes classes et 80 sous-classes de tumeurs différentes grâce à l'application aux profils de méthylation de l'ADN d'un algorithme de random forest calibré par régression logistique. Les auteurs de cette étude ont mis à disposition une plateforme pour le dépôt de fichiers de séquençage et ont ainsi pu récolter à ce jour les données de plus de 100 000 échantillons augmentant la fiabilité des prédictions. Plusieurs études de différents pays ont déjà montré l'important changement induit par l'arrivée de cette technique sur la prise en charge des patients dans la pratique clinique courante (Karimi et al. 2019; Pickles et al. 2020; Z. Wu et al. 2022). Par ailleurs cette équipe a plus récemment publié un classifieur des sarcomes (Koelsche et al. 2021) suivant une méthodologie identique à celle utilisée pour la classification des tumeurs cérébrales.

L'étude épigénétique de l'ADN circulant est aussi en pleine expansion. De nombreux algorithmes ont été publiés dans les dernières années à partir des profils des fragments ou de méthylation de l'ADN circulant pour prédire le diagnostic de cancers en partant essentiellement d'un échantillon de sang des patients (Cristiano et al. 2019; Jamshidi et al. 2022; Nassiri et al. 2020). Ces méthodes novatrices semblent cependant manquer de précision dans les stades localisés de maladie.

Enfin, l'ATAC-seq est une technique relativement récente par rapport aux études de méthylation de l'ADN, et est encore peu utilisée comme technique diagnostique. On retrouve tout de même une étude qui prédit le statut IGVH muté des LLC en fonction de leur profil épigénétique (Rendeiro et al. 2016). Un des freins à l'utilisation de l'ATAC-seq à grande échelle pour le diagnostic était la nécessité d'obtenir des échantillons frais ou congelés et non fixés en formol et inclus en paraffine ce qui limite le nombre de cas disponibles pour entraîner les modèles et la possibilité d'un accès large en routine diagnostique.

# OBJECTIFS

---

Comme nous l'avons vu en introduction, la classification des néoplasies matures T est une tâche extrêmement complexe. Actuellement le temps nécessaire pour poser de façon formelle un diagnostic de lymphome T mature en anatomopathologie oscille entre 1 à 4 semaines suivant les sous-types histologiques ce qui est difficilement compatible avec l'état clinique grave des patients qui nécessite une prise en charge rapide. L'ATAC-seq offre la possibilité d'une mesure rapide et économique de l'accessibilité chromatinienne à l'échelle du génome et a déjà démontré son potentiel pour discriminer différentes entités cancéreuses et différents stades du développement des lymphocytes T.

La question centrale de ce projet de thèse était la suivante : est-il possible de fournir une orientation diagnostique basée sur le profil d'accessibilité chromatinienne des néoplasies T matures ?

L'objectif principal était donc de définir une signature épigénétique basée sur l'ATAC-seq, spécifique des différentes sous-entités de lymphomes T, et permettant à terme une classification plus performante et plus rapide de ces pathologies.

Les objectifs secondaires consistaient à

- Etudier la possibilité de rattacher les lymphomes étiquetés PTCL-NOS aux entités déjà existantes de MTCN ou décrire l'existence réelle d'un ou plusieurs groupes distincts de PTCL-NOS ;
- Identifier la cellule tumorale d'origine de ces différents types de MTCN grâce à l'intégration de données d'ATAC-seq de cellules tumorales et de lymphocytes T normaux.

# ARTICLE

---



**Genome-wide chromatin accessibility landscape increases diagnostic accuracy and refines cell-of-origin classification of mature T-cell neoplasms.**

Edith Julia, Sylvain Mareschal, Amel Chebel, Sarah Sebaï, Camille Golfier, Camille Lours, Edoardo Missiaglia, Vimel Rattina, Marie Donzel, David Armisen, Sofiane Hadj Hamou, Till Braun, Nathalie Bissay, Manon Tschantz, Fanny Drieux, Michele Ceribelli, Dimitri Chartoire, Aurélie Verney, Sylvain Carras, David Sibon, Ambroise Marçais, Bruno Simon, Camille Laurent, Philippe Ruminy, Laurent Martin, Vahid Asnafi, Luc Xerri, Lea Payen-Gay, Virginie Fataccioli, Damien Sanlaville, Pierre Sujobert, Michaël Roussel, Vanessa Szablewski, Claire Bardel, Catherine Chassagne, Gilles Salles, Lucile Baseggio, Marco Herling, Louis M Staudt, Alexandra Traverse-Glehen, Philippe Gaulard, Laurence de Leval, Laurent Genestier\*, Emmanuel Bachy\*

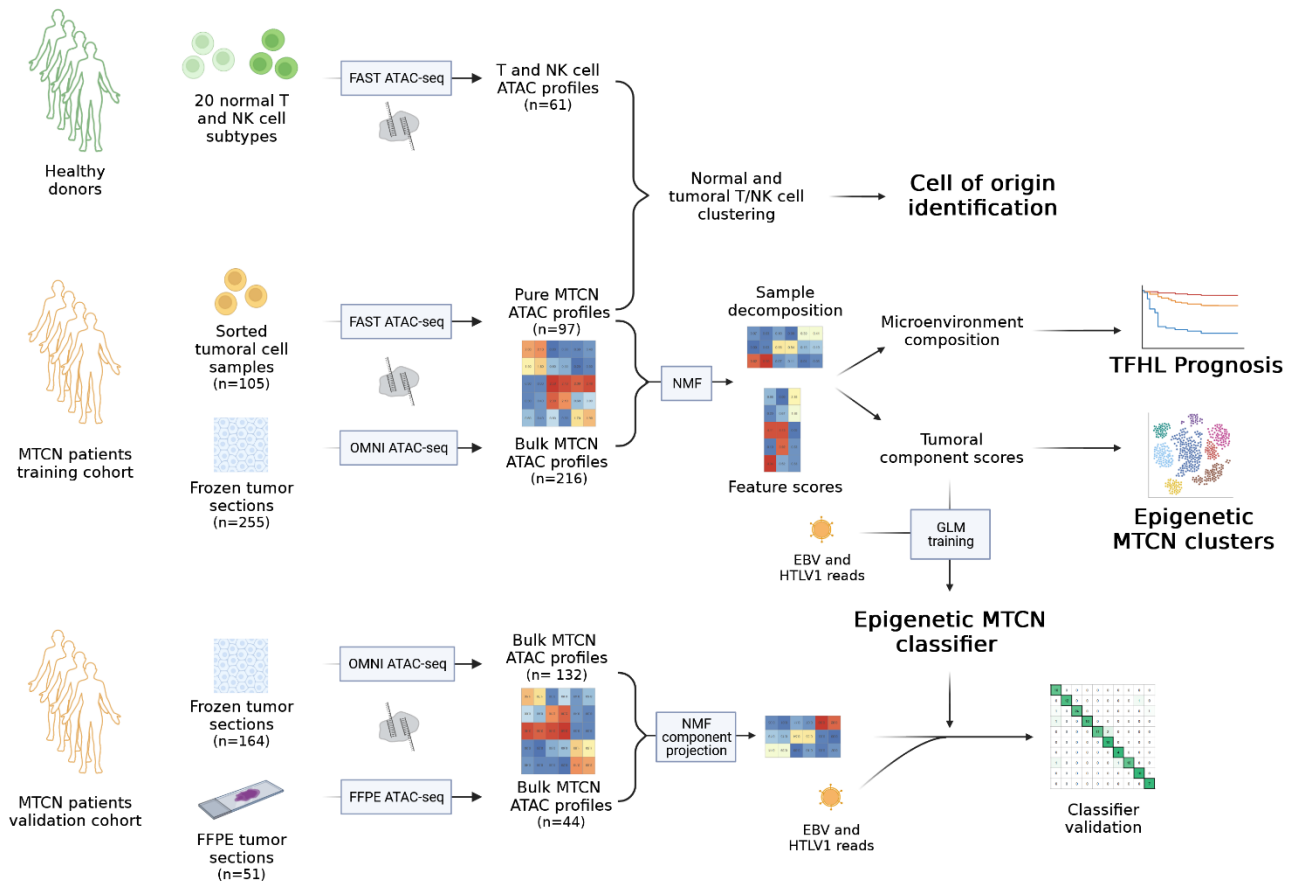
Correspondence: Prof. Emmanuel BACHY, Hematology Department, 165 chemin du grand Revoyet, Lyon Sud Hospital, 69410, Pierre-Bénite, France

Tel +33(0)4 78 86 22 05

Fax +33(0)4 78 86 43 47

Email: [mailto: emmanuel.bachy@chu-lyon.fr](mailto:emmanuel.bachy@chu-lyon.fr)

# GRAPHICAL ABSTRACT



ATAC-seq was applied to explore the epigenetic landscape of MTCN entities. We applied FAST-ATAC to several normal T and NK cell subsets of healthy donors and MTCN sorted samples and OMNI-ATAC seq to frozen MTCN sections. Clustering of sorted MTCN and normal T/NK cell epigenetic profiles was performed to find MTCN entities' cell of origin. Then, using NMF-based unsupervised deconvolution approaches on epigenetic profiles from sorted and frozen bulk samples, we were able to segregate the different MTCN subtypes in distinct epigenetic clusters. NMF microenvironment components showed interesting prognostic significance in TFH lymphomas. Eventually, we showed that ATAC-seq is feasible on formalin-fixed paraffin-embedded samples without batch effect compared to frozen samples. A generalized linear model classifier was trained to integrate NMF decomposition scores and viral read counts to predict MTCN diagnosis, showing high classification performance on an external validation cohort.



## INTRODUCTION

Mature lymphoid malignancies are neoplasms emerging from the transformation of a mature B- or T-cell<sup>1,2</sup>. They are highly heterogeneous with the last revision of the World Health Organization (WHO) classification encompassing more than 80 sub-entities<sup>3</sup>. Recent studies have shown that major or minor diagnosis reclassification rates by expert pathologists can reach as high as 30-50% in non-Hodgkin lymphoma (NHL), especially in peripheral T-cell lymphoma (PTCL)<sup>4</sup>. As a result of poorly delineated characteristics, 20-30% of PTCL are classified in a wastebasket as non-otherwise specified (NOS)<sup>3</sup>. Heterogeneity of the disease translates into a highly variable prognosis for patients ranging from 5-year overall survival (OS) of more than 70% in case of indolent anaplastic lymphoma kinase (ALK)-positive anaplastic large cell lymphoma (ALCL) to less than 10% in the most aggressive subtypes of T-cell malignancies like enteropathy associated T-cell lymphoma (EATL) or hepatosplenic T-cell lymphoma (HSTL)<sup>5</sup>.

Next generation sequencing (NGS) of large patient cohorts have pinpointed recurrent mutations and pathway alterations in NHL over the last decade<sup>6-13</sup>. In T-cell lymphoma, frequent and usually mutually exclusive mutations as well as copy number alterations have been identified in the T-cell receptor (TCR) pathway, in genes encoding epigenetic regulators, cell cycle control, genome integrity, or the cytoskeleton structure<sup>7-12</sup>. This genetic alteration landscape has provided insight into molecular drivers of PTCL but has been strikingly unable to segregate one disease from another<sup>7-12</sup>. Indeed, mutation frequencies strongly vary between subtypes, but most pathways are found repeatedly altered in different malignancies. For instance, activating mutations in the TCR pathway have been found in various peripheral T-cell lymphoma subtypes like angio-immunoblastic T-cell lymphoma

(AITL), HTLV-1 associated acute T-cell leukemia/lymphoma (ATLL) or Sézary syndrome (SS)<sup>14–</sup>

16.

Gene expression profile (GEP)-based classifiers using microarrays or RNA-sequencing (RNA-seq) have refined former histological classifications<sup>17–19</sup>. For instance, approximately one-third of PTCL-NOS have been reallocated to another molecularly predefined subtype, and high expression of GATA3 or TBX21 have further segregated remaining NOS cases with divergent prognosis<sup>18</sup>. More recently, in an effort to render GEP transposable for routine diagnosis classification on formalin-fixed and paraffin-embedded (FFPE) samples, reverse transcriptase multiplex ligation-dependent probe amplification (RT-MPLA) or the nCounter platform using the Nanostring® technology have been developed using a limited set of 20 or 153 genes, respectively<sup>20,21</sup>.

Assay for transposase-accessible chromatin using sequencing (ATAC-seq) is a recently developed technique allowing for genome-wide mapping of histone marks and transcription factors<sup>22</sup>. Compared to other epigenomic approaches like DNA methylation assay, DNase I hypersensitive sites sequencing (DNase-seq), or chromatin and immunoprecipitation sequencing (ChIP-seq), ATAC-seq can be performed on scarce tumoral material, with a rapid turn-over and at an affordable cost. Moreover, ATAC-seq has been shown to be highly reproducible across technical and biological replicates. Previous studies have demonstrated how ATAC-seq is able to outperform GEP by RNA-seq in discriminating normal hematopoietic lineages or chronic lymphoid leukemia (CLL) disease subtypes<sup>23,24</sup>.

In this study, we provided a whole-genome chromatin accessibility landscape of mature T-cell neoplasms (MTCN) and showed that machine-learning-based ATAC-seq profile classification allows for accurate diagnosis assignment readily transposable to FFPE samples, putative cell-

of-origin identification, and new subtypes delineation. This classification provides a new nosology for MTCN amenable to routine practice for accurate diagnosis and reliable patient stratification for future clinical trials.

## RESULTS

### *Chromatin accessibility landscape of mature T-cell neoplasms*

A total of 575 samples from patients with NK and T cell malignancies were sequenced with ATAC-seq. T cell acute lymphoblastic leukemia (T-ALL) cases were added to allow reclassification of potentially misdiagnosed PTCL-NOS. Among those, 360 samples were used to build an epigenetic-based classification derived from subtype-specific whole-genome chromatin accessibility landscapes. Sorted tumoral cells from dimethyl sulfoxide (DMSO) cryopreserved samples (n=105) and frozen sections from whole tissue biopsy (n=255) were sequenced. Three hundred and thirteen samples (88%) with sufficient number of mapping reads and high signal over background ratio (assessed by transcription starting site (TSS) enrichment metric) passed quality control (QC) check. A map of 1,150,630 non-overlapping accessible chromatin peaks across the genome was derived from the single-cell chromatin atlas accessibility study [Zhang cell 2021]<sup>25</sup>. For unsupervised projection, lowest infiltration samples (<30%) were removed from the training cohort and overrepresented entities (Tfh-PTCL, CTCL) were downsampled to avoid class imbalance. Non-negative matrix factorization (NMF) deconvolution was used to identify specific tumoral epigenetic signatures from sorted cells and bulk tissue and to remove key components from the tumor microenvironment (B-cell, plasma-cell, and macrophages cell signal) or poor-quality pattern that blurred actual tumor classification in case of low tumor content (Supplementary Figure S1A and S1B). Positive correlation between microenvironment signatures identified by ATAC-seq from bulk

tissue following deconvolution and single cell ATAC-seq (scATAC-seq) data from a publicly available atlas was high (rho coefficient > 0.75 for macrophage and B-cell pattern for instance, Supplementary Figure S1C)<sup>25</sup>. An unsupervised clustering approach using shared nearest neighbor (SNN) followed by uniform manifold approximation and projection (UMAP) visualization revealed a clear segregation between known histological subgroups for most subtypes (Figure 1A). Only a few samples at the border of epigenetic clusters (n=31 out of 200, 15.5%) could not be formally attributed with a definite epigenetic tumor label (Figure 1A). Subtype-specific opened chromatin regions (OCR) patterns were differentially accessible in genomic regions concordant with known biological genes and pathways involved in the disease pathophysiology. For instance, significantly more accessible OCR in *PDCD1* (coding for PD-1), *CXCL13*, *ICOS* or *CXCR5* that are highly expressed in T follicular helper (Tfh) lymphocytes, were found in Tfh-PTCL. A significantly more accessible OCR in *GRZMB* (coding for Granzyme B) was identified in large granular lymphocyte (LGL) leukemia samples known to express cytotoxic markers (Figure 1B). Specific OCR in *PAX5* and *CSF1R* were found in the microenvironment signatures of normal B-cell and macrophages, respectively (Figure 1C). Overall, distinct patterns of OCR could be specifically selected to characterize one MTCN entity versus the others (Figure 1C).

Epigenetic reclassification of histologically defined samples occurred in 47 cases (25%; Figure 1D). Virtually all PTCL-NOS shared another specific PTCL epigenetic signature. Most PTCL-NOS were refined as Tfh-PTCL, likely reflecting World Health Organization (WHO) classification evolution now recognizing previously unclassified PTCL expressing at least 2 Tfh markers as Tfh-PTCL. Others were epigenetically grouped with natural killer T-cell lymphomas (NKTCL) due to a similar cytotoxic profile but were further distinguished from true NKTCL by EBV read counts available from ATAC-seq data (highly positive EBV counts for NKTCL, low or

negative counts for cytotoxic PTCL-NOS, Figure 1E). The remaining PTCL-NOS were allocated to various other subgroups. Only 12 PTCL-NOS samples formed a separate epigenetic cluster that could not be formally labelled by a known histological subgroup (cluster 10; Figure 1D). Further immunohistological, cytogenetic and biomolecular analyses with spare material were performed whenever possible for discordant cases between epigenetic and histological diagnosis. For instance, HTLV-1 read counts were found in one PTCL-NOS case later confirmed to be an ATLL by *HBZ* positive expression. A rearrangement of the *DUSP22-IRF4* locus on 6p25.3 (usually associated with ALK-negative ALCL in 1/3 of cases) was found in one PTCL-NOS correctly classified as an ALK-negative ALCL sample by ATAC-seq. Two PTCL-NOS samples classified as T-cell acute lymphoblastic leukemia by ATAC-seq were further found to express the immature TdT marker with subsequent immunohistological staining. One patient with a prior history of cutaneous T-cell lymphoma (CTCL) but with nodal involvement by PTCL cells expressing Tfh markers were classified as Tfh-PTCL by the pathologist but reclassified as CTCL by ATAC-seq, likely reflecting Tfh-phenotype acquisition by CTCL cells migrating from the skin. Because of insufficient number of cases to define specific sub-clusters, the MEITL cases, harboring a mixed epigenetic profile between NKTCL and enteropathy-associated T cell lymphoma (EATL) were partially included in the NKTCL/cytotoxic lymphoma group and two cases of transformed mycosis fungoides (tMF) could not be distinguished from ALK-positive ALCL.

### ***Epigenetic profiling by ATAC-seq identifies a new subentity among PTCL-NOS***

As previously stated, we identified a small cluster of 12 samples including a majority of PTCL-NOS that could not be formally allocated to another specific epigenetic subgroup of MTCN. All these lymphomas were CD4<sup>+</sup> and had lost CD7 expression. Differential OCR analysis



revealed that peaks in specific genes like *TLK1*, *ARID1B*, *EZH2* or *IRF4* were found to be recurrently more open in this subgroup compared to all other entities (Figure 2A). Transcription factor (TF) footprint analysis found specific signatures associated with various GATA family members, ZEB1, TCF4, MEF members and NFKB1 to be enriched (Figure 2B).

Since PTCL-NOS have been recently divided into PTCL-TBX21 and PTCL-GATA3 subentities (provisional status in the latest WHO classification), we specifically focused on these two TF motif accessibility across epigenetic clusters. The cluster of 12 PTCL-NOS samples was clearly skewed toward a GATA3 footprint but, overall, other PTCL-NOS with high GATA3 activity were distributed among various epigenetic clusters (mainly CTCL, T-PLL or T-ALL) meaning that overlap with this previously defined PTCL-NOS subgroup was only marginal (Figure 2C and D). Inferred copy-number variations (CNV) using background reads from ATAC-seq showed that this PTCL-NOS subgroup was characterized by frequent genomic gains and losses across the genome reflecting high genomic instability (Figure 2E). In particular, 17p deletion was observed in more than 50% of samples. Whole-exome sequencing revealed recurrent mutations in 8 out of 12 patients (66%) in the gene *NCOR1*, a nuclear receptor corepressor with known tumor suppressor role in various cancers (Figure 2F). One patient presented with mutations both in *NCOR1* and *NCOR2*, an important paralog of *NCOR1*. Similarly, *TP53* was found mutated in 66% of patients. Interestingly, *NCOR1* is located on the 17p chromosomal arm and was found commonly disrupted by the recurrent 17p deletion encompassing the *TP53* locus (Figure 2G), likely reflecting a mechanism by which both copies of *NCOR1* and *TP53* are inactivated by deletion and inactivating mutation for most patients of this PTCL-NOS subgroup, referred to as the PTCL-NCOR subgroup hereafter. Most *NCOR1* mutations were frameshift or nonsense, and 2 out of 3 missense mutations were predicted to be deleterious for the protein function in the AlphaMissense database (Figure 2H)<sup>26</sup>.

### ***The chromatin landscape of MTCN is reminiscent of their putative cell-of-origin***

A compendium of ATAC-seq signatures from most normal mature T-cell subtypes was generated based on specific phenotypic-based flow cytometry cell-sorted populations (Supplementary Table 1). Data were retrieved from public datasets for innate lymphoid cells (ILC) and immature thymic subtypes and added to the atlas. A map of a thousand peaks with the highest variance across normal T-cell populations was generated. The chromatin openness of sorted tumoral samples along this map of peaks was assessed. Clustering of normal and tumoral cells altogether confirmed the epigenetic similarity between normal Tfh cells and Tfh-PTCL, confirming the known cell-of-origin (COO) of this particular MTCN subtype (Figure 3A). Differentially opened peaks were identified in the promoters of *CD4* and *CXCR5*, known to be highly expressed in both normal Tfh and Tfh-PTCL. Likewise, a T-ALL sample was perfectly clustered amongst normal thymic progenitor cells and opened peaks in the promoters of the recombinases *RAG1* and *RAG2* were found. CTCL were found to share a common branch origin with memory T cells in the dendrogram, reminiscent of the putative COO determined by immunophenotypic proximity. Interestingly, all T-cell prolymphocytic leukemia (T-PLL) were found to cluster near normal naïve CD4 and CD8 T cells with which they shared common OCR in the promoters of *CR2*, *CD248*, *NRCAM* and distant from memory T cells although they shared some memory-associated OCR in genes like *NFATC2* and *BCL2*. All MTCN now recognized as belonging to an innate-like T-cell lymphoma or leukemia subgroup (LGL, enteropathy-associated T cell lymphoma [EATL], hepatosplenic T-cell lymphoma [HSTL] and NKTCL) in the most recent WHO classification clustered with innate-like T-cell lymphocytes (NK cells, and invariant NKT and gamma/delta T cells) or CD8+

effector-memory and effector-memory reexpressing CD45RA T cells (TEM and TEMRA) known to share an innate-like epigenetic phenotype (Figure 3A). Indeed, many OCR in the promoter regions of cytotoxic (*CD8A* and *CD8B*, *FYN*, *IFNG*) and NK-specific (*SYK* or *NCR1* [coding for NKp46]) genes were shared between normal T cells and their tumoral counterparts.

To assess if deeper knowledge translatable to routine practice could be learned from cross-comparison of the epigenetic landscape between MTCN and their putative COO, we performed differential OCR analysis between normal Tfh cells and Tfh-PTCL sorted cells. A peak located at the promoter region of a long non-coding RNA (lncRNA, ENSG284966) and the 3'-UTR region of the *EFNB2* gene in opposite strand directions was found to be highly opened in Tfh-PTCL compared with all other MTCN ( $\log_2$  fold change [ $\log_2FC$ ] of 2.79 and false discovery rate [FDR]  $q$ -value of  $1.3 \times 10^{-18}$ , Figure 3B). Total stranded RNA-sequencing demonstrated that this opened peak correlated with a specific expression of the lncRNA-ENSG284966 in Tfh-PTCL while no difference was observed regarding *EFNB2* read coverage (Supplementary Fig 2). Strikingly, this highly specific OCR was reminiscent of the Tfh COO epigenetic profile since the peak was found similarly opened in normal Tfh cells while closed in all other normal T-cell subsets (Figure 3C). Relative expression of the lncRNA-ENSG284966 by stranded total RNA-seq readily segregated Tfh-PTCL from all other MTCN (Figure 3D), advocating for the use of this specific marker to help for Tfh-PTCL diagnosis in routine practice. *In situ* hybridization using BaseScope™ assay confirmed that staining of the lncRNA-ENSG284966 on 5 $\mu$ m FFPE sections stained numerous cells in a case of Tfh-PTCL cells and was scarcely positive in reactive adenopathies (Figure 3E).

### ***Microenvironment deconvolution by ATAC-seq predicts survival in TFH lymphomas***

The tumor microenvironment has been demonstrated to be of paramount importance in lymphomagenesis and treatment resistance. ATAC-seq was able to partially deconvolute tumoral and microenvironment cell content from bulk whole tumor section (Figure 1B and Supplementary Figure S1A and S1C). In Tfh-PTCL, B-cell infiltration appears to consistently correlate with a better prognosis<sup>18,21</sup>. To assess if ATAC-seq could segregate patient prognosis according to the characteristics of the microenvironment signatures, deconvolution using NMF was applied to bulk Tfh-PTCL samples (N=118). Four cellular patterns were identified, one corresponding to the tumor content, the three others corresponding to various microenvironment cellular components made of B-, plasma- and T-cell (Figure 4A). Four different clusters of Tfh-PTCL could be separated, two with a high B- (cluster #4) or plasma-cell (cluster #2) signature; and two with a higher T-cell infiltration (cluster #1 and #3). Two prototypic Tfh-PTCL from clusters #2 and #3 were sequenced using scATAC-seq from frozen tumor sections to assess cell-type content as predicted with bulk ATAC-seq compared with ground truth as quantified using scATAC-seq (Figure 4B). Higher plasma cell infiltration was found in cluster #2 sample and higher tumoral TFH cell infiltration in cluster #3 sample confirming NMF pattern repartition (Figure 4C). Significantly higher gammaglobulin level and older age were observed in patients within the cluster #2 (plasma-cell rich) and a significantly higher proportion of tumors with TET2 mutations were found in patients from clusters #1 and #3 (T-cell rich) (Supplementary Table 2). Of note, no significant correlation between other known recurrent mutations in Tfh-PTCL (in genes like *RHOA*, *DNMT3A* or *IDH2*) and cellular pattern was observed (Supplementary Table 2). Patients with B-cell rich Tfh-PTCL (cluster #4) had the most favorable prognosis with prolonged overall survival (OS) compared with patients with high tumor and T-cell rich microenvironment (cluster #1 and #3) (Figure

4I). Patients with a plasma-cell enriched microenvironment (cluster #2) had an intermediate prognosis (Figure 4I). In multivariate analysis including common prognostic factors (age, sex, ECOG and LDH elevation), only belonging to cluster #2 (plasma-cell high) still conferred better prognosis (HR = 0.45 (0.24-0.82)). Altogether, ATAC-seq from bulk tissue allows for tumor and microenvironment signal deconvolution and provides a powerful approach recapitulating and refining known prognostic correlations according to microenvironment content in Tfh-PTCL.

### ***An epigenetic classifier of mature T-cell neoplasms***

ATAC-seq can be performed on scarce frozen material, with short hands-on time and at an affordable price compared to other genome-wide sequencing techniques. ATAC-seq data generated from FFPE samples is considered to be of poor quality due to a high degree of DNA damage during reverse cross-linking resulting in short fragment length<sup>27</sup>. Despite these limitations, we evaluated if ATAC-seq genome-wide OCR profiles obtained from 10µm FFPE sections could be used to accurately predict MTCN subtypes based on the large training set generated from high-quality frozen samples. We confirmed that only nucleosome-free fragments could be sequenced from FFPE samples while the usual mono- and bi-nucleosome pattern observed with frozen tissue was completely lost (Figure 5A). However, very high correlations were observed for read counts in peaks and TSS enrichment scores between frozen and FFPE samples from the same specimen despite fewer total mapping reads ( $\rho=0.81$ , Figure 5B and supplementary Figure 4A-C). Except for a few outliers, NMF decomposition followed by UMAP visualization of frozen and FFPE duplicates demonstrated a high epigenetic profile similarity between the two approaches (Figure 5C).

A validation cohort of 215 frozen and FFPE samples from patients with various MTCN subtypes was collected and sequenced with ATAC-seq. A machine-learning (ML) approach was developed using a generalized linear model (GLM) on NMF patterns defined from the training cohort (Figure 1B) and normalized viral read counts as explanatory variables and epigenetic subtypes as the response variable. Although the small number of MEITL samples in the training cohort prevented the constitution of a distinct group through NMF and SNN clustering, MEITL were considered as a separate class for training. Prediction was considered as possible only if the likelihood of a predicted tumoral epigenetic subtype was over 0.5. Out of 215 samples, 176 passed QC (*i.e.*, number of aligned reads over 3 million and TSS enrichment over 4) (Figure 5D). No difference was observed between frozen and FFPE with 80% and 86% of samples passing QC, respectively. Epigenetic subtype prediction by ML was inconclusive in only 15 samples out of 176 (8.5%). Overall, concordance between histological diagnosis and diagnosis predicted by ML was 87.3% without significant difference between prediction of frozen or FFPE samples (Figure 5E and Supplementary Figure 4). Most discordant cases were ALK-negative ALCL samples predicted as ALK-positive ALCL according to their chromatin accessibility landscape likely due to a shared cytotoxic profile (7 cases out of 11 leading to an overall poor 36% accuracy in ALK-negative ALCL). Despite few MEITL samples for ML training as previously stated, all 3 samples from the validation cohort were correctly predicted. Agreement between histological classification and ML was not computed for PTCL-NOS since, in the training cohort, all of them had been reallocated to a more specific epigenetic cluster or to the new PTCL-NCOR subentity previously defined (Figure 5F). After review of discordant cases (15 out of 118, Figure 5G), 7 were due to ALK-negative ALCL misclassification as ALK-positive ALCL despite new immunohistological staining confirming ALK negativity and negative ALK breakapart FISH. Surprisingly, one sample presented with 4

copies of ALK, and another with a t(2;5) that did not involve ALK. One diagnosis was corrected from EATL to ALK-negative ALCL diagnosis by ML prediction. Patient presented an intestinal CD30-positive lymphoma but with an anaplastic morphology and without any sign of villous atrophy on the biopsy. One Tfh-PTCL predicted to belong to the new PTCL-NCOR subtype was confirmed to be CD4+CD7-GATA3+, harbored a TP53 mutation and presented no typical TFH mutation in DNMT3A, RHOA nor TET2. Misclassification by ATAC-seq was confirmed in 6 samples. One ALCL-ALK+ sample predicted as T-PLL with an intermediate confidence score of 0.61 presented millimetric infiltration on hematoxylin eosin-stained section, one Tfh-PTCL was predicted NKTCCL because of an unusually high number of EBV reads. Two Tfh-PTCL were predicted as LGL and cytotoxic PTCL, one EATL as Tfh-PTCL and one NKTCCL as ALCL ALK-. The algorithm allowed reclassification of PTCL-NOS diagnosis in 39/43 cases. Interestingly, two nodal PTCL-EBV+ cases from the validation cohort were predicted as NKTCCL/cytotoxic subtype hinting at a common epigenetic signature among these entities.

## DISCUSSION

The genetic complexity and phenotypic heterogeneity of MTCN have hampered a proper and unambiguous classification. Among cancer in general, and lymphoma in particular, the higher misclassification rate has been observed in MTCN, usually ranging between 30 to 50%<sup>4</sup>. Although most misclassifications are minor and would not lead to treatment change, recent refinements in therapy like the benefit of epigenetic-targeted agents in Tfh-PTCL are probably becoming of meaningful prognostic impact<sup>28-30</sup>. Here, we demonstrate that epigenetic-based classification with ATAC-seq provides a blueprint for rapid and accurate classification of MTCN. Unsupervised clustering produced an epigenetic classifier mirroring the most recent WHO classification. Aside from the concordance with well-defined

histopathological diagnoses, ATAC-seq further allocated virtually all PTCL-NOS, a known wastebasket where PTCL are classified by default, to other predefined epigenetic subgroups. For instance, part of PTCL-NOS was hence epigenetically defined as Tfh-PTCL, a subgroup that has been formally referred to as a definite category only in the latest WHO classification in 2022<sup>3</sup>. More intriguing was the observation of a unique independent epigenetic cluster of 12 PTCL-NOS samples. Subsequent phenotypic, transcription factor and molecular analyses substantiated the relevance of isolating this ATAC-seq defined cluster as a specific subentity. First, recurrent inactivating mutations in *NCOR1*, a known tumor suppressor in solid tumor, were found in 2/3 of samples and pointed out to some specific underlying genetic drivers of the disease<sup>31–34</sup>. To date, *NCOR1* mutations have been identified at an extremely low rate in CTCL only but never recurrently in PTCL-NOS<sup>7,8</sup>. *NCOR1* serves as a co-repressor of multiple histone deacetylases (HDACs) to mediate transcriptional repression of gene expression<sup>35,36</sup>. *TP53* mutations were found to be extremely frequent (67%) as well in this peculiar subgroup. While a generic GATA3 TF footprint signature was found, it was also observed in CTCL, ATLL, T-PLL and the few T-ALL samples of the cohort. This new PTCL-NCOR entity encompassing ~10% of PTCL-NOS samples does not correspond to the recently identified GATA3-expressing PTCL-NOS subgroup usually representing 30 to 45% of PTCL-NOS cases<sup>18,37–39</sup>. Here we found that GATA3 PTCL-NOS largely overspread many other subtypes of our epigenetic classification without constituting a subgroup *per se*. Same findings were observed with the reciprocal *TBX21* signature reported in PTCL-NOS since it was distributed across all innate-like PTCL<sup>18,39,39</sup>.

Finding the COO from which a lymphoma cell is derived has been the center of intensive research in the hematology field. Apart from the theoretical understanding of lymphomagenesis, it has been successfully linked to prognosis assessment, therapeutic



vulnerability identification and drug design development in aggressive B-cell lymphomas<sup>13,40,41</sup>. Here, we confirmed that Tfh-PTCL derive from Tfh normal cells and that CTCL mirror an epigenetic program found in memory T cells<sup>17,42</sup>. We also pinpointed that T-PLL likely derive from the transformation of naïve T cells rather than effector memory cells as their phenotype would suggest based on cell surface expression markers<sup>43</sup>. The utility of a COO-based classification was confirmed by the identification of a specific opened region in the promoter of a new lncRNA found to be expressed both in Tfh-PTCL and normal Tfh cells. This highly specific marker readily quantifiable on FFPE tumor section by RNA *in situ* hybridization will help routine diagnosis of Tfh-PTCL, a usual challenge even for trained hematopathologists. Further studies will be necessary to elucidate the potential function of this lncRNA in the physiology of normal Tfh cells.

*In silico* deconvolution of signal from mixed cell types without an *a priori* knowledge of the tumor and microenvironment precise components has been increasingly used<sup>44,45</sup>. Here, we show that ATAC-seq not only robustly classifies samples by epigenetic tumor features but also identify various components of the microenvironment to refine subgroup prognosis for instance. We therefore confirmed that a specific B-cell infiltration (as well as a non-overlapping plasma-cell specific signature) in Tfh-PTCL was associated with a better prognosis<sup>18,21</sup>.

Finally, one of the most significant limits of ATAC-seq that is considered unexploitable without cumbersome library reconstruction steps on FFPE samples, was overcome and unleashed its full potential for routine diagnostic use<sup>27</sup>. We confirmed that DNA fragment length was severely altered after paraffin removal and decrosslinking steps but that peak counts in specific OCR defined on frozen tissue samples were sufficient for ML-based

diagnosis prediction. Altogether, the algorithm was able to concordantly predict MTCN diagnosis made by expert pathologists in 87% of cases, while most discrepancies were due to the inability to distinguish some ALK-negative ALCL with cytotoxic features from ALK-positive samples. Further work is ongoing with additional samples to overcome this limit.

Other sequencing technics appear as promising for what is now designated as artificial intelligence-based diagnosis. Several studies have shown that GEP can finely distinguish between various entities of MTCN or that whole-genome DNA methylation sequencing can leverage epigenetic patterns for solid tumor classification<sup>18,46,47</sup>. However, most epigenome sequencing methods require cumbersome base conversion chemistries while RNA-seq is highly dependent upon tissue quality. Besides, both whole-methylome sequencing and RNA-seq need far longer hands-on technical time than ATAC-seq. Other technics based on RT-MLPA or the nCounter® technology have been proposed using a reduced-size gene expression and/or mutation/translocation panel to provide ML-derived diagnosis in PTCL on FFPE samples<sup>20,21</sup>. Further work will help understand how these different approaches can improve routine-based diagnosis in the hematology field.

In conclusion, we show that the whole-genome chromatin accessibility landscape of MTCN assessed by ATAC-seq refines the current classification and can be used for routine ML-based diagnosis.

## METHODS

### *Samples*

Cryopreserved cell suspensions of MTCN biopsies were obtained from the CeVI collection, Lyon Sud Hospital Biologic Resource Center (CRB Lyon), Necker, Henri Mondor and Grenoble hospitals. Frozen T cell malignancies sections for the training cohort were obtained from the French T-cell lymphoma Consortium TENOMIC. Additional training cases and samples for the validation cohort were obtained from Lyon Sud Hospital and collaborating centers (Centre Léon Bérard, Necker, Dijon, Toulouse, and Lausanne University Hospitals). Normal T- and NK-cell subsets were obtained from apheresis blood unit delivered by Etablissement Français du sang (EFS) except from TFH cells that were sorted from cryopreserved reactive lymph node cell suspensions from CRB Lyon Sud. All patients gave their written consent for the use of their biopsies in research protocols.

### ***Cell preparation and sorting***

For cryopreserved cell suspensions (tumoral and reactive lymph nodes), freshly thawed cells were directly stained with fluorochrome-conjugated antibodies and FACS-sorted without prior T-cell enrichment. Peripheral blood mononuclear cells (PBMCs) from each sample were purified using a density gradient with Ficoll-Paque medium (Eurobio). T cells were then enriched from PBMCs by depletion of non-T cells using paramagnetic MACS beads with pan T-Cell Isolation Kit (Miltenyi Biotech) according to the manufacturer's protocol. Markers used to stain each cell sample are described for normal cell types and tumoral samples in Supplementary Tables 1 and 3, respectively. For T cell malignancies, entity specific panel could be adapted for a specific sample to include known positive markers according to the original diagnosis made by the pathologist (using IHC or flow cytometry). Ten thousand cells per sample were sorted into 50% RMPI-50% FBS, washed with PBS 1X, and immediately transposed as described below.

### ***ATAC-seq processing***

Genome-wide chromatin accessibility profile was obtained using the previously published protocol (called FAST-ATAC) for live sorted cells with a few adaptations described hereafter<sup>24</sup>. Ten thousand cells were used for each transposition reaction. If less than 5,000 cells were sorted, Illumina Tn5 transposase quantity was divided by two (i.e. 1.25 $\mu$ L). The number of PCR amplification cycles was standardized to 13. For frozen sections, the OMNI-ATAC seq protocol described previously was used without any modification, except for adapting the number of PCR cycles with an intermediate qPCR to avoid over-amplification of libraries<sup>48</sup>. Five  $\mu$ L of the eluted DNA were subjected to qPCR in a volume of 15  $\mu$ L using SYBR Green I dye (final 0.6x SYBR GreenI, Life Technologies) with the respective primers (0.41  $\mu$ M each), used in the first round of PCR. Following qPCR [98°C for 30 s, followed by 30 cycles (98°C for 10 s, 63°C for 30 s and 72°C for 60 s), amplification curves were analyzed and the optimal number of PCR cycles (5 to 13 cycles) for each sample was estimated with cycle thresholds reaching  $\frac{1}{4}$  of the maximum. The libraries were quantified by qPCR using Library Quantification Primer plus KAPA SYBR FAST (Roche Diagnostics) and pooled at 1 nM. Samples were sequenced using single-index paired-end (75+75 bp) reads on an Illumina<sup>®</sup> NextSeq 500 instrument in high-output mode.

### ***ATAC-seq of FFPE samples***

Human FFPE tissue blocks were sectioned into 10- $\mu$ m-thick sections using a microtome and placed onto a slide. One slide with tissue section was deparaffined with 1 mL of xylene (VWR, 28975.30), 5 min, twice. Rehydration was performed by sequential ethanol washes, starting with 100% ethanol, 5 min twice, and 95%, 70%, 50%, 30% ethanol, 5 min each. After deparaffinization and rehydration, tissue was washed with 1 mL water. Then the tissue was

scrapped from the slide, washed, and digested overnight at 37°C in a 1 mL enzymatic cocktail (3 mg/mL of collagenase; Merck C9891 and 300 U/mL of hyaluronidase (Sigma, H3506) with adding 100 µg of ampicillin (Sigma A5354) and 50 µg of sodium azide (Merck Millipore 26628-22-8). After incubation, NST buffer (146mM NaCl [Sigma, S5150], 10 mM Tris at pH 7.8 [Sigma T2569], 1 mM CaCl<sub>2</sub> [Sigma21115], 21 mM of MgCl<sub>2</sub> [Invitrogen AM9530G], 0.05% BSA [Miltenyi Biotec MACS 130-091-376], 0.2% IGEPAL CA-360 [Sigma-Aldrich 13021-50]) was added. The mixture was passed through a 27-G needle syringe 30 times and filtered with a 30µM filter (Miltenyi Biotec MACS 130-098-458), centrifugated at 3000g for 10 min, and the nuclei were resuspended in phosphate-buffered saline (PBS) and counted. Fifty thousand isolated FFPE nuclei were used in each transposition reaction. Fifty microliters of lysis buffer (10 mM Tris-HCl at pH 7.4 [Sigma, T2194], 10 mM NaCl [Sigma, S5150], 3 mM MgCl<sub>2</sub> [Invitrogen AM9530G], 0.1% IGEPAL CA-360 [Sigma-Aldrich 13021-50]) were added to the nuclei pellet, and the nuclei suspension was immediately centrifuged at 3000g for 10 min at room temperature. After the supernatant was discarded, the nuclei pellet was resuspended in 50 µL of transposase master mixture (25 µL of 2× TD buffer, 22.5 µL nuclease-free water, and 2.5 µL of Tn5) and then incubated for 30 min at 37°C with 300 rpm shaking. After the incubation, 50 µL of 2× reverse cross-linking solution (100 mM Tris at pH 8.0 [Invitrogen AM9855G], 2 mM EDTA [Invitrogen AM9260G], 2% SDS [Invitrogen 15553-035], 0.4M NaCl [Invitrogen AM9759]) and 1µL Proteinase K (Thermo Fisher Scientific EO0491) were directly added into the tagmentation reaction mixture, and the mixture was incubated overnight at 65°C with 1200 rpm shaking. Following purification with a MinElute PCR purification kit (Qiagen 28006), amplification was performed using the same protocol as described in the OMNI-ATAC protocol but with a fixed total of 13 cycles. Final purification, library quantification and sequencing were similar to the OMNI-ATAC protocol.

### ***Single cell ATAC-seq***

Two angioimmunoblastic lymphoma samples were subjected to single-cell ATAC sequencing. Nuclei were isolated from flash frozen sections by incubation in lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>, 1% BSA, 0.1% Tween-20, 0.1% Nonidet P40 Substitute, 0.01% Digitonin) for 15 minutes on ice, filtered, and centrifuged. Nuclei were pelleted in a fixed-angle rotor at 500 r.c.f. for 5 min at 4 °C. The supernatant was removed, and the pellet was resuspended in Diluted Nuclei Buffer (10x Genomics). The nuclei stock concentration was determined for each sample using trypan blue, and a total of 16000 nuclei from each sample were pooled together and loaded into the 10x Genomics scATAC-seq (version 1.1) transposition reaction. Subsequent steps for scATAC-seq library preparation were performed as recommended by the manufacturer. Libraries were pooled and sequenced on an Illumina HiSeqX using paired-end 100-bp reads. Raw sequencing data were converted to FastQ format using the 'cellranger-atac mkfastq' pipeline (10x Genomics, version 2.0.0).

scATAC-seq reads were aligned to the GrCH38 reference genome and quantified using the 'cellranger-atac count' pipeline (10x Genomics, version 2.0.0). ArchR software<sup>49</sup> was used to make Arrowfiles from fragment files, filter doublets and low and very high-quality cells (TSS enrichment under 5 or above 30, fragment number under 8000 or above 50000). Fragments were counted in the peak set from a published single-cell ATAC-seq chromatin atlas<sup>25</sup>. Iterative LSI dimension reduction and clustering were done on peak matrix. UMAP projection parameters were slightly modified (nNeighbors = 30, minDist = 0.5). Clusters were identified using gene scores and findMarkers function. Tumoral TFH cell clusters were identified for each sample according to high CD3, CD4, PDCD1, CXCL13, BCL6, ICOS and/or CXCR5 gene scores. To integrate single cell and bulk data, pseudobulk profiles were obtained by summing

cell counts per clusters and samples in each peak. Pseudobulks with total number of counts under 1e6 were not kept for further analysis.

Single cell peak matrix from the same human atlas of chromatin accessibility were downloaded<sup>25</sup>. For each cell cluster, pseudobulk profiles were made by summing the signal of 200 cells. Cell types of interest were then selected and grouped according to large categories (epithelial cells, fibroblasts, endothelial cells, macrophages, memory B cells, plasma cells) (Supplementary Table 4). Five pseudobulk profiles were randomly chosen for each category for further analysis.

### ***ATAC-seq data preprocessing***

ATAC-seq data was processed with custom Nextflow pipelines. Sequencing adapters were trimmed with cutadapt version 3.6<sup>50</sup>, reads were aligned to GRCh38 (GCA\_000001405.15), HTLV1 (J02029.1) with BWA-MEM<sup>51</sup> version 2.2.1 and duplicates were filtered using samblaster<sup>52</sup> version 0.1.26. Chromatin availability peaks were called running MACS2 version 2.2.7.1 on tagmentation sites extended by 75 bp in each direction, with the --nolambda parameter. Peaks were post-processed as previously described<sup>53</sup> to enforce a consistent 500-bp width and to resolve overlaps. Peak maps were built as previously described<sup>53</sup>, retaining peaks overlapped by at least two distinct samples. Minimal quality thresholds enforced for map building and counting are presented in Supplementary Table 5. Peaks were annotated with Rgb<sup>54</sup> version 1.7.4 using Ensembl Genes 106 annotation.

### ***Unsupervised projection and clustering of MTCN ATAC-seq profiles***

For unsupervised analysis of sorted and frozen MTCN sections, matrix of insertion events in Open Chromatin Region derived from the single cell atlas of chromatin accessibility was used

excluding chromosome Y, TCR and BCR locus<sup>25</sup>. Tagmentation events per OCR were counted with bedtools<sup>55</sup> version 2.30.0. Matrix of counts were log-transformed and normalized using EdgeR<sup>56</sup> version 3.35.3 (TMM algorithm, prior.count 4). Further quantile normalization was done using preprocessCore function normalize.quantiles. Ten thousand most variant OCR were selected before dimension reduction using genome-wide CoGAPS non-negative matrix factorization (NMF) algorithm with 10 to 20 patterns, and nSet = 8<sup>57</sup>. Final pattern scores for each sample were determined by clustering the 880 patterns obtained from NMF at different k, selecting most recurrent patterns (correlated over more than 40 iterations). A new NMF was then seeded with the final sample pattern matrix to get the peak pattern scores. NMF sample patterns correlated to TSS enrichment scores (Pearson's  $\rho < -0.5$ ) were removed. Microenvironment patterns were determined by correlating peak pattern scores and single-cell pseudobulk log counts derived from the single-cell chromatin accessibility landscape. To identify epigenetic signature of MTCN entities, shared nearest neighbours clustering was done on remaining NMF patterns using the function SNNclust from the dbscan packages. Clusters were visualized with Uniform Manifold Approximation and Projection (UMAP). Each cluster was named according to its most frequent pathology entity. Top features describing each pattern were found by normalizing the peak pattern matrix so that the sum of pattern scores for each peak equals to one and selecting the top 100 peaks for each pattern.

### ***Classifier training and prediction***

Using the projectR function, NMF peak patterns were projected on the matrix of most variant OCR per samples. Scaled EBV and HTLV1 counts were concatenated to the pattern scores excluding microenvironment, and quality patterns. The resulting matrix was used as input into a generalized linear model multinomial classifier (R package glmnet). Optimal



lambda was determined by maximizing class accuracy on the training cohort using a leave-one-out approach.

For prediction on the validation cohort, matrix of tagmentation events in OCR per samples were TMM-normalized and log-transformed as previously described. Quantile normalization was applied to each sample in order to fit the distribution of the training cohort (preprocessCore: normalize.quantiles.use.target function). And the pattern decomposition was obtained using projectR and submitted together with scaled viral counts to the glm classifier prediction. Discordant cases and PTCL-NOS were submitted to pathology review.

### ***Copy number variation analysis***

### ***Cell of origin identification***

To increase the diversity of our atlas of normal T/NK cells ATAC-seq profiles, innate lymphoid cell and thymocyte ATAC-seq FASTQ files from previous publications<sup>58,59</sup> were downloaded and processed with our custom pipeline. Peaks from all normal cell subtypes were used to generate a map of normal chromatin OCRs by iterative filtering as described previously<sup>24</sup>. Differential analysis using edgeR was then performed between all normal cell subtype pairs (one vs. one) to define subset-specific peaks. The top 100 peaks that were over or under-accessible in each comparison were retained for further analysis. Transposition events were counted in this set of peaks for normal cells and sorted tumor cell samples. The matrix of counts was then subjected to Ward clustering on Pearson's correlation distance.

### ***Angioimmunoblastic microenvironment deconvolution and prognostic assessment***

To assess the capacity of NMF deconvolution to identify microenvironment components, 118 TFH lymphomas ATAC-seq profiles from the training class were used for a new NMF projection. CoGAPS package was used with 2 to 10 patterns on 8 partitions (nSets) of the 10,000 most variant OCRs. For each feature partition, a connectivity matrix  $C$  of size  $m \times m$  was generated where  $m$  is the number of samples, with entry  $c_{ij} = 1$  if samples  $i$  and  $j$  have the same maximal component, and  $c_{ij} = 0$  if they have different maximal components. The consensus matrix was defined as the average connectivity matrix obtained from each OCR partition. Cophenetic coefficients were derived from the consensus matrix as defined in Brunet et al. <sup>60</sup>. The final rank of  $k=5$  was chosen to minimize the cophenetic coefficient. Quality component was recognized as being highly correlated with the inverse of sample TSS enrichment scores and was removed for downstream analyses. To identify cell type contribution to each pattern Pearson correlation of the OCR patterns with the single cell pseudobulk profiles was calculated. Samples were clustered based on patterns using Ward Clustering and Pearson's distance and 4 clusters were chosen based on visual inspection. Survival was compared using Cox regression model to correct for confounding variables (R packages survival and survminer).

### ***Exome sequencing and analysis***

Total DNA was extracted from 50 $\mu$ m frozen sections using Nucleospin Tissue Kit (Macherey-Nagel, Cat# 740952). Libraries were prepared using the Twist Human Comprehensive Exome kit and sequenced on a Novaseq 6000 to generate 150 base paired-end reads to a target depth of 300X. Sequences were aligned to hg19 assembly using BWA-MEM. Duplicate reads were filtered with MarkDuplicates (Picard). Realignment around indels and base recalibration

was done using GATK tools (v3.7 and v4.1). Single nucleotide and indel variant calling were performed using VarScan (v2.4.4) and MuTect2 (GATK v4.1). Pass variants were filtered if they filled the following criteria: > 30 read depth, > 10 read supporting alteration, allele frequency between 10 and 90%, low orientation bias, present in less than 4 samples, present in more than 6 patients in GnomAD database.

### ***RNA-seq wet lab***

Total RNA was extracted from dry cell pellets or 50µm frozen sections using the Qiagen AllPrep DNA/RNA micro kit (Qiagen Cat# 80284). Libraries were prepared using the combination of 2 kits: QIAseqFastSelect –rRNA HMR Kit from QiaGen, and NEBNext Ultra II Directional RNA Library Prep Kit for Illumina protocol according to suppliers' recommendations. Briefly, the key stages of this protocol were the removal of ribosomal RNA fraction from 100ng of total RNA using QIAseq FastSelect removal reagent. The reagent was directly combined with total RNA and the library preparation-specific buffers, a fragmentation under elevated temperature was used to obtain approximately 300bp pieces followed by double-strand cDNA synthesis and Illumina® adapters ligation before cDNA library amplification by PCR for sequencing. Libraries were sequenced using 100 bases paired-end run on an Illumina® NovaSeq 6000. Image analysis and base calling was performed using Illumina Real Time Analysis (3.4.4) with default parameters.

### ***RNA-seq data analysis***

RNA-sequencing data was processed with the SAMI pipeline version 0.9.1 (<https://bioinformatics.ovsa.fr/SAMI>). Briefly, it consisted of a two-pass alignment to GRCh38 with STAR version 2.7.9a, read pair counting at the gene level using the “strand” option from featureCounts (Rsubreads version 2.8.0) in GENCODE version 41 annotation and TMM

normalization with edgeR version 3.35.3. Differential analysis was conducted using edgeR exactTest.

### ***BaseScope Detection Assay***

Branched RNA ISH was performed in vitro and on lymph node tissue with the BaseScope assays in accordance with guidelines provided by the supplier (Advanced Cell Diagnostics, Newark, CA). Sections were taken at 5 µm thickness onto Superfrost plus slides (Fisher Scientific, Loughborough, UK) and allowed to dry overnight at room temperature (RT). Sections were then baked at 60 °C for 1 h before deparaffinizing in xylene (2 × 5 min) and ethanol (2 × 2 min), then drying by baking at 60 °C for 5 min. Pretreat 1 (hydrogen peroxide) was applied for 10 min at RT, Pretreat 2 (target retrieval) for 15 min at 100 °C and Pretreat 3 (protease) for 30 min (tissue sections) at 40 °C in a HybEZ oven (ACD HybEZ™ II Hybridization System), with two rinses in distilled water between pretreatments. BaseScope probes were then applied for 2 h at 40 °C in a HybEZ oven before incubation with reagents AMP1 (30 min at 40 °C), AMP2 (30 min at 40 °C), AMP3 (15 min at 40 °C), AMP4 (30 min at 40 °C), AMP5 (30 min at 40 °C), AMP6 (15min at 40°C), AMP7 (30 min at RT) and AMP8 (15min at RT). Slides were rinsed with wash buffer (2 × 2 min) between each AMP incubation. Finally slides were incubated with Fast Red for 10 min at room temperature in the dark and then counterstained with Gill's Hematoxylin I. Slides were dried for 15 min at 60 °C before mounting in EcoMount permanent mounting medium (Biocare Medical). Sections were examined under a standard bright field microscope at 20–40X magnification.

## **ACKNOWLEDGEMENTS**

This work was supported by a PhD grant from the Association de la Recherche Contre le Cancer. The authors thank the Institut Carnot Calym for providing access to the CeVi

collection. We would like to thank the Centre de Ressources Biologiques (CRB Sud) from Hospices Civils de Lyon for samples. We acknowledge the contribution from SFR Biosciences and cytometry facility Lyon Sud for providing access to cell sorting.

## REFERENCES

1. Shaffer AL, Young RM, Staudt LM. Pathogenesis of human B cell lymphomas. *Annu Rev Immunol*. 2012;30:565-610. doi:10.1146/annurev-immunol-020711-075027
2. Van Arnam JS, Lim MS, Elenitoba-Johnson KSJ. Novel insights into the pathogenesis of T-cell lymphomas. *Blood*. 2018;131(21):2320-2330. doi:10.1182/blood-2017-11-764357
3. Alaggio R, Amador C, Anagnostopoulos I, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia*. 2022;36(7):1720-1748. doi:10.1038/s41375-022-01620-2
4. Laurent C, Baron M, Amara N, et al. Impact of Expert Pathologic Review of Lymphoma Diagnosis: Study of Patients From the French Lymphopath Network. *J Clin Oncol*. 2017;35(18):2008-2017. doi:10.1200/JCO.2016.71.2083
5. Smith A, Crouch S, Lax S, et al. Lymphoma incidence, survival and prevalence 2004–2014: sub-type analyses from the UK's Haematological Malignancy Research Network. *Br J Cancer*. 2015;112(9):1575-1584. doi:10.1038/bjc.2015.94
6. Chapuy B, Stewart C, Dunford AJ, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nature Medicine*. 2018;24(5):679. doi:10.1038/s41591-018-0016-8
7. Choi J, Goh G, Walradt T, et al. Genomic landscape of cutaneous T cell lymphoma. *Nature Genetics*. 2015;47(9):1011-1019. doi:10.1038/ng.3356
8. da Silva Almeida AC, Abate F, Khiabani H, et al. The mutational landscape of cutaneous T cell lymphoma and Sézary syndrome. *Nature Genetics*. 2015;47(12):1465-1470. doi:10.1038/ng.3442
9. Jiang L, Gu ZH, Yan ZX, et al. Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma. *Nature Genetics*. 2015;47(9):1061-1066. doi:10.1038/ng.3358
10. Kataoka K, Nagata Y, Kitanaka A, et al. Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nat Genet*. 2015;47(11):1304-1315. doi:10.1038/ng.3415
11. Palomero T, Couronné L, Khiabani H, et al. Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nature Genetics*. 2014;46(2):166-170. doi:10.1038/ng.2873

12. Sakata-Yanagimoto M, Enami T, Yoshida K, et al. Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. *Nature Genetics*. 2014;46(2):171-175. doi:10.1038/ng.2872
13. Schmitz R, Wright GW, Huang DW, et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *New England Journal of Medicine*. Published online April 11, 2018. doi:10.1056/NEJMoa1801445
14. Kogure Y, Kameda T, Koya J, et al. Whole-genome landscape of adult T-cell leukemia/lymphoma. *Blood*. 2022;139(7):967-982. doi:10.1182/blood.2021013568
15. Park J, Yang J, Wenzel AT, et al. Genomic analysis of 220 CTCLs identifies a novel recurrent gain-of-function alteration in RLTPR (p.Q575E). *Blood*. 2017;130(12):1430-1440. doi:10.1182/blood-2017-02-768234
16. Vallois D, Dobay MPD, Morin RD, et al. Activating mutations in genes related to TCR signaling in angioimmunoblastic and other follicular helper T-cell-derived lymphomas. *Blood*. 2016;128(11):1490-1502. doi:10.1182/blood-2016-02-698977
17. de Leval L, Rickman DS, Thielen C, et al. The gene expression profile of nodal peripheral T-cell lymphoma demonstrates a molecular link between angioimmunoblastic T-cell lymphoma (AITL) and follicular helper T (TFH) cells. *Blood*. 2007;109(11):4952-4963. doi:10.1182/blood-2006-10-055145
18. Iqbal J, Wright G, Wang C, et al. Gene expression signatures delineate biological and prognostic subgroups in peripheral T-cell lymphoma. *Blood*. 2014;123(19):2915-2923. doi:10.1182/blood-2013-11-536359
19. Piccaluga PP, Fuligni F, De Leo A, et al. Molecular Profiling Improves Classification and Prognostication of Nodal Peripheral T-Cell Lymphomas: Results of a Phase III Diagnostic Accuracy Study. *JCO*. 2013;31(24):3019-3025. doi:10.1200/JCO.2012.42.5611
20. Drieux F, Ruminy P, Abdel-Sater A, et al. Defining signatures of peripheral T-cell lymphoma with a targeted 20-marker gene expression profiling assay. *Haematologica*. 2020;105(6):1582-1592. doi:10.3324/haematol.2019.226647
21. Amador C, Bouska A, Wright G, et al. Gene Expression Signatures for the Accurate Diagnosis of Peripheral T-Cell Lymphoma Entities in the Routine Clinical Practice. *Journal of Clinical Oncology*. Published online July 15, 2022. doi:10.1200/JCO.21.02707
22. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol*. 2015;109:21.29.1-21.29.9. doi:10.1002/0471142727.mb2129s109
23. Rendeiro AF, Schmidl C, Strefford JC, et al. Chromatin accessibility maps of chronic lymphocytic leukaemia identify subtype-specific epigenome signatures and transcription regulatory networks. *Nat Commun*. 2016;7(1):11938. doi:10.1038/ncomms11938

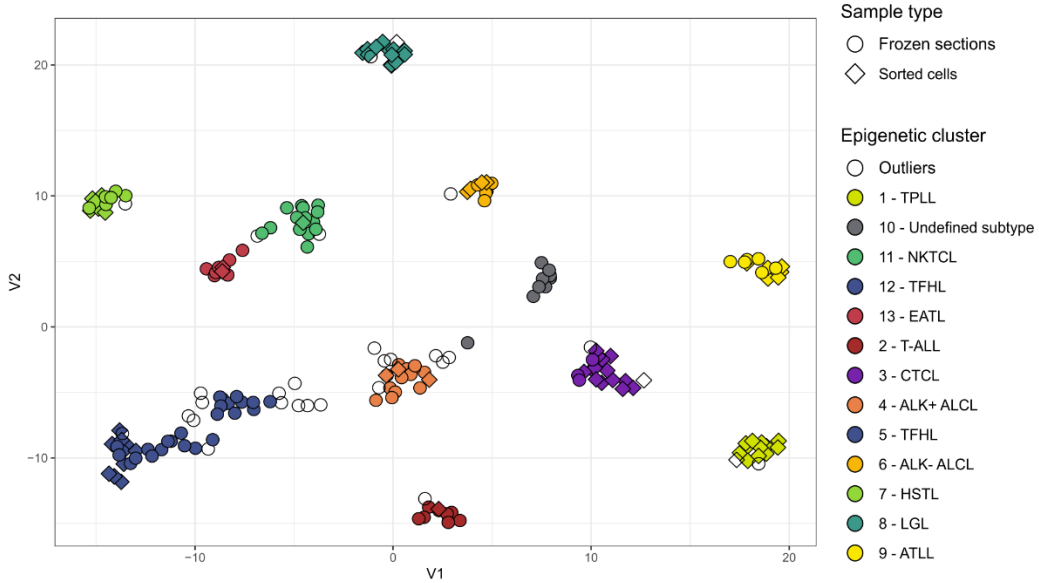
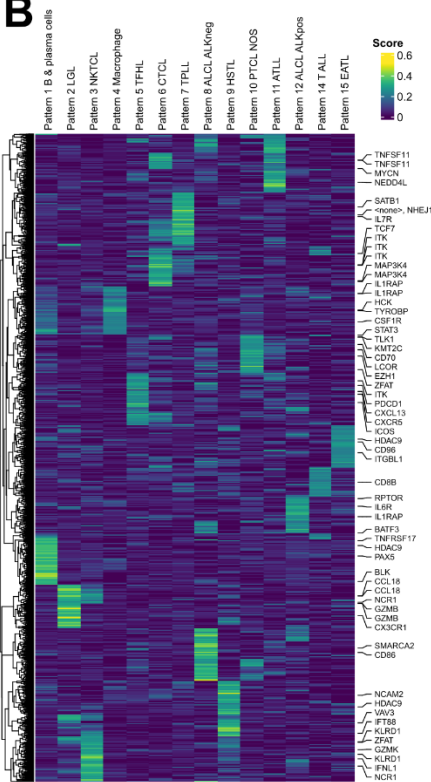
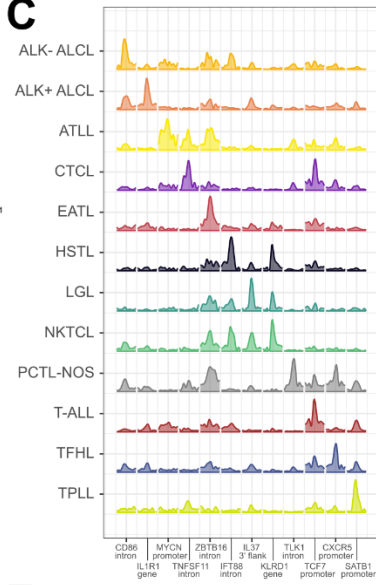
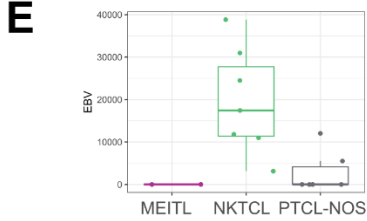
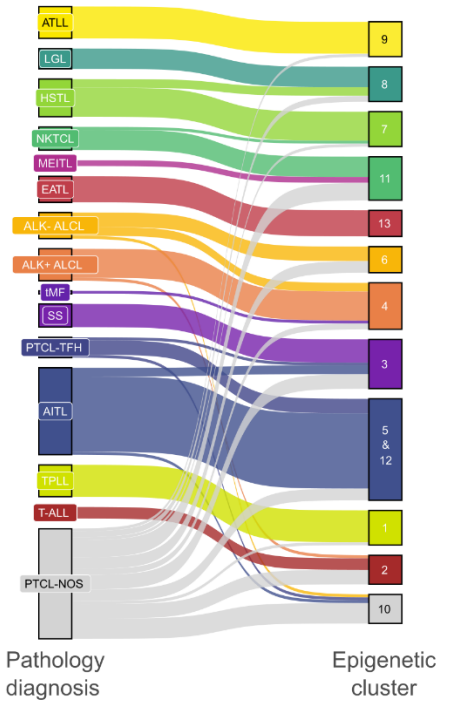
24. Corces MR, Buenrostro JD, Wu B, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet.* 2016;48(10):1193-1203. doi:10.1038/ng.3646
25. Zhang K, Hocker JD, Miller M, et al. A single-cell atlas of chromatin accessibility in the human genome. *Cell.* 2021;184(24):5985-6001.e19. doi:10.1016/j.cell.2021.10.024
26. Cheng J, Novati G, Pan J, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science.* 2023;381(6664):eadg7492. doi:10.1126/science.adg7492
27. Yadav RP, Polavarapu VK, Xing P, Chen X. FFPE-ATAC: A Highly Sensitive Method for Profiling Chromatin Accessibility in Formalin-Fixed Paraffin-Embedded Samples. *Curr Protoc.* 2022;2(8):e535. doi:10.1002/cpz1.535
28. O'Connor OA, Falchi L, Lue JK, et al. Oral 5-azacytidine and romidepsin exhibit marked activity in patients with PTCL: a multicenter phase 1 study. *Blood.* 2019;134(17):1395-1405. doi:10.1182/blood.2019001285
29. Ghione P, Faruque P, Mehta-Shah N, et al. T follicular helper phenotype predicts response to histone deacetylase inhibitors in relapsed/refractory peripheral T-cell lymphoma. *Blood Advances.* 2020;4(19):4640-4647. doi:10.1182/bloodadvances.2020002396
30. Bachy E, Brice P, Delarue R, et al. Long-Term Follow-Up of Patients With Newly Diagnosed Follicular Lymphoma in the Prerituximab Era: Effect of Response Quality on Survival—A Study From the Groupe d'Etude des Lymphomes de l'Adulte. *Journal of Clinical Oncology.* 2010;28(5):822-829. doi:10.1200/JCO.2009.22.7819
31. Banerji S, Cibulskis K, Rangel-Escareno C, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature.* 2012;486(7403):405-409. doi:10.1038/nature11154
32. Stephens PJ, Tarpey PS, Davies H, et al. The landscape of cancer genes and mutational processes in breast cancer. *Nature.* 2012;486(7403):400-404. doi:10.1038/nature11017
33. Gui Y, Guo G, Huang Y, et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet.* 2011;43(9):875-878. doi:10.1038/ng.907
34. Ivanov I, Lo KC, Hawthorn L, Cowell JK, Ionov Y. Identifying candidate colon cancer tumor suppressor genes using inhibition of nonsense-mediated mRNA decay in colon cancer cells. *Oncogene.* 2007;26(20):2873-2884. doi:10.1038/sj.onc.1210098
35. Guenther MG, Barak O, Lazar MA. The SMRT and N-CoR corepressors are activating cofactors for histone deacetylase 3. *Mol Cell Biol.* 2001;21(18):6091-6101. doi:10.1128/MCB.21.18.6091-6101.2001

36. You SH, Lim HW, Sun Z, Broache M, Won KJ, Lazar MA. Nuclear receptor co-repressors are required for the histone-deacetylase activity of HDAC3 in vivo. *Nat Struct Mol Biol.* 2013;20(2):182-187. doi:10.1038/nsmb.2476
37. Wang T, Feldman AL, Wada DA, et al. GATA-3 expression identifies a high-risk subset of PTCL, NOS with distinct molecular and clinical features. *Blood.* 2014;123(19):3007-3015. doi:10.1182/blood-2013-12-544809
38. Amador C, Greiner TC, Heavican TB, et al. Reproducing the molecular subclassification of peripheral T-cell lymphoma-NOS by immunohistochemistry. *Blood.* 2019;134(24):2159-2170. doi:10.1182/blood.2019000779
39. Heavican TB, Bouska A, Yu J, et al. Genetic drivers of oncogenic pathways in molecular subgroups of peripheral T-cell lymphoma. *Blood.* 2019;133(15):1664-1676. doi:10.1182/blood-2018-09-872549
40. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature.* 2000;403(6769):503-511. doi:10.1038/35000501
41. Palmer AC, Kurtz DM, Alizadeh AA. Cell-of-Origin Subtypes and Therapeutic Benefit from Polatuzumab Vedotin. *N Engl J Med.* 2023;389(8):764-766. doi:10.1056/NEJMc2306105
42. Campbell JJ, Clark RA, Watanabe R, Kupper TS. Sézary syndrome and mycosis fungoides arise from distinct T-cell subsets: a biologic rationale for their distinct clinical behaviors. *Blood.* 2010;116(5):767-771. doi:10.1182/blood-2009-11-251926
43. Oberbeck S, Schrader A, Warner K, et al. Non-canonical effector functions of the T-memory-like T-PLL cell are shaped by cooperative TCL1A and TCR signaling. *Blood.* Published online September 21, 2020: blood.2019003348. doi:10.1182/blood.2019003348
44. Newman AM, Bratman SV, To J, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med.* 2014;20(5):548-554. doi:10.1038/nm.3519
45. Chu T, Wang Z, Pe'er D, Danko CG. Cell type and gene expression deconvolution with BayesPrism enables Bayesian integrative analysis across bulk and single-cell RNA sequencing in oncology. *Nat Cancer.* 2022;3(4):505-517. doi:10.1038/s43018-022-00356-3
46. Capper D, Jones DTW, Sill M, et al. DNA methylation-based classification of central nervous system tumours. *Nature.* 2018;555(7697):469-474. doi:10.1038/nature26000
47. Koelsche C, Schrimpf D, Stichel D, et al. Sarcoma classification by DNA methylation profiling. *Nat Commun.* 2021;12(1):498. doi:10.1038/s41467-020-20603-4
48. Corces MR, Trevino AE, Hamilton EG, et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods.* 2017;14(10):959-962. doi:10.1038/nmeth.4396



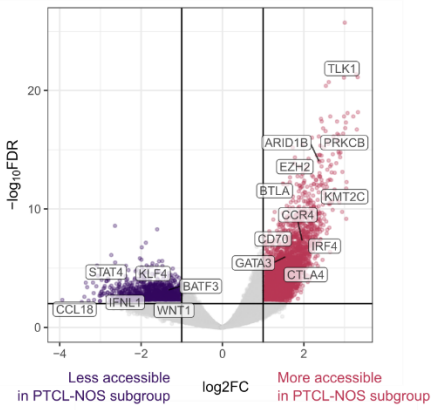
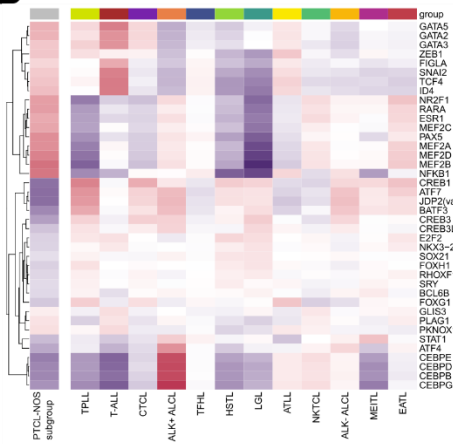
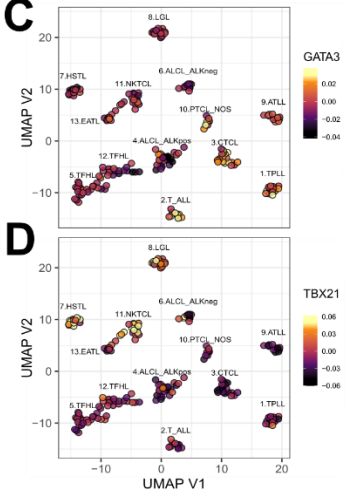
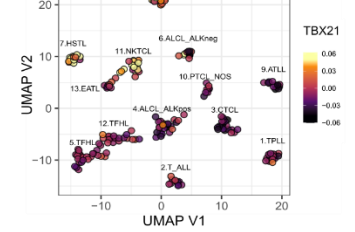
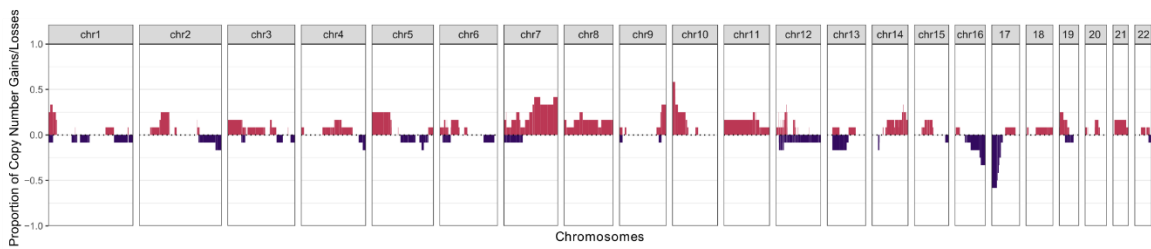
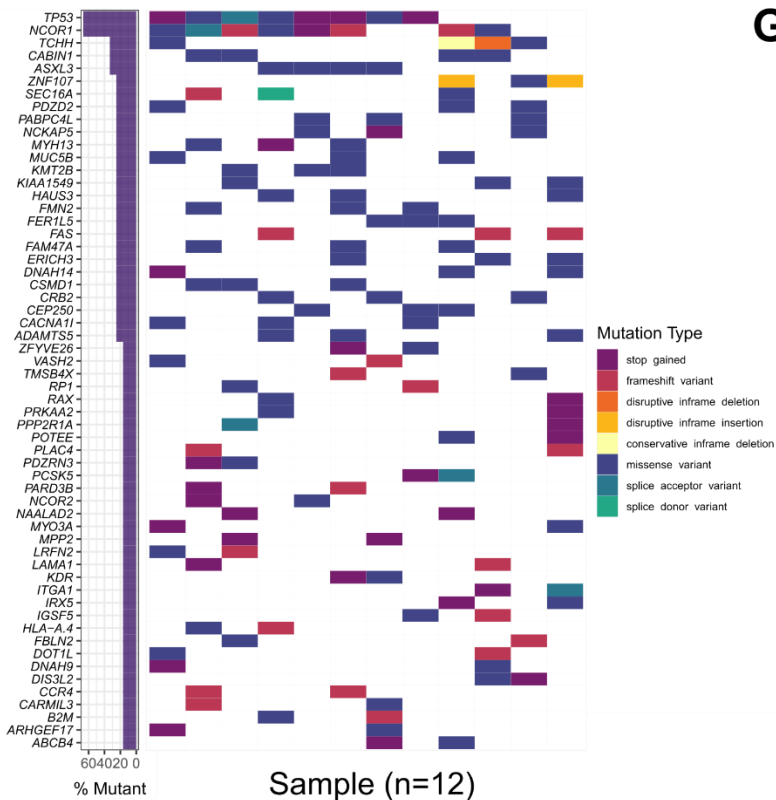
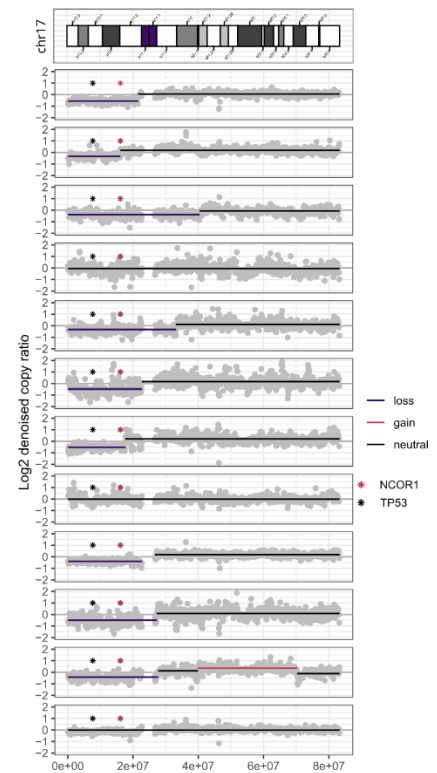
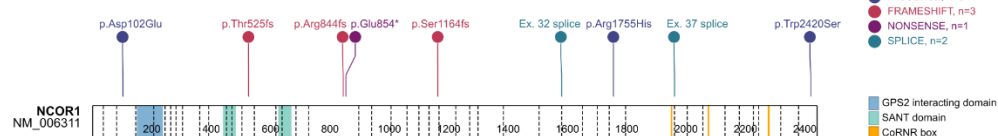
49. Granja JM, Corces MR, Pierce SE, et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet.* 2021;53(3):403-411. doi:10.1038/s41588-021-00790-6
50. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* 2011;17(1):10-12. doi:10.14806/ej.17.1.200
51. Vasimuddin Md, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In: *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS).* ; 2019:314-324. doi:10.1109/IPDPS.2019.00041
52. Faust GG, Hall IM. SAMBLASTER : fast duplicate marking and structural variant read extraction. *Bioinformatics.* 2014;30(17):2503-2505. doi:10.1093/bioinformatics/btu314
53. Corces MR, Granja JM, Shams S, et al. The chromatin accessibility landscape of primary human cancers. *Science.* 2018;362(6413). doi:10.1126/science.aav1898
54. Mareschal S, Dubois S, Lecroq T, Jardin F. Rgb: a scriptable genome browser for R. *Bioinformatics.* 2014;30(15):2204-2205. doi:10.1093/bioinformatics/btu185
55. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841-842. doi:10.1093/bioinformatics/btq033
56. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26(1):139-140. doi:10.1093/bioinformatics/btp616
57. Stein-O'Brien GL, Carey JL, Lee WS, et al. PatternMarkers & GWCoGAPS for novel data-driven biomarkers via whole transcriptome NMF. *Bioinformatics.* 2017;33(12):1892-1894. doi:10.1093/bioinformatics/btx058
58. Koues OI, Collins PL, Cella M, et al. Distinct Gene Regulatory Pathways for Human Innate versus Adaptive Lymphoid Cells. *Cell.* 2016;165(5):1134-1146. doi:10.1016/j.cell.2016.04.014
59. Roels J, Van Hulle J, Lavaert M, et al. Transcriptional dynamics and epigenetic regulation of E and ID protein encoding genes during human T cell development. *Front Immunol.* 2022;13:960918. doi:10.3389/fimmu.2022.960918
60. Brunet JP, Tamayo P, Golub TR, Mesirov JP. Metagenes and molecular pattern discovery using matrix factorization. *Proc Natl Acad Sci U S A.* 2004;101(12):4164-4169. doi:10.1073/pnas.0308531101

## FIGURES

**A****B****C****E****D**

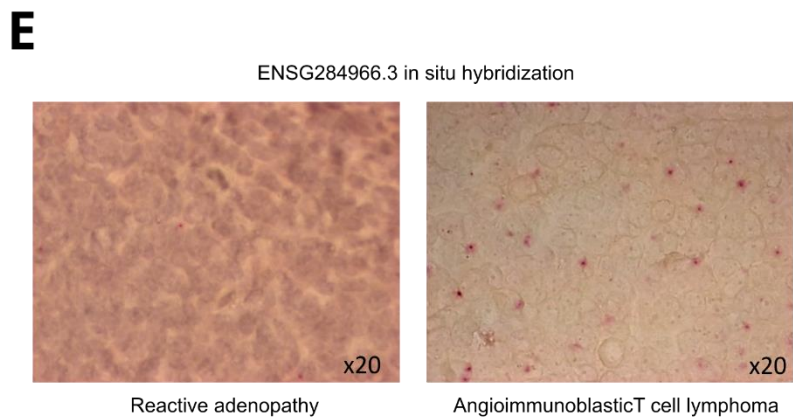
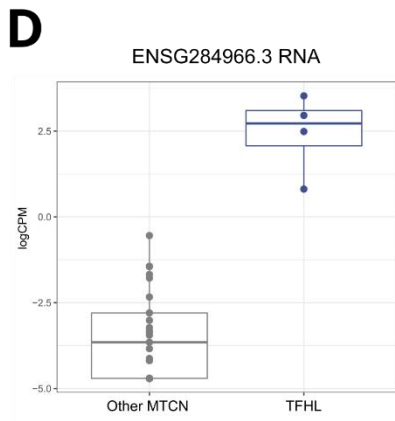
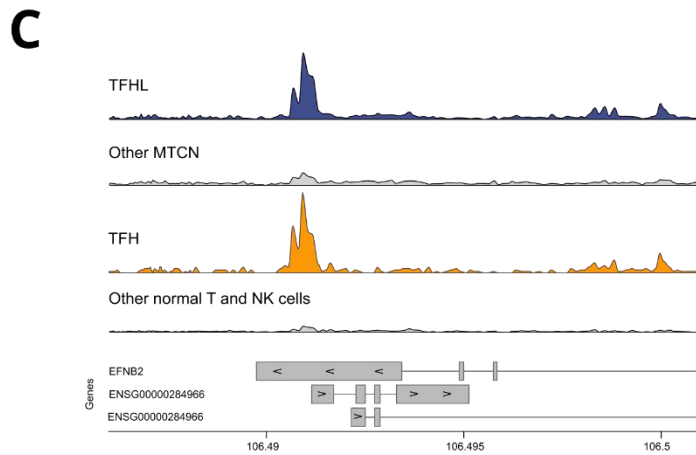
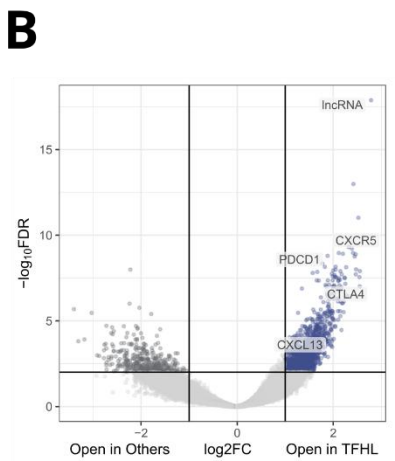
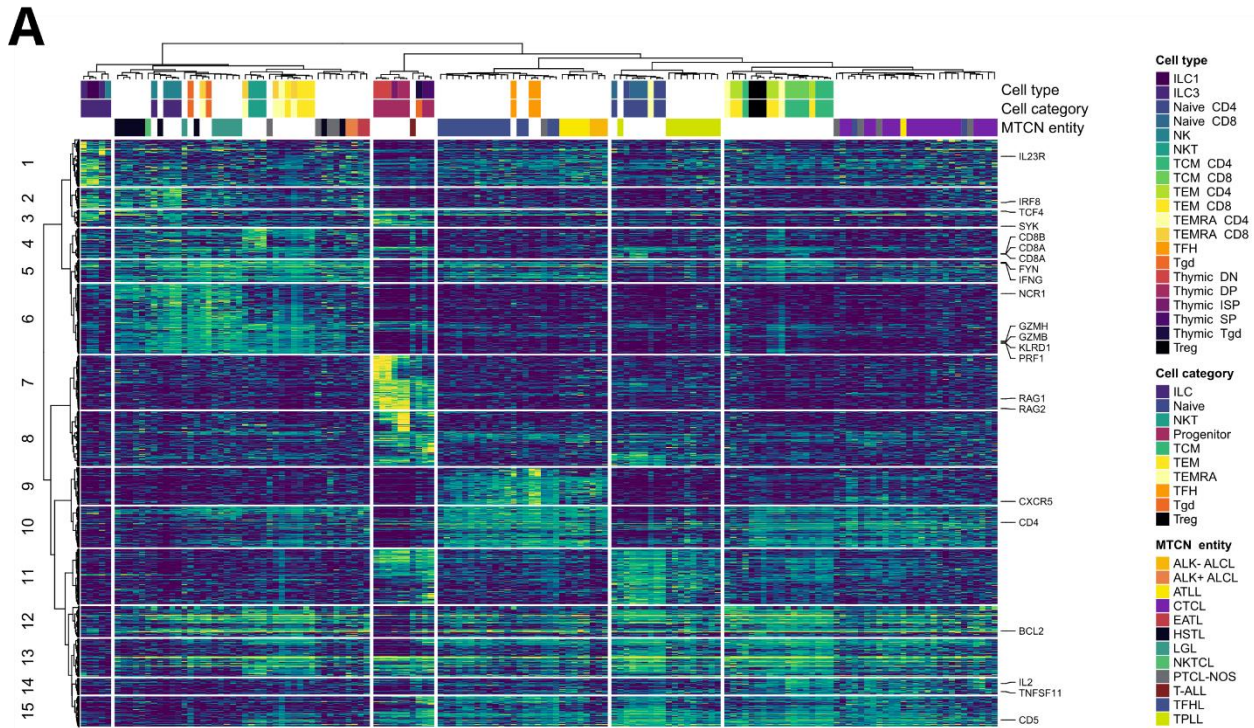
## Figure 1

- A. Diagram of samples included in the study and main processing steps.
- B. Unsupervised UMAP projection on NMF patterns of MTCN samples from the training cohort (excluding microenvironment and quality patterns). Individual samples are color-coded based on SNN clustering. Light grey is used for samples without formal cluster assignment.
- C. Heatmap of chromatin accessibility in pattern-specific peaks. The top 200 peaks per pattern are shown.
- D. Average chromatin accessibility in cluster specific peaks.
- E. Sankey diagram showing the correspondence between anatomopathological diagnosis and epigenetic clustering.
- F. EBV counts from ATAC-sequencing allow for the differentiation of EBV positive and negative cytotoxic lymphomas among cluster #11.

**A****B****C****D****E****F****G****H**

## Figure 2

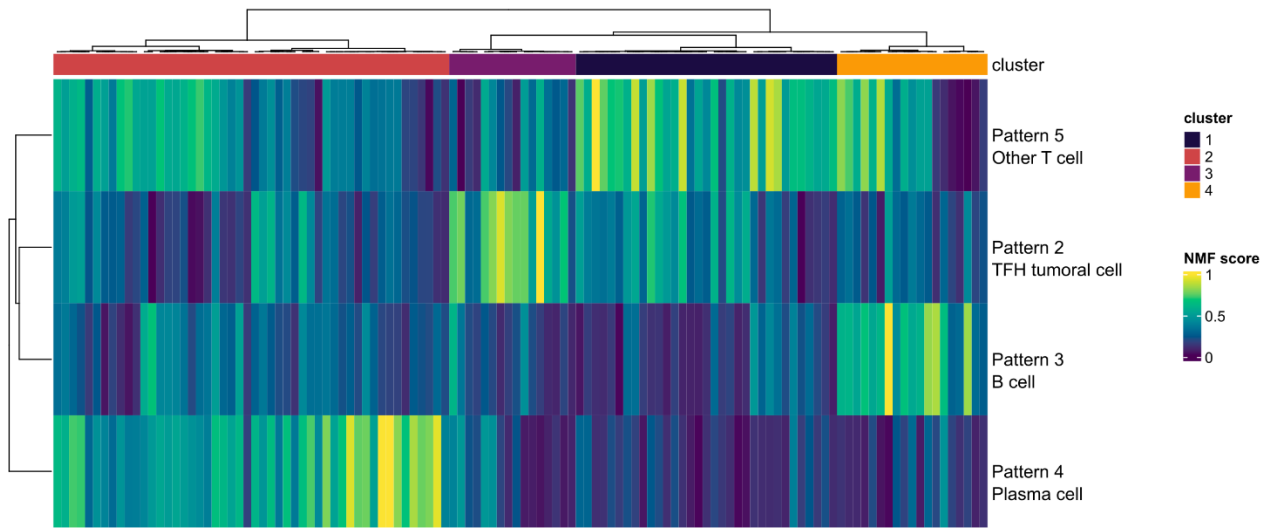
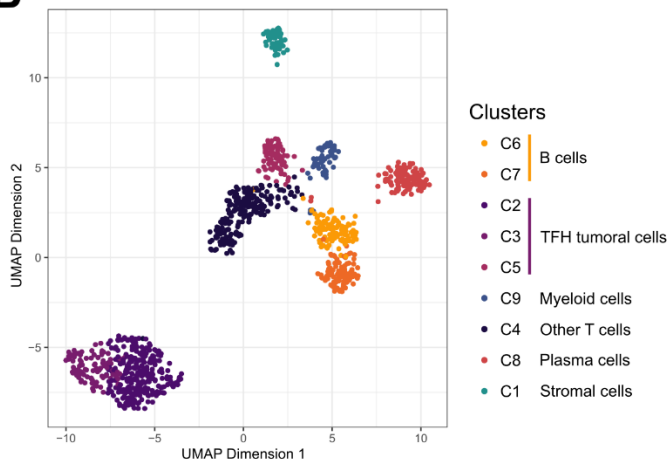
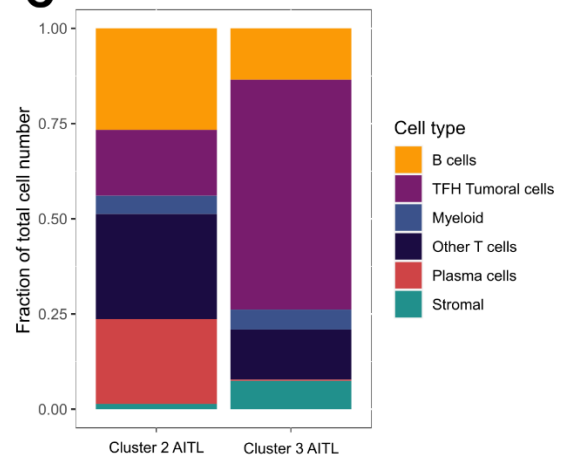
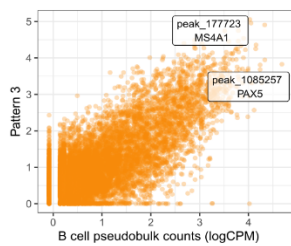
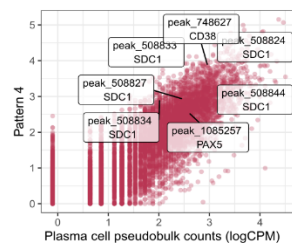
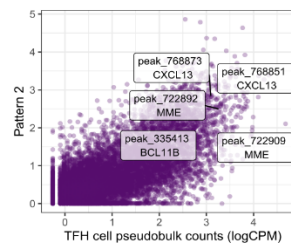
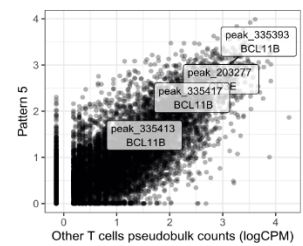
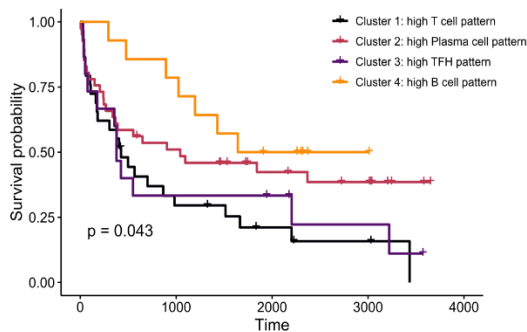
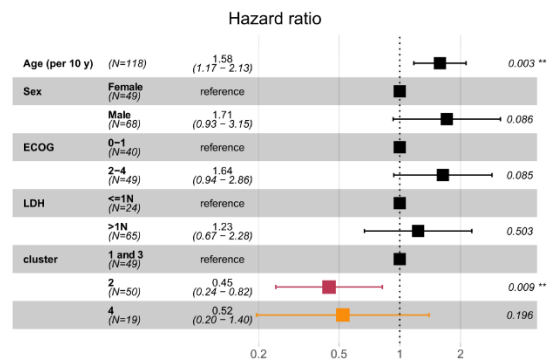
- A. Volcano plot of differentially accessible peaks in PTCL samples from the cluster #10 vs. all other MTCN samples. Top peaks located in genes of interest are labelled.
- B. ChromVAR TF deviation heatmap average over each MTCN entity. The top 20 most accessible and least accessible TF motifs in the cluster #10 PTCL subgroup vs all other samples are shown.
- C. UMAP of MTCN entities showing high GATA3 transcription factor activity in CTCL, T-PLL, T-ALL as well as in the PTCL cluster #10.
- D. UMAP of MTCN entities showing high TBX21 transcription factor activity in LGL, HSTL and NKTCL.
- E. Proportion of copy number variations derived from ATAC-seq data shows a high rate of 17p deletion in the PTCL cluster #10 subgroup.
- F. Heatmap of single nucleotide variants and small insertion/deletions found in the cluster #10 subgroup by whole exome-sequencing showing high rate of *TP53* and *NCOR1* mutations.
- G. Individual copy number ratios showing deleted 17p regions encompassing not only the *TP53* but also the *NCOR1* gene.



### Figure 3

- A. Heatmap of T-cell subset specific OCR showing close proximity of MTCN tumoral sorted cells and normal T and NK cell counterparts.
- B. Differential accessibility analysis of Tfh-PTCL vs all other entities uncovers a specific opened peak in a new long non coding RNA. \*ENSG00000284966
- C. Mean chromatin accessibility at ENSG00000284966 promoter is high in Tfh lymphomas but also in normal Tfh cells.
- D. RNA seq expression of ENSG00000284966 is found significantly higher in Tfh-lymphoma than in other MTCN
- E. Basescope in situ hybridization of ENSG00000284966 in a reactive lymph node (left) and an AITL (right).

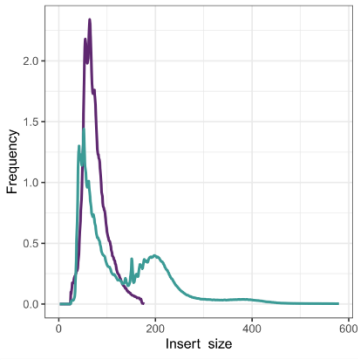


**A****B****C****D****E****F****G****H****I**

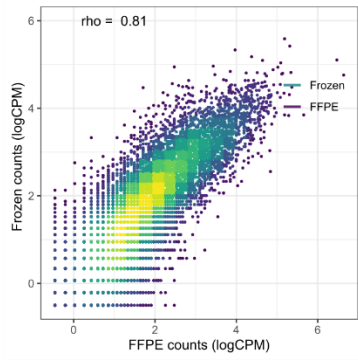
#### Figure 4

- A. Heatmap showing four clusters of T follicular helper cell lymphomas based on NMF component scores.
- B. UMAP projection of scATAC-seq profiles of two Tfh-PTCL samples with various microenvironment composition.
- C. Single cell composition of two samples showing a predominant plasma cell infiltration in the sample from cluster #2 and a predominant tumoral Tfh infiltration in the sample from cluster #3.
- D-G. Correlation between pseudobulk profiles from single-cell ATAC-seq and corresponding NMF patterns from Tfh-PTCL shown in A. (D: B-cell, E: plasma cells, F: CD4 TFH cell, G: other T cells).
- H. Survival curves showing different overall survival according to microenvironment composition
- I. Forest plot for multivariate analysis of overall survival

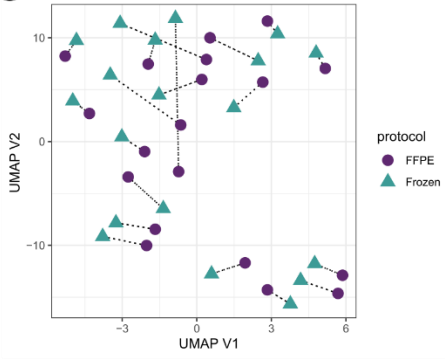
**A**



**B**



**C**



**D**

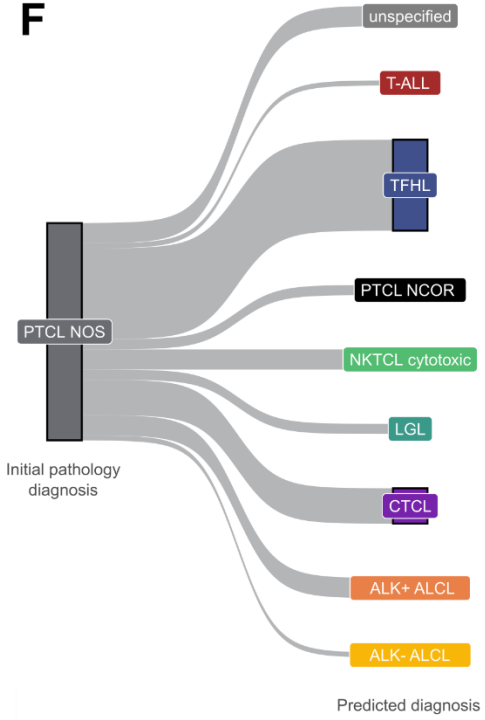


Prediction

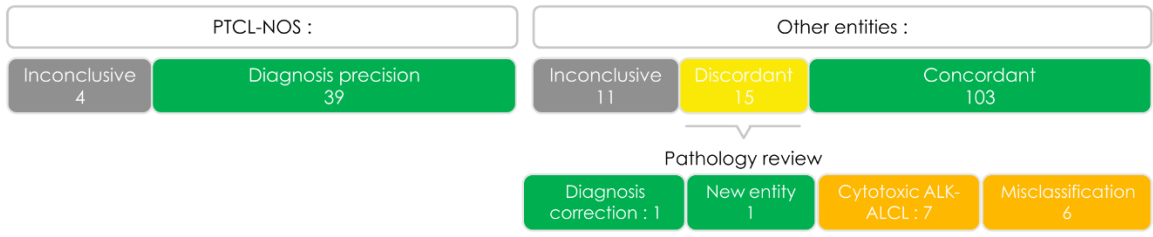
**E**

Initial Pathology Diagnosis	Predicted diagnosis												Class Acc.	
	ALK-ALCL	ALK+ALCL	ATLL	CTCL	EATL	HSTL	LGL	MEITL	NKTCL & cytotoxic	PTCL-NCOR	T-ALL	TFHL		TPLL
ALK-ALCL	4	7	0	0	0	0	0	0	0	0	0	0	0	36%
ALK+ALCL	0	7	0	0	0	0	0	0	0	0	0	0	1	88%
ATLL	0	0	1	0	0	0	0	0	0	0	0	0	0	100%
CTCL	0	0	0	5	0	0	0	0	0	0	0	0	0	100%
EATL	1	0	0	0	4	0	0	0	0	0	0	1	0	67%
HSTL	0	0	0	0	0	1	0	0	0	0	0	0	0	100%
LGL	0	0	0	0	0	0	0	0	0	0	0	0	0	NaN%
MEITL	0	0	0	0	0	0	0	3	0	0	0	0	0	100%
NKTCL & cytotoxic	1	0	0	0	0	0	0	0	3	0	0	0	0	75%
T-ALL	0	0	0	0	0	0	0	0	0	6	0	0	0	100%
TFHL	0	0	0	0	0	0	1	0	2	1	0	68	0	94%
TPLL	0	0	0	0	0	0	0	0	0	0	0	0	1	100%

**F**



**G**

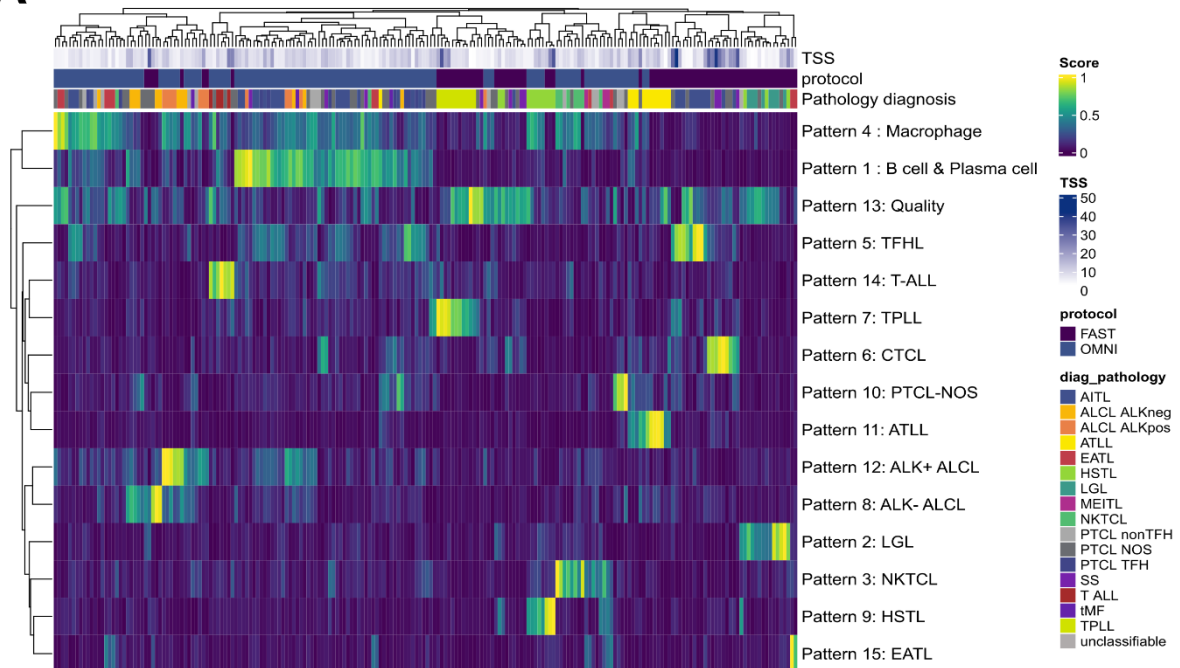


## Figure 5

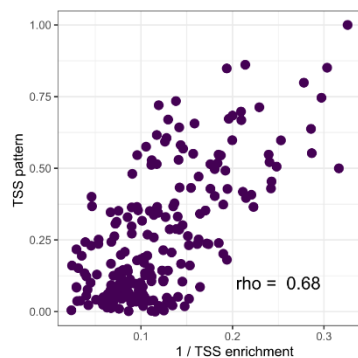
- A. Fragment size distribution in ATAC-seq from frozen and FFPE samples
- B. Correlation of ATAC-seq counts in top variant peaks between duplicate frozen and FFPE samples from the same patient biopsy
- C. UMAP after NMF dimension reduction showing very similar chromatin accessibility profiles obtained by ATAC-seq from frozen and FFPE sample pairs (linked by dashed line).
- D. Diagram of samples included in the validation cohort
- E. Confusion matrix of pathology diagnosis and predicted epigenetic class for all conclusive cases in the validation cohort excluding PTCL-NOS
- F. Reclassification of PTCL-NOS according to epigenetic prediction
- G. Diagram of reclassifications and pathology review concordance

# SUPPLEMENTARY DATA

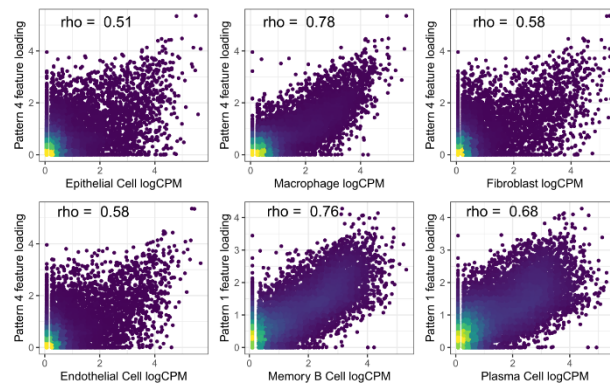
**A**



**B**

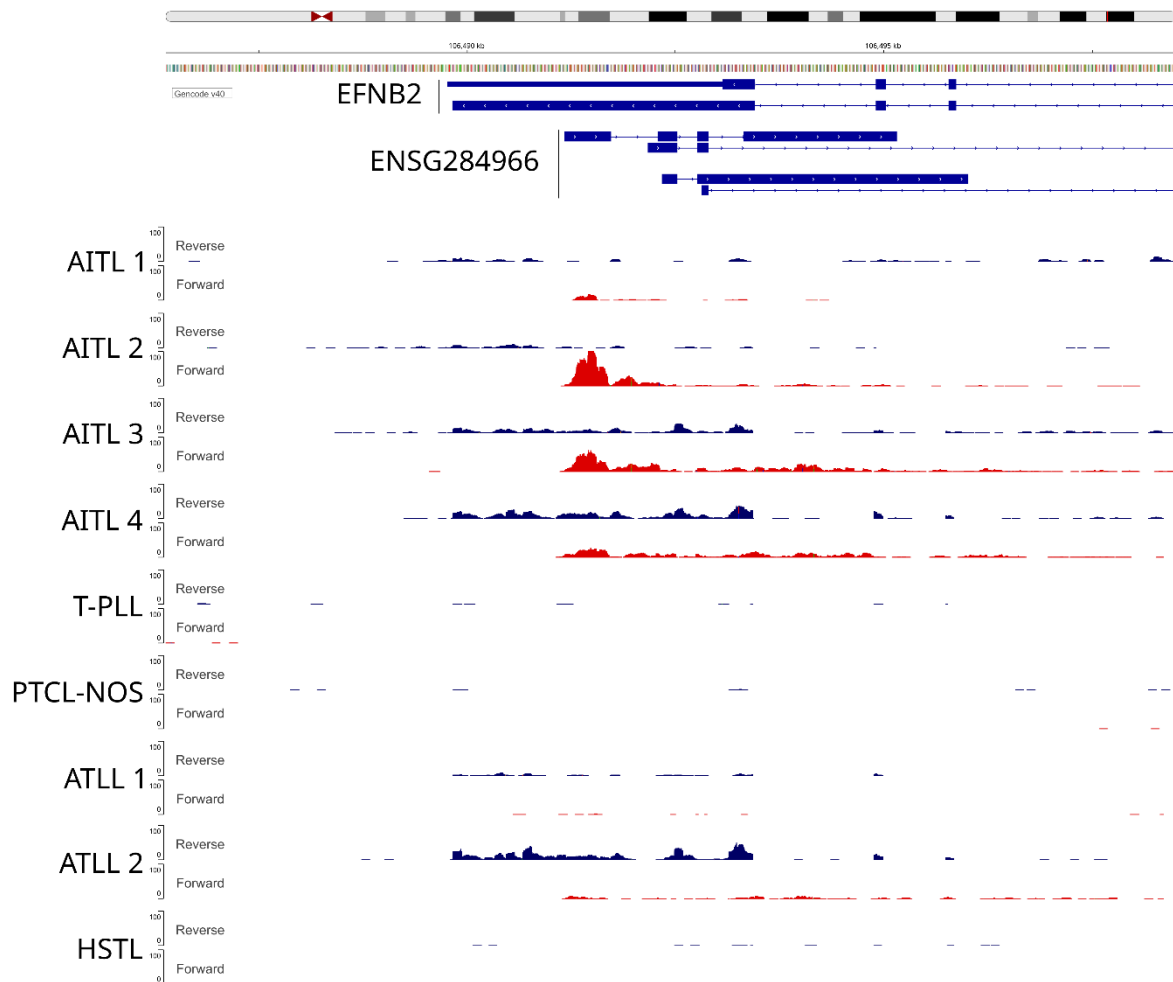


**C**



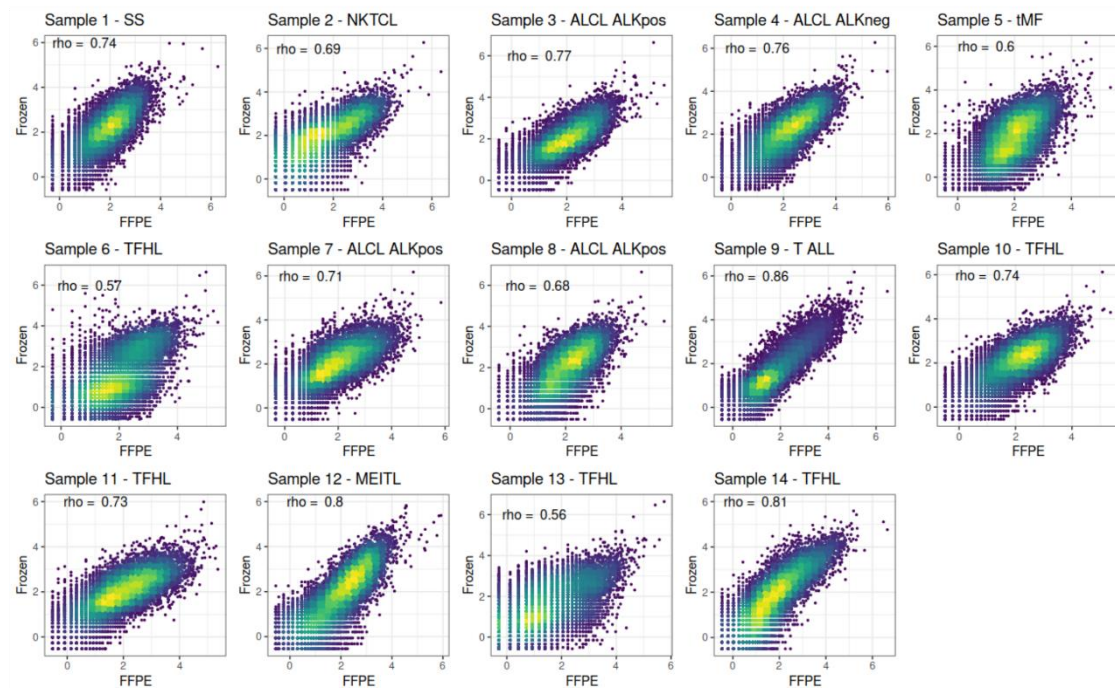
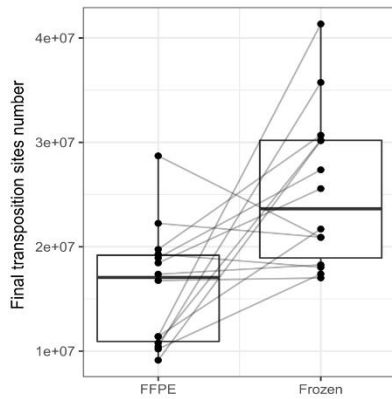
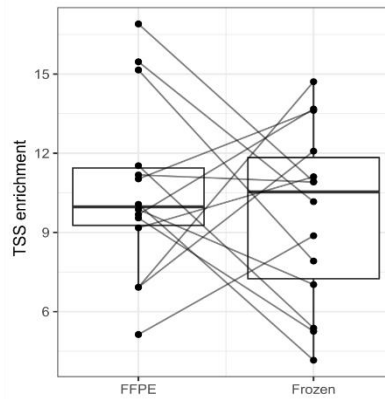
**Supplementary Figure 1**

- Heatmap of NMF pattern scores shows microenvironment and quality pattern alters bias MTCN clustering
- Pattern 13 is highly correlated with the noise to signal ratio (inverse of the TSS enrichment score)
- Strong correlation of pattern #4 with normal macrophage epigenetic profile and of pattern #1 with memory B cells and plasma cells epigenetic profiles derived from scATAC-seq pseudobulk.



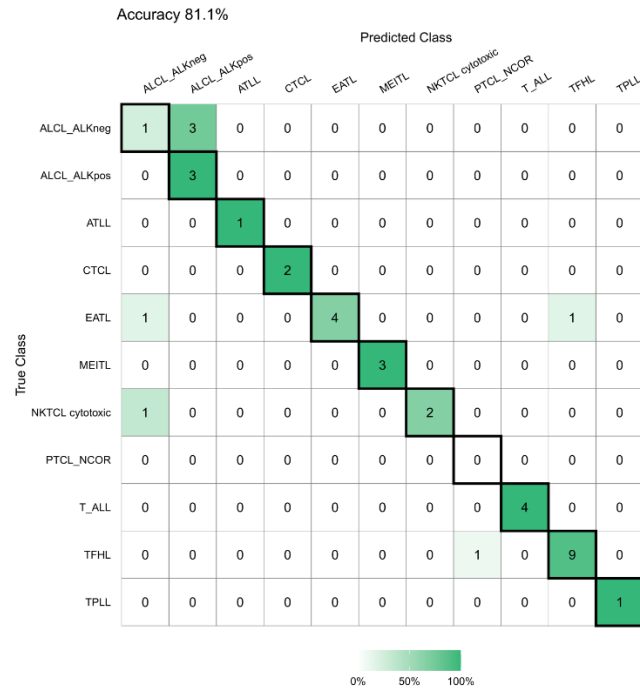
**Supplementary Figure 2**

RNA sequencing forward and reverse strand coverage at ENSG00000284966 locus in sorted lymphoma samples

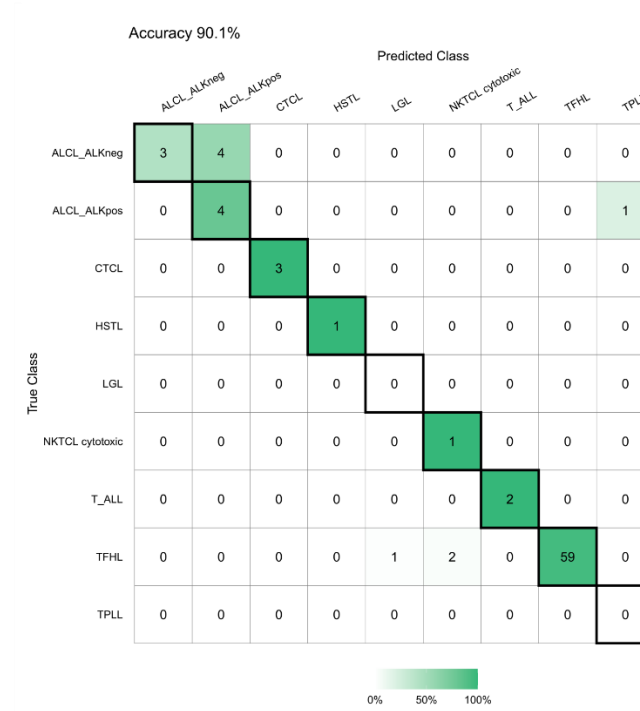
**A****B****C****Supplementary Figure 3**

- A. Assessment of FFPE-ATAC and OMNI-ATAC correlation on paired samples from the same biopsy
- B. Final library sizes from FFPE-ATAC are lower than OMNI-ATAC
- C. TSS enrichment is comparable between OMNI-ATAC and FFPE-ATAC samples.

**A**



**B**



**Supplementary Figure 4**

- Confusion matrix of FFPE samples from the validation cohort comparing predicted diagnosis and pre-review pathology diagnosis
- Confusion matrix of frozen samples from the validation cohort comparing predicted diagnosis and pre-review pathology diagnosis



**Supplementary table 1 : T-cell immunophenotypes used to sort normal NK and T-cell subpopulations**

<b>Cell type</b>	<b>Markers</b>
CD4 naïve CD4 T cells	CD4+ CD45RA+ CCR7+ CD62L+ CD95-
CD4 central memory T cells	CD4+ CD45RA- CCR7+ CD62L+ CD95+
CD4 effector memory T cells	CD4+ CD45RA- CCR7- CD62L- CD95+
CD4 effector memory T cells expressing CD45RA	CD4+ CD45RA+CCR7-
CD4 regulatory T cells	CD4+ CD25+ CD127low(FoxP3high*)
CD4 follicular helper T cells	CD3+ CD4+ CXCR5+ PD1+ ICOShigh (BCL6+*)
CD8 naïve T cells	CD8+ CD45RA+ CCR7+ CD62L+ CD95-
CD8 central memory T cells	CD8+ CD45RA- CCR7+ CD62L+ CD95+
CD8 effector memory T cells	CD8+ CD45RA- CCR7- CD62L- CD95+
CD8 effector memory T cells expressing CD45RA	CD8+ CD45RA+CCR7-
Natural killer cells	CD3- CD16+ CD56+
Gamma-delta T cells	CD3+Vdelta1+
NKT cells	CD3+ CD1d-tetramer+
(*) intracytoplasmic antibodies used for post-sorting purity analysis only	

**Supplementary table 2: Clinical characteristics of TFHL samples according to microenvironment-defined clusters**

Clusters	Cluster 1 and 3	Cluster 2	Cluster 4	p
<b>n</b>	49	50	19	
<b>Age (mean (SD))</b>	64.33 (13.01)	69.06 (9.22)	58.54 (13.59)	0.01
<b>SEX = MALE (%)</b>	32 (65.3)	24 (49.0)	12 (63.2)	0.232
<b>ANN ARBOR STAGE (%)</b>				0.375
<b>2</b>	0 (0.0)	1 (2.6)	1 (7.1)	
<b>3</b>	11 (27.5)	13 (33.3)	2 (14.3)	
<b>4</b>	29 (72.5)	25 (64.1)	11 (78.6)	
<b>ECOG = 2-4 (%)</b>	24 (60.0)	24 (61.5)	1 (10.0)	0.01
<b>LDH = &gt;1N (%)</b>	29 (76.3)	31 (77.5)	5 (45.5)	0.088
<b>PLATELET (MEAN (SD))</b>	225.26 (97.91)	217.91 (138.76)	256.05 (114.90)	0.633
<b>DNMT3A MUTATION = MUTATED (%)</b>	5 (23.8)	3 (20.0)	0 (0.0)	0.482
<b>IDH2 MUTATION = MUTATED (%)</b>	5 (23.8)	2 (11.8)	1 (20.0)	0.635
<b>RHOA MUTATION = MUTATED (%)</b>	15 (65.2)	9 (47.4)	2 (33.3)	0.282
<b>TET2 MUTATION = MUTATED (%)</b>	17 (77.3)	7 (43.8)	1 (20.0)	0.022
<b>GAMMAGLOBULIN LEVEL (MEAN (SD))</b>	12.44 (7.62)	20.87 (11.98)	13.82 (3.61)	0.004
<b>EBV SERUM COPY NUMBER (MEAN (SD))</b>	669.00 (1914.06)	11241.50 (16065.73)	2292.50 (2009.33)	0.091
<b>EBV SERUM POSITIVITY (%)</b>				0.109
<b>NE</b>	20 (57.1)	18 (56.2)	3 (33.3)	
<b>Negative</b>	8 (22.9)	2 (6.2)	1 (11.1)	
<b>Positive</b>	7 (20.0)	12 (37.5)	5 (55.6)	

### Supplementary table 3: Flow cytometry markers used to sort T cell malignancies

PTCL-NOS	ALCL	AITL	CTCL	T-PLL
CD4 APC	CD30 APC	CD4 APC	CD26 APC	CD4 APC
CD3 APC-R700	CD3 APC-R700	CD3 APC-R700	CD3 APC-R700	CD3 APC-R700
CD8 APC-H7	CD19 APC-H7	CD8 APC-H7	CD19 APC-H7	CD8 APC-H7
CD5 PE	CD2 FITC	CD5 PE	KIR3DL2 PE	Viability efluor 506
CD7 PE-Cy7	CD5 PE	CD10 PE-CF594	CD7 PE-Cy7	
DAPI	CD8 PE-CF594	PD1 PE-Cy7	Viability efluor 506	
	CD7 PE-Cy7	ICOS BV650	CD4 BV711	
	Viability efluor 506	Viability efluor 506		
	CD4 BV711			

EATL	LGL	HSTL	NKTCL
CD3 BV786	CD4 BV711	CD4 BV711	CD56 APC
CD4 BV711	CD8 PE CF594	CD8 PE CF594	CD7 APC-R700
CD8 PE CF594	CD3 BV786	CD3 BV786	CD45 APC-H7
CD56 APC	CD57 PE-Cy7	CD57 PE-Cy7	CD2 FITC
CD103 FITC	CD5 PE	CD5 PE	CD19 PerCP Cy 5.5
CD5 PE	CD7 APC-R700	CD7 APC-R700	CD4 BV711
CD7 PE Cy7	CD16 BB515	CD16 BB515	CD3 BV786
Viability efluor 506	CD56 APC	CD56 APC	CD8 PE-CF594
	Viability efluor 506	Viability efluor 506	Viability efluor 506

**Supplementary Table 4: Pseudobulk cell type annotations**

<b>Cell cluster</b>	<b>Cell category</b>
Endothelial Cell General 1	Endothelial Cell
Endothelial Cell General 2	Endothelial Cell
Endothelial Cell General 3	Endothelial Cell
Lymphatic Endothelial Cell	Endothelial Cell
Airway Goblet Cell	Epithelial Cell
Alveolar Type 1 AT1 Cell	Epithelial Cell
Alveolar Type 2 AT2 Cell	Epithelial Cell
Alveolar Type 2 Immune	Epithelial Cell
Basal Epidermal Skin	Epithelial Cell
Chief Cell	Epithelial Cell
Club Cell	Epithelial Cell
Colon Epithelial Cell 1	Epithelial Cell
Colon Epithelial Cell 2	Epithelial Cell
Colon Epithelial Cell 3	Epithelial Cell
Colonic Goblet Cell	Epithelial Cell
Esophageal Epithelial Cell	Epithelial Cell
Foveolar Cell	Epithelial Cell
Granular Epidermal Skin	Epithelial Cell
Hepatocyte	Epithelial Cell
Keratinocyte 1	Epithelial Cell
Keratinocyte 2	Epithelial Cell
Parietal Cell	Epithelial Cell
Small Intestinal Enterocyte	Epithelial Cell
Small Intestinal Goblet Cell	Epithelial Cell
Tuft Cell	Epithelial Cell
Fibroblast Epithelial	Fibroblast
Fibroblast Gastrointestinal	Fibroblast
Fibroblast General	Fibroblast
Fibroblast Liver Adrenal	Fibroblast
Fibroblast Peripheral Nerve	Fibroblast
Fibroblast Sk Muscle Associated	Fibroblast
Macrophage General	Macrophage
Macrophage General Alveolar	Macrophage
Memory B Cell	Memory B Cell
Plasma Cell	Plasma Cell

**Supplementary Table 5: Quality thresholds for sample inclusion**

<b>Cohort</b>	<b>Analysis step</b>	<b>Filtered reads</b>	<b>Signal enrichment in TSS</b>
<b>FAST-ATAC</b>	Map building	5 million	5
<b>FAST-ATAC</b>	Counting	3 million	3
<b>OMNI-ATAC</b>	Counting	3 million	4
<b>FFPE-ATAC</b>	Counting	3 million	4

# DISCUSSION

---

Les hémopathies matures T sont un groupe de maladies rares, extrêmement hétérogènes et difficiles à diagnostiquer. Les distinctions entre ces entités sont basées sur l'intégration des caractéristiques cliniques, morphologiques, de combinaisons complexes de marqueurs immunohistochimiques et d'anomalies génétiques. L'arrivée des techniques de séquençage à haut débit en omique a permis l'avènement de la classification non supervisée de différents sous-types de cancers notamment grâce à des méthodes épigénomiques. L'ATAC-seq est une méthode rapide et peu coûteuse de caractérisation de l'accessibilité chromatinienne à l'échelle du génome. Dans cette étude, nous avons pu décrire le paysage d'accessibilité chromatinienne des différentes hémopathies matures T les plus fréquentes. Nous avons démontré que la plupart des entités de la classification OMS présentent une signature d'accessibilité chromatinienne intrinsèquement spécifique exploitable pour une aide à la classification diagnostique en routine.

## 1 Eclairage apporté par les points de divergence entre classification OMS et clusters épigénétiques

Nous démontrons qu'il existe une grande concordance entre les clusters épigénétiques obtenus en ATAC-seq de manière non supervisée et les diagnostics anatomopathologiques d'experts (Figure 1D). De nombreuses discordances ont pu être corrigées par une relecture anatomopathologique. Il est néanmoins important de chercher à expliquer les discordances persistantes que nous retrouvons.

### 1.1 MEITL

Les MEITL ne représentent pas une classe à part dans notre clustering car la factorisation de matrice non négative n'arrive pas à leur attribuer une composante spécifique. On observe en effet que les MEITL ont un score médian à la fois sur la composante des EATL mais aussi sur la composante des NKTCL. Ils présentent donc un profil d'accessibilité chromatinienne qui est intermédiaire entre ces deux entités. Ceci n'est pas surprenant et rappelle les caractéristiques immunohistochimiques connues dans les MEITL : positivité du CD56 comme certains NKTCL et, positivité du CD103 démontrant leur provenance de l'épithélium intestinal comme les EATL. Lors de la classification supervisée, les cas de MEITL sont néanmoins prédits de manière très fiable grâce à cette signature mixte (Figure 5). Il est fort possible que le faible

nombre d'échantillons inclus dans la cohorte de training initiale (n=4) limite la capacité de la NMF à identifier une signature particulière aux MEITL, mais qu'elle apparaîtrait de manière plus évidente dans des études avec un nombre d'échantillons plus grand.

## 1.2 ALCL

Le clustering semble aussi mettre en évidence des groupes d'ALCL qui ne correspondent pas parfaitement à la distinction habituelle entre lymphomes anaplasiques ALK+ et ALK-. En effet on retrouve un cluster comprenant des ALCL-ALK- et ALK+, alors que le second cluster ne contient que des ALCL ALK-. La signature qui définit le premier groupe comprend notamment des régions chromatinienne localisées au niveau des gènes des récepteurs aux interleukines IL1R1, IL1RAP, IL1RL, IL6R, et les facteurs de transcription RORC et STAT4. Le deuxième groupe qui est, lui, caractérisé par une plus forte accessibilité de SMARCA2, HLA-DRB5 et CD86, contient tous les cas de réarrangement DUSP22, et aucun ALCL cytotoxique. En revanche, le premier cluster ne contient que des échantillons cytotoxiques.

L'absence de signature commune aux lymphomes T anaplasiques ALK- a déjà été décrite dans les études transcriptomiques antérieures (Iqbal et al. 2014; Drieux et al. 2020; Luchtel et al. 2018). Dans l'analyse sur panel de RT-MLPA, on retrouve en effet deux groupes de lymphomes anaplasiques ALK- distincts en fonction de leur profil cytotoxique. Les lymphomes ALCL ALK+ sont connus pour partager cette caractéristique cytotoxique. Dans cette étude, seul le niveau d'expression d'ALK permet de distinguer les ALCL cytoxiqes des ALCL ALK+. Dans une large étude transcriptomique sur les lymphomes anaplasiques, cette même séparation est retrouvée entre deux types d'ALCL, l'un incluant les ALCL DUSP22 et l'autre incluant les ALCL ALK- cytotoxique et les ALCL-ALK+ qui partagent un enrichissement de l'expression des gènes de la voie IL6-JAK-STAT3 (Luchtel et al. 2018). D'autres études ont par ailleurs montré que certains ALCL ALK- cytotoxiques présentaient des mutations activatrices de STAT3 ou des fusions de kinases autres que ALK dont TYK2, ROS1 et FRK ayant la capacité d'induire la voie STAT3 (Crescenzo et al. 2015; Velusamy et al. 2014; Hu et al. 2018). Ces lymphomes présenteraient des profils transcriptionnels similaires aux ALCL ALK+. Il serait donc intéressant de rechercher ces anomalies génétiques et de confronter les résultats aux clusters que nous obtenons. Au total, notre étude renforce l'hypothèse d'un groupe incluant les ALCL DUSP22 peu cytotoxiques et distinct des autres ALCL ALK-.



## 2 Caractérisation épigénétique des PTCL-NOS en ATAC-seq

Nos résultats incluant 38 échantillons de PTCL-NOS ne montrent pas un cluster spécifique des PTCL-NOS, mais une redistribution de la majorité des cas des PTCL-NOS au sein d'autres entités. La même diversité des PTCL-NOS est d'ailleurs retrouvée de manière non supervisée dans les clusterings sur données transcriptomiques par microarray (Iqbal et al. 2010). La classification actuelle de l'OMS propose une distinction des PTCL-NOS en fonction de l'expression des facteurs de transcription TBX21 et GATA3. Cette classification est la conséquence de plusieurs publications démontrant des différences de profils transcriptomiques et moléculaires entre ces deux groupes (Heavican et al. 2019; Iqbal et al. 2014; Amador et al. 2019). Nous observons certes une différence d'activité des facteurs de transcription TBX21 et GATA3, mais celle-ci se distribue au sein des différents clusters épigénétiques et ne permet pas d'affirmer l'existence de 2 groupes uniques de PTCL-NOS GATA3 ou TBX21.

En effet les PTCL-NOS ayant une forte activité de TBX21 sont reclassifiés dans différents clusters avec les HSTL, les LGL ou les NKTCL. De manière intéressante, le cluster des NKTCL regroupe aussi des lymphomes cytotoxiques EBV-. Par ailleurs, les lymphomes nodaux EBV+ de notre cohorte de validation présentent aussi un profil similaire aux NKTCL d'un point de vue épigénétique. Alors même que les études moléculaires tendent à distinguer ces entités en raison de différences d'anomalies génétiques (Climent et al. 2023; Nicolae et al. 2022), il semble donc exister une signature épigénétique commune entre certains PTCL cytotoxiques, les NKTCL et les PTCL-EBV+ nodaux.

De même, les cas de PTCL-NOS présentant une forte accessibilité au niveau des motifs de GATA3 peuvent être épigénétiquement proches des T-PLL, des CTCL ou appartenir à un sous-groupe distinct.

### 2.1 *Les PTCL NCOR, une nouvelle entité ?*

Nous avons en effet démontré que seule une petite proportion de PTCL-NOS se regroupaient avec un ALCL et un AITL au sein d'un cluster de phénotype CD4<sup>+</sup>CD7<sup>-</sup>GATA3<sup>+</sup> qui présentent un grand nombre de réarrangements chromosomiques. Les analyses d'exomes de ce sous-

groupe ont mis en évidence une forte proportion de mutations de TP53 ce qui a déjà été rapporté dans la littérature pour les PTCL-GATA3<sup>+</sup> (Heavican et al. 2019). Mais de manière plus surprenante, nous retrouvons un nombre important de mutations du nuclear receptor co-repressor 1 (*NCOR1*).

*NCOR1* fait partie d'un complexe corépresseur qui comprend l'histone désacétylase 3 (*HDAC3*), les transducin  $\beta$ -like 1/transducin  $\beta$ -like related protein 1 (*TBL1* et *TBL1XR1*) et la G-protein pathway suppressor 2 (*GPS2*). Ce complexe peut aussi contenir la protéine *NCOR2* analogue de *NCOR1* (alias *SMRT*) (Emmett and Lazar 2019). Le complexe *NCOR* est impliqué dans la répression des gènes cibles des récepteurs nucléaires (récepteurs à l'acide rétinoïque, aux hormones thyroïdiennes, à la progestérone, etc ...). Mais *NCOR1* peut aussi interagir avec d'autres histones désacétylases comme *HDAC4*, *HDAC5* et *HDAC7* et de nombreux facteurs de transcription dont *NF-KB* et *BCL6* (Ahad et al. 2019; Huynh and Bardwell 1998). Ce complexe a aussi été impliqué dans la régulation de la stabilité génomique ce qui fait écho aux nombreuses altérations chromosomiques retrouvées dans notre étude (Bhaskara et al. 2010).

Des mutations de *NCOR1* ont été peu décrites dans les lymphomes T. *NCOR1* est très fréquemment délété avec *p53* (80%) dans les syndromes de Sézary mais moins fréquemment muté (3%) (Iżykowska et al. 2017; Kiel et al. 2015; Park et al. 2017b, 220). On ne retrouve pas de mutation de *NCOR1* dans les études génomiques des PTCL-NOS GATA3 mais plusieurs mutations invalidantes de *NCOR2* (Heavican et al. 2019). Des mutations missenses de *NCOR2* ont aussi été décrites dans des NKTCL (Xiong et al. 2020). En revanche, on retrouve de nombreuses mutations dans les leucémies aiguës lymphoblastiques B et des cancers solides (mélanome, carcinome de vessie, cancer du sein) (Gallardo et al. 2015; Gui et al. 2011; R. D. Lee et al. 2022; Bertucci et al. 2019). De plus, il a été démontré que la déplétion de *NCOR1* dans un modèle de tumeur hépatique favorise la croissance tumorale et les métastases (Martínez-Iglesias et al. 2016). De même le knock-down de *NCOR1* promeut le développement de leucémies aiguës lymphoblastiques B dans un modèle murin d'activation constitutionnelle de *STAT5B* (R. D. Lee et al. 2022) et de leucémies myéloïdes dans le modèle *MLL-AF9* (Wan et al. 2019). Par comparaison, une seule étude évoque une action plutôt protumorale du gène *NCOR1* dans un modèle de rhabdomyosarcome, où le complexe *HDAC3/NCOR1/2* semble promouvoir un état dédifférencié des myocytes (Phelps et al.

2016). De manière intéressante, la perte de NCOR1 dans des lymphocytes T régulateurs induit une augmentation de l'expression de MYC (Stolz et al. 2022). Dans notre étude, la prévalence de délétions du gène et la forte proportion de mutations tronquantes évoquent plutôt un caractère suppresseur de tumeur du gène NCOR1 dans les lymphomes T. Des études fonctionnelles seront nécessaires pour confirmer cette hypothèse et en étudier les mécanismes.

Par ailleurs, il est intéressant de voir que ces lymphomes présentent une proximité épigénétique avec les ATLL avec qui ils forment un seul cluster lorsque la factorisation de matrice non négative est réalisée avec un nombre faible de composantes (données non présentées). Ils ont notamment en commun une ouverture importante de certaines régions du locus d'EZH2. L'activité des inhibiteurs d'EZH1/2 a récemment été démontrée dans le traitement des ATLL (Izutsu et al. 2022). Il serait intéressant d'évaluer si cette proximité épigénétique des PTCL-NCOR offre la possibilité d'un ciblage thérapeutique similaire.

### 3 Découverte d'un nouveau marqueur des lymphomes TFH ?

L'ATAC-seq permet la découverte de régions chromatiniennees ouvertes spécifiquement dans chaque entité et retrouve des pics au sein de certains marqueurs diagnostics habituels comme CXCR5, PDCD1, GZMB en cohérence avec les profils d'expression connus des hémopathies T. De nombreux autres pics spécifiques de chaque entité ont été retrouvés dans notre étude et notamment un potentiel marqueur des lymphomes TFH. En effet, nous avons trouvé un pic d'accessibilité dans la région 3' du gène EFNB2 qui est extrêmement spécifique des lymphomes d'origine TFH et lymphocytes TFH normaux. Cette région est caractérisée dans les bases d'annotation ENCODE (Encyclopedia of DNA Elements) comme un enhancer car situé à distance d'un promoteur et portant les marques H3K27Ac et H3K4me3. L'accessibilité des régions chromatiniennees distales a été décrite comme étant plus spécifique des types cellulaires que celle des régions promotrices proximales et que la quantification des transcrits détectés en 3'RNA-seq (Corces et al. 2018; Satpathy et al. 2018; Corces et al. 2016). Nous pensons ainsi confirmer ici la possibilité donnée par l'ATAC-seq de découvrir de nouvelles régions distales dont l'ouverture est très spécifique d'un type cellulaire. Cependant, d'après les bases de données de transcrits récentes, ce pic se situe au

niveau du promoteur d'un transcrit antisens à EFNB2 qui correspond à un ARN non codant. En quantifiant cet ARN non codant grâce à des méthodes de RNA-seq qui prennent en compte l'orientation de la transcription ('stranded'), nous avons pu démontrer qu'il est nettement plus spécifique des lymphomes TFH que l'expression du gène EFNB2 rapporté par une étude antérieure (Iqbal et al. 2014). La fonction de ce transcrit n'est pas décrite et comme la plupart des ARN non codant, il pourrait être considéré comme un sous-produit de l'activité de transcription sans fonction spécifique. Cependant certains ARN non codants jouent un rôle majeur dans la régulation transcriptionnelle et il serait intéressant d'étudier son rôle dans les lymphocytes TFH (Statello et al. 2021). De manière intéressante, nous avons pu démontrer à l'aide de la technologie Basescope une expression très localisée du transcrit dans certaines cellules d'un AITL. La localisation nucléaire punctiforme évoque une transcription locale de l'ARN non codant comme il peut être observé pour cette catégorie de transcrit (Statello et al. 2021). Cette caractéristique offre la possibilité de combiner cette technique avec d'autres marquages protéiques ou ARN qui pourraient être utiles lors du diagnostic de lymphome TFH. La valeur diagnostique de cette technique reste cependant à valider sur une large cohorte de lymphomes T et de différentes pathologies réactionnelles.

## 4 ATAC-seq, microenvironnement et pronostic des lymphomes TFH

Nous avons aussi démontré que, comme pour la transcriptomique, une des premières sources de variabilité entre échantillons est la différence de microenvironnement et que celle-ci altère le clustering non supervisé des MTCN. Cette différence a pu être un facteur confondant lors d'études antérieures réalisant des clusterings non supervisés basés sur la transcriptomique (Maura et al. 2019; Iqbal et al. 2010; Piccaluga et al. 2013). Notre approche de déconvolution des profils épigénétiques par NMF favorisée par l'ajout d'échantillons triés de lymphomes a permis de séparer le signal tumoral des autres composantes cellulaires du microenvironnement. Elle donne ainsi la possibilité de regrouper les échantillons d'une même entité même s'ils présentent des niveaux d'infiltrations parfois très variables. Des AITL non triés sont ainsi rapprochés d'échantillons de cellules de PTCL-TFH triées. Il serait intéressant de mesurer la limite de détection de notre méthode en réalisant des pseudo-

échantillons plus ou moins infiltrés à partir de profils single-cell pour estimer l'infiltration minimale nécessaire à la détection de la population tumorale en ATAC-seq.

Les composantes du microenvironnement obtenues par NMF ne sont pas utilisées dans notre étude pour prédire le diagnostic mais elles peuvent présenter un intérêt pronostic. En effet, nous avons montré l'existence de 4 clusters au sein des lymphomes TFH qui varient selon la composition de leur microenvironnement. Nous démontrons l'impact pronostic favorable d'une forte infiltration B, déjà décrite dans des études supervisées qui retrouvent un effet du niveau d'expression des gènes de lymphocytes B (CD19, BLK, MS4A1) sur la survie (Iqbal et al. 2014; Rodríguez et al. 2021). Nous retrouvons aussi un deuxième cluster présentant un score élevé dans la composante plasmocytaire et une survie globale supérieure à celle des patients présentant une forte infiltration T malgré un âge moyen plus élevé. Une des limites de notre étude est l'utilisation d'une déconvolution semi-supervisée qui permet uniquement de donner un score sur chaque composante du microenvironnement et non de quantifier exactement les proportions cellulaires présentes. Une approche de déconvolution supervisée pourrait donner des résultats complémentaires intéressants. L'ATAC-seq permet donc d'obtenir en plus du diagnostic, une stratification pronostique des patients atteints de lymphomes TFH basée sur la composition de leur microenvironnement tumoral.

## 5 Identification de la cellule d'origine des lymphomes T par ATAC-seq

L'identification de la cellule d'origine des différentes hémopathies a permis une amélioration de la stratification thérapeutique et pronostique des patients (Chiorazzi and Ferrarini 2011; Scott et al. 2014; Hamblin et al. 1999; Seifert et al. 2012). Comme pour les pathologies B, cette tâche est rendue difficile dans les lymphomes T/NK par la complexité du développement normal des lymphocytes T et NK. Nous démontrons dans cette étude que le profil d'accessibilité chromatinienne des différentes entités de lymphomes T se rapprochent de celui de leur contrepartie normale. Ainsi les plus proches voisins normaux des lymphomes TFH sont les lymphocytes TFH conformément à ce qui a été publié (de Leval et al. 2007; Rodríguez-Pinilla et al. 2008).

De même les plus proches voisins des lymphomes cutanés sont les lymphocytes T centraux mémoires comme identifié précédemment (Campbell et al. 2010). Il a en effet été démontré que les syndromes de Sézary, entité majoritaire dans notre étude, présentent un phénotype de lymphocytes T centraux mémoires, mais aussi que les mycosis fongoïdes dériveraient de lymphocytes T résidents mémoires (Campbell et al. 2010). Cependant, cette dichotomie a été remise en question par une étude de single-cell RNA-seq démontrant que le profil transcriptomique des cellules de Sézary localisées au niveau de la peau était similaire à celui des cellules de mycosis fongoïdes et différent de celui des cellules de Sézary circulantes (Herrera et al. 2021). Notre étude n'ayant pu inclure d'échantillons de mycosis fongoïdes issus de la peau, ni de lymphocytes T résidents mémoires en raison de la rareté de cette population, nous n'avons pu tester cette hypothèse qu'il serait intéressant d'explorer en épigénomique.

Une découverte surprenante de notre étude est la proximité des leucémies prolymphocytaires T (T-PLL) avec les lymphocytes T naïfs. En effet, cette entité est connue pour présenter des marqueurs immunophénotypiques de lymphocytes T centraux mémoires (Oberbeck et al. 2020). De plus *in vitro*, les T-PLL ont une réponse cytokinique à la stimulation CD3/CD28 qui est similaire à celle des lymphocytes T mémoires. Cependant nos données semblent indiquer que les T-PLL sont issus d'une population de cellules naïves qui n'a pas encore subi le remaniement épigénétique majeur qui accompagne la transition naïve-mémoire (Moskowitz et al. 2017). Nous supposons que la surexpression de TCL1A qui favorise une signalisation TCR accrue pourrait être à l'origine du phénotype mémoire alors même que les cellules de T-PLL n'auraient pas encore rencontré d'antigène (Oberbeck et al. 2020).

De manière intéressante, les lymphomes hépatospléniques semblent avoir un profil épigénétique plus proche des lymphocytes NK que des lymphocytes  $\gamma\delta$  ce qui apparaît en contradiction avec leur phénotype majoritaire  $\gamma\delta$  (V $\delta$ 1) ou  $\alpha\beta$ . Cependant une étude antérieure a déjà démontré qu'il existe une surexpression de gènes NK (Killer-cell immunoglobulin-like receptors (KIRs), SYK) au sein des lymphomes hépatospléniques par rapport aux lymphocytes T  $\gamma\delta$  sanguins (Travert et al. 2012). Il est possible que le profil épigénétique des lymphocytes T V $\delta$ 1 spléniques supposés à l'origine des HSTL soit différent de celui des lymphocytes T V $\delta$ 1 du sang ou alors que les lymphocytes T V $\delta$ 1 du sang

présentent une hétérogénéité, auquel cas nous aurions pu ne pas avoir obtenu le profil épigénétique de la cellule d'origine correcte. Ce sont des limites de notre étude pour laquelle nous n'avons pas pu inclure des lymphocytes T issues de tissus périphériques et de notre méthode qui nécessite l'utilisation de cellules en bulk. Des approches single-cell permettraient peut-être de résoudre cette question.

Une autre explication à cette apparente divergence pourrait être le mécanisme de transdifférenciation décrit au cours de la lymphomagenèse T qui favorise l'émergence d'un phénotype innate-like dans de nombreuses entités (Carras et al. 2021). Notre approche de clustering permettant uniquement de mesurer la proximité entre deux profils épigénétiques, elle pourrait être induite en erreur par une reprogrammation cellulaire modifiant de façon notable le profil épigénétique de la cellule tumorale.

Dans ces cas de transdifférenciation cellulaire, le concept de cellule d'origine est difficile à définir. De plus, il est clairement démontré dans plusieurs types de lymphomes que les anomalies moléculaires sont acquises à divers stades de la différenciation cellulaire T. Ainsi les anomalies TET2, DNMT3A sont souvent acquises au niveau d'un précurseur hématopoïétique avant l'engagement dans la lignée T (Quivoron et al. 2011; Lewis et al. 2020; Sakata-Yanagimoto et al. 2014). De même l'hypothèse a été émise que les lymphomes anaplasiques ALK+ acquièrent la translocation NPM-ALK avant la sélection thymique (Malcolm et al. 2016).

Cependant la recherche d'une contrepartie cellulaire normale présente un intérêt important. En effet, dans notre étude comme dans des études précédentes sur la LLC ou les LAM (Corces et al. 2016; Rendeiro et al. 2016), la première source de régions différentiellement ouvertes entre deux néoplasies T ou entre néoplasies T et contreparties normales est liée à des régions impliquées dans la différenciation T et NK. Il est donc nécessaire lors des analyses omiques de comparer les lymphomes T/NK à des cellules qui se situent à un stade de différenciation proche pour permettre de trouver les gènes impliqués dans la lymphomagenèse et non dans le processus de différenciation. Une approche intéressante développée par R. Corces consiste à créer des contreparties normales artificielles en combinant les différents profils des cellules normales proches, ce qui permettrait une découverte plus efficace des gènes impliqués dans la transformation (Corces et al. 2016).

## 6 Vers une utilisation de l'ATAC-seq en routine ?

### 6.1 *Faisabilité*

En routine diagnostique, une proportion importante des biopsies réalisées pour recherche de lymphome sont uniquement disponibles en FFPE. Il était donc nécessaire pour une utilisation généralisée d'adapter l'ATAC-seq à ces échantillons. Nous avons modifié le protocole d'ATAC-seq FFPE publié précédemment (Yadav et al. 2022) et obtenu des profils d'accessibilité chromatinienne de qualité similaire aux échantillons congelés, d'après leur score d'enrichissement au TSS et malgré une taille courte des fragments. Le taux d'échec de la technique (14%) est principalement dû à un trop faible nombre de reads uniques qui peut être attribué à deux causes : au crosslinking de l'ADN et aux contaminations bactériennes. En effet, le crosslinking entre ADN et FFPE induit des cassures de l'ADN qui empêchent l'amplification des fragments d'ADN transposés. Il est connu que le crosslinking se renforce avec l'ancienneté des échantillons. Nous avons utilisé des échantillons datant des années 2011 à 2023 et les échantillons les plus récents présentaient en moyenne un nombre de reads plus élevé. Il est probable que si la technique devait être utilisée en prospectif, le taux d'échec serait donc moindre. L'autre facteur jouant sur la qualité de l'ATAC FFPE est la présence d'ADN bactérien lié à la contamination des blocs. Nous avons pu les éliminer au fur et à mesure des différentes expériences grâce à l'ajout d'antibiotiques au tampon de lyse (amoxicilline, polymyxine) et grâce à l'ajout d'une étape de digestion de l'ADN extranucléaire à la DNase. Le problème des contaminations bactériennes devrait ainsi être minime pour les études prospectives.

En ce qui concerne la rapidité de la technique, notre protocole inclut deux incubations sur la nuit, mais le temps technique nécessaire est très court. De plus les étapes de déparaffinage et de construction des librairies sont facilement automatisables dans un laboratoire d'anatomopathologie. Enfin la question du flux de patients est un facteur souvent limitant pour programmer les runs de séquençage à haut débit en routine diagnostique. L'ATAC-seq n'étant pas une technique ciblée, elle pourrait être appliquée à d'autres hémopathies qui pourront être séquencées ensemble, pour permettre un rendu rapide des résultats.



## 6.2 Validité

Nous avons démontré que notre méthode permet d'obtenir une « accuracy » de 87%. La prédiction est accompagnée d'une probabilité d'appartenance à une classe qui permet d'estimer la confiance avec laquelle le diagnostic est posé et la nécessité pour les cliniciens et anatomopathologistes d'ajouter des techniques complémentaires diagnostiques. Les erreurs de prédiction constatées sont en grande partie liées à la problématique des ALCL ALK- décrite ci-dessus et il sera nécessaire d'augmenter le nombre de cas d'ALCL ALK- cytotoxiques de notre cohorte d'entraînement pour mieux caractériser ce sous-groupe et permettre de les classer comme une entité séparée des ALCL ALK+ et du groupe des ALCL ALK- DUSP22-R non cytotoxiques. En dehors de ce cas, le taux d'erreur est minime dans les entités de PTCL, et notre technique pourrait apporter une aide diagnostique utile. La validité de la reclassification des PTCL-NOS est plus difficile à objectiver. Des techniques complémentaires (marquages immunohistochimiques, séquençage d'exomes) ont déjà permis de reclasser certains PTCL-NOS (réarrangement DUSP22 orientant vers un ALCL, marquage positif à la TdT reclassifiant une T-ALL, présence d'EBV orientant vers un PTCL-EBV+ nodal). Mais il n'existe pas toujours de marqueurs spécifiques d'entité permettant de corriger le diagnostic et il est alors seulement possible de vérifier que le profil immunohistochimique est compatible avec l'entité prédite.

Un nombre non négligeable de cas obtiennent un score de prédiction inférieur au seuil minimal de certitude ce qui ne permet pas de conclure à un diagnostic. Ce taux est similaire à celui obtenu dans d'autres études (Capper et al. 2018; Koelsche et al. 2021). Les probabilités données par l'algorithme sont particulièrement basses pour les entités pour lesquelles nous avons peu d'échantillons. L'augmentation du nombre d'échantillons dans les classes rares de lymphomes T grâce à l'ajout de nouveaux cas FFPE permettra d'améliorer notre modèle et de diminuer le nombre de cas sans diagnostic final.

Enfin, notre classifieur a été entraîné uniquement à reconnaître des lymphomes T/NK et requiert que ce diagnostic soit déjà posé par un anatomopathologiste. Il serait donc important d'évaluer les possibilités diagnostiques de l'ATAC-seq dans les hémopathies B et des pathologies ganglionnaires réactionnelles pour que cette technique puisse être utilisée de manière plus étendue pour le diagnostic des hémopathies lymphoïdes.



# CONCLUSION

---

Notre étude a permis de décrire le paysage d'accessibilité chromatinienne de diverses entités de néoplasies matures T. Après déconvolution du microenvironnement, le clustering non supervisé des profils d'accessibilité définit des classes épigénétiques correspondant à la classification actuelle des hémopathies T matures. Les PTCL-NOS font cependant exceptions et peuvent être en majorité rattachés à une entité de MTCN spécifique. Un seul sous-groupe de PTCL-NOS présente une signature épigénétique particulière et pourrait constituer une classe de MTCN à part. On y retrouve un grand nombre d'altérations chromosomiques et d'inactivations des gènes *TP53* et *NCOR1*, un gène suppresseur de tumeur encore peu décrit dans les MTCN. Par ailleurs l'ATAC-seq a permis de décrire le stade de différenciation des différents lymphomes T en les comparant à des cellules non tumorales, et de découvrir un potentiel nouveau marqueur des lymphocytes et lymphomes TFH. Enfin nous avons démontré la possibilité d'utiliser l'ATAC-seq pour prédire le diagnostic des MTCN à partir d'échantillons congelés ou FFPE. Sous réserve de confirmation dans des études prospectives de plus grande ampleur, cette technique pourrait offrir un nouvel outil diagnostique transposable à la routine.

# SOURCES

---

- Abe, Masaki, Kinya Uchihashi, Tsuruda Kazuto, Akemi Osaka, Katsunori Yanagihara, Kunihiro Tsukasaki, Hiroo Hasegawa, Yasuaki Yamada, and Shimeru Kamihira. 2008. "Foxp3 Expression on Normal and Leukemic CD4+CD25+ T Cells Implicated in Human T-Cell Leukemia Virus Type-1 Is Inconsistent with Treg Cells." *European Journal of Haematology* 81 (3): 209–17. <https://doi.org/10.1111/j.1600-0609.2008.01105.x>.
- Adey, Andrew, Hilary G. Morrison, null Asan, Xu Xun, Jacob O. Kitzman, Emily H. Turner, Bethany Stackhouse, et al. 2010. "Rapid, Low-Input, Low-Bias Construction of Shotgun Fragment Libraries by High-Density in Vitro Transposition." *Genome Biology* 11 (12): R119. <https://doi.org/10.1186/gb-2010-11-12-r119>.
- Advani, Ranjana, Tetiana Skrypets, Monica Civallero, Michael A Spinner, Martina Manni, Wonseog Kim, Andrei Shustov, et al. 2021. "Outcomes and Prognostic Factors in Angioimmunoblastic T Cell Lymphoma: Final Report from the International TCell Project." *Blood*, no. blood.2020010387 (March). <https://doi.org/10.1182/blood.2020010387>.
- Aggarwal, Charu C., Alexander Hinneburg, and Daniel A. Keim. 2001. "On the Surprising Behavior of Distance Metrics in High Dimensional Space." In *Database Theory — ICDT 2001*, edited by Jan Van Den Bussche and Victor Vianu, 1973:420–34. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/3-540-44503-X\\_27](https://doi.org/10.1007/3-540-44503-X_27).
- Agrawal, B., M. J. Krantz, J. Parker, and B. M. Longenecker. 1998. "Expression of MUC1 Mucin on Activated Human T Cells: Implications for a Role of MUC1 in Normal Immune Regulation." *Cancer Research* 58 (18): 4079–81.
- Ahad, Abdul, Mathias Stevanin, Shuchi Smita, Gyan Prakash Mishra, Dheerendra Gupta, Sebastian Waszak, Uday Aditya Sarkar, et al. 2019. "NCoR1: Putting the Brakes on the Dendritic Cell Immune Tolerance." *iScience* 19 (September): 996–1011. <https://doi.org/10.1016/j.isci.2019.08.024>.
- Alaggio, Rita, Catalina Amador, Ioannis Anagnostopoulos, Ayoma D. Attygalle, Iguaracyra Barreto De Oliveira Araujo, Emilio Berti, Govind Bhagat, et al. 2022. "The 5th Edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms." *Leukemia* 36 (7): 1720–48. <https://doi.org/10.1038/s41375-022-01620-2>.
- Amador, Catalina, Alyssa Bouska, George Wright, Dennis D. Weisenburger, Andrew L. Feldman, Timothy C. Greiner, Waseem Lone, et al. 2022. "Gene Expression Signatures for the Accurate Diagnosis of Peripheral T-Cell Lymphoma Entities in the Routine Clinical Practice." *Journal of Clinical Oncology*, July. <https://doi.org/10.1200/JCO.21.02707>.
- Amador, Catalina, and Andrew L. Feldman. 2021. "How I Diagnose Anaplastic Large Cell Lymphoma." *American Journal of Clinical Pathology* 155 (4): 479–97. <https://doi.org/10.1093/ajcp/aqab012>.
- Amador, Catalina, Timothy C. Greiner, Tayla B. Heavican, Lynette M. Smith, Karen Tatiana Galvis, Waseem Lone, Alyssa Bouska, et al. 2019. "Reproducing the Molecular

- Subclassification of Peripheral T-Cell Lymphoma-NOS by Immunohistochemistry." *Blood* 134 (24): 2159–70. <https://doi.org/10.1182/blood.2019000779>.
- Ambroise, Christophe, and Geoffrey J. McLachlan. 2002. "Selection Bias in Gene Extraction on the Basis of Microarray Gene-Expression Data." *Proceedings of the National Academy of Sciences of the United States of America* 99 (10): 6562–66. <https://doi.org/10.1073/pnas.102102699>.
- Andersson, Emma I., Takahiro Tanahashi, Nodoka Sekiguchi, Vanessa Rebecca Gasparini, Sabrina Bortoluzzi, Toru Kawakami, Kazuyuki Matsuda, et al. 2016. "High Incidence of Activating STAT5B Mutations in CD4-Positive T-Cell Large Granular Lymphocyte Leukemia." *Blood* 128 (20): 2465–68. <https://doi.org/10.1182/blood-2016-06-724856>.
- Ando, Y., S. Nakano, K. Saito, I. Shimamoto, M. Ichijo, T. Toyama, and Y. Hinuma. 1987. "Transmission of Adult T-Cell Leukemia Retrovirus (HTLV-I) from Mother to Child: Comparison of Bottle- with Breast-Fed Babies." *Japanese Journal of Cancer Research: Gann* 78 (4): 322–24.
- Aran, Dvir, Marina Sirota, and Atul J. Butte. 2015. "Systematic Pan-Cancer Analysis of Tumour Purity." *Nature Communications* 6 (1): 8971. <https://doi.org/10.1038/ncomms9971>.
- Aref-Eshghi, Erfan, Laila C. Schenkel, Peter Ainsworth, Hanxin Lin, David I. Rodenhiser, Jean-Claude Cutz, and Bekim Sadikovic. 2018. "Genomic DNA Methylation-Derived Algorithm Enables Accurate Detection of Malignant Prostate Tissues." *Frontiers in Oncology* 8. <https://www.frontiersin.org/articles/10.3389/fonc.2018.00100>.
- Attikum, Haico van, Olivier Fritsch, Barbara Hohn, and Susan M. Gasser. 2004. "Recruitment of the INO80 Complex by H2A Phosphorylation Links ATP-Dependent Chromatin Remodeling with DNA Double-Strand Break Repair." *Cell* 119 (6): 777–88. <https://doi.org/10.1016/j.cell.2004.11.033>.
- Au, W.-Y., S.-Y. Ma, C.-S. Chim, C. Choy, F. Loong, A. K. W. Lie, C. C. K. Lam, et al. 2005. "Clinicopathologic Features and Treatment Outcome of Mature T-Cell and Natural Killer-Cell Lymphomas Diagnosed According to the World Health Organization Classification Scheme: A Single Center Experience of 10 Years." *Annals of Oncology: Official Journal of the European Society for Medical Oncology* 16 (2): 206–14. <https://doi.org/10.1093/annonc/mdi037>.
- Bachy, Emmanuel, Mirjam Urb, Shilpi Chandra, Rémy Robinot, Gabriel Bricard, Simon de Bernard, Alexandra Traverse-Glehen, et al. 2016. "CD1d-Restricted Peripheral T Cell Lymphoma in Mice and Humans." *The Journal of Experimental Medicine* 213 (5): 841–57. <https://doi.org/10.1084/jem.20150794>.
- Ballard, D. W., E. Böhnlein, J. W. Lowenthal, Y. Wano, B. R. Franza, and W. C. Greene. 1988. "HTLV-I Tax Induces Cellular Proteins That Activate the Kappa B Element in the IL-2 Receptor Alpha Gene." *Science (New York, N.Y.)* 241 (4873): 1652–55. <https://doi.org/10.1126/science.241.4873.1652>.
- Banerji, J., S. Rusconi, and W. Schaffner. 1981. "Expression of a Beta-Globin Gene Is Enhanced by Remote SV40 DNA Sequences." *Cell* 27 (2 Pt 1): 299–308. [https://doi.org/10.1016/0092-8674\(81\)90413-x](https://doi.org/10.1016/0092-8674(81)90413-x).
- Bao, Yunhe, and Xuetong Shen. 2007. "SnapShot: Chromatin Remodeling Complexes." *Cell* 129 (3): 632. <https://doi.org/10.1016/j.cell.2007.04.018>.
- Bartholomew, C., N. Jack, J. Edwards, W. Charles, D. Corbin, F. R. Cleghorn, and W. A. Blattner. 1998. "HTLV-I Serostatus of Mothers of Patients with Adult T-Cell Leukemia and HTLV-I-Associated Myelopathy/Tropical Spastic Paraparesis." *Journal of Human Virology* 1 (4): 302–5.

- Basha, Basma M., Sandra C. Bryant, Karen L. Rech, Andrew L. Feldman, Julie A. Vrana, Min Shi, Katelyn A. Reed, and Rebecca L. King. 2019. "Application of a 5 Marker Panel to the Routine Diagnosis of Peripheral T-Cell Lymphoma With T-Follicular Helper Phenotype." *The American Journal of Surgical Pathology* 43 (9): 1282–90. <https://doi.org/10.1097/PAS.0000000000001315>.
- Bell, A. C., A. G. West, and G. Felsenfeld. 1999. "The Protein CTCF Is Required for the Enhancer Blocking Activity of Vertebrate Insulators." *Cell* 98 (3): 387–96. [https://doi.org/10.1016/s0092-8674\(00\)81967-4](https://doi.org/10.1016/s0092-8674(00)81967-4).
- Bellanger, D., V. Jacquemin, M. Chopin, G. Pierron, O. A. Bernard, J. Ghysdael, and M.-H. Stern. 2014. "Recurrent JAK1 and JAK3 Somatic Mutations in T-Cell Prolymphocytic Leukemia." *Leukemia* 28 (2): 417–19. <https://doi.org/10.1038/leu.2013.271>.
- Bentsen, Mette, Philipp Goymann, Hendrik Schultheis, Kathrin Klee, Anastasiia Petrova, René Wiegandt, Annika Fust, et al. 2020. "ATAC-Seq Footprinting Unravels Kinetics of Transcription Factor Binding during Zygotic Genome Activation." *Nature Communications* 11 (1): 4267. <https://doi.org/10.1038/s41467-020-18035-1>.
- Berest, Ivan, Christian Arnold, Armando Reyes-Palomares, Giovanni Palla, Kasper Dindler Rasmussen, Holly Giles, Peter-Martin Bruch, et al. 2019. "Quantification of Differential Transcription Factor Activity and Multiomics-Based Classification into Activators and Repressors: diffTF." *Cell Reports* 29 (10): 3147–3159.e12. <https://doi.org/10.1016/j.celrep.2019.10.106>.
- Berge, R. L. ten, F. G. Snijdwint, S. von Mensdorff-Pouilly, R. J. Poort-Keesom, J. J. Oudejans, J. W. Meijer, R. Willemze, J. Hilgers, and C. J. Meijer. 2001. "MUC1 (EMA) Is Preferentially Expressed by ALK Positive Anaplastic Large Cell Lymphoma, in the Normally Glycosylated or Only Partly Hypoglycosylated Form." *Journal of Clinical Pathology* 54 (12): 933–39. <https://doi.org/10.1136/jcp.54.12.933>.
- Bertucci, François, Charlotte K. Y. Ng, Anne Patsouris, Nathalie Droin, Salvatore Piscuoglio, Nadine Carbuca, Jean Charles Soria, et al. 2019. "Genomic Characterization of Metastatic Breast Cancers." *Nature* 569 (7757): 560–64. <https://doi.org/10.1038/s41586-019-1056-z>.
- Bhaskara, Srividya, Sarah K. Knutson, Guochun Jiang, Mahesh B. Chandrasekharan, Andrew J. Wilson, Siyuan Zheng, Ashwini Yenamandra, et al. 2010. "Hdac3 Is Essential for the Maintenance of Chromatin Structure and Genome Stability." *Cancer Cell* 18 (5): 436–47. <https://doi.org/10.1016/j.ccr.2010.10.022>.
- Biton, Anne, Isabelle Bernard-Pierrot, Yinjun Lou, Clémentine Krucker, Elodie Chapeaublanc, Carlota Rubio-Pérez, Nuria López-Bigas, et al. 2014. "Independent Component Analysis Uncovers the Landscape of the Bladder Tumor Transcriptome and Reveals Insights into Luminal and Basal Subtypes." *Cell Reports* 9 (4): 1235–45. <https://doi.org/10.1016/j.celrep.2014.10.035>.
- Bonzheim, Irina, Martin Irmeler, Margit Klier-Richter, Julia Steinhilber, Nataša Anastasov, Sabine Schäfer, Patrick Adam, et al. 2013. "Identification of C/EBP $\beta$  Target Genes in ALK+ Anaplastic Large Cell Lymphoma (ALCL) by Gene Expression Profiling and Chromatin Immunoprecipitation." *PloS One* 8 (5): e64544. <https://doi.org/10.1371/journal.pone.0064544>.
- Bossard, Céline, Maria Pamela Dobay, Marie Parrens, Laurence Lamant, Edoardo Missiaglia, Corinne Haioun, Antoine Martin, et al. 2014. "Immunohistochemistry as a Valuable Tool to Assess CD30 Expression in Peripheral T-Cell Lymphomas: High Correlation with

- mRNA Levels." *Blood* 124 (19): 2983–86. <https://doi.org/10.1182/blood-2014-07-584953>.
- Boulesteix, A.-L., C. Strobl, T. Augustin, and M. Daumer. 2008. "Evaluating Microarray-Based Classifiers: An Overview." *Cancer Informatics* 6 (January): CIN.S408. <https://doi.org/10.4137/CIN.S408>.
- Bourgault-Rouxel, A. S., T. P. Loughran, R. Zambello, P. K. Epling-Burnette, G. Semenzato, J. Donadieu, L. Amiot, T. Fest, and T. Lamy. 2008. "Clinical Spectrum of Gammadelta+ T Cell LGL Leukemia: Analysis of 20 Cases." *Leukemia Research* 32 (1): 45–48. <https://doi.org/10.1016/j.leukres.2007.04.011>.
- Boyle, Alan P., Sean Davis, Hennady P. Shulha, Paul Meltzer, Elliott H. Margulies, Zhiping Weng, Terrence S. Furey, and Gregory E. Crawford. 2008. "High-Resolution Mapping and Characterization of Open Chromatin across the Genome." *Cell* 132 (2): 311–22. <https://doi.org/10.1016/j.cell.2007.12.014>.
- Brenchley, Jason M., Nitin J. Karandikar, Michael R. Betts, David R. Ambrozak, Brenna J. Hill, Laura E. Crotty, Joseph P. Casazza, et al. 2003. "Expression of CD57 Defines Replicative Senescence and Antigen-Induced Apoptotic Death of CD8+ T Cells." *Blood* 101 (7): 2711–20. <https://doi.org/10.1182/blood-2002-07-2103>.
- Brivanlou, Ali H., and James E. Darnell. 2002. "Signal Transduction and the Control of Gene Expression." *Science* 295 (5556): 813–18. <https://doi.org/10.1126/science.1066355>.
- Brunet, Jean-Philippe, Pablo Tamayo, Todd R. Golub, and Jill P. Mesirov. 2004. "Metagenes and Molecular Pattern Discovery Using Matrix Factorization." *Proceedings of the National Academy of Sciences of the United States of America* 101 (12): 4164–69. <https://doi.org/10.1073/pnas.0308531101>.
- Buenrostro, Jason D., Paul G. Giresi, Lisa C. Zaba, Howard Y. Chang, and William J. Greenleaf. 2013. "Transposition of Native Chromatin for Fast and Sensitive Epigenomic Profiling of Open Chromatin, DNA-Binding Proteins and Nucleosome Position." *Nature Methods* 10 (12): 1213–18. <https://doi.org/10.1038/nmeth.2688>.
- Buenrostro, Jason D., Beijing Wu, Ulrike M. Litzenburger, Dave Ruff, Michael L. Gonzales, Michael P. Snyder, Howard Y. Chang, and William J. Greenleaf. 2015. "Single-Cell Chromatin Accessibility Reveals Principles of Regulatory Variation." *Nature* 523 (7561): 486–90. <https://doi.org/10.1038/nature14590>.
- Bulger, Michael, and Mark Groudine. 2011. "Functional and Mechanistic Diversity of Distal Transcription Enhancers." *Cell* 144 (3): 327–39. <https://doi.org/10.1016/j.cell.2011.01.024>.
- Calderon, Diego, Michelle L. T. Nguyen, Anja Mezger, Arwa Kathiria, Fabian Müller, Vinh Nguyen, Ninnia Lescano, et al. 2019. "Landscape of Stimulation-Responsive Chromatin across Diverse Human Immune Cells." *Nature Genetics* 51 (10): 1494–1505. <https://doi.org/10.1038/s41588-019-0505-9>.
- Campbell, James J., Rachael A. Clark, Rei Watanabe, and Thomas S. Kupper. 2010. "Sézary Syndrome and Mycosis Fungoides Arise from Distinct T-Cell Subsets: A Biologic Rationale for Their Distinct Clinical Behaviors." *Blood* 116 (5): 767–71. <https://doi.org/10.1182/blood-2009-11-251926>.
- Campo, Elias, Elaine S. Jaffe, James R. Cook, Leticia Quintanilla-Martinez, Steven H. Swerdlow, Kenneth C. Anderson, Pierre Brousset, et al. 2022. "The International Consensus Classification of Mature Lymphoid Neoplasms: A Report from the Clinical Advisory Committee." *Blood* 140 (11): 1229–53. <https://doi.org/10.1182/blood.2022015851>.

- Cao, F., H. Zhao, Y. Li, S. Dai, and C. Wang. 2015. "Clinicopathological and Phenotypic Features of Chronic NK Cell Lymphocytosis Identified among Patients with Asymptomatic Lymphocytosis." *International Journal of Laboratory Hematology* 37 (6): 783–90. <https://doi.org/10.1111/ijlh.12410>.
- Cao, Junyue, Darren A. Cusanovich, Vijay Ramani, Delasa Aghamirzaie, Hannah A. Pliner, Andrew J. Hill, Riza M. Daza, et al. 2018. "Joint Profiling of Chromatin Accessibility and Gene Expression in Thousands of Single Cells." *Science* 361 (6409): 1380–85. <https://doi.org/10.1126/science.aau0730>.
- Capper, David, David T. W. Jones, Martin Sill, Volker Hovestadt, Daniel Schrimpf, Dominik Sturm, Christian Koelsche, et al. 2018. "DNA Methylation-Based Classification of Central Nervous System Tumours." *Nature* 555 (7697): 469–74. <https://doi.org/10.1038/nature26000>.
- Carbone, A., A. Gloghini, G. Gaidano, R. Dalla-Favera, and B. Falini. 1997. "BCL-6 Protein Expression in Human Peripheral T-Cell Neoplasms Is Restricted to CD30+ Anaplastic Large-Cell Lymphomas." *Blood* 90 (6): 2445–50.
- Carninci, Piero, Albin Sandelin, Boris Lenhard, Shintaro Katayama, Kazuro Shimokawa, Jasmina Ponjavic, Colin A. M. Semple, et al. 2006. "Genome-Wide Analysis of Mammalian Promoter Architecture and Evolution." *Nature Genetics* 38 (6): 626–35. <https://doi.org/10.1038/ng1789>.
- Carras, Sylvain, Dimitri Chartoire, Sylvain Mareschal, Maël Heiblig, Antoine Marçais, Rémy Robinot, Mirjam Urb, et al. 2021. "Chronic T Cell Receptor Stimulation Unmasks NK Receptor Signaling in Peripheral T Cell Lymphomas via Epigenetic Reprogramming." *The Journal of Clinical Investigation* 131 (13): e139675. <https://doi.org/10.1172/JCI139675>.
- Catovsky, D., J. Galetto, A. Okos, D. A. Galton, E. Wiltshaw, and G. Stathopoulos. 1973. "Prolymphocytic Leukaemia of B and T Cell Type." *Lancet (London, England)* 2 (7823): 232–34. [https://doi.org/10.1016/s0140-6736\(73\)93135-8](https://doi.org/10.1016/s0140-6736(73)93135-8).
- Catovsky, D., M. F. Greaves, M. Rose, D. A. Galton, A. W. Goolden, D. R. McCluskey, J. M. White, et al. 1982. "Adult T-Cell Lymphoma-Leukaemia in Blacks from the West Indies." *Lancet (London, England)* 1 (8273): 639–43. [https://doi.org/10.1016/s0140-6736\(82\)92200-0](https://doi.org/10.1016/s0140-6736(82)92200-0).
- Cetinözman, Fatma, Patty M. Jansen, Maarten H. Vermeer, and Rein Willemze. 2012. "Differential Expression of Programmed Death-1 (PD-1) in Sézary Syndrome and Mycosis Fungoides." *Archives of Dermatology* 148 (12): 1379–85. <https://doi.org/10.1001/archdermatol.2012.2089>.
- Chan, J. K., W. Y. Tsang, and C. S. Ng. 1996. "Clarification of CD3 Immunoreactivity in Nasal T/Natural Killer Cell Lymphomas: The Neoplastic Cells Are Often CD3 Epsilon+." *Blood* 87 (2): 839–41.
- Chang, Li-Wei, Christina Chung Patrone, Wei Yang, Raquel Rabionet, Fernando Gallardo, Blanca Espinet, Mukesh K. Sharma, et al. 2018. "An Integrated Data Resource for Genomic Analysis of Cutaneous T-Cell Lymphoma." *Journal of Investigative Dermatology* 138 (12): 2681–83. <https://doi.org/10.1016/j.jid.2018.06.176>.
- Chatterjee, Nilanjana, Divya Sinha, Mekonnen Lemma-Dechassa, Song Tan, Michael A. Shogren-Knaak, and Blaine Bartholomew. 2011. "Histone H3 Tail Acetylation Modulates ATP-Dependent Remodeling through Multiple Mechanisms." *Nucleic Acids Research* 39 (19): 8378–91. <https://doi.org/10.1093/nar/gkr535>.



- Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. 2002. "SMOTE: Synthetic Minority Over-Sampling Technique." *Journal of Artificial Intelligence Research* 16 (June): 321–57. <https://doi.org/10.1613/jair.953>.
- Cheminant, Morgane, Julie Bruneau, Georgia Malamut, David Sibon, Nicolas Guegan, Tom van Gils, Sascha Cording, et al. 2019. "NKp46 Is a Diagnostic Biomarker and May Be a Therapeutic Target in Gastrointestinal T-Cell Lymphoproliferative Diseases: A CELAC Study." *Gut* 68 (8): 1396–1405. <https://doi.org/10.1136/gutjnl-2018-317371>.
- Chen, Song, Blue B. Lake, and Kun Zhang. 2019. "High-Throughput Sequencing of the Transcriptome and Chromatin Accessibility in the Same Cell." *Nature Biotechnology* 37 (12): 1452–57. <https://doi.org/10.1038/s41587-019-0290-0>.
- Chen, Yongheng, Darren L. Bates, Raja Dey, Po-Han Chen, Ana Carolina Dantas Machado, Ite A. Laird-Offringa, Remo Rohs, and Lin Chen. 2012. "DNA Binding by GATA Transcription Factor Suggests Mechanisms of DNA Looping and Long-Range Gene Regulation." *Cell Reports* 2 (5): 1197–1206. <https://doi.org/10.1016/j.celrep.2012.10.012>.
- Cheon, HeeJin, Jeffrey C. Xing, Katharine B. Moosic, Johnson Ung, Vivian W. Chan, David S. Chung, Mariella F. Toro, et al. 2022. "Genomic Landscape of TCR $\alpha\beta$  and TCR $\gamma\delta$  T-Large Granular Lymphocyte Leukemia." *Blood* 139 (20): 3058–72. <https://doi.org/10.1182/blood.2021013164>.
- Chiorazzi, Nicholas, and Manlio Ferrarini. 2011. "Cellular Origin(s) of Chronic Lymphocytic Leukemia: Cautionary Notes and Additional Considerations and Possibilities." *Blood* 117 (6): 1781–91. <https://doi.org/10.1182/blood-2010-07-155663>.
- Choi, Jaehyuk, Gerald Goh, Trent Walradt, Bok S. Hong, Christopher G. Bunick, Kan Chen, Robert D. Bjornson, et al. 2015. "Genomic Landscape of Cutaneous T Cell Lymphoma." *Nature Genetics* 47 (9): 1011–19. <https://doi.org/10.1038/ng.3356>.
- Chu, Tinyi, Zhong Wang, Dana Pe'er, and Charles G. Danko. 2022. "Cell Type and Gene Expression Deconvolution with BayesPrism Enables Bayesian Integrative Analysis across Bulk and Single-Cell RNA Sequencing in Oncology." *Nature Cancer* 3 (4): 505–17. <https://doi.org/10.1038/s43018-022-00356-3>.
- Chung, Dongjun, and Sunduz Keles. 2010. "Sparse Partial Least Squares Classification for High Dimensional Data." *Statistical Applications in Genetics and Molecular Biology* 9 (1). <https://doi.org/10.2202/1544-6115.1492>.
- Cichocki, Frank, Bartosz Grzywacz, and Jeffrey S. Miller. 2019. "Human NK Cell Development: One Road or Many?" *Frontiers in Immunology* 10. <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02078>.
- Ciofani, Maria, Gisèle C. Knowles, David L. Wiest, Harald Von Boehmer, and Juan Carlos Zúñiga-Pflücker. 2006. "Stage-Specific and Differential Notch Dependency at the A $\beta$  and  $\Gamma\delta$  T Lineage Bifurcation." *Immunity* 25 (1): 105–16. <https://doi.org/10.1016/j.immuni.2006.05.010>.
- Clapier, Cedric R., Janet Iwasa, Bradley R. Cairns, and Craig L. Peterson. 2017. "Mechanisms of Action and Regulation of ATP-Dependent Chromatin-Remodelling Complexes." *Nature Reviews. Molecular Cell Biology* 18 (7): 407–22. <https://doi.org/10.1038/nrm.2017.26>.
- Climent, Fina, Alina Nicolae, Laurence de Leval, Stefan Dirnhofer, Lorenzo Leoncini, Sarah L. Ondrejka, Lorinda Soma, et al. 2023. "Cytotoxic Peripheral T-Cell Lymphomas and EBV-Positive T/NK-Cell Lymphoproliferative Diseases: Emerging Concepts, Recent Advances, and the Putative Role of Clonal Hematopoiesis. A Report of the 2022

- EA4HP/SH Lymphoma Workshop." *Virchows Archiv: An International Journal of Pathology* 483 (3): 333–48. <https://doi.org/10.1007/s00428-023-03616-4>.
- Cohen, Carla J., Wynne M. Lock, and Dixie L. Mager. 2009. "Endogenous Retroviral LTRs as Promoters for Human Genes: A Critical Assessment." *Gene* 448 (2): 105–14. <https://doi.org/10.1016/j.gene.2009.06.020>.
- Coiffier, Bertrand, Barbara Pro, H. Miles Prince, Francine Foss, Lubomir Sokol, Matthew Greenwood, Dolores Caballero, et al. 2012. "Results from a Pivotal, Open-Label, Phase II Study of Romidepsin in Relapsed or Refractory Peripheral T-Cell Lymphoma after Prior Systemic Therapy." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 30 (6): 631–36. <https://doi.org/10.1200/JCO.2011.37.4223>.
- Congras, Annabelle, Coralie Hoareau-Aveilla, Nina Caillet, Marie Tosolini, Patrick Villarese, Agata Cieslak, Laura Rodriguez, et al. 2020. "ALK-Transformed Mature T Lymphocytes Restore Early Thymus Progenitor Features." *Journal of Clinical Investigation*, November, 10.1172/JCI134990. <https://doi.org/10.1172/JCI134990>.
- Cooke, C. B., L. Krenacs, M. Stetler-Stevenson, T. C. Greiner, M. Raffeld, D. W. Kingma, L. Abruzzo, C. Frantz, M. Kaviani, and E. S. Jaffe. 1996. "Hepatosplenic T-Cell Lymphoma: A Distinct Clinicopathologic Entity of Cytotoxic Gamma Delta T-Cell Origin." *Blood* 88 (11): 4265–74.
- Corces, M. Ryan, Jason D. Buenrostro, Beijing Wu, Peyton G. Greenside, Steven M. Chan, Julie L. Koenig, Michael P. Snyder, et al. 2016. "Lineage-Specific and Single-Cell Chromatin Accessibility Charts Human Hematopoiesis and Leukemia Evolution." *Nature Genetics* 48 (10): 1193–1203. <https://doi.org/10.1038/ng.3646>.
- Corces, M. Ryan, Jeffrey M. Granja, Shadi Shams, Bryan H. Louie, Jose A. Seoane, Wanding Zhou, Tiago C. Silva, et al. 2018. "The Chromatin Accessibility Landscape of Primary Human Cancers." *Science (New York, N.Y.)* 362 (6413). <https://doi.org/10.1126/science.aav1898>.
- Corces, M. Ryan, Alexandro E. Trevino, Emily G. Hamilton, Peyton G. Greenside, Nicholas A. Sinnott-Armstrong, Sam Vesuna, Ansuman T. Satpathy, et al. 2017. "An Improved ATAC-Seq Protocol Reduces Background and Enables Interrogation of Frozen Tissues." *Nature Methods* 14 (10): 959–62. <https://doi.org/10.1038/nmeth.4396>.
- Cording, Sascha, Ludovic Lhermitte, Georgia Malamut, Sofia Berrabah, Amélie Trinquand, Nicolas Guegan, Patrick Villarese, et al. 2022. "Oncogenetic Landscape of Lymphomagenesis in Coeliac Disease." *Gut* 71 (3): 497–508. <https://doi.org/10.1136/gutjnl-2020-322935>.
- Core, Leighton J., André L. Martins, Charles G. Danko, Colin T. Waters, Adam Siepel, and John T. Lis. 2014. "Analysis of Nascent RNA Identifies a Unified Architecture of Initiation Regions at Mammalian Promoters and Enhancers." *Nature Genetics* 46 (12): 1311–20. <https://doi.org/10.1038/ng.3142>.
- Cosma, Maria Pia. 2002. "Ordered Recruitment: Gene-Specific Mechanism of Transcription Activation." *Molecular Cell* 10 (2): 227–36. [https://doi.org/10.1016/S1097-2765\(02\)00604-4](https://doi.org/10.1016/S1097-2765(02)00604-4).
- Couronné, Lucile, Christian Bastard, and Olivier A. Bernard. 2012. "TET2 and DNMT3A Mutations in Human T-Cell Lymphoma." *The New England Journal of Medicine* 366 (1): 95–96. <https://doi.org/10.1056/NEJMc1111708>.
- Coustan-Smith, Elaine, Charles G. Mullighan, Mihaela Onciu, Frederick G. Behm, Susana C. Raimondi, Deqing Pei, Cheng Cheng, et al. 2009. "Early T-Cell Precursor Leukaemia: A

- Subtype of Very High-Risk Acute Lymphoblastic Leukaemia." *The Lancet. Oncology* 10 (2): 147–56. [https://doi.org/10.1016/S1470-2045\(08\)70314-0](https://doi.org/10.1016/S1470-2045(08)70314-0).
- Crawford, Gregory E., Sean Davis, Peter C. Scacheri, Gabriel Renaud, Mohamad J. Halawi, Michael R. Erdos, Roland Green, Paul S. Meltzer, Tyra G. Wolfsberg, and Francis S. Collins. 2006. "DNase-Chip: A High-Resolution Method to Identify DNase I Hypersensitive Sites Using Tiled Microarrays." *Nature Methods* 3 (7): 503–9. <https://doi.org/10.1038/nmeth888>.
- Cremer, Thomas, and Marion Cremer. 2010. "Chromosome Territories." *Cold Spring Harbor Perspectives in Biology* 2 (3): a003889. <https://doi.org/10.1101/cshperspect.a003889>.
- Crescenzo, Ramona, Francesco Abate, Elena Lasorsa, Fabrizio Tabbo', Marcello Gaudiano, Nicoletta Chiesa, Filomena Di Giacomo, et al. 2015. "Convergent Mutations and Kinase Fusions Lead to Oncogenic STAT3 Activation in Anaplastic Large Cell Lymphoma." *Cancer Cell* 27 (4): 516–32. <https://doi.org/10.1016/j.ccell.2015.03.006>.
- Creyghton, Menno P., Albert W. Cheng, G. Grant Welstead, Tristan Kooistra, Bryce W. Carey, Eveline J. Steine, Jacob Hanna, et al. 2010. "Histone H3K27ac Separates Active from Poised Enhancers and Predicts Developmental State." *Proceedings of the National Academy of Sciences of the United States of America* 107 (50): 21931–36. <https://doi.org/10.1073/pnas.1016071107>.
- Cristiano, Stephen, Alessandro Leal, Jillian Phallen, Jacob Fiksel, Vilmos Adleff, Daniel C. Bruhm, Sarah Østrup Jensen, et al. 2019. "Genome-Wide Cell-Free DNA Fragmentation in Patients with Cancer." *Nature* 570 (7761): 385–89. <https://doi.org/10.1038/s41586-019-1272-6>.
- Cusanovich, Darren A., Riza Daza, Andrew Adey, Hannah A. Pliner, Lena Christiansen, Kevin L. Gunderson, Frank J. Steemers, Cole Trapnell, and Jay Shendure. 2015. "Multiplex Single-Cell Profiling of Chromatin Accessibility by Combinatorial Cellular Indexing." *Science* 348 (6237): 910–14. <https://doi.org/10.1126/science.aab1601>.
- Dankowski, Theresa, and Andreas Ziegler. 2016. "Calibrating Random Forests for Probability Estimation." *Statistics in Medicine* 35 (22): 3949–60. <https://doi.org/10.1002/sim.6959>.
- Diamandidou, E., M. Colome-Grimmer, L. Fayad, M. Duvic, and R. Kurzrock. 1998. "Transformation of Mycosis Fungoides/Sezary Syndrome: Clinical Characteristics and Prognosis." *Blood* 92 (4): 1150–59.
- Díaz-Uriarte, Ramón, and Sara Alvarez de Andrés. 2006. "Gene Selection and Classification of Microarray Data Using Random Forest." *BMC Bioinformatics* 7 (1): 3. <https://doi.org/10.1186/1471-2105-7-3>.
- Dinga, Richard, Lianne Schmaal, Brenda W.J.H. Penninx, Dick Veltman, and Andre Marquand. 2020. "Controlling for Effects of Confounding Variables on Machine Learning Prediction." <https://doi.org/10.1101/2020.08.17.255034>.
- Dobay, Maria Pamela, Francois Lemonnier, Edoardo Missiaglia, Christian Bastard, David Vallois, Jean-Philippe Jais, Laurianne Scourzic, et al. 2017. "Integrative Clinicopathological and Molecular Analyses of Angioimmunoblastic T-Cell Lymphoma and Other Nodal Lymphomas of Follicular Helper T-Cell Origin." *Haematologica* 102 (4): e148–51. <https://doi.org/10.3324/haematol.2016.158428>.
- Dogan, Ahmet, Ayoma D. Attygalle, and Chara Kyriakou. 2003. "Angioimmunoblastic T-Cell Lymphoma." *British Journal of Haematology* 121 (5): 681–91. <https://doi.org/10.1046/j.1365-2141.2003.04335.x>.

- Dong, Kevin B, Irina A Maksakova, Fabio Mohn, Danny Leung, Ruth Appanah, Sandra Lee, Hao W Yang, et al. 2008. "DNA Methylation in ES Cells Requires the Lysine Methyltransferase G9a but Not Its Catalytic Activity." *The EMBO Journal* 27 (20): 2691–2701. <https://doi.org/10.1038/emboj.2008.193>.
- Dorfman, David M., Julia A. Brown, Aliakbar Shahsafaei, and Gordon J. Freeman. 2006. "Programmed Death-1 (PD-1) Is a Marker of Germinal Center-Associated T Cells and Angioimmunoblastic T-Cell Lymphoma." *The American Journal of Surgical Pathology* 30 (7): 802–10. <https://doi.org/10.1097/01.pas.0000209855.28282.ce>.
- Drieux, Fanny, Philippe Ruminy, Ahmad Abdel-Sater, François Lemonnier, Pierre-Julien Vially, Virginie Fataccioli, Vinciane Marchand, et al. 2020. "Defining Signatures of Peripheral T-Cell Lymphoma with a Targeted 20-Marker Gene Expression Profiling Assay." *Haematologica* 105 (6): 1582–92. <https://doi.org/10.3324/haematol.2019.226647>.
- Edinger, James T., Beth Z. Clark, Brian E. Pucevich, Larisa J. Geskin, and Steven H. Swerdlow. 2009. "CD30 Expression and Proliferative Fraction in Nontransformed Mycosis Fungoides." *The American Journal of Surgical Pathology* 33 (12): 1860–68. <https://doi.org/10.1097/PAS.0b013e3181bf677d>.
- Emmett, Matthew, and Mitchell Lazar. 2019. "Integrative Regulation of Physiology by Histone Deacetylase 3." *Nature Reviews Molecular Cell Biology* 20 (February): 102–15. <https://doi.org/10.1038/s41580-018-0076-0>.
- Ertöz, Levent, Michael Steinbach, and Vipin Kumar. 2003. "Finding Clusters of Different Sizes, Shapes, and Densities in Noisy, High Dimensional Data." In *Proceedings of the 2003 SIAM International Conference on Data Mining*, 47–58. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9781611972733.5>.
- Ester, M., H. Kriegel, J. Sander, and Xiaowei Xu. 1996. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." In . <https://www.semanticscholar.org/paper/A-Density-Based-Algorithm-for-Discovering-Clusters-Ester-Kriegel/5c8fe9a0412a078e30eb7e5eeb0068655b673e86>.
- Ettersperger, Julien, Nicolas Montcuquet, Georgia Malamut, Nicolas Guegan, Silvia Lopez-Lastra, Ségolène Gayraud, Christian Reimann, et al. 2016. "Interleukin-15-Dependent T-Cell-like Innate Intraepithelial Lymphocytes Develop in the Intestine and Transform into Lymphomas in Celiac Disease." *Immunity* 45 (3): 610–25. <https://doi.org/10.1016/j.immuni.2016.07.018>.
- Falini, B., B. Bigerna, M. Fizzotti, K. Pulford, S. A. Pileri, G. Delsol, A. Carbone, et al. 1998. "ALK Expression Defines a Distinct Group of T/Null Lymphomas ('ALK Lymphomas') with a Wide Morphological Spectrum." *The American Journal of Pathology* 153 (3): 875–86. [https://doi.org/10.1016/S0002-9440\(10\)65629-5](https://doi.org/10.1016/S0002-9440(10)65629-5).
- Faulkner, Geoffrey J., Yasumasa Kimura, Carsten O. Daub, Shivangi Wani, Charles Plessy, Katharine M. Irvine, Kate Schroder, et al. 2009. "The Regulated Retrotransposon Transcriptome of Mammalian Cells." *Nature Genetics* 41 (5): 563–71. <https://doi.org/10.1038/ng.368>.
- Fei, Jia, Sharon E. Torigoe, Christopher R. Brown, Mai T. Khuong, George A. Kassavetis, Hinrich Boeger, and James T. Kadonaga. 2015. "The Prenucleosome, a Stable Conformational Isomer of the Nucleosome." *Genes & Development* 29 (24): 2563–75. <https://doi.org/10.1101/gad.272633.115>.
- Feldman, Andrew L., Ahmet Dogan, David I. Smith, Mark E. Law, Stephen M. Ansell, Sarah H. Johnson, Julie C. Porcher, et al. 2011. "Discovery of Recurrent t(6;7)(P25.3;Q32.3) Translocations in ALK-Negative Anaplastic Large Cell Lymphomas by Massively Parallel

- Genomic Sequencing." *Blood* 117 (3): 915–19. <https://doi.org/10.1182/blood-2010-08-303305>.
- Feldman, Andrew L., Mark Law, Karen L. Grogg, Erik C. Thorland, Stephanie Fink, Paul J. Kurtin, William R. Macon, Ellen D. Remstein, and Ahmet Dogan. 2008. "Incidence of *TCR* and *TCL1* Gene Translocations and Isochromosome 7q in Peripheral T-Cell Lymphomas Using Fluorescence In Situ Hybridization." *American Journal of Clinical Pathology* 130 (2): 178–85. <https://doi.org/10.1309/PNXUKA1CFJMVGCN1>.
- Ferreiro, Julio Finalet, Leila Rouhigharabaei, Helena Urbankova, Jo-Anne van der Krogt, Lucienne Michaux, Shashirekha Shetty, Laszlo Krenacs, et al. 2014. "Integrative Genomic and Transcriptomic Analysis Identified Candidate Genes Implicated in the Pathogenesis of Hepatosplenic T-Cell Lymphoma." *PLOS ONE* 9 (7): e102977. <https://doi.org/10.1371/journal.pone.0102977>.
- Fertig, Elana J., Jie Ding, Alexander V. Favorov, Giovanni Parmigiani, and Michael F. Ochs. 2010. "CoGAPS: An R/C++ Package to Identify Patterns and Biological Process Activity in Transcriptomic Data." *Bioinformatics* 26 (21): 2792–93. <https://doi.org/10.1093/bioinformatics/btq503>.
- Foss, H. D., I. Anagnostopoulos, I. Araujo, C. Assaf, G. Demel, J. A. Kummer, M. Hummel, and H. Stein. 1996. "Anaplastic Large-Cell Lymphomas of T-Cell and Null-Cell Phenotype Express Cytotoxic Molecules." *Blood* 88 (10): 4005–11.
- Fridman, Wolf Herman, Franck Pagès, Catherine Sautès-Fridman, and Jérôme Galon. 2012. "The Immune Contexture in Human Tumours: Impact on Clinical Outcome." *Nature Reviews. Cancer* 12 (4): 298–306. <https://doi.org/10.1038/nrc3245>.
- Frith, Martin C., Eivind Valen, Anders Krogh, Yoshihide Hayashizaki, Piero Carninci, and Albin Sandelin. 2008. "A Code for Transcription Initiation in Mammalian Genomes." *Genome Research* 18 (1): 1–12. <https://doi.org/10.1101/gr.6831208>.
- Fujisawa, Manabu, Tran B. Nguyen, Yoshiaki Abe, Yasuhito Suehara, Kota Fukumoto, Sakurako Suma, Kenichi Makishima, et al. 2022. "Clonal Germinal Center B Cells Function as a Niche for T-Cell Lymphoma." *Blood* 140 (18): 1937–50. <https://doi.org/10.1182/blood.2022015451>.
- Gallardo, Fernando, Andreina Padrón, Ricard Garcia-Carbonell, Cristina Rius, Abel Gonzalez-Perez, Montserrat Arumí-Uria, Mar Iglesias, et al. 2015. "Cytoplasmic Accumulation of NCoR in Malignant Melanoma: Consequences of Altered Gene Repression and Prognostic Significance." *Oncotarget* 5 (March). <https://doi.org/10.18632/oncotarget.3252>.
- Galton, D. A., J. M. Goldman, E. Wiltshaw, D. Catovsky, K. Henry, and G. J. Goldenberg. 1974. "Prolymphocytic Leukaemia." *British Journal of Haematology* 27 (1): 7–23. <https://doi.org/10.1111/j.1365-2141.1974.tb06769.x>.
- Gao, Yuan, and George Church. 2005. "Improving Molecular Cancer Class Discovery through Sparse Non-Negative Matrix Factorization." *Bioinformatics* 21 (21): 3970–75. <https://doi.org/10.1093/bioinformatics/bti653>.
- Gaujoux, Renaud, and Cathal Seoighe. 2012. "Semi-Supervised Nonnegative Matrix Factorization for Gene Expression Deconvolution: A Case Study." *Infection, Genetics and Evolution* 12 (5): 913–21. <https://doi.org/10.1016/j.meegid.2011.08.014>.
- . 2013. "CellMix: A Comprehensive Toolbox for Gene Expression Deconvolution." *Bioinformatics (Oxford, England)* 29 (17): 2211–12. <https://doi.org/10.1093/bioinformatics/btt351>.

- Gessain, Antoine, and Olivier Cassar. 2012. "Epidemiological Aspects and World Distribution of HTLV-1 Infection." *Frontiers in Microbiology* 3: 388. <https://doi.org/10.3389/fmicb.2012.00388>.
- Giles, Josephine R., Sasikanth Manne, Elizabeth Freilich, Derek A. Oldridge, Amy E. Baxter, Sangeeth George, Zeyu Chen, et al. 2022. "Human Epigenetic and Transcriptional T Cell Differentiation Atlas for Identifying Functional T Cell-Specific Enhancers." *Immunity* 55 (3): 557-574.e7. <https://doi.org/10.1016/j.immuni.2022.02.004>.
- Giresi, Paul G., Jonghwan Kim, Ryan M. McDaniell, Vishwanath R. Iyer, and Jason D. Lieb. 2007. "FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) Isolates Active Regulatory Elements from Human Chromatin." *Genome Research* 17 (6): 877–85. <https://doi.org/10.1101/gr.5533506>.
- Goryshin, I. Y., and W. S. Reznikoff. 1998. "Tn5 in Vitro Transposition." *The Journal of Biological Chemistry* 273 (13): 7367–74. <https://doi.org/10.1074/jbc.273.13.7367>.
- Granja, Jeffrey M., M. Ryan Corces, Sarah E. Pierce, S. Tansu Bagdatli, Hani Choudhry, Howard Y. Chang, and William J. Greenleaf. 2021. "ArchR Is a Scalable Software Package for Integrative Single-Cell Chromatin Accessibility Analysis." *Nature Genetics* 53 (3): 403–11. <https://doi.org/10.1038/s41588-021-00790-6>.
- Grogg, Karen L., Ayoma D. Attygalle, William R. Macon, Ellen D. Remstein, Paul J. Kurtin, and Ahmet Dogan. 2005. "Angioimmunoblastic T-Cell Lymphoma: A Neoplasm of Germinal-Center T-Helper Cells?" *Blood* 106 (4): 1501–2. <https://doi.org/10.1182/blood-2005-03-1083>.
- Gui, Yaoting, Guangwu Guo, Yi Huang, Xueda Hu, Aifa Tang, Shengjie Gao, Renhua Wu, et al. 2011. "Frequent Mutations of Chromatin Remodeling Genes in Transitional Cell Carcinoma of the Bladder." *Nature Genetics* 43 (9): 875–78. <https://doi.org/10.1038/ng.907>.
- Hamblin, T. J., Z. Davis, A. Gardiner, D. G. Oscier, and F. K. Stevenson. 1999. "Unmutated Ig V(H) Genes Are Associated with a More Aggressive Form of Chronic Lymphocytic Leukemia." *Blood* 94 (6): 1848–54.
- Hapgood, Greg, Susana Ben-Neriah, Anja Mottok, Derrick G. Lee, Kridel Robert, Diego Villa, Laurie H. Sehn, et al. 2019. "Identification of High-Risk DUSP22-Rearranged ALK-Negative Anaplastic Large Cell Lymphoma." *British Journal of Haematology* 186 (3): e28–31. <https://doi.org/10.1111/bjh.15860>.
- Heavican, Tayla B., Alyssa Bouska, Jiayu Yu, Waseem Lone, Catalina Amador, Qiang Gong, Weiwei Zhang, et al. 2019. "Genetic Drivers of Oncogenic Pathways in Molecular Subgroups of Peripheral T-Cell Lymphoma." *Blood* 133 (15): 1664–76. <https://doi.org/10.1182/blood-2018-09-872549>.
- Heintzman, Nathaniel D., Gary C. Hon, R. David Hawkins, Pouya Kheradpour, Alexander Stark, Lindsey F. Harp, Zhen Ye, et al. 2009. "Histone Modifications at Human Enhancers Reflect Global Cell-Type-Specific Gene Expression." *Nature* 459 (7243): 108–12. <https://doi.org/10.1038/nature07829>.
- Herling, Marco, Kaushali A. Patel, Michael A. Teitell, Marina Konopleva, Farhad Ravandi, Ryuji Kobayashi, and Dan Jones. 2008. "High TCL1 Expression and Intact T-Cell Receptor Signaling Define a Hyperproliferative Subset of T-Cell Prolymphocytic Leukemia." *Blood* 111 (1): 328–37. <https://doi.org/10.1182/blood-2007-07-101519>.
- Herrera, Alberto, Anthony Cheng, Eleni P. Mimitou, Angelina Seffens, Dean George, Michal Bar-Natan, Adriana Heguy, et al. 2021. "Multimodal Single-Cell Analysis of Cutaneous

- T-Cell Lymphoma Reveals Distinct Subclonal Tissue-Dependent Signatures." *Blood* 138 (16): 1456–64. <https://doi.org/10.1182/blood.2020009346>.
- Hesselberth, Jay R., Xiaoyu Chen, Zhihong Zhang, Peter J. Sabo, Richard Sandstrom, Alex P. Reynolds, Robert E. Thurman, et al. 2009. "Global Mapping of Protein-DNA Interactions in Vivo by Digital Genomic Footprinting." *Nature Methods* 6 (4): 283–89. <https://doi.org/10.1038/nmeth.1313>.
- Hewish, D. R., and L. A. Burgoyne. 1973. "Chromatin Sub-Structure. The Digestion of Chromatin DNA at Regularly Spaced Sites by a Nuclear Deoxyribonuclease." *Biochemical and Biophysical Research Communications* 52 (2): 504–10. [https://doi.org/10.1016/0006-291x\(73\)90740-7](https://doi.org/10.1016/0006-291x(73)90740-7).
- Hnisz, Denes, Brian J. Abraham, Tong Ihn Lee, Ashley Lau, Violaine Saint-André, Alla A. Sigova, Heather A. Hoke, and Richard A. Young. 2013. "Super-Enhancers in the Control of Cell Identity and Disease." *Cell* 155 (4): 934–47. <https://doi.org/10.1016/j.cell.2013.09.053>.
- Hotelling, H. 1933. "Analysis of a Complex of Statistical Variables into Principal Components." *Journal of Educational Psychology* 24 (6): 417–41. <https://doi.org/10.1037/h0071325>.
- Houseman, E. Andres, Molly L. Kile, David C. Christiani, Tan A. Ince, Karl T. Kelsey, and Carmen J. Marsit. 2016. "Reference-Free Deconvolution of DNA Methylation Data and Mediation by Cell Composition Effects." *BMC Bioinformatics* 17 (1): 259. <https://doi.org/10.1186/s12859-016-1140-4>.
- Hu, G., S. Dasari, Y. W. Asmann, P. T. Greipp, R. A. Knudson, H. K. Benson, Y. Li, et al. 2018. "Targetable Fusions of the FRK Tyrosine Kinase in ALK-Negative Anaplastic Large Cell Lymphoma." *Leukemia* 32 (2): 565–69. <https://doi.org/10.1038/leu.2017.309>.
- Huang, Yenlin, Anne Moreau, Jehan Dupuis, Berthold Streubel, Barbara Petit, Steven Le Gouill, Nadine Martin-Garcia, et al. 2009. "Peripheral T-Cell Lymphomas with a Follicular Growth Pattern Are Derived from Follicular Helper T Cells (TFH) and May Show Overlapping Features with Angioimmunoblastic T-Cell Lymphomas." *The American Journal of Surgical Pathology* 33 (5): 682–90. <https://doi.org/10.1097/PAS.0b013e3181971591>.
- Huynh, Khanh D., and Vivian J. Bardwell. 1998. "The BCL-6 POZ Domain and Other POZ Domains Interact with the Co-Repressors N-CoR and SMRT." *Oncogene* 17 (19): 2473–84. <https://doi.org/10.1038/sj.onc.1202197>.
- Iqbal, Javeed, Dennis D. Weisenburger, Timothy C. Greiner, Julie M. Vose, Timothy McKeithan, Can Kucuk, Huimin Geng, et al. 2010. "Molecular Signatures to Improve Diagnosis in Peripheral T-Cell Lymphoma and Prognostication in Angioimmunoblastic T-Cell Lymphoma." *Blood* 115 (5): 1026–36. <https://doi.org/10.1182/blood-2009-06-227579>.
- Iqbal, Javeed, George Wright, Chao Wang, Andreas Rosenwald, Randy D. Gascoyne, Dennis D. Weisenburger, Timothy C. Greiner, et al. 2014. "Gene Expression Signatures Delineate Biological and Prognostic Subgroups in Peripheral T-Cell Lymphoma." *Blood* 123 (19): 2915–23. <https://doi.org/10.1182/blood-2013-11-536359>.
- Isakova, Alina, Romain Groux, Michael Imbeault, Pernille Rainer, Daniel Alpern, Riccardo Dainese, Giovanna Ambrosini, Didier Trono, Philipp Bucher, and Bart Deplancke. 2017. "SMiLE-Seq Identifies Binding Motifs of Single and Dimeric Transcription Factors." *Nature Methods* 14 (3): 316–22. <https://doi.org/10.1038/nmeth.4143>.
- Izutsu, Koji, Shinichi Makita, Kisato Nosaka, Makoto Yoshimitsu, Atae Utsunomiya, Shigeru Kusumoto, Satoko Morishima, et al. 2022. "An Open-Label, Single-Arm, Phase 2 Trial

- of Valemetostat in Relapsed or Refractory Adult T-Cell Leukemia/Lymphoma.” *Blood*, September, blood.2022016862. <https://doi.org/10.1182/blood.2022016862>.
- Iżykowska, Katarzyna, Grzegorz K. Przybylski, Claudia Gand, Floriane C. Braun, Piotr Grabarczyk, Andreas W. Kuss, Karolina Olek-Hrab, et al. 2017. “Genetic Rearrangements Result in Altered Gene Expression and Novel Fusion Transcripts in S&#x00E9;Zary Syndrome.” *Oncotarget* 8 (24): 39627–39. <https://doi.org/10.18632/oncotarget.17383>.
- Jaffe, E. S., W. A. Blattner, D. W. Blayney, P. A. Bunn, J. Cossman, M. Robert-Guroff, and R. C. Gallo. 1984. “The Pathologic Spectrum of Adult T-Cell Leukemia/Lymphoma in the United States. Human T-Cell Leukemia/Lymphoma Virus-Associated Lymphoid Malignancies.” *The American Journal of Surgical Pathology* 8 (4): 263–75. <https://doi.org/10.1097/00000478-198404000-00003>.
- Jaffe, E. S., J. K. Chan, I. J. Su, G. Frizzera, S. Mori, A. C. Feller, and F. C. Ho. 1996. “Report of the Workshop on Nasal and Related Extranodal Angiocentric T/Natural Killer Cell Lymphomas. Definitions, Differential Diagnosis, and Epidemiology.” *The American Journal of Surgical Pathology* 20 (1): 103–11. <https://doi.org/10.1097/00000478-199601000-00012>.
- Jamshidi, Arash, Minetta C. Liu, Eric A. Klein, Oliver Venn, Earl Hubbell, John F. Beausang, Samuel Gross, et al. 2022. “Evaluation of Cell-Free DNA Approaches for Multi-Cancer Early Detection.” *Cancer Cell* 40 (12): 1537-1549.e12. <https://doi.org/10.1016/j.ccell.2022.10.022>.
- Jerez, Andres, Michael J. Clemente, Hideki Makishima, Hanna Koskela, Francis Leblanc, Kwok Peng Ng, Thomas Olson, et al. 2012. “STAT3 Mutations Unify the Pathogenesis of Chronic Lymphoproliferative Disorders of NK Cells and T-Cell Large Granular Lymphocyte Leukemia.” *Blood* 120 (15): 3048–57. <https://doi.org/10.1182/blood-2012-06-435297>.
- Jiang, Lu, Zhao-Hui Gu, Zi-Xun Yan, Xia Zhao, Yin-Yin Xie, Zi-Guan Zhang, Chun-Ming Pan, et al. 2015. “Exome Sequencing Identifies Somatic Mutations of DDX3X in Natural Killer/T-Cell Lymphoma.” *Nature Genetics* 47 (9): 1061–66. <https://doi.org/10.1038/ng.3358>.
- Johnson, W. Evan, Cheng Li, and Ariel Rabinovic. 2007. “Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods.” *Biostatistics* 8 (1): 118–27. <https://doi.org/10.1093/biostatistics/kxj037>.
- Jones, Christine L., Andrea Degasperi, Vieri Grandi, Tauanne D. Amarante, Tracey J. Mitchell, Serena Nik-Zainal, and Sean J. Whittaker. 2021. “Spectrum of Mutational Signatures in T-Cell Lymphoma Reveals a Key Role for UV Radiation in Cutaneous T-Cell Lymphoma.” *Scientific Reports* 11 (1): 3962. <https://doi.org/10.1038/s41598-021-83352-4>.
- Kagdi, Huseini H., Maria A. Demontis, Paul A. Fields, Juan Carlos Ramos, Charles R. M. Bangham, and Graham P. Taylor. 2017. “Risk Stratification of Adult T-Cell Leukemia/Lymphoma Using Immunophenotyping.” *Cancer Medicine* 6 (1): 298–309. <https://doi.org/10.1002/cam4.928>.
- Karimi, Shirin, Jeffrey A. Zuccato, Yasin Mamatjan, Sheila Mansouri, Suganth Suppiah, Farshad Nassiri, Phedias Diamandis, David G. Munoz, Kenneth D. Aldape, and Gelareh Zadeh. 2019. “The Central Nervous System Tumor Methylation Classifier Changes Neuro-Oncology Practice for Challenging Brain Tumor Diagnoses and Directly Impacts Patient Care.” *Clinical Epigenetics* 11 (1): 185. <https://doi.org/10.1186/s13148-019-0766-2>.



- Karube, Kennosuke, Koichi Ohshima, Takeshi Tsuchiya, Takahiro Yamaguchi, Riko Kawano, Junji Suzumiya, Atae Utsunomiya, Mine Harada, and Masahiro Kikuchi. 2004. "Expression of FoxP3, a Key Molecule in CD4CD25 Regulatory T Cells, in Adult T-Cell Leukaemia/Lymphoma Cells." *British Journal of Haematology* 126 (1): 81–84. <https://doi.org/10.1111/j.1365-2141.2004.04999.x>.
- Kataoka, Keisuke, Yasunobu Nagata, Akira Kitanaka, Yuichi Shiraishi, Teppei Shimamura, Jun-ichirou Yasunaga, Yasushi Totoki, et al. 2015. "Integrated Molecular Analysis of Adult T Cell Leukemia/Lymphoma." *Nature Genetics* 47 (11): 1304–15. <https://doi.org/10.1038/ng.3415>.
- Kato, Seiichi, Naoko Asano, Tomoko Miyata-Takata, Katsuyoshi Takata, Ahmed Ali Elsayed, Akira Satou, Emiko Takahashi, Tomohiro Kinoshita, and Shigeo Nakamura. 2015. "T-Cell Receptor (TCR) Phenotype of Nodal Epstein-Barr Virus (EBV)-Positive Cytotoxic T-Cell Lymphoma (CTL): A Clinicopathologic Study of 39 Cases." *The American Journal of Surgical Pathology* 39 (4): 462–71. <https://doi.org/10.1097/PAS.0000000000000323>.
- Kaufman, Rdsuseun. 1987. "Clustering by Means of Medoids." *Proceedings of the Statistical Data Analysis Based on the L1 Norm Conference, Neuchatel, Switzerland* 31.
- Kawano, Fumio, Hiroyuki Tsuda, Kazunari Yamaguchi, Hiromichi Nishimura, Isao Sanada, Hiromitsu Matsuzaki, Masao Ishii, and Kiyoshi Takatsuki. 1984. "Unusual Clinical Courses of Adult T-Cell Leukemia in Siblings." *Cancer* 54 (1): 131–34. [https://doi.org/10.1002/1097-0142\(19840701\)54:1<131::AID-CNCR2820540126>3.0.CO;2-Q](https://doi.org/10.1002/1097-0142(19840701)54:1<131::AID-CNCR2820540126>3.0.CO;2-Q).
- Kiel, Mark J., Anagh A. Sahasrabudde, Delphine C. M. Rolland, Thirunavukkarasu Velusamy, Fuzon Chung, Matthew Schaller, Nathanael G. Bailey, et al. 2015. "Genomic Analyses Reveal Recurrent Mutations in Epigenetic Modifiers and the JAK–STAT Pathway in Sézary Syndrome." *Nature Communications* 6 (1): 8470. <https://doi.org/10.1038/ncomms9470>.
- Kiel, Mark J., Thirunavukkarasu Velusamy, Delphine Rolland, Anagh A. Sahasrabudde, Fuzon Chung, Nathanael G. Bailey, Alexandra Schrader, et al. 2014. "Integrated Genomic Sequencing Reveals Mutational Landscape of T-Cell Prolymphocytic Leukemia." *Blood* 124 (9): 1460–72. <https://doi.org/10.1182/blood-2014-03-559542>.
- Kim, Philip M., and Bruce Tidor. 2003. "Subsystem Identification Through Dimensionality Reduction of Large-Scale Gene Expression Data." *Genome Research* 13 (7): 1706–18. <https://doi.org/10.1101/gr.903503>.
- Kim, Tae-Kyung, Martin Hemberg, Jesse M. Gray, Allen M. Costa, Daniel M. Bear, Jing Wu, David A. Harmin, et al. 2010. "Widespread Transcription at Neuronal Activity-Regulated Enhancers." *Nature* 465 (7295): 182–87. <https://doi.org/10.1038/nature09033>.
- Kim, Youn H., Howard L. Liu, Serena Mraz-Gernhard, Anna Varghese, and Richard T. Hoppe. 2003. "Long-Term Outcome of 525 Patients With Mycosis Fungoides and Sézary Syndrome: Clinical Prognostic Factors and Risk for Disease Progression." *Archives of Dermatology* 139 (7): 857–66. <https://doi.org/10.1001/archderm.139.7.857>.
- Kinney, M. C., R. D. Collins, J. P. Greer, J. A. Whitlock, N. Sioutos, and M. E. Kadin. 1993. "A Small-Cell-Predominant Variant of Primary Ki-1 (CD30)+ T-Cell Lymphoma." *The American Journal of Surgical Pathology* 17 (9): 859–68. <https://doi.org/10.1097/0000478-199309000-00001>.
- Klemke, C. D., N. Booken, C. Weiss, J. P. Nicolay, S. Goerdts, M. Felcht, C. Géraud, et al. 2015. "Histopathological and Immunophenotypical Criteria for the Diagnosis of Sézary

- Syndrome in Differentiation from Other Erythrodermic Skin Diseases: A European Organisation for Research and Treatment of Cancer (EORTC) Cutaneous Lymphoma Task Force Study of 97 Cases." *The British Journal of Dermatology* 173 (1): 93–105. <https://doi.org/10.1111/bjd.13832>.
- Koch, Meghan A., Gladys Tucker-Heard, Nikole R. Perdue, Justin R. Killebrew, Kevin B. Urdahl, and Daniel J. Campbell. 2009. "The Transcription Factor T-Bet Controls Regulatory T Cell Homeostasis and Function during Type 1 Inflammation." *Nature Immunology* 10 (6): 595–602. <https://doi.org/10.1038/ni.1731>.
- Koelsche, Christian, Daniel Schrimpf, Damian Stichel, Martin Sill, Felix Sahm, David E. Reuss, Mirjam Blattner, et al. 2021. "Sarcoma Classification by DNA Methylation Profiling." *Nature Communications* 12 (1): 498. <https://doi.org/10.1038/s41467-020-20603-4>.
- Kogure, Yasunori, Takuro Kameda, Junji Koya, Makoto Yoshimitsu, Kisato Nosaka, Jun-ichirou Yasunaga, Yoshitaka Imaizumi, et al. 2022. "Whole-Genome Landscape of Adult T-Cell Leukemia/Lymphoma." *Blood* 139 (7): 967–82. <https://doi.org/10.1182/blood.2021013568>.
- Koskela, Hanna L.M., Samuli Eldfors, Pekka Ellonen, Arjan J. van Adrichem, Heikki Kuusanmäki, Emma I. Andersson, Sonja Lagström, et al. 2012. "Somatic STAT3 Mutations in Large Granular Lymphocytic Leukemia." *The New England Journal of Medicine* 366 (20): 1905–13. <https://doi.org/10.1056/NEJMoa1114885>.
- Kossenkov, Andrew V., and Michael F. Ochs. 2009. "Matrix Factorization for Recovery of Biological Processes from Microarray Data." *Methods in Enzymology* 467: 59–77. [https://doi.org/10.1016/S0076-6879\(09\)67003-8](https://doi.org/10.1016/S0076-6879(09)67003-8).
- Kouzarides, Tony. 2007. "Chromatin Modifications and Their Function." *Cell* 128 (4): 693–705. <https://doi.org/10.1016/j.cell.2007.02.005>.
- Krenacs, Laszlo, Patrick Schaerli, Gyongyi Kis, and Eniko Bagdi. 2006. "Phenotype of Neoplastic Cells in Angioimmunoblastic T-Cell Lymphoma Is Consistent with Activated Follicular B Helper T Cells." *Blood* 108 (3): 1110–11. <https://doi.org/10.1182/blood-2006-01-0394>.
- Kristjánisdóttir, Katla, Alexis Dziubek, Hyun Min Kang, and Hojoong Kwak. 2020. "Population-Scale Study of eRNA Transcription Reveals Bipartite Functional Enhancer Architecture." *Nature Communications* 11 (1): 5963. <https://doi.org/10.1038/s41467-020-19829-z>.
- Kruppa, Jochen, Yufeng Liu, Hans-Christian Diener, Theresa Holste, Christian Weimar, Inke R. König, and Andreas Ziegler. 2014. "Probability Estimation with Machine Learning Methods for Dichotomous and Multicategory Outcome: Applications." *Biometrical Journal* 56 (4): 564–83. <https://doi.org/10.1002/bimj.201300077>.
- Lai, Anne Y., and Paul A. Wade. 2011. "Cancer Biology and NuRD: A Multifaceted Chromatin Remodelling Complex." *Nature Reviews. Cancer* 11 (8): 588–96. <https://doi.org/10.1038/nrc3091>.
- Lamant, L., N. Dastugue, K. Pulford, G. Delsol, and B. Mariamé. 1999. "A New Fusion Gene TPM3-ALK in Anaplastic Large Cell Lymphoma Created by a (1;2)(Q25;P23) Translocation." *Blood* 93 (9): 3088–95.
- Lamy, T., J. H. Liu, T. H. Landowski, W. S. Dalton, and T. P. Loughran. 1998. "Dysregulation of CD95/CD95 Ligand-Apoptotic Pathway in CD3(+) Large Granular Lymphocyte Leukemia." *Blood* 92 (12): 4771–77.

- Lamy, Thierry, Aline Moignet, and Thomas P. Loughran Jr. 2017. "LGL Leukemia: From Pathogenesis to Treatment." *Blood* 129 (9): 1082–94. <https://doi.org/10.1182/blood-2016-08-692590>.
- Lando, Z., P. Sarin, M. Megson, W. C. Greene, T. A. Waldman, R. C. Gallo, and S. Broder. 1983. "Association of Human T-Cell Leukaemia/Lymphoma Virus with the Tac Antigen Marker for the Human T-Cell Growth Factor Receptor." *Nature* 305 (5936): 733–36. <https://doi.org/10.1038/305733a0>.
- Laurent, Camille, Marine Baron, Nadia Amara, Corinne Haioun, Mylène Dandoit, Marc Maynadié, Marie Parrens, et al. 2017. "Impact of Expert Pathologic Review of Lymphoma Diagnosis: Study of Patients From the French Lymphopath Network." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 35 (18): 2008–17. <https://doi.org/10.1200/JCO.2016.71.2083>.
- Laurini, Javier A., Anamarija M. Perry, Eugene Boilesen, Jacques Diebold, Kenneth A. MacLennan, H. Konrad Müller-Hermelink, Bharat N. Nathwani, James O. Armitage, and Dennis D. Weisenburger. 2012. "Classification of Non-Hodgkin Lymphoma in Central and South America: A Review of 1028 Cases." *Blood* 120 (24): 4795–4801. <https://doi.org/10.1182/blood-2012-07-440073>.
- Lee, Gap Ryol, Patrick E. Fields, Thomas J. Griffin, and Richard A. Flavell. 2003. "Regulation of the Th2 Cytokine Locus by a Locus Control Region." *Immunity* 19 (1): 145–53. [https://doi.org/10.1016/S1074-7613\(03\)00179-1](https://doi.org/10.1016/S1074-7613(03)00179-1).
- Lee, Jeeyun, Yeon Hee Park, Won Seog Kim, Seung-Sook Lee, Baek-Yeol Ryoo, Sung Hyun Yang, Keon Woo Park, et al. 2005. "Extranodal Nasal Type NK/T-Cell Lymphoma: Elucidating Clinical Prognostic Factors for Risk-Based Stratification of Therapy." *European Journal of Cancer* 41 (10): 1402–8. <https://doi.org/10.1016/j.ejca.2005.03.010>.
- Lee, Robin D., Todd P. Knutson, Sarah A. Munro, Jeffrey T. Miller, Lynn M. Heltemes-Harris, Charles G. Mullighan, Kristen Jepsen, and Michael A. Farrar. 2022. "Nuclear Corepressors NCOR1/NCOR2 Regulate B Cell Development, Maintain Genomic Integrity and Prevent Transformation." *Nature Immunology* 23 (12): 1763–76. <https://doi.org/10.1038/s41590-022-01343-7>.
- Leek, Jeffrey T., W. Evan Johnson, Hilary S. Parker, Andrew E. Jaffe, and John D. Storey. 2012. "The Sva Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments." *Bioinformatics (Oxford, England)* 28 (6): 882–83. <https://doi.org/10.1093/bioinformatics/bts034>.
- Lemonnier, François, Lucile Couronné, Marie Parrens, Jean-Philippe Jaïs, Marion Travert, Laurence Lamant, Olivier Tournillac, et al. 2012. "Recurrent TET2 Mutations in Peripheral T-Cell Lymphomas Correlate with TFH-like Features and Adverse Clinical Parameters." *Blood* 120 (7): 1466–69. <https://doi.org/10.1182/blood-2012-02-408542>.
- Lenhard, Boris, Albin Sandelin, and Piero Carninci. 2012. "Metazoan Promoters: Emerging Characteristics and Insights into Transcriptional Regulation." *Nature Reviews. Genetics* 13 (4): 233–45. <https://doi.org/10.1038/nrg3163>.
- Lettice, Laura A., Simon J.H. Heaney, Lorna A. Purdie, Li Li, Philippe de Beer, Ben A. Oostra, Debbie Goode, Greg Elgar, Robert E. Hill, and Esther de Graaff. 2003. "A Long-Range Shh Enhancer Regulates Expression in the Developing Limb and Fin and Is Associated with Preaxial Polydactyly." *Human Molecular Genetics* 12 (14): 1725–35. <https://doi.org/10.1093/hmg/ddg180>.

- Leval, Laurence de, David S. Rickman, Caroline Thielen, Aurélien de Reynies, Yen-Lin Huang, Georges Delsol, Laurence Lamant, et al. 2007. "The Gene Expression Profile of Nodal Peripheral T-Cell Lymphoma Demonstrates a Molecular Link between Angioimmunoblastic T-Cell Lymphoma (AITL) and Follicular Helper T (TFH) Cells." *Blood* 109 (11): 4952–63. <https://doi.org/10.1182/blood-2006-10-055145>.
- Lewis, Natasha E., Kseniya Petrova-Drus, Sarah Huet, Zachary D. Epstein-Peterson, Qi Gao, Allison E. Sigler, Jeeyeon Baik, et al. 2020. "Clonal Hematopoiesis in Angioimmunoblastic T-Cell Lymphoma with Divergent Evolution to Myeloid Neoplasms." *Blood Advances* 4 (10): 2261–71. <https://doi.org/10.1182/bloodadvances.2020001636>.
- Li, Shaoying, Xiaoli Feng, Ting Li, Shuang Zhang, Zhuang Zuo, Pei Lin, Sergej Konoplev, et al. 2013. "Extranodal NK/T-Cell Lymphoma, Nasal Type: A Report of 73 Cases at MD Anderson Cancer Center." *The American Journal of Surgical Pathology* 37 (1): 14–23. <https://doi.org/10.1097/PAS.0b013e31826731b5>.
- Lieberman-Aiden, Erez, Nynke L. van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragozy, Agnes Telling, Ido Amit, et al. 2009. "Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome." *Science (New York, N.Y.)* 326 (5950): 289–93. <https://doi.org/10.1126/science.1181369>.
- Lima, Margarida, Julia Almeida, Maria Dos Anjos Teixeira, Maria del Carmen Alguero Md, Ana Helena Santos, Ana Balanzategui, Maria Luís Queirós, et al. 2003. "TCRalpha+beta+/CD4+ Large Granular Lymphocytosis: A New Clonal T-Cell Lymphoproliferative Disorder." *The American Journal of Pathology* 163 (2): 763–71. [https://doi.org/10.1016/s0002-9440\(10\)63703-0](https://doi.org/10.1016/s0002-9440(10)63703-0).
- Listgarten, Jennifer, Carl Kadie, Eric E. Schadt, and David Heckerman. 2010. "Correction for Hidden Confounders in the Genetic Analysis of Gene Expression." *Proceedings of the National Academy of Sciences of the United States of America* 107 (38): 16465–70. <https://doi.org/10.1073/pnas.1002425107>.
- Liu, Yu, John Easton, Ying Shao, Jamie Maciaszek, Zhaoming Wang, Mark R. Wilkinson, Kelly McCastlain, et al. 2017. "The Genomic Landscape of Pediatric and Young Adult T-Lineage Acute Lymphoblastic Leukemia." *Nature Genetics* 49 (8): 1211–18. <https://doi.org/10.1038/ng.3909>.
- Luca, Bogdan A., Chloé B. Steen, Magdalena Matusiak, Armon Azizi, Sushama Varma, Chunfang Zhu, Joanna Przybyl, et al. 2021. "Atlas of Clinically Distinct Cell States and Ecosystems across Human Solid Tumors." *Cell* 184 (21): 5482-5496.e28. <https://doi.org/10.1016/j.cell.2021.09.014>.
- Luchtel, Rebecca A., Surendra Dasari, Naoki Oishi, Martin Bjerregård Pedersen, Guangzhen Hu, Karen L. Rech, Rhett P. Ketterling, et al. 2018. "Molecular Profiling Reveals Immunogenic Cues in Anaplastic Large Cell Lymphomas with DUSP22 Rearrangements." *Blood* 132 (13): 1386–98. <https://doi.org/10.1182/blood-2018-03-838524>.
- Luo, Chongyuan, Petra Hajkova, and Joseph R. Ecker. 2018. "Dynamic DNA Methylation: In the Right Place at the Right Time." *Science (New York, N.Y.)* 361 (6409): 1336–40. <https://doi.org/10.1126/science.aat6806>.
- Luoma, Adrienne M., Caitlin D. Castro, Toufic Mayassi, Leslie A. Bembinster, Li Bai, Damien Picard, Brian Anderson, et al. 2013. "Crystal Structure of Vδ1 T Cell Receptor in Complex with CD1d-Sulfatide Shows MHC-like Recognition of a Self-Lipid by Human

- $\gamma\delta$  T Cells." *Immunity* 39 (6): 1032–42. <https://doi.org/10.1016/j.immuni.2013.11.001>.
- Ma, Guangyong, Jun-Ichirou Yasunaga, and Masao Matsuoka. 2016. "Multifaceted Functions and Roles of HBZ in HTLV-1 Pathogenesis." *Retrovirology* 13 (March): 16. <https://doi.org/10.1186/s12977-016-0249-x>.
- Macqueen, Macqueen. 1967. *Some Methods for Classification and Analysis of Multivariate Observations*. Vol. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press.
- Malamut, Georgia, Olivia Chandesris, Virginie Verkarre, Bertrand Meresse, Céline Callens, Elizabeth Macintyre, Yoram Bouhnik, et al. 2013. "Enteropathy Associated T Cell Lymphoma in Celiac Disease: A Large Retrospective Study." *Digestive and Liver Disease: Official Journal of the Italian Society of Gastroenterology and the Italian Association for the Study of the Liver* 45 (5): 377–84. <https://doi.org/10.1016/j.dld.2012.12.001>.
- Malcolm, Tim I. M., Patrick Villarese, Camilla J. Fairbairn, Laurence Lamant, Amélie Trinquand, C. Elizabeth Hook, G. A. Amos Burke, et al. 2016. "Anaplastic Large Cell Lymphoma Arises in Thymocytes and Requires Transient TCR Expression for Thymic Egress." *Nature Communications* 7 (1): 10087. <https://doi.org/10.1038/ncomms10087>.
- Malik, Sohail, and Robert G. Roeder. 2010. "The Metazoan Mediator Co-Activator Complex as an Integrative Hub for Transcriptional Regulation." *Nature Reviews. Genetics* 11 (11): 761–72. <https://doi.org/10.1038/nrg2901>.
- Marçais, Ambroise, Ludovic Lhermitte, Maria Artesi, Cécile Laurent, Keith Durkin, Vincent Hahaut, Nicolas Rosewick, et al. 2021. "Targeted Deep Sequencing Reveals Clonal and Subclonal Mutational Signatures in Adult T-Cell Leukemia/Lymphoma and Defines an Unfavorable Indolent Subtype." *Leukemia* 35 (3): 764–76. <https://doi.org/10.1038/s41375-020-0900-3>.
- Margueron, Raphaël, and Danny Reinberg. 2011. "The Polycomb Complex PRC2 and Its Mark in Life." *Nature* 469 (7330): 343–49. <https://doi.org/10.1038/nature09784>.
- Maros, Máté E., David Capper, David T. W. Jones, Volker Hovestadt, Andreas von Deimling, Stefan M. Pfister, Axel Benner, Manuela Zucknick, and Martin Sill. 2020. "Machine Learning Workflows to Estimate Class Probabilities for Precision Cancer Diagnostics on DNA Methylation Microarray Data." *Nature Protocols* 15 (2): 479–512. <https://doi.org/10.1038/s41596-019-0251-6>.
- Martínez-Iglesias, Olaia A., Elvira Alonso-Merino, Sara Gómez-Rey, Juan Pedro Velasco-Martín, Rosa Martín Orozco, Enrique Luengo, Rosa García Martín, et al. 2016. "Autoregulatory Loop of Nuclear Corepressor 1 Expression Controls Invasion, Tumor Growth, and Metastasis." *Proceedings of the National Academy of Sciences* 113 (3): E328–37. <https://doi.org/10.1073/pnas.1520469113>.
- Matutes, E., V. Brito-Babapulle, J. Swansbury, J. Ellis, R. Morilla, C. Dearden, A. Sempere, and D. Catovsky. 1991. "Clinical and Laboratory Features of 78 Cases of T-Prolymphocytic Leukemia." *Blood* 78 (12): 3269–74.
- Maura, Francesco, Luca Agnelli, Daniel Leongamornlert, Niccolò Bolli, Wing C. Chan, Anna Doderò, Cristiana Carniti, et al. 2019. "Integration of Transcriptional and Mutational Data Simplifies the Stratification of Peripheral T-Cell Lymphoma." *American Journal of Hematology* 94 (6): 628–34. <https://doi.org/10.1002/ajh.25450>.

- McKinney, Matthew, Andrea B. Moffitt, Philippe Gaulard, Marion Travert, Laurence De Leval, Alina Nicolae, Mark Raffeld, et al. 2017. "The Genetic Basis of Hepatosplenic T Cell Lymphoma." *Cancer Discovery* 7 (4): 369–79. <https://doi.org/10.1158/2159-8290.CD-16-0330>.
- Melenhorst, J. J., L. Sorbara, M. Kirby, N. F. Hensel, and A. J. Barrett. 2001. "Large Granular Lymphocyte Leukaemia Is Characterized by a Clonal T-Cell Receptor Rearrangement in Both Memory and Effector CD8(+) Lymphocyte Populations." *British Journal of Haematology* 112 (1): 189–94. <https://doi.org/10.1046/j.1365-2141.2001.02509.x>.
- Messi, Mara, Isabella Giacchetto, Kinya Nagata, Antonio Lanzavecchia, Gioacchino Natoli, and Federica Sallusto. 2003. "Memory and Flexibility of Cytokine Gene Expression as Separable Properties of Human TH1 and TH2 Lymphocytes." *Nature Immunology* 4 (1): 78–86. <https://doi.org/10.1038/ni872>.
- Mimitou, Eleni P., Caleb A. Lareau, Kelvin Y. Chen, Andre L. Zorzetto-Fernandes, Yusuke Takeshima, Wendy Luo, Tse-Shun Huang, et al. 2020. "Scalable, Multimodal Profiling of Chromatin Accessibility and Protein Levels in Single Cells." *bioRxiv*, September, 2020.09.08.286914. <https://doi.org/10.1101/2020.09.08.286914>.
- Misteli, Tom. 2020. "The Self-Organizing Genome: Principles of Genome Architecture and Function." *Cell* 183 (1): 28–45. <https://doi.org/10.1016/j.cell.2020.09.014>.
- Mitchell, Pamela J., and Robert Tjian. 1989. "Transcriptional Regulation in Mammalian Cells by Sequence-Specific DNA Binding Proteins." *Science* 245 (4916): 371–78. <https://doi.org/10.1126/science.2667136>.
- Mittal, Priya, and Charles W. M. Roberts. 2020. "The SWI/SNF Complex in Cancer — Biology, Biomarkers and Therapy." *Nature Reviews Clinical Oncology* 17 (7): 435–48. <https://doi.org/10.1038/s41571-020-0357-3>.
- Moffitt, Andrea B., Sarah L. Ondrejka, Matthew McKinney, Rachel E. Rempel, John R. Goodlad, Chun Huat Teh, Sirpa Leppä, et al. 2017. "Enteropathy-Associated T Cell Lymphoma Subtypes Are Characterized by Loss of Function of SETD2." *The Journal of Experimental Medicine* 214 (5): 1371–86. <https://doi.org/10.1084/jem.20160894>.
- Morris, Stephan W., Mark N. Kirstein, Marcus B. Valentine, Kristopher G. Dittmer, David N. Shapiro, David L. Saltman, and A. Thomas Look. 1994. "Fusion of a Kinase Gene, ALK, to a Nucleolar Protein Gene, NPM, in Non-Hodgkin's Lymphoma." *Science* 263 (5151): 1281–84. <https://doi.org/10.1126/science.8122112>.
- Mortaz, Ebrahim. 2020. "Imbalance Accuracy Metric for Model Selection in Multi-Class Imbalance Classification Problems." *Knowledge-Based Systems* 210 (December): 106490. <https://doi.org/10.1016/j.knosys.2020.106490>.
- Moskowitz, David M., David W. Zhang, Bin Hu, Sabine Le Saux, Rolando E. Yanes, Zhongde Ye, Jason D. Buenrostro, Cornelia M. Weyand, William J. Greenleaf, and Jörg J. Goronzy. 2017. "Epigenomics of Human CD8 T Cell Differentiation and Aging." *Science Immunology* 2 (8): eaag0192. <https://doi.org/10.1126/sciimmunol.aag0192>.
- Nagai, Yuya, Masahiro Kawahara, Masakatsu Hishizawa, Yayoi Shimazu, Noriko Sugino, Sumie Fujii, Norimitsu Kadowaki, and Akifumi Takaori-Kondo. 2015. "T Memory Stem Cells Are the Hierarchical Apex of Adult T-Cell Leukemia." *Blood* 125 (23): 3527–35. <https://doi.org/10.1182/blood-2014-10-607465>.
- Nagy, Peter L., Michael L. Cleary, Patrick O. Brown, and Jason D. Lieb. 2003. "Genomewide Demarcation of RNA Polymerase II Transcription Units Revealed by Physical Fractionation of Chromatin." *Proceedings of the National Academy of Sciences of the*

- Najidh, Safa, Cornelis P. Tensen, Alita J. van der Sluijs-Gelling, Cristina Teodosio, Davy Cats, Hailing Mei, Thomas B. Kuipers, et al. 2021. “Improved Sézary Cell Detection and Novel Insights into Immunophenotypic and Molecular Heterogeneity in Sézary Syndrome.” *Blood* 138 (24): 2539–54. <https://doi.org/10.1182/blood.2021012286>.
- Nassiri, Farshad, Ankur Chakravarthy, Shengrui Feng, Shu Yi Shen, Romina Nejad, Jeffrey A. Zuccato, Mathew R. Voisin, et al. 2020. “Detection and Discrimination of Intracranial Tumors Using Plasma Cell-Free DNA Methylomes.” *Nature Medicine* 26 (7): 1044–47. <https://doi.org/10.1038/s41591-020-0932-2>.
- Nazarov, Petr V., Anke K. Wienecke-Baldacchino, Andrei Zinovyev, Urszula Czerwińska, Arnaud Muller, Dorothée Nashan, Gunnar Dittmar, Francisco Azuaje, and Stephanie Kreis. 2019. “Deconvolution of Transcriptomes and miRNomes by Independent Component Analysis Provides Insights into Biological Processes and Clinical Outcomes of Melanoma Patients.” *BMC Medical Genomics* 12 (1): 132. <https://doi.org/10.1186/s12920-019-0578-4>.
- Newman, Aaron M., Chih Long Liu, Michael R. Green, Andrew J. Gentles, Weiguo Feng, Yue Xu, Chuong D. Hoang, Maximilian Diehn, and Ash A. Alizadeh. 2015. “Robust Enumeration of Cell Subsets from Tissue Expression Profiles.” *Nature Methods* 12 (5): 453–57. <https://doi.org/10.1038/nmeth.3337>.
- Nicolae, Alina, Justine Bouilly, Diane Lara, Virginie Fataccioli, François Lemonnier, Fanny Drieux, Marie Parrens, et al. 2022. “Nodal Cytotoxic Peripheral T-Cell Lymphoma Occurs Frequently in the Clinical Setting of Immunodysregulation and Is Associated with Recurrent Epigenetic Alterations.” *Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc* 35 (8): 1126–36. <https://doi.org/10.1038/s41379-022-01022-w>.
- Oberbeck, Sebastian, Alexandra Schrader, Kathrin Warner, Dennis Jungherz, Giuliano Crispantu, Jana von Jan, Markus Chmielewski, et al. 2020. “Non-Canonical Effector Functions of the T-Memory-like T-PLL Cell Are Shaped by Cooperative TCL1A and TCR Signaling.” *Blood*, September, blood.2019003348. <https://doi.org/10.1182/blood.2019003348>.
- Odejide, Oreofe, Oliver Weigert, Andrew A. Lane, Dan Toscano, Matthew A. Lunning, Nadja Kopp, Sunhee Kim, et al. 2014. “A Targeted Mutational Landscape of Angioimmunoblastic T-Cell Lymphoma.” *Blood* 123 (9): 1293–96. <https://doi.org/10.1182/blood-2013-10-531509>.
- Oishi, Naoki, Akira Satou, Masashi Miyaoka, Ichiro Kawashima, Takahiro Segawa, Kunio Miyake, Kunio Mochizuki, et al. 2023. “Genetic and Immunohistochemical Profiling of NK/T-Cell Lymphomas Reveals Prognostically Relevant BCOR-MYC Association.” *Blood Advances* 7 (1): 178–89. <https://doi.org/10.1182/bloodadvances.2022007541>.
- Olsen, Elise A., Youn H. Kim, Timothy M. Kuzel, Theresa R. Pacheco, Francine M. Foss, Sareeta Parker, Stanley R. Frankel, et al. 2007. “Phase IIb Multicenter Trial of Vorinostat in Patients with Persistent, Progressive, or Treatment Refractory Cutaneous T-Cell Lymphoma.” *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 25 (21): 3109–15. <https://doi.org/10.1200/JCO.2006.10.2434>.
- Onaindia, Arantza, Sonia González de Villambrosía, Lucía Prieto-Torres, Socorro M. Rodríguez-Pinilla, Santiago Montes-Moreno, Carmen González-Vela, and Miguel A. Piris. 2019. “DUSP22-Rearranged Anaplastic Lymphomas Are Characterized by

- Specific Morphological Features and a Lack of Cytotoxic and JAK/STAT Surrogate Markers." *Haematologica* 104 (4): e158–62. <https://doi.org/10.3324/haematol.2018.205880>.
- Ørom, Ulf Andersson, Thomas Derrien, Malte Beringer, Kiranmai Gumireddy, Alessandro Gardini, Giovanni Bussotti, Fan Lai, et al. 2010. "Long Noncoding RNAs with Enhancer-like Function in Human Cells." *Cell* 143 (1): 46–58. <https://doi.org/10.1016/j.cell.2010.09.001>.
- Palomero, Teresa, Lucile Couronné, Hossein Khiabani, Mi-Yeon Kim, Alberto Ambesi-Impiombato, Arianne Perez-Garcia, Zachary Carpenter, et al. 2014. "Recurrent Mutations in Epigenetic Regulators, RHOA and FYN Kinase in Peripheral T Cell Lymphomas." *Nature Genetics* 46 (2): 166–70. <https://doi.org/10.1038/ng.2873>.
- Pang, Baoxu, Jan H. van Weerd, Feija L. Hamoen, and Michael P. Snyder. 2023. "Identification of Non-Coding Silencer Elements and Their Regulation of Gene Expression." *Nature Reviews Molecular Cell Biology* 24 (6): 383–95. <https://doi.org/10.1038/s41580-022-00549-9>.
- Park, Joonhee, Jingyi Yang, Alexander T. Wenzel, Akshaya Ramachandran, Wung J. Lee, Jay C. Daniels, Juhyun Kim, et al. 2017a. "Genomic Analysis of 220 CTCLs Identifies a Novel Recurrent Gain-of-Function Alteration in RLTPR (p.Q575E)." *Blood* 130 (12): 1430–40. <https://doi.org/10.1182/blood-2017-02-768234>.
- . 2017b. "Genomic Analysis of 220 CTCLs Identifies a Novel Recurrent Gain-of-Function Alteration in RLTPR (p.Q575E)." *Blood* 130 (12): 1430–40. <https://doi.org/10.1182/blood-2017-02-768234>.
- Parrilla Castellar, Edgardo R., Elaine S. Jaffe, Jonathan W. Said, Steven H. Swerdlow, Rhett P. Ketterling, Ryan A. Knudson, Jagmohan S. Sidhu, et al. 2014. "ALK-Negative Anaplastic Large Cell Lymphoma Is a Genetically Heterogeneous Disease with Widely Disparate Clinical Outcomes." *Blood* 124 (9): 1473–80. <https://doi.org/10.1182/blood-2014-04-571091>.
- Pedersen, Martin Bjerregård, Stephen Jacques Hamilton-Dutoit, Knud Bendix, Rhett P. Ketterling, Patrick P. Bedroske, Ivy M. Luoma, Christopher A. Sattler, et al. 2017. "DUSP22 and TP63 Rearrangements Predict Outcome of ALK-Negative Anaplastic Large Cell Lymphoma: A Danish Cohort Study." *Blood* 130 (4): 554–57. <https://doi.org/10.1182/blood-2016-12-755496>.
- Pellicci, Daniel G., Hui-Fern Koay, and Stuart P. Berzins. 2020. "Thymic Development of Unconventional T Cells: How NKT Cells, MAIT Cells and  $\Gamma\delta$  T Cells Emerge." *Nature Reviews Immunology* 20 (12): 756–70. <https://doi.org/10.1038/s41577-020-0345-y>.
- Peng, Xianlu Laura, Richard A. Moffitt, Robert J. Torphy, Keith E. Volmar, and Jen Jen Yeh. 2019. "De Novo Compartment Deconvolution and Weight Estimation of Tumor Samples Using DECODER." *Nature Communications* 10 (1): 4729. <https://doi.org/10.1038/s41467-019-12517-7>.
- Phair, Robert D., Paola Scaffidi, Cem Elbi, Jaromíra Vecerová, Anup Dey, Keiko Ozato, David T. Brown, Gordon Hager, Michael Bustin, and Tom Misteli. 2004. "Global Nature of Dynamic Protein-Chromatin Interactions in Vivo: Three-Dimensional Genome Scanning and Dynamic Interaction Networks of Chromatin Proteins." *Molecular and Cellular Biology* 24 (14): 6393–6402. <https://doi.org/10.1128/MCB.24.14.6393-6402.2004>.
- Phelps, Michael P., Jenna N. Bailey, Terra Vleeshouwer-Neumann, and Eleanor Y. Chen. 2016. "CRISPR Screen Identifies the NCOR/HDAC3 Complex as a Major Suppressor of



- Differentiation in Rhabdomyosarcoma." *Proceedings of the National Academy of Sciences* 113 (52): 15090–95. <https://doi.org/10.1073/pnas.1610270114>.
- Piccaluga, Pier Paolo, Fabio Fuligni, Antonio De Leo, Clara Bertuzzi, Maura Rossi, Francesco Bacci, Elena Sabbatini, et al. 2013. "Molecular Profiling Improves Classification and Prognostication of Nodal Peripheral T-Cell Lymphomas: Results of a Phase III Diagnostic Accuracy Study." *Journal of Clinical Oncology* 31 (24): 3019–25. <https://doi.org/10.1200/JCO.2012.42.5611>.
- Pickles, Jessica C., Amy R. Fairchild, Thomas J. Stone, Lorelle Brownlee, Ashirwad Merve, Shireena A. Yasin, Aimee Avery, et al. 2020. "DNA Methylation-Based Profiling for Paediatric CNS Tumour Diagnosis and Treatment: A Population-Based Study." *The Lancet Child & Adolescent Health* 4 (2): 121–30. [https://doi.org/10.1016/S2352-4642\(19\)30342-6](https://doi.org/10.1016/S2352-4642(19)30342-6).
- Pittaluga, S., I. Wlodarska, K. Pulford, E. Campo, S. W. Morris, H. Van den Berghe, and C. De Wolf-Peeters. 1997. "The Monoclonal Antibody ALK1 Identifies a Distinct Morphological Subtype of Anaplastic Large Cell Lymphoma Associated with 2p23/ALK Rearrangements." *The American Journal of Pathology* 151 (2): 343–51.
- Pongpruttipan, Tawatchai, Sanya Sukpanichnant, Thamathorn Assanasen, Pongsak Wannakrairot, Paisarn Boonsakan, Wasana Kanoksil, Kanita Kayasut, et al. 2012. "Extranodal NK/T-Cell Lymphoma, Nasal Type, Includes Cases of Natural Killer Cell and A $\beta$ ,  $\Gamma\delta$ , and A $\beta$ / $\Gamma\delta$  T-Cell Origin: A Comprehensive Clinicopathologic and Phenotypic Study." *The American Journal of Surgical Pathology* 36 (4): 481–99. <https://doi.org/10.1097/PAS.0b013e31824433d8>.
- Pro, Barbara, Pamela Allen, and Amir Behdad. 2020. "Hepatosplenic T-Cell Lymphoma: A Rare but Challenging Entity." *Blood* 136 (18): 2018–26. <https://doi.org/10.1182/blood.2019004118>.
- Przybylski, G. K., H. Wu, W. R. Macon, J. Finan, D. G. Leonard, R. E. Felgar, J. A. DiGiuseppe, et al. 2000. "Hepatosplenic and Subcutaneous Panniculitis-like Gamma/Delta T Cell Lymphomas Are Derived from Different Vdelta Subsets of Gamma/Delta T Lymphocytes." *The Journal of Molecular Diagnostics: JMD* 2 (1): 11–19. [https://doi.org/10.1016/s1525-1578\(10\)60610-1](https://doi.org/10.1016/s1525-1578(10)60610-1).
- Quivoron, Cyril, Lucile Couronné, Véronique Della Valle, Cécile K. Lopez, Isabelle Plo, Oriane Wagner-Ballon, Marcio Do Cruzeiro, et al. 2011. "TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis." *Cancer Cell* 20 (1): 25–38. <https://doi.org/10.1016/j.ccr.2011.06.003>.
- Rada-Iglesias, Alvaro, Ruchi Bajpai, Tomek Swigut, Samantha A. Brugmann, Ryan A. Flynn, and Joanna Wysocka. 2011. "A Unique Chromatin Signature Uncovers Early Developmental Enhancers in Humans." *Nature* 470 (7333): 279–83. <https://doi.org/10.1038/nature09692>.
- Rahmani, Elinor, Regev Schweiger, Brooke Rhead, Lindsey A. Criswell, Lisa F. Barcellos, Eleazar Eskin, Saharon Rosset, Sriram Sankararaman, and Eran Halperin. 2019. "Cell-Type-Specific Resolution Epigenetics without the Need for Cell Sorting or Single-Cell Biology." *Nature Communications* 10 (1): 3417. <https://doi.org/10.1038/s41467-019-11052-9>.
- Ree, H. J., M. E. Kadin, M. Kikuchi, Y. H. Ko, J. Suzumiya, and J. H. Go. 1999. "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." *Human Pathology* 30 (4): 403–11. [https://doi.org/10.1016/s0046-8177\(99\)90115-6](https://doi.org/10.1016/s0046-8177(99)90115-6).
- Reinhardt, R. Lee, Hong-Erh Liang, and Richard M. Locksley. 2009. "Cytokine-Secreting Follicular T Cells Shape the Antibody Repertoire." *Nature Immunology* 10 (4): 385–93. <https://doi.org/10.1038/ni.1715>.
- Rendeiro, André F., Christian Schmidl, Jonathan C. Strefford, Renata Walewska, Zadie Davis, Matthias Farlik, David Oscier, and Christoph Bock. 2016. "Chromatin Accessibility Maps of Chronic Lymphocytic Leukaemia Identify Subtype-Specific Epigenome Signatures and Transcription Regulatory Networks." *Nature Communications* 7 (1): 11938. <https://doi.org/10.1038/ncomms11938>.
- Roberti, Annalisa, Maria Pamela Dobay, Bettina Bisig, David Vallois, Cloé Boéchat, Eviropidis Lanitis, Brigitte Bouchindhomme, et al. 2016. "Type II Enteropathy-Associated T-Cell Lymphoma Features a Unique Genomic Profile with Highly Recurrent SETD2 Alterations." *Nature Communications* 7 (1): 12602. <https://doi.org/10.1038/ncomms12602>.
- Rodríguez, Marta, Ruth Alonso-Alonso, Laura Tomás-Roca, Socorro M. Rodríguez-Pinilla, Rebeca Manso-Alonso, Laura Cereceda, Jennifer Borregón, et al. 2021. "Peripheral T-Cell Lymphoma: Molecular Profiling Recognizes Subclasses and Identifies Prognostic Markers." *Blood Advances* 5 (24): 5588–98. <https://doi.org/10.1182/bloodadvances.2021005171>.
- Rodríguez-Pinilla, Socorro María, Lidia Atienza, Cristina Murillo, Alberto Pérez-Rodríguez, Santiago Montes-Moreno, Giovanna Roncador, Carlos Pérez-Seoane, Purificación Domínguez, Francisca I. Camacho, and Miguel A. Piris. 2008. "Peripheral T-Cell Lymphoma with Follicular T-Cell Markers." *The American Journal of Surgical Pathology* 32 (12): 1787–99. <https://doi.org/10.1097/PAS.0b013e31817f123e>.
- Roelens, Marie, Marc Delord, Caroline Ram-Wolff, Anne Marie-Cardine, Antonio Alberdi, Guitta Maki, Laurence Homyrda, et al. 2017. "Circulating and Skin-Derived Sézary Cells: Clonal but with Phenotypic Plasticity." *Blood* 130 (12): 1468–71. <https://doi.org/10.1182/blood-2017-03-772996>.
- Ruben, Steven, Harry Poteat, Tse-Hua Tan, Kyoshi Kawakami, Robert Roeder, William Haseltine, and Craig A. Rosen. 1988. "Cellular Transcription Factors and Regulation of IL-2 Receptor Gene Expression by HTLV-I Tax Gene Product." *Science* 241 (4861): 89–92. <https://doi.org/10.1126/science.2838905>.
- Sabo, Peter J., Michael S. Kuehn, Robert Thurman, Brett E. Johnson, Ericka M. Johnson, Hua Cao, Man Yu, et al. 2006. "Genome-Scale Mapping of DNase I Sensitivity in Vivo Using Tiling DNA Microarrays." *Nature Methods* 3 (7): 511–18. <https://doi.org/10.1038/nmeth890>.
- Saeyns, Yvan, Iñaki Inza, and Pedro Larrañaga. 2007. "A Review of Feature Selection Techniques in Bioinformatics." *Bioinformatics* 23 (19): 2507–17. <https://doi.org/10.1093/bioinformatics/btm344>.
- Sakata-Yanagimoto, Mamiko, Terukazu Enami, Kenichi Yoshida, Yuichi Shiraishi, Ryohei Ishii, Yasuyuki Miyake, Hideharu Muto, et al. 2014. "Somatic RHOA Mutation in Angioimmunoblastic T Cell Lymphoma." *Nature Genetics* 46 (2): 171–75. <https://doi.org/10.1038/ng.2872>.
- Sallusto, Federica, Jens Geginat, and Antonio Lanzavecchia. 2004. "Central Memory and Effector Memory T Cell Subsets: Function, Generation, and Maintenance." *Annual*

- Review of Immunology* 22 (1): 745–63.  
<https://doi.org/10.1146/annurev.immunol.22.012703.104702>.
- Sathaliyawala, Taheri, Masaru Kubota, Naomi Yudanin, Damian Turner, Philip Camp, Joseph J. C. Thome, Kara L. Bickham, et al. 2013. “Distribution and Compartmentalization of Human Circulating and Tissue-Resident Memory T Cell Subsets.” *Immunity* 38 (1): 187–97. <https://doi.org/10.1016/j.immuni.2012.09.020>.
- Satpathy, Ansuman T., Naresha Saligrama, Jason D. Buenrostro, Yuning Wei, Beijing Wu, Adam J. Rubin, Jeffrey M. Granja, et al. 2018. “Transcript-Indexed ATAC-Seq for Precision Immune Profiling.” *Nature Medicine* 24 (5): 580–90. <https://doi.org/10.1038/s41591-018-0008-8>.
- Scarisbrick, Julia J., H. Miles Prince, Maarten H. Vermeer, Pietro Quaglino, Steven Horwitz, Pierluigi Porcu, Rudolf Stadler, et al. 2015. “Cutaneous Lymphoma International Consortium Study of Outcome in Advanced Stages of Mycosis Fungoides and Sézary Syndrome: Effect of Specific Prognostic Markers on Survival and Development of a Prognostic Model.” *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 33 (32): 3766–73. <https://doi.org/10.1200/JCO.2015.61.7142>.
- Schep, Alicia N., Beijing Wu, Jason D. Buenrostro, and William J. Greenleaf. 2017. “chromVAR: Inferring Transcription Factor-Associated Accessibility from Single-Cell Epigenomic Data.” *Nature Methods* 14 (10): 975–78. <https://doi.org/10.1038/nmeth.4401>.
- Scott, David W., George W. Wright, P. Mickey Williams, Chih-Jian Lih, William Walsh, Elaine S. Jaffe, Andreas Rosenwald, et al. 2014. “Determining Cell-of-Origin Subtypes of Diffuse Large B-Cell Lymphoma Using Gene Expression in Formalin-Fixed Paraffin-Embedded Tissue.” *Blood* 123 (8): 1214–17. <https://doi.org/10.1182/blood-2013-11-536433>.
- Seifert, Marc, Ludger Sellmann, Johannes Bloehdorn, Frederik Wein, Stephan Stilgenbauer, Jan Dürig, and Ralf Küppers. 2012. “Cellular Origin and Pathophysiology of Chronic Lymphocytic Leukemia.” *The Journal of Experimental Medicine* 209 (12): 2183–98. <https://doi.org/10.1084/jem.20120833>.
- Shen-Orr, Shai S., Robert Tibshirani, Purvesh Khatri, Dale L. Bodian, Frank Staedtler, Nicholas M. Perry, Trevor Hastie, Minnie M. Sarwal, Mark M. Davis, and Atul J. Butte. 2010. “Cell Type–Specific Gene Expression Differences in Complex Tissues.” *Nature Methods* 7 (4): 287–89. <https://doi.org/10.1038/nmeth.1439>.
- Shih, Han-Yu, Giuseppe Sciumè, Yohei Mikami, Liying Guo, Hong-Wei Sun, Stephen R. Brooks, Joseph F. Urban, Fred P. Davis, Yuka Kanno, and John J. O’Shea. 2016. “Developmental Acquisition of Regulomes Underlies Innate Lymphoid Cell Functionality.” *Cell* 165 (5): 1120–33. <https://doi.org/10.1016/j.cell.2016.04.029>.
- Shopland, Lindsay S., Christopher R. Lynch, Kevin A. Peterson, Kathleen Thornton, Nick Kepper, Johann von Hase, Stefan Stein, et al. 2006. “Folding and Organization of a Contiguous Chromosome Region According to the Gene Distribution Pattern in Primary Genomic Sequence.” *Journal of Cell Biology* 174 (1): 27–38. <https://doi.org/10.1083/jcb.200603083>.
- Sibon, David, Bettina Bisig, Christophe Bonnet, Elsa Poullot, Emmanuel Bachy, Doriane Cavalieri, Virginie Fataccioli, et al. 2023. “ALK-Negative Anaplastic Large Cell Lymphoma with DUSP22 Rearrangement Has Distinctive Disease Characteristics with Better Progression-Free Survival: A LYSA Study.” *Haematologica* 108 (6): 1590–1603. <https://doi.org/10.3324/haematol.2022.281442>.

- Sibon, David, Marion Fournier, Josette Brière, Laurence Lamant, Corinne Haioun, Bertrand Coiffier, Serge Bologna, et al. 2012. "Long-Term Outcome of Adults with Systemic Anaplastic Large-Cell Lymphoma Treated within the Groupe d'Etude Des Lymphomes de l'Adulte Trials." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 30 (32): 3939–46. <https://doi.org/10.1200/JCO.2012.42.2345>.
- Sibon, David, Sherine Khater, Julie Bruneau, Chantal Brouzes, Ludovic Lhermitte, Thierry Jo Molina, Guillaume Cartron, et al. 2021. "The Eat1-001 Trial: Results of a Phase 2 Study of Brentuximab Vedotin and CHP Followed By Consolidation with High-Dose Therapy - Autologous Stem-Cell Transplantation (HDT-ASCT) in the Frontline Treatment of Patients with Enteropathy-Associated T-Cell Lymphoma." *Blood* 138 (Supplement 1): 136. <https://doi.org/10.1182/blood-2021-153709>.
- Sibon, David, Dinh-Phong Nguyen, Norbert Schmitz, Ritsuro Suzuki, Andrew L. Feldman, Rémy Gressin, Laurence Lamant, et al. 2019. "ALK-Positive Anaplastic Large-Cell Lymphoma in Adults: An Individual Patient Data Pooled Analysis of 263 Patients." *Haematologica* 104 (12): e562–65. <https://doi.org/10.3324/haematol.2018.213512>.
- Silva Almeida, Ana Carolina da, Francesco Abate, Hossein Khiabani, Estela Martinez-Escala, Joan Guitart, Cornelis P. Tensen, Maarten H. Vermeer, Raul Rabadan, Adolfo Ferrando, and Teresa Palomero. 2015. "The Mutational Landscape of Cutaneous T Cell Lymphoma and Sézary Syndrome." *Nature Genetics* 47 (12): 1465–70. <https://doi.org/10.1038/ng.3442>.
- Simon, Richard. 2014. "Class Probability Estimation for Medical Studies." *Biometrical Journal. Biometrische Zeitschrift* 56 (4): 597–600. <https://doi.org/10.1002/bimj.201300296>.
- Slattery, Matthew, Todd Riley, Peng Liu, Namiko Abe, Pilar Gomez-Alcala, Iris Dror, Tianyin Zhou, et al. 2011. "Cofactor Binding Evokes Latent Differences in DNA Binding Specificity between Hox Proteins." *Cell* 147 (6): 1270–82. <https://doi.org/10.1016/j.cell.2011.10.053>.
- Song, Yuqin, Yanyan Liu, Zhi-Ming Li, Lanfang Li, Hang Su, Zhengming Jin, Xuelan Zuo, et al. 2022. "SHR2554, an EZH2 Inhibitor, in Relapsed or Refractory Mature Lymphoid Neoplasms: A First-in-Human, Dose-Escalation, Dose-Expansion, and Clinical Expansion Phase 1 Trial." *The Lancet Haematology* 9 (7): e493–503. [https://doi.org/10.1016/S2352-3026\(22\)00134-X](https://doi.org/10.1016/S2352-3026(22)00134-X).
- Statello, Luisa, Chun-Jie Guo, Ling-Ling Chen, and Maite Huarte. 2021. "Gene Regulation by Long Non-Coding RNAs and Its Biological Functions." *Nature Reviews Molecular Cell Biology* 22 (2): 96–118. <https://doi.org/10.1038/s41580-020-00315-9>.
- Statnikov, Alexander, Lily Wang, and Constantin F. Aliferis. 2008. "A Comprehensive Comparison of Random Forests and Support Vector Machines for Microarray-Based Cancer Classification." *BMC Bioinformatics* 9 (1): 1–10. <https://doi.org/10.1186/1471-2105-9-319>.
- Stax, Annelein M., Jessica Tuengel, Enrico Girardi, Naoki Kitano, Lenka L. Allan, Victor Liu, Dongjun Zheng, et al. 2017. "Autoreactivity to Sulfatide by Human Invariant NKT Cells." *Journal of Immunology (Baltimore, Md.: 1950)* 199 (1): 97–106. <https://doi.org/10.4049/jimmunol.1601976>.
- Stolz, Valentina, Ramona Rica, Ci Zhu, Teresa Preglej, Patricia Hamminger, Daniela Hainberger, Marlis Alteneder, et al. 2022. "Nuclear Receptor Corepressor 1 Controls Regulatory T Cell Subset Differentiation and Effector Function." bioRxiv. <https://doi.org/10.1101/2022.03.24.485609>.

- Takeshita, M., M. Akamatsu, K. Ohshima, S. Kobari, M. Kikuchi, J. Suzumiya, N. Uike, and T. Okamura. 1995. "CD30 (Ki-1) Expression in Adult T-Cell Leukaemia/Lymphoma Is Associated with Distinctive Immunohistological and Clinical Characteristics." *Histopathology* 26 (6): 539–46. <https://doi.org/10.1111/j.1365-2559.1995.tb00272.x>.
- Tamaki, Tomoko, Kennosuke Karube, Shugo Sakihama, Yuma Tsuruta, Ryoko Awazawa, Masaki Hayashi, Norihiro Nakada, et al. 2023. "A Comprehensive Study of the Immunophenotype and Its Clinicopathologic Significance in Adult T-Cell Leukemia/Lymphoma." *Modern Pathology* 36 (8): 100169. <https://doi.org/10.1016/j.modpat.2023.100169>.
- Tan, Benji Jy, Kenji Sugata, Omnia Reda, Misaki Matsuo, Kyosuke Uchiyama, Paola Miyazato, Vincent Hahaut, et al. 2021. "HTLV-1 Infection Promotes Excessive T Cell Activation and Transformation into Adult T Cell Leukemia/Lymphoma." *The Journal of Clinical Investigation* 131 (24): e150472. <https://doi.org/10.1172/JCI150472>.
- Tan, S-Y, S-S Chuang, T Tang, L Tan, Y-H Ko, K-L Chuah, S-B Ng, et al. 2013. "Type II EATL (Epitheliotropic Intestinal T-Cell Lymphoma): A Neoplasm of Intra-Epithelial T-Cells with Predominant CD8 $\alpha\alpha$  Phenotype." *Leukemia* 27 (8): 1688–96. <https://doi.org/10.1038/leu.2013.41>.
- Tarbell, Evan D, and Tao Liu. 2019. "HMMRATAC: A Hidden Markov Modeler for ATAC-Seq." *Nucleic Acids Research* 47 (16): e91. <https://doi.org/10.1093/nar/gkz533>.
- Taylor, A. M., J. A. Metcalfe, J. Thick, and Y. F. Mak. 1996. "Leukemia and Lymphoma in Ataxia Telangiectasia." *Blood* 87 (2): 423–38.
- Tembhare, Prashant R., Gaurav Chatterjee, Twinkle Khanka, Sitaram Ghogale, Yajamanam Badrinath, Nilesh Deshpande, Devasis Panda, et al. 2021. "Eleven-Marker 10-Color Flow Cytometric Assessment of Measurable Residual Disease for T-Cell Acute Lymphoblastic Leukemia Using an Approach of Exclusion." *Cytometry. Part B, Clinical Cytometry* 100 (4): 421–33. <https://doi.org/10.1002/cyto.b.21939>.
- Thurman, Robert E., Eric Rynes, Richard Humbert, Jeff Vierstra, Matthew T. Maurano, Eric Haugen, Nathan C. Sheffield, et al. 2012. "The Accessible Chromatin Landscape of the Human Genome." *Nature* 489 (7414): 75–82. <https://doi.org/10.1038/nature11232>.
- Tian, Yuan, Mariana Babor, Jerome Lane, Veronique Schulten, Veena S. Patil, Grégory Seumois, Sandy L. Rosales, et al. 2017. "Unique Phenotypes and Clonal Expansions of Human CD4 Effector Memory T Cells Re-Expressing CD45RA." *Nature Communications* 8 (1): 1473. <https://doi.org/10.1038/s41467-017-01728-5>.
- Toh, Yasushi, and Garth L. Nicolson. 2009. "The Role of the MTA Family and Their Encoded Proteins in Human Cancers: Molecular Functions and Clinical Implications." *Clinical & Experimental Metastasis* 26 (3): 215–27. <https://doi.org/10.1007/s10585-008-9233-8>.
- Torchia, Joseph, Christopher Glass, and Michael G Rosenfeld. 1998. "Co-Activators and Co-Repressors in the Integration of Transcriptional Responses." *Current Opinion in Cell Biology* 10 (3): 373–83. [https://doi.org/10.1016/S0955-0674\(98\)80014-8](https://doi.org/10.1016/S0955-0674(98)80014-8).
- Travert, Marion, Yenlin Huang, Laurence De Leval, Nadine Martin-Garcia, Marie-Helene Delfau-Larue, Françoise Berger, Jacques Bosq, et al. 2012. "Molecular Features of Hepatosplenic T-Cell Lymphoma Unravels Potential Novel Therapeutic Targets." *Blood* 119 (24): 5795–5806. <https://doi.org/10.1182/blood-2011-12-396150>.
- Uchiyama, Takashi, Junji Yodoi, Kimitaka Sagawa, Kiyoshi Takatsuki, and Haruto Uchino. 1977. "Adult T-Cell Leukemia: Clinical and Hematologic Features of 16 Cases." *Blood* 50 (3): 481–92. <https://doi.org/10.1182/blood.V50.3.481.481>.

- Uldrich, Adam P., Jérôme Le Nours, Daniel G. Pellicci, Nicholas A. Gherardin, Kirsty G. McPherson, Ricky T. Lim, Onisha Patel, et al. 2013. "CD1d-Lipid Antigen Recognition by the  $\gamma\delta$  TCR." *Nature Immunology* 14 (11): 1137–45. <https://doi.org/10.1038/ni.2713>.
- Vallois, David, Maria Pamela D. Dobay, Ryan D. Morin, François Lemonnier, Edoardo Missiaglia, Mélanie Juilland, Justyna Iwaszkiewicz, et al. 2016. "Activating Mutations in Genes Related to TCR Signaling in Angioimmunoblastic and Other Follicular Helper T-Cell-Derived Lymphomas." *Blood* 128 (11): 1490–1502. <https://doi.org/10.1182/blood-2016-02-698977>.
- Vega, Francisco, L. Jeffrey Medeiros, and Philippe Gaulard. 2007. "Hepatosplenic and Other Gammadelta T-Cell Lymphomas." *American Journal of Clinical Pathology* 127 (6): 869–80. <https://doi.org/10.1309/LRKX8CE7GVPCR1FT>.
- Veloza, Luis, Doriane Cavalieri, Edoardo Missiaglia, Albane Ledoux-Pilon, Bettina Bisig, Bruno Pereira, Christophe Bonnet, et al. 2023. "Monomorphic Epitheliotropic Intestinal T-Cell Lymphoma Comprises Morphologic and Genomic Heterogeneity Impacting Outcome." *Haematologica* 108 (1): 181–95. <https://doi.org/10.3324/haematol.2022.281226>.
- Velusamy, Thirunavukkarasu, Mark J. Kiel, Anagh A. Sahasrabudhe, Delphine Rolland, Catherine A. Dixon, Nathanael G. Bailey, Bryan L. Betz, et al. 2014. "A Novel Recurrent NPM1-TYK2 Gene Fusion in Cutaneous CD30-Positive Lymphoproliferative Disorders." *Blood* 124 (25): 3768–71. <https://doi.org/10.1182/blood-2014-07-588434>.
- Vose, Julie, James Armitage, Dennis Weisenburger, and International T-Cell Lymphoma Project. 2008. "International Peripheral T-Cell and Natural Killer/T-Cell Lymphoma Study: Pathology Findings and Clinical Outcomes." *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 26 (25): 4124–30. <https://doi.org/10.1200/JCO.2008.16.4558>.
- Wai, Cho Mar Myint, Shangying Chen, The Phyu, Shuangyi Fan, Sai Mun Leong, Wenning Zheng, Louis Ching Yi Low, et al. 2022. "Immune Pathway Upregulation and Lower Genomic Instability Distinguish EBV-Positive Nodal T/NK-Cell Lymphoma from ENKTL and PTCL-NOS." *Haematologica* 107 (8): 1864–79. <https://doi.org/10.3324/haematol.2021.280003>.
- Wan, Xiaoling, Lulu Liu, Peipei Zhou, Xinhui Hui, Qiaomei He, Fangfang Yu, Wei Zhang, et al. 2019. "The Nuclear Receptor Corepressor NCoR1 Regulates Hematopoiesis and Leukemogenesis in Vivo." *Blood Advances* 3 (4): 644–57. <https://doi.org/10.1182/bloodadvances.2018022756>.
- Wano, Y., M. Feinberg, J. B. Hosking, H. Bogerd, and W. C. Greene. 1988. "Stable Expression of the Tax Gene of Type I Human T-Cell Leukemia Virus in Human T Cells Activates Specific Cellular Genes Involved in Growth." *Proceedings of the National Academy of Sciences of the United States of America* 85 (24): 9733–37. <https://doi.org/10.1073/pnas.85.24.9733>.
- Way, Gregory P., Michael Zietz, Vincent Rubinetti, Daniel S. Himmelstein, and Casey S. Greene. 2020. "Compressing Gene Expression Data Using Multiple Latent Space Dimensionalities Learns Complementary Biological Representations." *Genome Biology* 21 (1): 109. <https://doi.org/10.1186/s13059-020-02021-3>.
- Weng, Andrew P., Adolfo A. Ferrando, Woojoong Lee, John P. Morris, Lewis B. Silverman, Cheryll Sanchez-Irizarry, Stephen C. Blacklow, A. Thomas Look, and Jon C. Aster. 2004. "Activating Mutations of NOTCH1 in Human T Cell Acute Lymphoblastic Leukemia."

- Science* (New York, N.Y.) 306 (5694): 269–71.  
<https://doi.org/10.1126/science.1102160>.
- West, Adam G., Miklos Gaszner, and Gary Felsenfeld. 2002. “Insulators: Many Functions, Many Mechanisms.” *Genes & Development* 16 (3): 271–88.  
<https://doi.org/10.1101/gad.954702>.
- Wilson, Boris G., and Charles W. M. Roberts. 2011. “SWI/SNF Nucleosome Remodellers and Cancer.” *Nature Reviews. Cancer* 11 (7): 481–92. <https://doi.org/10.1038/nrc3068>.
- Wu, C., Y. C. Wong, and S. C. Elgin. 1979. “The Chromatin Structure of Specific Genes: II. Disruption of Chromatin Structure during Gene Activity.” *Cell* 16 (4): 807–14.  
[https://doi.org/10.1016/0092-8674\(79\)90096-5](https://doi.org/10.1016/0092-8674(79)90096-5).
- Wu, Zhichao, Zied Abdullaev, Drew Pratt, Hye-Jung Chung, Shannon Skarshaug, Valerie Zgonc, Candice Perry, et al. 2022. “Impact of the Methylation Classifier and Ancillary Methods on CNS Tumor Diagnostics.” *Neuro-Oncology* 24 (4): 571–81.  
<https://doi.org/10.1093/neuonc/noab227>.
- Xiong, Jie, Bo-Wen Cui, Nan Wang, Yu-Ting Dai, Hao Zhang, Chao-Fu Wang, Hui-Juan Zhong, et al. 2020. “Genomic and Transcriptomic Characterization of Natural Killer T Cell Lymphoma.” *Cancer Cell* 37 (3): 403–419.e6.  
<https://doi.org/10.1016/j.ccell.2020.02.005>.
- Yabe, Mariko, L. Jeffrey Medeiros, Yahya Daneshbod, Masoud Davanlou, Carlos E. Bueso-Ramos, Elisa J. Moran, Ken H. Young, and Roberto N. Miranda. 2017. “Hepatosplenic T-Cell Lymphoma Arising in Patients with Immunodysregulatory Disorders: A Study of 7 Patients Who Did Not Receive Tumor Necrosis Factor- $\alpha$  Inhibitor Therapy and Literature Review.” *Annals of Diagnostic Pathology* 26 (February): 16–22.  
<https://doi.org/10.1016/j.anndiagpath.2016.10.005>.
- Yabe, Mariko, L. Jeffrey Medeiros, Guilin Tang, Sa A. Wang, Sairah Ahmed, Yago Nieto, Shimin Hu, et al. 2016. “Prognostic Factors of Hepatosplenic T-Cell Lymphoma: Clinicopathologic Study of 28 Cases.” *The American Journal of Surgical Pathology* 40 (5): 676. <https://doi.org/10.1097/PAS.0000000000000614>.
- Yabe, Mariko, L. Jeffrey Medeiros, Sa A. Wang, Guilin Tang, Carlos E. Bueso-Ramos, Jeffrey L. Jorgensen, Govind Bhagat, et al. 2017. “Distinguishing Between Hepatosplenic T-Cell Lymphoma and  $\Gamma\delta$  T-Cell Large Granular Lymphocytic Leukemia: A Clinicopathologic, Immunophenotypic, and Molecular Analysis.” *The American Journal of Surgical Pathology* 41 (1): 82–93. <https://doi.org/10.1097/PAS.0000000000000743>.
- Yabe, Mariko, Roberto N. Miranda, and L. Jeffrey Medeiros. 2018. “Hepatosplenic T-Cell Lymphoma: A Review of Clinicopathologic Features, Pathogenesis, and Prognostic Factors.” *Human Pathology* 74 (April): 5–16.  
<https://doi.org/10.1016/j.humpath.2018.01.005>.
- Yadav, Ram Prakash, Vamsi Krishna Polavarapu, Pengwei Xing, and Xingqi Chen. 2022. “FFPE-ATAC: A Highly Sensitive Method for Profiling Chromatin Accessibility in Formalin-Fixed Paraffin-Embedded Samples.” *Current Protocols* 2 (8): e535.  
<https://doi.org/10.1002/cpz1.535>.
- Yamaguchi, K., H. Nishimura, H. Kohrogi, M. Jono, Y. Miyamoto, and K. Takatsuki. 1983. “A Proposal for Smoldering Adult T-Cell Leukemia: A Clinicopathologic Study of Five Cases.” *Blood* 62 (4): 758–66.
- Yamaguchi, K., R. Yoshioka, T. Kiyokawa, M. Seiki, M. Yoshida, and K. Takatsuki. 1986. “Lymphoma Type Adult T-Cell Leukemia--a Clinicopathologic Study of HTLV Related T-

- Cell Type Malignant Lymphoma." *Hematological Oncology* 4 (1): 59–65. <https://doi.org/10.1002/hon.2900040108>.
- Yoon, Ho-Geun, Doug W. Chan, Zhi-Qing Huang, Jiwen Li, Joseph D. Fondell, Jun Qin, and Jiemin Wong. 2003. "Purification and Functional Characterization of the Human N-CoR Complex: The Roles of HDAC3, TBL1 and TBLR1." *The EMBO Journal* 22 (6): 1336–46. <https://doi.org/10.1093/emboj/cdg120>.
- Yoshida, Hideyuki, Caleb A. Lareau, Ricardo N. Ramirez, Samuel A. Rose, Barbara Maier, Aleksandra Wroblewska, Fiona Desland, et al. 2019. "The Cis-Regulatory Atlas of the Mouse Immune System." *Cell* 176 (4): 897-912.e20. <https://doi.org/10.1016/j.cell.2018.12.036>.
- Yoshida, M., J. Inoue, J. Fujisawa, and M. Seiki. 1989. "Molecular Mechanisms of Regulation of HTLV-1 Gene Expression and Its Association with Leukemogenesis." *Genome* 31 (2): 662–67. <https://doi.org/10.1139/g89-121>.
- Zaitsev, Konstantin, Monika Bambouskova, Amanda Swain, and Maxim N. Artyomov. 2019. "Complete Deconvolution of Cellular Mixtures Based on Linearity of Transcriptional Signatures." *Nature Communications* 10 (1): 2209. <https://doi.org/10.1038/s41467-019-09990-5>.
- Zhan, He-Qin, Xiao-Qiu Li, Xiong-Zeng Zhu, Hong-Fen Lu, Xiao-Yan Zhou, and Yan Chen. 2011. "Expression of Follicular Helper T Cell Markers in Nodal Peripheral T Cell Lymphomas: A Tissue Microarray Analysis of 162 Cases." *Journal of Clinical Pathology* 64 (4): 319–24. <https://doi.org/10.1136/jcp.2010.084459>.
- Zhang, Yaohua, Yang Wang, Richard Yu, Yuanshen Huang, Mingwan Su, Cheng Xiao, Magdalena Martinka, et al. 2012. "Molecular Markers of Early-Stage Mycosis Fungoides." *The Journal of Investigative Dermatology* 132 (6): 1698–1706. <https://doi.org/10.1038/jid.2012.13>.
- Zhang, Yong, Tao Liu, Clifford A. Meyer, Jérôme Eeckhoute, David S. Johnson, Bradley E. Bernstein, Chad Nusbaum, et al. 2008. "Model-Based Analysis of CHIP-Seq (MACS)." *Genome Biology* 9 (9): R137. <https://doi.org/10.1186/gb-2008-9-9-r137>.
- Zheng, Chun-Hou, De-Shuang Huang, Xiang-Zhen Kong, and Xing-Ming Zhao. 2008. "Gene Expression Data Classification Using Consensus Independent Component Analysis." *Genomics, Proteomics & Bioinformatics* 6 (2): 74–82. [https://doi.org/10.1016/S1672-0229\(08\)60022-4](https://doi.org/10.1016/S1672-0229(08)60022-4).
- Zheng, Ye, Ashutosh Chaudhry, Arnold Kas, Paul deRoos, Jeong M. Kim, Tin-Tin Chu, Lynn Corcoran, Piper Treuting, Ulf Klein, and Alexander Y. Rudensky. 2009. "Regulatory T-Cell Suppressor Program Co-opts Transcription Factor IRF4 to Control TH2 Responses." *Nature* 458 (7236): 351–56. <https://doi.org/10.1038/nature07674>.



# LISTE DES TABLEAUX

---

Tableau 1 - Classification OMS 2022 des néoplasies T .....	16
Tableau 2 - Classification ICC 2022 des néoplasies matures T .....	17
Tableau 3 - marqueurs immunophénotypiques des principales néoplasies T .....	38
Tableau 4 - Principales anomalies chromosomiques ayant une valeur diagnostique dans les néoplasies matures T.....	39
Tableau 5 - Fréquence des mutations des protéines JAK et STAT retrouvées dans les néoplasies matures T.....	41
Tableau 6 - Principales origines cellulaires suspectées des différentes hémopathies matures T .....	50
Tableau 7 - Fonctionnalité des différentes classes de modification d'histones .....	56
Tableau 8 - Composition des principaux complexes remodeleurs de la chromatine.....	59
Tableau 9 - Comparaison des principales études d'algorithmes diagnostics des PTCL basés sur la transcriptomique .....	91

# LISTE DES FIGURES

---

Figure 1 - Répartition des nouveaux cas d'hémopathies matures B et T en France en 2018..	18
Figure 2 - Evolution du nombre de cas incidents de lymphomes T en France.....	18
Figure 3 - Distribution des sous-entités de lymphomes T périphériques diagnostiqués dans le réseau Lymphopath entre 2010 et 2013 (Laurent et al. JCO 2017) .....	19
Figure 4 - Cytokines et facteurs de transcription impliqués dans la différenciation T helper..	44
Figure 5 - Organisation du génome <i>d'après Misteli (Misteli 2020)</i> .....	54
Figure 6 - Exemples de motifs reconnus par les facteurs de transcription GATA3 d'après différentes méthodes (issu de la base JASPAR 2022) .....	64
Figure 7 - Préparation des noyaux et cellules pour l'ATAC-seq .....	68
Figure 8 - Transposition et amplification en ATAC-seq .....	69
Figure 9 - Traitement bioinformatique des données d'ATAC-seq .....	71
Figure 10 - Exemples de profils d'enrichissement au TSS pour un échantillon de mauvaise (à gauche) et bonne qualité (à droite).....	73
Figure 11 - Exemple de factorisation non négative de matrice de comptes à partir de profils transcriptomiques de lymphocytes B, T CD4 et T CD8 pour un rang $k=2$ et $k=3$ . .....	82

# LISTE DES GENES ET PROTEINES

---

ABCB1	ATP Binding Cassette Sub family B Member 1
AKT	AKT Serine/Threonine Kinase1
AP-1	Activator Protein1
ARID1A	AT-Rich Interaction Domain
ATIC	5-Aminoimidazole-4-Carboxamide Ribonucleotide Formyltransferase/IMP/Cyclohydrolase
ATM	Ataxia Telangiectasia Mutated
B2M	Beta-2-Microglobulin
Bcl11b	BCL11 Transcription Factor B
BCL6	BCL6 Transcription Repressor
BCOR	BCL6 Corepressor
BRDT	Bromodomain Testis Associated
cARM1	coActivator associated Arginine Methyltransferase
CBP/CREBBP	CREB-bind protein
CCR4	C-CMotifC hemokine Receptor 4
CCR7	C-CMotifC hemokine Receptor 7
CD	Cluster de Différenciation
CARD11	Caspase Recruitment Domain Family Member 11
CDKN	Cyclin Dependent Kinase Inhibitor
CHD	Chromodomain Hélicase DNAbinding
CHEK2	Checkpoint Kinase2
CREB	cAMP-Response Element Binding protein
CREBBP	CREB Binding Protein
CRTH2	Chemo Attractant Receptor-homologous Molecule Expressedon TH2cells
CHN2	Chimerin2
CLA	Cutaneous Lymphocyte Antigen
CTCF	CCCTC-Binding Factor
CTLA-4	Cytotoxic T-Lymphocyte Associated Protein 4
CXCL	C-X-C motif Chemokine Ligand
CXCR	C-X-C motif Chemokine Receptor
DDX3X	DEAD-Box Helicase3X-Linked
DNMT1	DNA méthyltransférase 1
DNMT3A	DNA Methyltransferase 3 Alpha
DUSP22	Dual Specificity Phosphatase 22
EEF1G	Eukaryotic Translation Elongation Factor 1 Gamma
Eomes	Eomesodermin
EP300	E1A Binding Protein P300
ER	Estrogen receptor
EZH2	Enhancer Of Zeste Homolog 2 (Polycomb Repressive Complex 2 Subunit)
FAS	Fas Cell Surface Death Receptor

FBXW7	F-Box And WD Repeat Domain Containing 7
FOXP3	Forkhead Box P3
FRA7H	Fragile Site Aphidicolin Type Common Fra(7)(Q32.3)
FYN	FYN Proto-Oncogene Src Family Tyrosine Kinase
GATA3	GATA Binding Protein 3
GR	Glucocorticoid receptor
GZMB	Granzyme B
HBZ	HTLV-1 bZIP factor
HDAC	Histone deacetylase 3
HLA-DRB5	Major Histocompatibility Complex, Class II, DR Beta5
HNF	Hepatocyte Nuclear Factor
HOX	Homeobox
HSF2	Heat Shock Factor Protein 2
ICOS	Inducible T Cell Costimulator
ICOSL	ICOS-Ligand
IDH2	Isocitrate Dehydrogenase (NADP(+)) 2
IDH2R172K/T	Isocitrate Dehydrogenase 2 Mutation Arginine 172
IFN	Interferon
IL2R	Interleukin-2 Receptor
IL	Interleukine
INFG	Interferon Gamma
INO80	INO80 Complex ATPase Subunit
IRF4	Interferon Regulatory Factor 4
ISWI	Imitation Switch
JAK	Janus Kinase
KIR3DL1	Killer Cell Immunoglobulin Like Receptor, Three Ig Domain and long Cytoplasmic Tail1
KLF	Krüppel Like Factor
KLR	Killer cell Lectin Like Receptor
KLRC2	Killer cell Lectin Like Receptor C2
KLRF1	Killer cell Lectin Like Receptor F1
KMT2C	Lysine Methyltransferase 2C
KMT2D	Lysine Methyltransferase 2D
LFA-1	Lymphocyte Function-Associated Antigen1
MadCAM-1	Mucosal Vascular Addressin Cell Adhesion Molecule 1
MAPK	Mitogen-Activated Protein Kinase
MGA	MAX Dimerization Protein MGA
MLL2	Lysine Methyltransferase 2D
MSN	Moesin
MTA	Metastasis Associated
MTCP1	Mature T Cell Proliferation 1
MYB	MYB Proto-Oncogene
MYC	MYC Proto-Oncogene
MYH9	Myosin Heavy Chain 9
NcoA/SRC	Nuclear Receptor coActivator

NCOR1/SMRT	Nuclear Receptor Corepressor
NFAT	Nuclear Factor of Activated T-cells
Nfil3	NuclearFactor Interleukin 3 Regulated
NFKB	Nuclear Factor-kappa B
NPM1	Nucleophosmin1
NuRD	Nucleosome Remodeling and Deacetylase
PABPC4	Poly(A) Binding Protein Cytoplasmic 4
PcG	Groupe polycomb
PD-1	Programmed Cell Death 1
PD-L1	Programmed Cell Death Ligand 1
PGP1	P Glycoprotein 1
PI3K	Phosphoinositide 3-kinase
PLCG1	Phospholipase C Gamma 1
POT1	Protection Of Telomeres 1
PPP1R9A	Protein Phosphatase 1 Regulatory Subunit 9A
PR	Progesteron receptor
PRC1	Protein Regulator of Cytokinesis 1
PRF	Perforine
PRKCB	Protein Kinase C Beta
PSGL1	P-Selectin Glycoprotein Ligand
PTEN	Phosphatase and Tensin Homolog
PTPN12	Protein Tyrosine Phosphatase Non-Receptor Type 12
Rag1	Recombination Activating 1
RARs/RXRs	Recepteur Acide Rétinoïque
RhoA	Ras Homolog Family Member A
RHOAG17V	Ras Homolog Family member A substitution position 17G→V
RING1A/B	Ring Finger Protein 1A/B
RLTPR	Capping Protein Regulator And Myosin 1 Linker 2
RNF213	Ring Finger Protein 213
RORC	Related Orphan Receptor C
ROS	ROSPProto-Oncogene1,ReceptorTyrosineKinase
Runx3	RUNX Family Transcription Factor 3
SETD2	SET Domain Containing2 Histone Lysine Methyltransferase
SETDB1	SET Domain bifurcated histone Lysine Methyltransferase 1
SMADs	Small Mothers against decapentaplegic
SMARCA2	SWI/SNF Related Matrix Associated Actin Dependent Regulator Of Chromatin Subfamily A Member 2
SMARCB1	SWI/SNF Related Matrix Associated Actin Dependent Regulator Of Chromatin Subfamily B Member 1
SOX4	SRY-box transcription factor 4
SREBP	Sterol Regulatory Element Binding Protein
STATs	Signal Transducer and Activator of Transcription
SWI/SNF	SWItch/Sucrose Non Fermentable
SWR1	Chromatin remodelling complex
TAL	T-Cell Acute Lymphocytic Leukemia Protein1
TAX	Transactivator of pX

Tbet	T-Box Transcription Factor 21235
TBL1X	Transducin betalike 1X-linked
TBX21	T-Box Transcription Factor 21
TCF1	T-CellFactor1
TCL1	T-Cell Leukemia/Lymphoma Protein 1
TET	Tet méthyl cytosine dioxygénase
TET2	TetMethylcytosineDioxygenase2
TFG	Trafficking From ER To Golgi Regulator
TFIID	Facteur de transcription IID
ThPOK	Zinc Finger And BTB Domain Containing 7B
TIA1	T-Cell-Restricted Intracellular Antigen-1
TLX	T Cell Leukemia Homeobox 1
TNF	Tumor Necrosis Factor
TNFAIP3	TNF Alpha Induced Protein 3
TNFRSF1B	TNF Receptor Superfamily Member 1 B
TNIP3	TNFAIP3 Interacting Protein 3
TP53	Tumor Protein P53
TPM3	Tropomyosin 3
TR	Thyroid hormone receptor
TRAF1	TNF Receptor Associated Factor 1
TRRAP	Transformation/Transcription Domain Associated Protein
TSIM	Tumor-suppressor/immune-modulator
TYK2	Tyrosine Kinase 2
VAV1	Vav Guanine Nucleotide Exchange Factor 1
VCAM-1	Vascular Cell Adhesion Molecule 1
VLA-4	Very Late Activation Protein 4