Université de Montréal

Tolérance au soi : rôle des éléments transposables dans les tissus somatiques et le thymus

Par

Jean-David Larouche

Programmes de biologie moléculaire, Faculté de médecine

Thèse présentée en vue de l'obtention du grade de Philosophiae Doctor

en biologie moléculaire, option générale

Août 2023

Université de Montréal

Institut de Recherche en Immunologie et Cancérologie, Programmes de biologie moléculaire,

Faculté de médecine

Cette thèse intitulée

Tolérance au soi : rôle des éléments transposables dans les tissus somatiques et le thymus

Présentée par

Jean-David Larouche

A été évaluée par un jury composé des personnes suivantes

Philippe Roux
Président-rapporteur

Claude Perreault
Directeur de recherche

Martin Sauvageau
Membre du jury

Guillaume Bourque
Examinateur externe

Résumé

Les éléments transposables (TE) sont des séquences répétitives représentant environ 45% des génomes humain et murin. Il est généralement assumé que leur expression est réprimée dans les cellules somatiques pour protéger l'intégrité du génome, et cette régulation épigénétique est fréquemment perdue dans les cancers, menant à la surexpression des TEs dans les tumeurs. Puisque l'expression aberrante des TEs est associée à l'infiltration de la tumeur par les cellules immunitaires, les TEs sont considérés comme des cibles prometteuses d'immunothérapies du cancer. Une meilleure description de l'expression des TEs dans les tissus somatiques ainsi que dans le thymus, l'organe responsable du développement de la tolérance au soi des lymphocytes T, est toutefois nécessaire pour évaluer la capacité des TEs d'induire des réponses immunitaires et déterminer si l'expression des TEs est belle et bien spécifique aux tumeurs. L'objectif de cette thèse est donc de brosser un portrait exhaustif de l'expression des TEs dans les tissus somatiques humains ainsi que dans le thymus. Pour ce faire, des données transcriptomiques et immunopeptidomiques ont été analysées pour mieux comprendre les interactions entre les TEs et les lymphocytes T à l'état basal. Nos résultats ont montré que l'expression des TEs est répandue dans les tissus somatiques humains, bien que leur niveau d'expression varie d'un tissu à l'autre et que plusieurs TEs sont exprimés de façon tissu-spécifique. De plus, les TEs peuvent être traduits et présentés par le CMH-I à la surface de cellules non-cancéreuses. Nous avons aussi déterminé que les TEs ont trois fonctions potentielles dans le thymus : ils pourraient fournir des sites de liaison à un grand nombre de facteurs de transcription dans toutes les populations cellulaires du thymus, ils stimuleraient la sécrétion d'IFN α/β par les pDCs thymiques, et ils contribuent aux sélections positive et négative des thymocytes. Nos travaux illustrent la complexité des interactions entre les TEs et le système immunitaire adaptatif. Finalement, étant donnée l'expression répandue des TEs dans les tissus somatiques, nos travaux soulignent l'importance d'établir la tolérance des lymphocytes T à l'égard des TEs pour éviter des réactions auto-immunes.

Mots-clés: Éléments transposables, peptides associés au CMH-I, thymus, immunologie, transcriptomique.

Summary

Transposable elements are repetitive sequences representing around 45% of the human and murine genomes. It is generally assumed that their expression is repressed in somatic cells to preserve genomic integrity, but this epigenetic regulation is frequently lost in cancer cells, leading to the aberrant expression of TEs in tumors. As aberrant TE expression is associated with tumor infiltration by immune cells, TEs are considered as promising cancer immunotherapy targets. However, a better description of TE expression in somatic tissues and in the thymus, the organ responsible of T cell self-tolerance induction, is required to evaluate the potential of TEs to induce immune responses as well as the tumor specificity of TE expression. Thus, this thesis' objective is to draw an exhaustive profile of TE expression in human somatic tissues and in the thymus. To do so, we analyzed transcriptomic and immunopeptidomic data to better understand interactions between TEs and T cells at steady state. Our work shows that TE expression is widespread in human somatic tissues, even though their expression level varies between tissues and many TEs are expressed in a tissue-specific manner. Additionally, TEs are translated and presented by the MHC-I on the surface of non-malignant cells. We also determined that TEs have three potential functions in the thymus: they could provide transcription factor binding sites in all cell populations of the thymus, they might induce the constitutive IFN α/β secretion of thymic pDCs, and they contribute to both positive and negative selections of thymocytes. Altogether, our work illustrates the complexity of the interactions between TEs and the vertebrate adaptive immune system. Given the widespread expression of TEs in somatic tissues, this thesis highlights the importance of establishing T cell tolerance towards TE sequences to avoid autoimmune reactions in peripheral tissues.

Keywords: Transposable elements, MHC I-associated peptides, thymus, immunology, transcriptomics.

Table des matières

Liste des figures

Liste des figures supplémentaires

Liste des sigles et abréviations

A-C

AIRE : régulateur auto-immun
ARNm : ARN messager
APC : cellules présentatrices d'antigènes
B2M : β_2-microglobuline
cDC1 : cellules dendritiques conventionnelles 1
cDC2 : cellules dendritiques conventionnelles 1
CHD4 : protéine hélicase à ADN à chromodomaine
cTEC : cellules épithéliales thymiques du cortex
CMH : complexe majeur d'histocompatibilité
CMH-I : complexe majeur d'histocompatibilité de classe I
CMH-II : complexe majeur d'histocompatibilité de classe II

D-E

DC : cellules dendritiques
DN : thymocytes double négatifs $CD4^-CD8^-$
DNMT : ADN méthyltransférases
DP : thymocytes double positifs $CD4^+CD8^+$
DRiPs: produits ribosomaux défectueux
dsRNA : ARN double brin
ESC : cellules souches embryonnaires
ER : réticulum endoplasmique
ERE : rétroéléments endogènes
ERV : rétrovirus endogènes
ereMAP : peptides dérivés des EREs et associés au CMH-I
ETP : progéniteurs des thymocytes

F-L

FEZF2 : protéine à doigt de zinc du prosencéphale embryonnaire 2
GTEx : Genotype-Tissue Expression
HLA : antigène des leucocytes humains
HSC : cellules souches hématopoïétiques
IFN : interféron
KZFP : protéines à doigt de zinc de la boîte associée Krüppel
LINE : éléments nucléaires dispersés longs
lncRNA : longs ARN non-codants
LTR : séquences terminales longues répétées

M-P

MAP : peptides associés au CMH-I
MS : Spectrométrie de masse
miRNA : micro ARN
mTEC : cellules épithéliales thymiques de la médulla
mTEC(I) : mTECs CMH-II^{lo} CD80^{lo} CCL21^{+}
mTEC(II) : mTECs CMH-II^{hi} CD80^{hi}
mTEC(III): mTECs CMH-II^{lo} CD80^{lo} KRT10^{+}
NK : cellules *natural killer*
NKT : lymphocytes tueur naturel T (*natural killer T cells*)
pDC : cellules dendritiques plasmacytoïdes
PGE : Expression génique promiscuitaire
piRNA : ARN interagissant avec Piwi

S-Z

scRNA-seq : séquençage à cellule unique
SINE : éléments nucléaires dispersés courts
SP : thymocytes simple positifs CD4^{+}CD8^{-} ou CD4^{-}CD8^{+}
TAA : antigène associé à la tumeur
TAP1/2 : transporteurs associés au traitement des antigènes 1 et 2
TCR : récepteur des lymphocytes T
TE : éléments transposables
TEC : cellules épithéliales thymiques
TIR : séquences terminales répétées inversées
TRE : ERE tissu-spécifiques
Treg : lymphocytes T régulateurs
TRG : gènes tissu-spécifiques
TSA : antigène tumeur-spécifique
TSS : sites d'initiation de la transcription
TSSP : sérine protéase thymus-spécifique
uORF : cadres de lecture ouverts en amont

« N'acceptez pas que l'on fixe, ni qui vous êtes, ni où rester. »

- Alain Damasio, *La Horde du Contrevent*

« It's a lovely day, des fois faut que je me rappelle, d'oublier tout le reste »

- Clay and Friends, AGUÀ EXTEND'EAU

« La volonté c'est comme l'acné : plus t'en as, et plus ça parait. »

- Capitaine Charles Patenaude

Remerciements

J'ai décidé d'amorcer l'écriture de ma thèse par la rédaction de cette section, en partie par procrastination, mais surtout parce qu'elle est pour moi la plus importante (désolé pour le comité évaluateur!). Ce doctorat, je n'aurais pas pu le réaliser sans l'aide et le soutien de l'ensemble des personnes mentionnées ci-bas. Des personnes qui m'ont supporté lorsque tout allait bien et que j'étais surexcité par mon projet de recherche, mais surtout dans les moments plus difficiles. Des personnes qui m'ont permis de grandir scientifiquement et humainement. Je ne pourrai jamais le dire assez, mais merci pour tout!

Premièrement, je voudrais prendre le temps de remercier mon superviseur, **Claude Perreault**, d'avoir pris une chance en me recrutant à l'automne 2016. Je me souviens être sorti de votre bureau, à la fin de l'entrevue, et de m'être dit que je n'avais aucune chance d'être sélectionné : je venais de passer 20 minutes à répondre à vos questions par oui ou non, complètement terrorisé, et la seule phrase complète que j'ai réussi à formuler était que je voulais « apprendre de nouvelles techniques » lorsque vous m'avez demandé ce que je cherchais pour un projet de maitrise. Je ne sais toujours pas ce qui a motivé votre décision, mais je suis incroyablement reconnaissant de l'opportunité que vous m'avez donnée. Au-delà de m'avoir permis de m'améliorer scientifiquement (du moins j'espère ;)), vous m'avez donné la liberté d'explorer les questions biologiques qui m'intéressaient et, par le fait même, vous m'avez transmis votre passion pour la recherche académique. On demande souvent aux doctorants ce qu'ils feraient différemment s'ils avaient à recommencer leur parcours; une chose que je ne changerais jamais est de vous avoir comme mentor pour mes études graduées. Merci encore pour tout! - Signé Rétroman ☺

Ensuite, je voudrais remercier les membres de mon comité de suivi, **Philippe Roux**, **Martin Sauvageau** et **Guillaume Bourque**, pour vos conseils et pour votre soutien tout au long de mon doctorat. J'ai eu la chance d'avoir des comités de suivi lors desquels on me challengeait, mais toujours dans une ambiance conviviale et des conversations extrêmement enrichissantes. Merci

aussi d'avoir pris le temps de répondre à mes questions sur la suite de mon parcours, ce fut grandement apprécié.

Finalement, merci à un grand nombre de personnes pour leur soutien au quotidien au laboratoire. Le laboratoire Perreault, ça a été pour moi une grande famille qui m'a épaulé dans les bons comme dans les mauvais moments. Tout d'abord merci à mon « quatuor des madames » : **Céline Laumont**, **Assya Trofimov**, **Krystel Vincent** et **Leslie Hesnard**. À Assya et Céline, merci de m'avoir pris sous vos ailes alors que je n'étais même pas encore un bébé bio-info, et merci de votre patience lorsque j'arrivais avec mes mille et une questions. Krystel et Leslie, je n'aurais pas pu passer au travers du *wetlab* sans vous! Merci d'avoir pris le temps de me former, et pour toutes les conversations sur le thymus et l'immunologie en général. Et Krystel, merci pour ton input tout au long de ces 5 années de thèse, à chaque fois que j'avais besoin de discuter d'un résultat ou de la méthodologie d'une analyse tu étais là! Finalement, merci à vous quatre pour les nombreux fous rires et pour les conversations éclairantes sur tout et sur rien! Merci à **Caroline Côté** de m'avoir aidé dans mes premiers pas au laboratoire, d'avoir eu une patience infinie lorsque j'oubliais pour la centième fois où les choses sont rangées dans le lab, et pour les chocolats du matin, du lunch ou de l'après-midi (quand tu sentais que j'en avais besoin quoi!). Merci aussi à **Lucyle Depoërs** et **Justine Mathé** (les *evil twins* du labo) ainsi qu'à **Eralda Kina** et **Nandita Noronha,** des amies que j'ai eu la chance de me faire dans le laboratoire! Avec vous, j'ai rencontré des personnes géniales, tant humainement que scientifiquement, et aussi d'imprimer pas mal de niaiseries que j'ai collées sur vos bureaux :P. Merci pour les conversations interminables dans le bureau, les 5 à 7, les chalets, les soupers, et tout le reste! Finalement, je veux remercier les autres membres du laboratoire avec qui j'ai eu la chance d'échanger : **Anca Apavaloaei, Mohamed Benhammadi**, **Maxime Cazuhac, Gabriel Ouellet Lavallée, Gregory Ehx**, **Qinchuan Zhao, Marie-Pierre Hardy**, **Sylvie Brochu et Catherine Thériault.**

Merci aussi à toutes les ressources de l'IRIC qui m'ont permis de réaliser mes expériences, analyses ou projets extracurriculaires (et dieu sait qu'ils étaient nombreux!). Merci à **Patrick**

Gendron, **Jean-Philippe Laverdure** et **Éric Audemard**, les génies de la bioinfo qui ont pris le temps de m'aider quand j'en avais besoin. Merci aussi de ne pas trop m'avoir jugé quand je rentrais dans votre bureau et que je ressortais immédiatement sans rien dire parce que j'avais trop honte de ma question. Je voudrais aussi remercier toutes les ressources de l'administration de l'IRIC pour leur soutien aux initiatives étudiantes auxquelles j'ai participé: **Lynda Landry, Judith Lafaille, Julie Mantovani**, **Pascale Le Thérizien**, **Virginie Mondin**, **Évelyne Muhire** et **Sébastien Roy**.

J'ai aussi eu la chance d'interagir avec des étudiants et étudiantes extraordinaires lors de mon passage à l'IRIC. Je pense qu'on sous-estime à quel point l'implication des étudiants gradués est essentielle à la recherche académique. Un doctorat c'est difficile par sa nature même, mais c'est beaucoup plus agréable quand on est entouré par les bonnes personnes. Merci à **Eugénie Goupil**, **Marjorie Lapouge**, **Kévin Leguay**, **Mathilde Soulez, Thomas Milan** et **Audrey Herrmann**. Merci à tous mes collègues de l'association étudiante de l'IRIC, de RéseauLAB, du CEBM et de la semaine de sensibilisation à la santé mentale de l'IRIC, ce furent des projets dans lesquels j'ai eu la chance de côtoyer des personnes passionnées et inspirantes.

Merci aussi aux autres amis, extérieurs à l'IRIC, qui ne comprenaient peut-être pas toujours ce que je traversais, mais qui ont toujours été là, que ce soit pour célébrer la publication d'un article, ou dans les moments de doutes quand je me demandais à quoi ça servait le doctorat (ceux qui y ont assisté s'en souviendront sûrement plus que moi). Merci aux sleux, mes amis de Sciences Lettres et Arts de Grasset, d'avoir gardé leur légendaire curiosité, et par le fait même de m'exposer à des œuvres, des concepts et des idées qui m'auraient échappé sans vous. Merci à **Philippe Rivière** pour les marches durant la covid et les soirées supposément tranquilles qui finissent un peu trop arrosées ;). Un merci tout spécial à **Jessica Youwakim** (#i will take that), que j'ai eu la chance de côtoyer au cégep, au baccalauréat en biochimie, lors de notre échange à Strasbourg, et tout au long du doctorat. Merci aussi aux amis de Strasbourg, on se recroise à l'occasion et c'est toujours un énorme plaisir d'avoir de vos nouvelles : **Philippa Lévesque-Damphousse**, **Vanessa Paré**, **Maude Vallancourt-Audet** et **Léa Garneau**. Merci à toute l'équipe

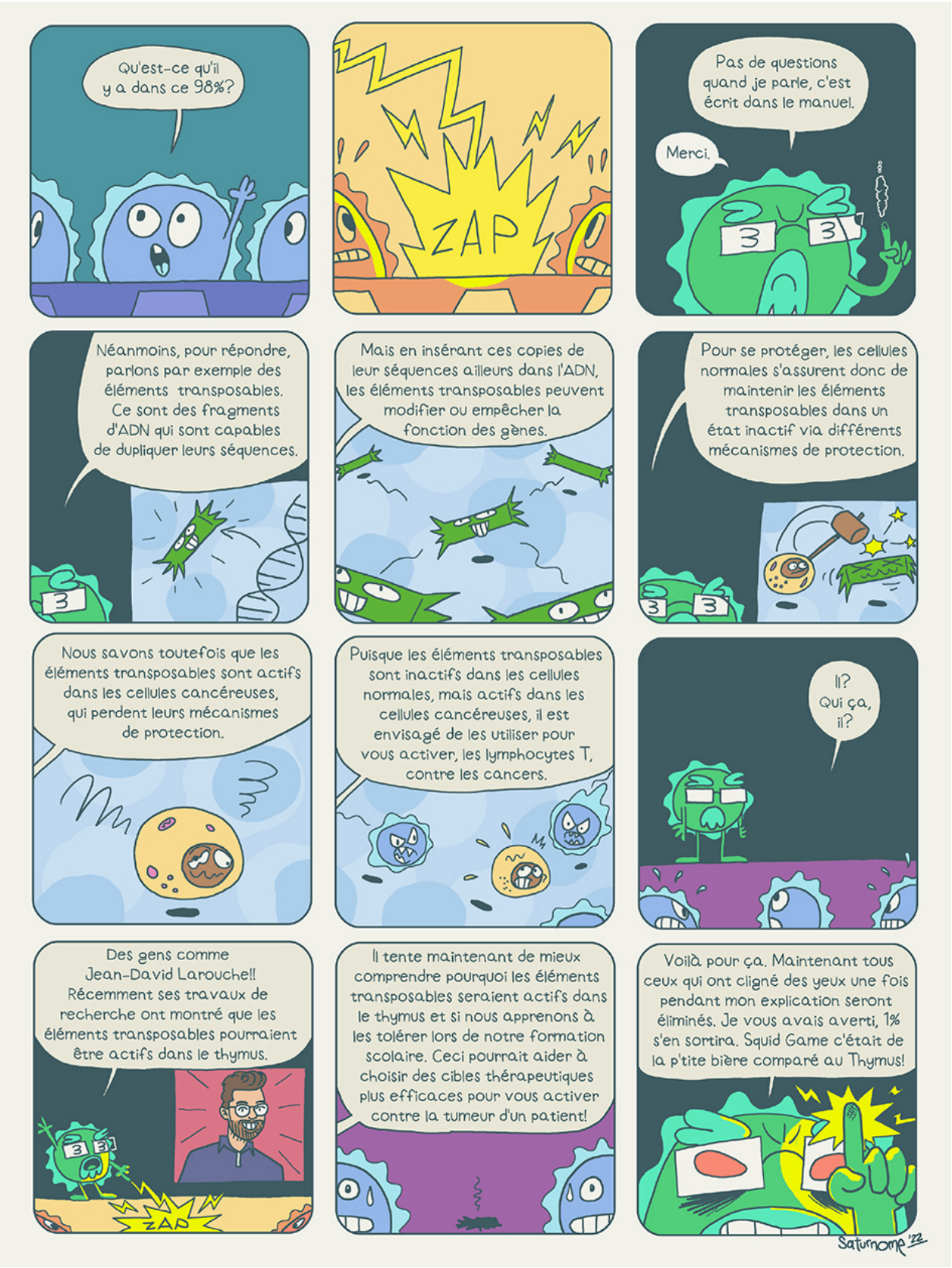
des Hurricanes, la meilleure équipe de dragon boat au monde (dans mon cœur à tout le moins) : **Sophie Viala** (Bambi), **Mathieu Nadeau** (McGuy), **Caroline Galipeau**, **Émilie Rainville**, **Sacha Lalonde**, **Dave et Jack**. Merci aux amis de la communauté yolo esti, ma deuxième famille. Merci de m'avoir accepté comme je suis, avec tous mes défauts et mes travers. J'ai la chance d'avoir un groupe d'amis ouverts, intelligents et attentionnés. Merci à **Claudia Guay** (claudichat), notre organisatrice en chef sans qui nous nous serions sûrement perdus de vue depuis longtemps. Merci pour les soupers, les journées/soirées jeux de société, les voyages, le volleyball et le soccer, les campings, les nombreuses discussions, les soirées de danse dans votre appart, les chalets, et j'en passe. Merci à **Maryline Laflamme** (linlin) et **Étienne Demeules** (Tintin), mes *partners in crime* qui sont parmi les personnes qui me font le plus rire au monde, mes mousquetaires et des amis en or. Merci à **Gabrielle Tremblay** (gabizibidi), **Olivier Dulude** (olivieri), **Francis Harel-Desgroseillers** (Franco Nuovo) et **Benjamin Hamel** (Benhur), mon équipe all-star sur laquelle je peux toujours compter. Merci aux boys du TMB : Ben et Franco, **Charles Couture-Lebrun** (Carlito) et **Étienne Pelletier-Gagné** (timou). Avec vous j'ai réalisé cette randonnée dont je chérirai les moments pour le restant de ma vie <3. J'ai la chance d'être entouré d'amis sensibles, drôles et intelligents, qui m'ont beaucoup appris sur moi et sur la persévérance. Merci aussi à tous les autres membres de la communauté pour tous les beaux moments: **Frédérique Legris** (ferd), **Sarah Ostiguy** (Rassa), **Rosalie Parent**, **Véronique Boucher,** et tous les autres.

Finalement, merci à ma vraie famille, celle qui me suit depuis mes tous débuts et qui m'a soutenu dans tous mes projets. J'ai la chance d'avoir une famille qui m'a offert un soutien inconditionnel et qui m'a toujours accepté comme je suis. Merci à mes parents, **Popo et Momo**, pour toutes les attentions, les appels et les petits plats quand vous saviez que j'étais dans le rush, cette thèse n'aurait pas été possible sans vous. Merci à **grand-maman Nicole**, un modèle incroyable d'altruisme, de résilience et d'ouverture à l'autre. Merci à ma sœur **Myreille**, Mimi (ou naine des montagnes :P), celle qui m'a donné envie d'aller découvrir ce qu'était la biochimie, ce qui m'a ouvert la porte à toute cette aventure! C'était cool d'avoir quelqu'un pour *debriefer* des meetings et ventiler de ce qui se passait avec le projet (ou autre) sans censure. Merci à **Nicolas Hardy** pour sa compréhension et sa bonne humeur éternelle. Merci de me remonter le moral quand j'en ai

besoin, et de me supporter dans mes projets. Ma dernière année de doc a été beaucoup plus facile à traverser grâce à toi, même si t'es un ptit bum ;).

Bon un dernier finalement (et un vrai cette fois!), mais je voulais aussi remercier certaines personnes qui, bien qu'elles n'aient aucune idée que j'existe, ont marqué mes années de doctorat. Merci à Luciano D'Orazio pour les plats réconfortants. Merci à Jean-Sébastien Girard, Olivier Niquet et Jean-Philippe Wauthier de *la Soirée est (encore) jeune* pour toutes les fois où j'ai dû m'arrêter durant ma course parce que j'éclatais de rire. Longue vie au théâtre Duceppe, toujours touchant et intelligent, et à la musica popular de Verdun!


Bienvenue les jeunes! Vous voici dans le thymus!
Pour vous, les lymphocytes T, cellules immunitaires, il s'agit de votre école, où l'on apprend comment faire partie du système immunitaire!
Trachée
Thymus
Poumons
cœur
Votre temps passé ici vous apprendra à tolérer les cellules normales, et d'ainsi pouvoir les différencier des cellules cancéreuses et des virus!
Une fois diplomé, vous pourrez quitter le thymus et faire la patrouille du corps humain!
Mais sachez bien que ce ne sera pas une mince tâche que d'obtenir votre diplôme...
Bla bla
Bla
... Car seul 1% d'entre vous arriveront jusqu'au bout.
ZAP
ZAP
Vous vous dites peut-être, mais c'est trop sévère! Hé bien NON!
Imaginez un lymphocyte T défectueux qui s'attaque à des cellules normales, cela serait catastrophique!!
Il faut être irréprochables, car vous allez être de puissants gardiens de notre écosystème!
Un lymphocyte T apprend à tolérer le produit des gènes contenus dans l'ADN des cellules normales.
CELLULE
L'ADN étant comme un grand livre de recettes contenant toute l'information nécessaire au fonctionnement des cellules.
Enfant aux yeux bleus: une recette de Ricardo
La calviti dès 32 ans selon Ricard
Mais les gènes, ce n'est que 2% de l'ADN! C'est comme si 98% du livre de recettes était incohérent à nos yeux.
Chou-banane d'oiseau mou Sauce Ricardo


Qu'est-ce qu'il y a dans ce 98%?
ZAP
Pas de questions quand je parle, c'est écrit dans le manuel.
Merci.
Néanmoins, pour répondre, parlons par exemple des éléments transposables. Ce sont des fragments d'ADN qui sont capables de dupliquer leurs séquences.
Mais en insérant ces copies de leur séquences ailleurs dans l'ADN, les éléments transposables peuvent modifier ou empêcher la fonction des gènes.
Pour se protéger, les cellules normales s'assurent donc de maintenir les éléments transposables dans un état inactif via différents mécanismes de protection.
Nous savons toutefois que les éléments transposables sont actifs dans les cellules cancéreuses, qui perdent leurs mécanismes de protection.
Puisque les éléments transposables sont inactifs dans les cellules normales, mais actifs dans les cellules cancéreuses, il est envisagé de les utiliser pour vous activer, les lymphocytes T, contre les cancers.
Il? Qui ça, il?
Des gens comme Jean-David Larouche!! Récemment ses travaux de recherche ont montré que les éléments transposables pourraient être actifs dans le thymus.
ZAP
Il tente maintenant de mieux comprendre pourquoi les éléments transposables seraient actifs dans le thymus et si nous apprenons à les tolérer lors de notre formation scolaire. Ceci pourrait aider à choisir des cibles thérapeutiques plus efficaces pour vous activer contre la tumeur d'un patient!
Voilà pour ça. Maintenant tous ceux qui ont cligné des yeux une fois pendant mon explication seront éliminés. Je vous avais averti, 1% s'en sortira. Squid Game c'était de la p'tite bière comparé au Thymus!
Saturnome '22

Chapitre 1 : Introduction

1.1 Le thymus et le développement des lymphocytes T

Le thymus est l'organe lymphoïde primaire responsable de la génération d'un répertoire de lymphocytes T fonctionnels et tolérants au soi (1). La fonction du thymus est d'apprendre aux lymphocytes T à distinguer le soi immunitaire, c'est-à-dire l'ensemble des antigènes présentés par les cellules saines de notre corps, du non-soi (2, 3). Au niveau cellulaire, le thymus contient deux populations principales : les cellules hématopoïétiques et les cellules stromales. La lignée hématopoïétique comprend majoritairement des lymphocytes T, mais également des lymphocytes B, des cellules dendritiques (DC) et des macrophages (4-9). Le compartiment stromal contient quant à lui principalement des cellules épithéliales, endothéliales et des fibroblastes (10-12).

Le thymus est séparé en deux régions principales : le cortex et la médulla thymiques (13). Au cours de leur développement, les thymocytes (lymphocytes T immatures) passent d'abord environ deux semaines dans le cortex (14), puis ils migrent vers la médulla (15) pour un séjour d'environ 4-5 jours (16). (**Figure 1.1**). Suivant leur entrée dans le thymus, les progéniteurs des thymocytes (ETP) qui migrent depuis la moelle osseuse se différencient en thymocytes double négatifs $CD4^-CD8^-$ (DN) (17, 18), stade au cours duquel ils initient l'expression de la chaine β du TCR (*T cell receptor*) (**Figure 1.1**) (19). Après la sélection β, qui assure que la chaine β du TCR est fonctionnelle, les thymocytes DN acquièrent un phénotype double positif $CD4^+CD8^+$ (DP) et commencent à exprimer la chaine α du TCR, produisant ainsi un récepteur TCRαβ complet (20). Les thymocytes DP subissent alors la sélection positive dans le cortex, qui a pour objectif d'assurer que la conformation du TCRαβ exprimé par le thymocyte est appropriée. Les thymocytes ayant survécu à la sélection positive se commettent à un statut simple positif $CD4^-CD8^+$ ou $CD4^+CD8^-$ (SP) et migrent vers la médulla thymique, où les thymocytes auto-réactifs sont éliminés lors de la sélection négative (**Figure 1.1**) (21). Plus de 95% des thymocytes seront éliminés par apoptose au cours de ces vagues de sélection pour assurer un répertoire de lymphocytes T fonctionnels et tolérants du soi (22, 23). Finalement, les thymocytes ayant survécu à la sélection négative peuvent compléter leur maturation en lymphocytes T CD4 ou CD8 naïfs qui migrent à l'extérieur du thymus

pour accomplir leur fonction de surveillance dans la périphérie (**Figure 1.1**) (16). Les sections suivantes décriront plus en détails les sélections positive et négative des thymocytes.

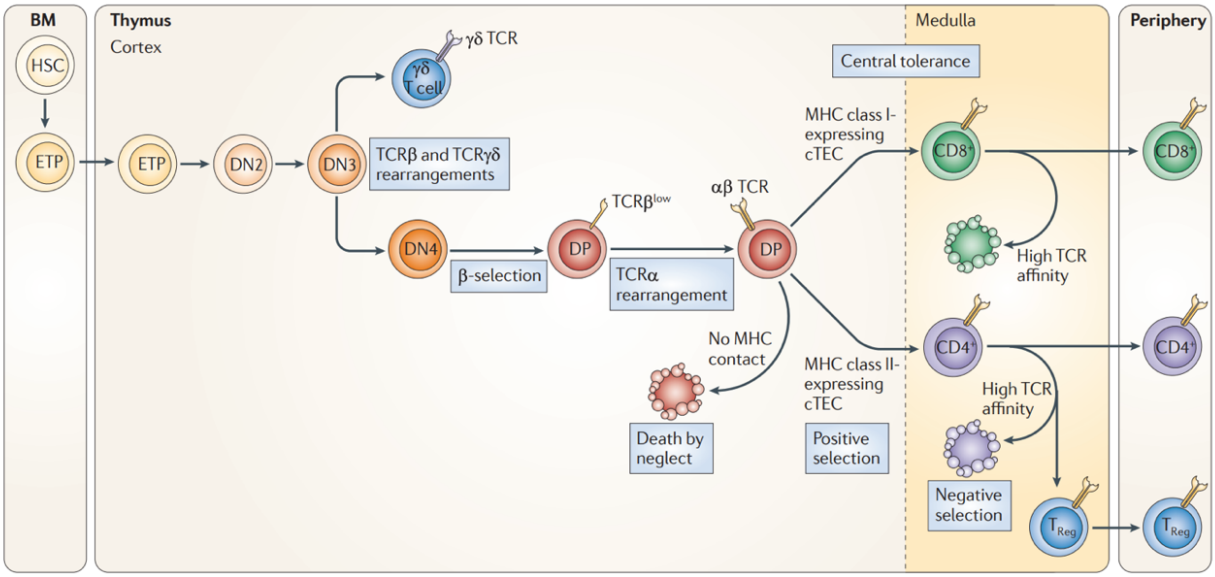


Figure 1.1 - Étapes de maturation des lymphocytes T dans le thymus.

Les précurseurs des lymphocytes T (ETP) entrent dans le cortex thymique et se différencient en thymocytes doubles négatifs $CD4^-CD8^-$ (DN). Durant leur séjour dans le cortex, les thymocytes DN amorcent l'expression de la chaine β du TCR et subissent la sélection β, réarrangent la chaine α du TCR pour exprimer un TCRα/β complet, transitent vers le stade double positif $CD4^+CD8^+$ (DP) et subissent la sélection positive. Les thymocytes DP sélectionnés peuvent alors migrer vers la médulla thymique, où ils se différencient en thymocytes simples positifs (SP) $CD4^+CD8^-$ ou $CD4^-CD8^+$ et les thymocytes auto-réactifs sont éliminés lors de la sélection positive. Les lymphocytes T matures fonctionnels et tolérants au soi peuvent émigrer du thymus vers la périphérie pour accomplir leur fonction de surveillance. Reproduit avec la permission de *Springer Nature* : Nature Reviews Immunology (Miller JFAP) ©2011 (24).

1.1.1 Présentation antigénique dans le thymus

La présentation antigénique par le complexe majeur d'histocompatibilité (CMH) est déterminante pour le développement des thymocytes (25). En effet, au cours de leur développement les thymocytes interagissent avec une grande diversité d'antigènes du soi présentés par le CMH à la

surface de cellules présentatrices d'antigènes (APC) thymiques. Ces interactions avec les APCs thymiques permettent de confirmer que le TCR exprimé par un thymocyte est fonctionnel, c'est-à-dire que sa conformation permet d'interagir avec les complexes CMH-peptides des cellules du corps, tout en ayant une faible affinité pour les antigènes du soi.

1.1.1.1 Voie de présentation classique du CMH

1.1.1.1.1 Molécules du CMH

La présentation antigénique se fait par l'entremise de deux catégories de molécules du CMH, qui sont des protéines transmembranaires aussi connues sous le nom d'antigènes des leukocytes humains (HLA) : les molécules du CMH de classe I (CMH-I), reconnues par les lymphocytes T CD8, et celles de classe II (CMH-II) (26) reconnues par les lymphocytes T CD4. Les molécules du CMH ont deux caractéristiques essentielles à leur fonction : elles sont polygéniques et hautement polymorphiques. En effet, les molécules du CMH-I sont encodées par trois gènes chez l'humain (HLA-A, HLA-B et HLA-C) et chez la souris (H2-K, H2-D et H2-L), alors que les molécules du CMH-II sont encodées par trois gènes chez l'humain (HLA-DR, HLA-DP et HLA-DQ) et deux gènes chez la souris (I-A et I-E) (27-30). Les séquences des molécules du CMH sont aussi hautement variables, avec plus de 36 000 allèles annotés chez l'humain (31). Ces variations dans la séquence des molécules du CMH permettent de diversifier le répertoire d'antigènes présentés par les cellules d'un individu (32-34). Cette grande diversité d'allèles HLA a un avantage évolutif : puisque les allèles HLA exprimés par un individu impactent sa propension aux infections par des pathogènes, augmenter le nombre total d'allèles HLA exprimés au sein de la population diminue la probabilité que la population soit décimée par un nouveau pathogène (35, 36).

Tel que mentionné plus tôt, il existe des distinctions importantes entre le CMH-I et le CMH-II, tant au niveau de leurs propriétés que des antigènes qu'ils présentent. Premièrement, alors que le CMH-I est exprimé à la surface de la quasi-totalité des cellules nuclées – bien que son niveau d'expression varie grandement entre types cellulaires (37, 38) – l'expression du CMH-II est restreinte aux cellules présentatrices d'antigènes (39). Deuxièmement, des différences importantes existent dans la structure des molécules du CMH-I et du CMH-II : le CMH-I est composé d'une chaine α ainsi que d'une sous-unité invariante β_2-microglobuline (B2M) (40), alors

que le CMH-II est un hétérodimère composé des chaines lourdes α et β (41). Ces différences structurelles ont un impact sur la nature des peptides qui peuvent être présentés par les molécules du CMH-I et du CMH-II. En effet, la région de liaison des peptides du CMH-I étant fermée à ses deux extrémités, les peptides pouvant lier le CMH-I ont généralement une longueur de 8 à 12 acides aminés (**Figure 1.2B,C**) (42). La région de liaison des peptides du CMH-II est quant à elle ouverte, permettant aux peptides de dépasser aux deux extrémités de la poche de liaison; les peptides liant le CMH-II peuvent ainsi atteindre des longueurs de 25 acides aminés (**Figure 1.2E,F**) (42). Finalement, une des différences les plus importantes entre les molécules du CMH-I et du CMH-II est l'origine des peptides qu'elles présentent : les peptides présentés par le CMH-I résultent de la dégradation des protéines endogènes à la cellule (43, 44), alors que les peptides présentés par le CMH-II proviennent de protéines exogènes captées par endocytose (**Figure 1.2A,D**) (45, 46).

Cette thèse se concentrera sur la présentation antigénique par le CMH-I pour deux raisons principales. Premièrement, le CMH-I est exprimé par presque toutes les cellules nuclées. Deuxièmement, les peptides présentés par le CMH-I sont endogènes à la cellule, et fournissent donc une représentation de son état interne (47).

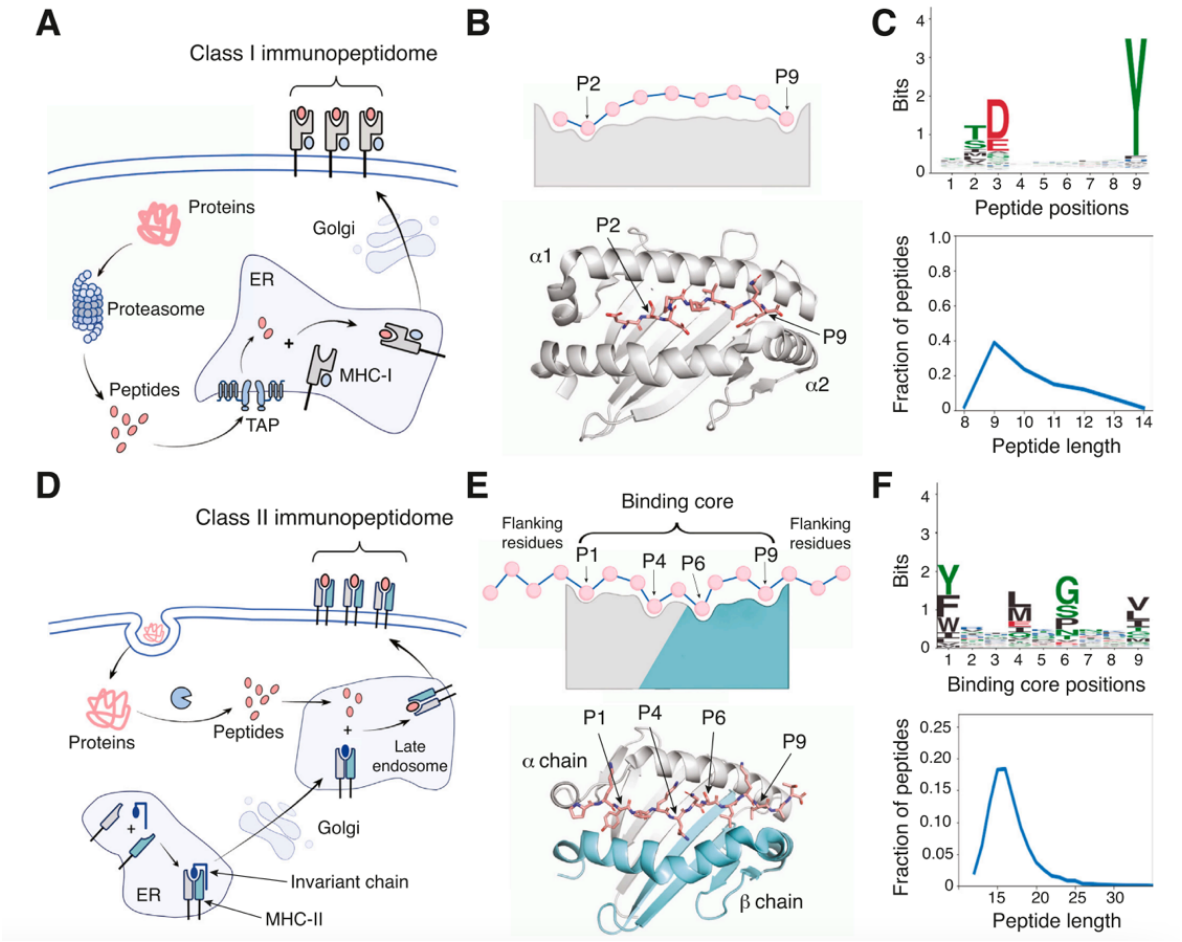


Figure 1.2 - Présentation antigénique par le CMH-I et le CMH-II

(A) Les peptides présentés par le CMH-I proviennent de la dégradation des protéines endogènes à la cellule par le complexe du protéasome. **(B)** Schématisation et structure de la poche de liaison du peptide au CMH-I. **(C)** Motifs de la séquence et longueur typique des peptides présentés par le CMH-I. **(D)** Les peptides présentés par le CMH-II proviennent de la capture de protéines exogènes par la cellule. **(E)** Schématisation et structure de la poche de liaison du peptide au CMH-II. **(F)** Motifs de la séquence et longueur typique des peptides présentés par le CMH-II. Reproduit avec la permission d'*Elsevier* : Seminars in Immunology (Gfeller D, Liu Y, Racle J) ©2023 (48).

1.1.1.1.2 Biogenèse des peptides associés au CMH-I

Les peptides associés au CMH-I (MAP) proviennent de protéines de tous les compartiments cellulaires (49-51). Ce ne sont toutefois pas toutes les protéines qui génèrent des MAPs. En effet, seul un faible pourcentage du protéome génère des MAPs : l'ensemble des MAPs couvre

seulement 10% des séquences protéiques exprimées par une cellule, et ces MAPs proviennent d'environ 60% des gènes exprimés par la cellule (52) qui sont enrichis en motifs de régulation par les micro ARNs (miRNA) (53). De plus, des études récentes ont montré que 3-10% des MAPs proviennent de régions non-codantes du génome comme les longs ARN non-codants (lncRNA), les introns, les régions non-traduites (5'UTR et 3'UTR) ou les régions intergéniques (54, 55). L'immunopeptidome, c'est-à-dire l'ensemble des MAPs présentés à la surface d'une cellule, est donc complexe puisqu'il contient des MAPs générés par une grande diversité de régions génomiques.

Traditionnellement, il a été assumé que la plupart des MAPs proviennent de la dégradation en peptides des protéines stables exprimées par la cellule par le complexe du protéasome (**Figure 1.3**) (43). Il a toutefois été démontré qu'une proportion importante des MAPs proviennent de polypeptides produits lors de la ronde pionnière de traduction (56), qui sert de contrôle de la qualité aux ARN messagers, ainsi que par des produits ribosomaux défectueux (DRiPs), c'est-à-dire des protéines mal traduites, mal repliées ou encore contenant des signaux de terminaison prématurés (**Figure 1.3**) (57, 58). Ceci permet à la cellule de présenter rapidement des MAPs aux lymphocytes T CD8 dans le contexte d'infections virales et la mise en place rapide de réponses immunitaires (59).

Les peptides générés par le protéasome sont ensuite acheminés dans le réticulum endoplasmique (ER) par les transporteurs associés au traitement des antigènes 1 et 2 (TAP1/2) (60, 61). Une fois entrés dans l'ER, l'ajout des peptides au dimère de la chaine α du CMH-I et de la sous-unité B2M est facilité par trois chaperonnes : la tapasine, la calréticuline et Erp57 (**Figure 1.3**) (62-64). Si la longueur du peptide ne permet pas sa liaison au CMH-I, sa séquence peut être clivée en N-terminal par l'aminopeptidase associée au traitement des antigènes dans l'ER (ERAAP) (65). Les peptides ne parvenant pas à lier le CMH-I ainsi que les molécules du CMH-I ayant un repliement inapproprié sont retournés dans le cytoplasme par le système de dégradation protéique associé à l'ER (ERAD) (66, 67). Les complexes CMH-peptides correctement assemblés sont quant à eux acheminés à la membrane plasmique pour être présentés aux lymphocytes T CD8 (**Figure 1.3**).

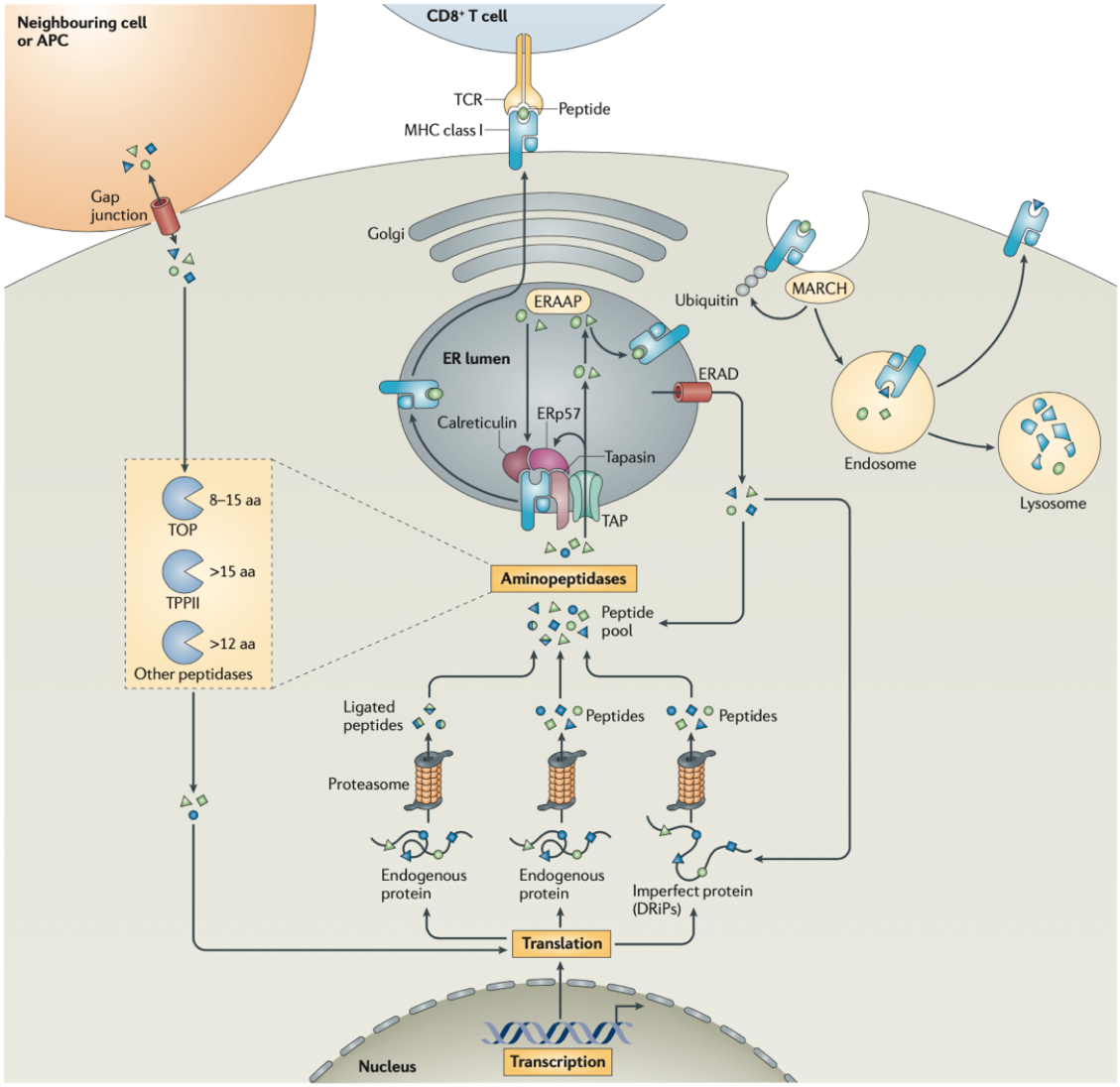


Figure 1.3 - Voie de présentation antigénique par le CMH-I.

Les protéines exprimées par la cellule ainsi que les produits ribosomaux défectueux (DRiPs) sont dégradés en peptides par le complexe du protéasome. Les peptides sont ensuite acheminés dans le réticulum endoplasmique (ER) par les transporteurs TAP1/2, et incorporés aux molécules du CMH-I stabilisées par les chaperonnes tapasine, ERp57 et calréticuline. Les peptides n'ayant pas une longueur appropriée pour lier le CMH-I peuvent être clivés en N-terminal par des aminopeptidases dans le cytoplasme ou dans l'ER (ERAAP). Les complexes CMH-peptides sont acheminés à la membrane plasmique, alors que les peptides ne parvenant toujours pas à lier le CMH-I sont retournés dans le cytoplasme par le système de dégradation protéique ERAD.

Reproduit avec la permission de *Springer Nature* : Nature Reviews Immunology (Neefjes J, Jongsma MLM, Paul P & Bakke O) ©2011 (68).

1.1.2 Vagues de sélection des lymphocytes T

1.1.2.1 Sélection positive dans le cortex

Suite à la sélection β, la première vague de sélection que les thymocytes exprimant un TCRαβ complet doivent traverser est la sélection positive, qui a lieu dans le cortex et est médiée par les cellules épithéliales thymiques du cortex (cTEC) (**Figure 1.4**) (69-71). La sélection positive assure que les TCRαβ exprimés par les thymocytes ont une affinité suffisante pour les complexes CMH-peptides du soi. Les thymocytes ayant un TCRαβ capable d'interagir avec le CMH-I se différencient en thymocytes $CD8^+$ SP, alors que ceux ayant un TCRαβ restreint au CMH-II deviennent des thymocytes $CD4^+$ SP. Les thymocytes dont le TCRαβ a peu ou pas d'affinité pour les complexes CMH-peptides entrent en apoptose, un phénomène nommé mort par négligence (**Figure 1.5**) (72).

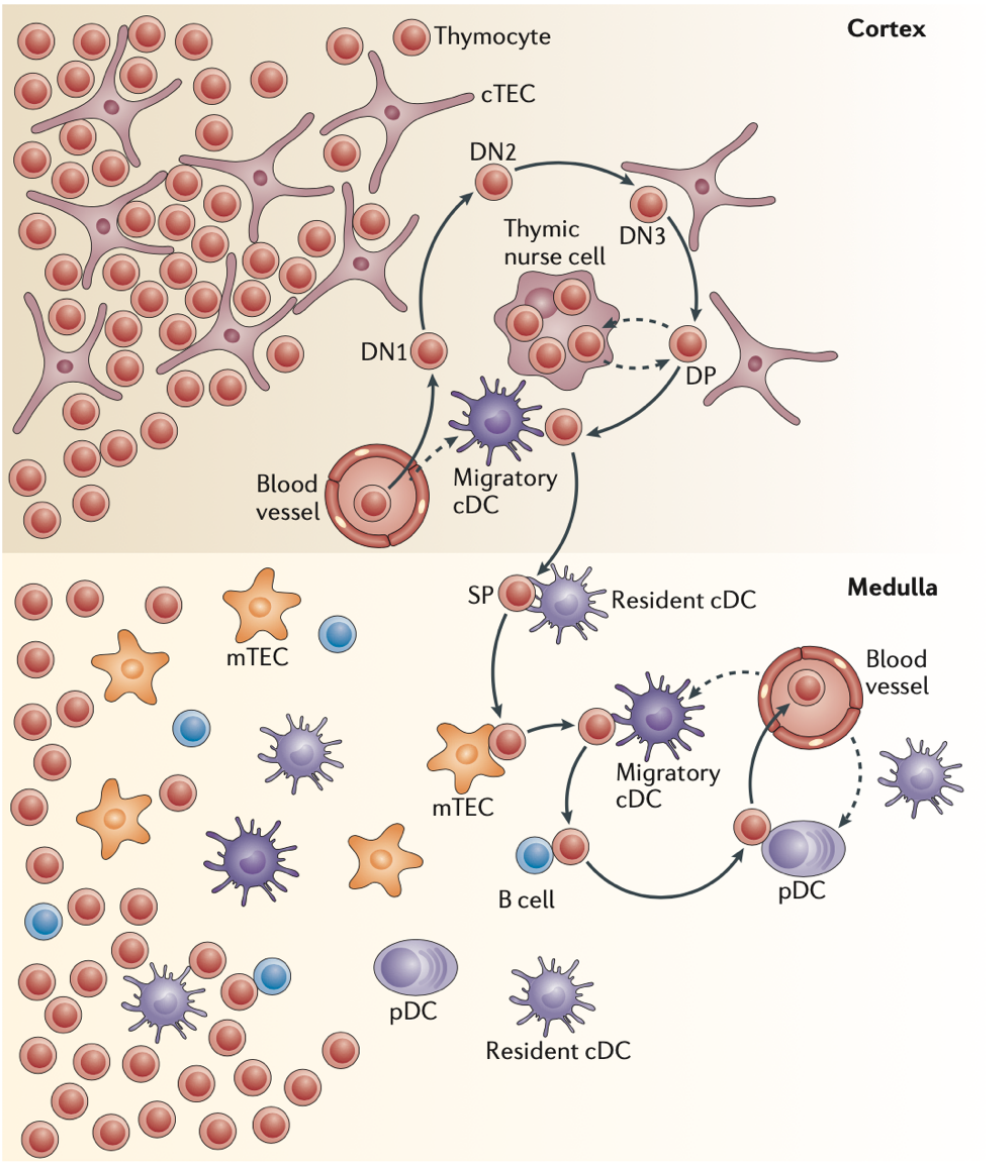


Figure 1.4 - Sélections positive et négative des thymocytes.

Suite à leur entrée par les vaisseaux sanguins de la jonction cortico-médullaire, les thymocytes DP exprimant un TCRα/β subissent la sélection positive dans le cortex en formant des interactions avec les molécules du CMH exprimées par les cTECs pour valider la fonctionnalité de leur TCR. Les thymocytes sélectionnés peuvent ensuite migrer vers la médulla et se différencier en thymocytes CD4 SP ou CD8 SP. Dans la médulla, les thymocytes scannent les antigènes du soi présentés par les différentes populations d'APCs médullaires (mTECs, DCs, lymphocytes B et fibroblastes), et les thymocytes auto-réactifs sont éliminés lors de la sélection négative. Reproduit avec la permission

de *Springer Nature* : Nature Reviews Immunology (Klein L, Kyewski B, Allen PM & Hogquist KA) ©2014 (21).

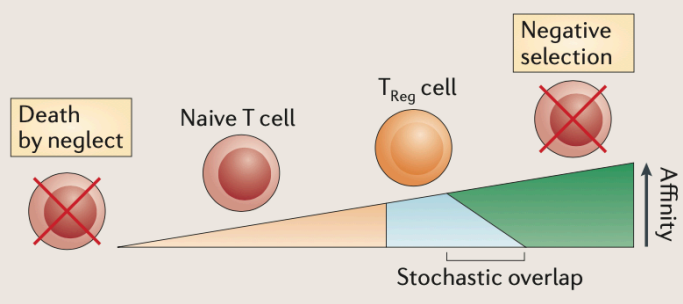


Figure 1.5 - Modèle de sélection des thymocytes par affinité du TCR

Dans ce modèle, le développement des thymocytes est déterminé par l'affinité de leur TCR pour les complexes CMH-peptides des APCs thymiques. Les thymocytes dont le TCR a une affinité trop faible meurent par négligence, alors que ceux ayant une affinité forte pour le soi sont éliminés lors de la sélection négative ou redirigés vers un état de lymphocytes T régulateurs (Tregs). Les thymocytes ayant une affinité faible ou intermédiaire pour le soi peuvent quant à eux compléter leur maturation en lymphocytes T naïfs. Reproduit avec la permission de *Springer Nature* : Nature Reviews Immunology (Klein L, Kyewski B, Allen PM & Hogquist KA) ©2014 (21).

Une caractéristique unique des cTECs est la présentation d'un répertoire distinct d'antigènes du soi par le CMH. Ceci est dû à l'expression de trois protéines impliquées dans la voie de présentation antigénique: la sous-unité β5t du thymoprotéasome (CMH-I), la cathepsine L et la sérine protéase thymus-spécifique (TSSP) (CMH-II) (73-76). La délétion de ces trois protéines entraine des défauts de différenciation en thymocytes SP, démontrant leur implication dans la sélection positive des thymocytes (77-79). La fréquence et l'abondance des peptides présentés CMH spécifiques aux cTECs, qui sont générés par ces trois protéines, restent cependant obscures. En effet, lorsque des peptides présentés par le CMH-II à la surface des cTECs ont été séquencés, il a été observé que ces antigènes étaient aussi présentés par des APCs spléniques (80). Il est toutefois à noter que seulement 12 antigènes parmi les plus abondants – ces 12 peptides occupaient environ 20% de toutes les molécules du CMH-II exprimées à la surface des cTECs – ont

été séquencés en raison de limitations techniques; les peptides spécifiques aux cTECs générés par le thymoprotéasome, la cathepsine L ou la TSSP pourraient être moins abondants et ne pas avoir été séquencés.

1.1.2.2 Sélection négative dans la médulla

Les thymocytes migrent ensuite vers la médulla thymique pour y subir la sélection négative, qui a pour but d'établir la tolérance au soi des lymphocytes T, ou tolérance centrale (81). Cette étape repose sur un mécanisme inverse à la sélection positive : les thymocytes dont le TCRα/β a une affinité trop grande pour les antigènes du soi sont éliminés par apoptose, alors que ceux ayant une affinité faible ou intermédiaire, c'est-à-dire les thymocytes tolérants au soi, survivent et peuvent compléter leur maturation (**Figure 1.5**) (82-85). Il a toutefois été montré que certains thymocytes ayant une grande affinité pour le soi peuvent éviter l'apoptose et se différencier en lymphocytes T régulateurs (Treg), des cellules ayant une fonction immunosuppressive (**Figure 1.5**) (86-89). L'établissement de la tolérance centrale se ferait donc via deux mécanismes : l'élimination par apoptose des thymocytes auto-réactifs et la génération de Tregs.

Contrairement à la sélection positive, qui est médiée uniquement par les cTECs, plusieurs types cellulaires différents sont impliqués dans la sélection négative des thymocytes. Les deux principaux types cellulaires dirigeant la sélection négative sont les cellules épithéliales thymiques de la médulla (mTEC) ainsi que les cellules dendritiques (DC) (**Figure 1.4**) (90-93). Deux autres populations de cellules, les lymphocytes B et les fibroblastes, ont des rôles plus limités dans la sélection négative des thymocytes CD4 SP et CD8 SP, respectivement (94, 95). Néanmoins, leur contribution est importante à l'établissement de la tolérance centrale puisque les souris dont la présentation antigénique est perturbée dans les populations thymiques de lymphocytes B et de fibroblastes développent des maladies auto-immunes.

Les mTECs sont une population présentant une grande hétérogénéité tant au niveau phénotypique que fonctionnel. De fait, il existe plusieurs sous-populations de mTECs ayant des fonctions distinctes dans le développement des thymocytes (**Figure 1.6**). Premièrement, les mTECs CMH-II^{lo} CD80^{lo} CCL21^{+} (mTEC(I)) sont une sous-population de mTECs immatures qui sécrètent la chimiokine CCL21, dont la fonction est de stimuler la migration des thymocytes DP

du cortex vers la médulla (96). Les mTEC(I) poursuivent ensuite leur maturation et forment la sous-population de mTECs CMH-II^{hi} CD80^{hi} (mTEC(II)), des mTECs matures qui sont essentiels à l'établissement de la tolérance au soi (97). Les mTEC(II) présentent une grande diversité d'antigènes du soi grâce à l'expression génique promiscuitaire (PGE), un mécanisme unique de régulation génique. Finalement, des études récentes ont montré que différentes populations de mTECs mimétiques, c'est-à-dire des mTECs présentant des similarités phénotypiques avec des cellules différenciées de la périphérie, peuvent aussi se développer à partir des mTEC(II) (**Figure 1.6**): une population principale de mTECs CMH-II^{lo} CD80^{lo} KRT10^{+} (mTEC(III)) présentant des similitudes avec des cornéocytes (98, 99) ainsi que plusieurs autres populations plus rares (100). Ces populations rares de cellules mimétiques comprennent des cellules similaires aux cellules basales du poumon, aux cellules *tuft* intestinales, aux cellules M, aux cellules musculaires, aux cellules neuroendocrines, aux cellules ciliées et aux ionocytes (101-103). Ces populations de cellules mimétiques permettent la présentation d'antigènes spécifiques à certains types cellulaires de la périphérie aux thymocytes.

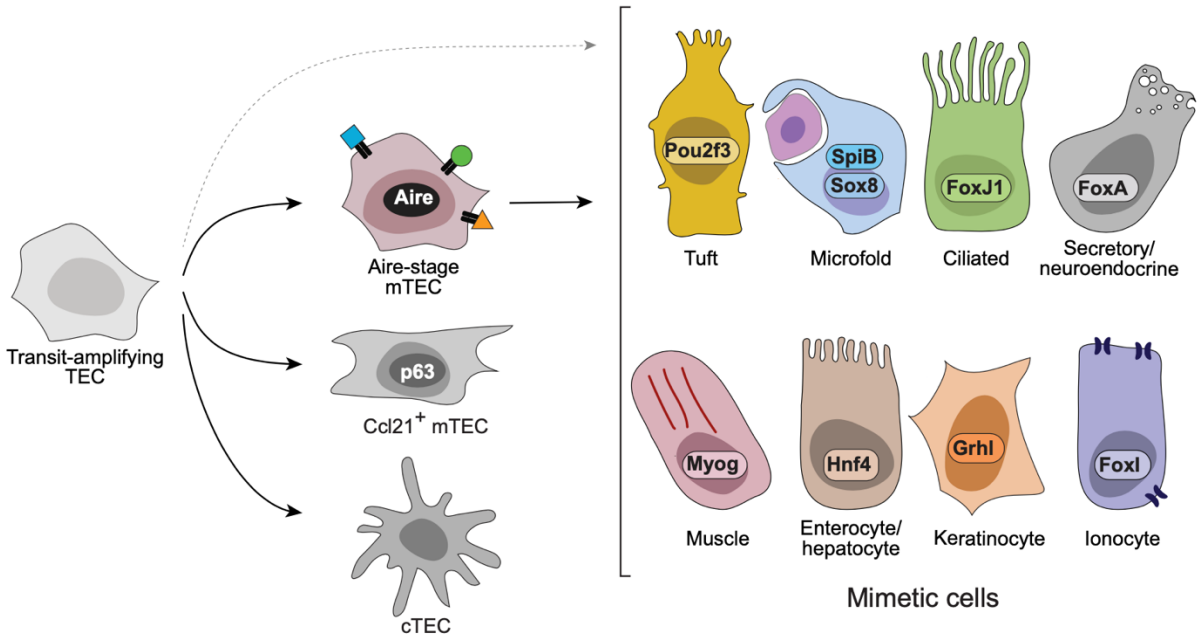


Figure 1.6 - Hétérogénéité des cellules épithéliales thymiques.

Les progéniteurs des cellules épithéliales thymiques (TEC) (*Transit-amplifying TEC*) se différencient en cTECs ou en mTECs. La population de mTECs est divisée en plusieurs sous-

populations : les mTEC(I) CCL21^{+} et les mTEC(II) (*Aire-stage mTEC*). Les mTEC(II) peuvent quant à elles poursuivre leur maturation en cellules mimétiques présentant des similarités avec des cellules différenciées retrouvées dans les tissus périphériques. Reproduit avec la permission de *Cell Press* : Trends in Immunology (Michelson DA, Mathis D) ©2023 (104).

Les cellules dendritiques (DC) sont le deuxième groupe de cellules centrales à l'établissement de la tolérance centrale. On retrouve trois sous-populations de DCs thymiques, qui sont localisées dans la médulla thymique : les DC conventionnelles de type 1 (cDC1) et de type 2 (cDC2) ainsi que les DCs plasmacytoïdes (pDC). Les pDCs et les cDC1 se développent dans le thymus, alors que les cDC2 migrent à partir de la périphérie (105-108). Les DCs thymiques peuvent présenter des antigènes captés dans les vaisseaux sanguins à la jonction cortico-médullaire du thymus qui ne sont pas nécessairement exprimés par les mTECs (109); les DCs thymiques jouent donc un rôle essentiel dans l'établissement de la tolérance centrale. Finalement, il a été montré que des antigènes exprimés par les mTECs peuvent être transférés aux DCs thymiques par transport vésiculaire (110-112). Les pDCs, bien qu'elles contribuent elles aussi à la présentation antigénique dans le thymus, sécrètent de l'interféron (IFN) α/β façon constitutive (113). Cette propriété est unique aux pDCs thymiques, puisque les pDCs extrathymiques ne sécrètent de l'IFNα/β qu'en présence de pathogènes. Dans le thymus, il a été démontré que l'IFNα/β régule les stades de maturation tardive des thymocytes en stimulant la génération de Tregs et de lymphocytes T CD8 innés (114-118).

1.1.4 Expression génique promiscuitaire

1.1.4.1 Expression de gènes tissu-spécifiques et tolérance centrale

La fonction principale du thymus est d'induire la tolérance des lymphocytes T à l'endroit des protéines exprimées par les différents tissus de la périphérie, c'est-à-dire hors du thymus. Pour ce faire, les mTECs matures (principalement les mTEC(II), et dans une moindre mesure les mTECs mimétiques) expriment une grande diversité de séquences, un phénomène nommé expression génique promiscuitaire (PGE) qui est essentiel à l'établissement de la tolérance au soi des

lymphocytes T. De fait, les mTECs murines expriment près de 20 000 gènes, ce qui représente environ 85% du gènes codant pour des protéines, un nombre bien plus élevé que les 12 000-14 000 gènes typiquement exprimés dans les tissus périphériques (119-121). De façon intéressante, le répertoire des gènes exprimés par les mTECs inclue un grand nombre de gènes et d'isoformes tissu-spécifiques (122-125); il est présumé que l'immunopeptidome des mTECs récapitule ainsi le répertoire des antigènes du soi que les lymphocytes T rencontreront dans la périphérie. Il a toutefois été montré que la PGE ne cible pas les gènes de façon stochastique : les gènes spécifiques aux tissus peu ou pas exposés à la surveillance immunitaire comme le cerveau ou les testicules sont sous-représentés parmi les gènes exprimés par les mTECs (121), suggérant que la tolérance centrale peut être plus permissive envers les antigènes de certains tissus. De plus, des mécanismes de tolérance périphérique sont en place pour limiter l'activation des lymphocytes T contre des antigènes du soi puisque des lymphocytes T autoréactifs sont présents dans le sang d'individus sains sans qu'ils ne développent de maladies auto-immunes (126, 127).

1.1.4.2 Régulateurs de la PGE

Trois protéines ont été identifiées comme les principales régulatrices de la PGE : le régulateur autoimmun (AIRE), la protéine à doigt de zinc du prosencéphale embryonnaire 2 (FEZF2), et la protéine hélicase à ADN à chromodomaine (CHD4). Ces trois protéines ont des rôles non-redondants dans l'établissement de la tolérance centrale en stimulant l'expression de répertoires de gènes majoritairement distincts (128-130). Des données récentes de séquençage à cellule unique (scRNA-seq) ont montré que l'expression des gènes tissu-spécifiques par les mTECs est fortement associée à l'expression d'AIRE et de FEZF2 (101, 131). Ces trois protéines régulent toutefois l'expression génique via des mécanismes distincts (**Figure 1.7**). De fait, alors que FEZF2 agit comme un facteur de transcription classique (132), AIRE ne lie pas une séquence d'ADN précise, mais plutôt des marques épigénétiques répressives. En effet, les gènes ciblés par AIRE sont associés avec des marques de chromatine inactive, telles que la triméthylation de la lysine 27 de l'histone 3 (H3K27me3), H3K9me3, ou encore la méthylation de l'ADN (133-135) (136). Finalement, CHD4 agit en remodelant la chromatine au niveau des promoteurs des gènes cibles de FEZF2 et au niveau des super-activateurs (*super-enhancers*) des cibles d'AIRE, ce qui permet l'induction de l'expression des gènes tissus-spécifiques par FEZF2 et AIRE (**Figure 1.7**) (130).

Chacun de ces trois régulateurs est essentiel à l'établissement de la tolérance centrale; la perte de la fonction d'une de ces protéines est suffisante pour causer des réactions auto-immunes affectant plusieurs organes (128-130).

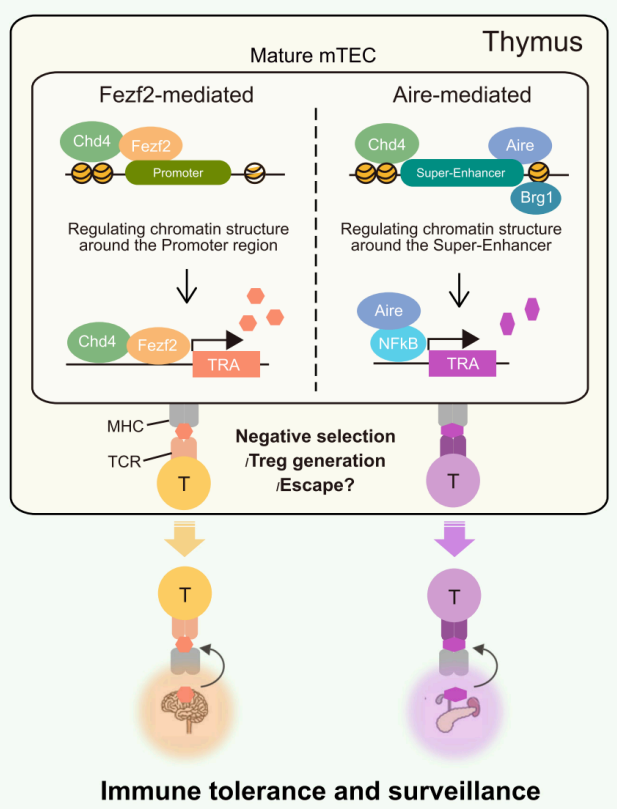


Figure 1.7 - Induction de l'expression des gènes tissus-spécifiques par AIRE et FEZF2.

AIRE, FEZF2 et CHD4 stimulent l'expression de gènes tissus-spécifiques dans les mTECs matures (mTEC(II)) dans le cadre de l'expression génique promiscuitaire via des mécanismes distincts. FEZF2 lie la séquence de son motif de liaison à l'ADN au promoteur de ses gènes cibles en formant des interactions avec CHD4, un régulateur épigénétique. AIRE et CHD4 lient tous les deux la séquence des super-activateurs (*super-enhancers*) des gènes cibles d'AIRE sans former d'interaction directe. Le remodelage de la chromatine aux régions des super-activateurs permet à AIRE d'initier la transcription de ses gènes cibles sans implication de CHD4. Reproduit avec la permission de *Springer Nature* : Inflammation and Regeneration (Benlaribi R, Gou Q & Takaba H) ©2022 (137).

1.1.4.3 Mosaïcisme de la médulla thymique

L'expression de près de 85% des gènes pose cependant un dilemme pour les mTECs : l'expression d'une aussi grande diversité de protéines pourrait induire un stress protéotoxique délétère pour les cellules. Il a cependant été montré que l'expression des gènes tissu-spécifiques (TRG), et plus spécifiquement des TRGs induits par AIRE, est mosaïque au sein des mTECs, avec seulement 3-5% des mTECs exprimant un TRG donné (**Figure 1.8**) (119, 130). L'expression des TRGs ne semble toutefois pas être stochastique, puisque des sous-groupes de gènes colocalisés dans des régions chromosomiques précises sont coexprimés par les mTECs (138, 139). De plus, le complexe du protéasome semble jouer un rôle central dans la prévention du stress protéotoxique en dégradant les protéines exprimées par les mTECs : l'inhibition de son activité cause une diminution du nombre de progéniteurs mTECs, une atrophie thymique ainsi que de l'auto-immunité (140).

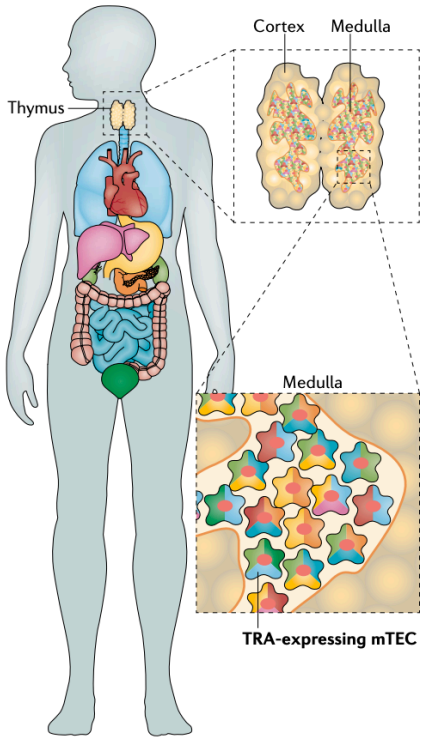


Figure 1.8 - Mosaïcité de la médulla thymique.

En tant que population, les mTEC(II) expriment environ 85% des gènes du génome humain, ce qui permet de présenter aux thymocytes un répertoire d'antigènes représentatif des différents organes périphériques (représentés par les différentes couleurs). Les gènes tissus-spécifiques ne sont toutefois exprimés que par 3-5% des mTECs, créant une grande hétérogénéité au niveau de l'expression génique au sein de la médulla thymique. Reproduit avec la permission de *Springer Nature* : Nature Reviews Immunology (Kadouri N, Nevo S, Goldfarb Y & Abramsom J) ©2020 (141).

La dégradation protéique par le protéasome ne semble toutefois pas affecter toutes les protéines exprimées dans le cadre de la PGE. En effet, il a été montré que l'expression de certains facteurs de transcription importants pour la différenciation cellulaire mène à la formation des populations

rares de mTECs mimétiques (100). Bien que des expériences de lignage cellulaire aient montré que les mTECs mimétiques émergent des mTECs exprimant AIRE, l'expression de AIRE ne semble pas essentielle au développement des mTECs mimétiques puisque ces populations sont présentes, bien qu'en nombre restreint, lorsqu'AIRE est déplété (100). Ces mTECs mimétiques possèdent des caractéristiques phénotypiques de cellules périphériques différenciées (par exemple, de kératinocytes, de cellules musculaires, ou de cellules ciliées) (103, 142-144). Ces cellules conservent tout de même une signature génique propre aux mTECs, révélant que leur différenciation en cellules périphériques est seulement partielle (100). Les mTECs mimétiques semblent néanmoins importantes pour l'établissement de la tolérance centrale, puisqu'induire l'expression d'antigènes dans ces cellules est suffisant pour causer la délétion des lymphocytes T spécifiques à ces antigènes (100).

1.2 Les éléments transposables

Les éléments transposables (TE) sont des séquences répétitives ayant la capacité de dupliquer leur séquence dans l'ADN qui représentent entre le tiers et la moitié des génomes des mammifères (145, 146). Bien qu'ils soient considérés comme des séquences non-codantes de l'ADN, les TEs contiennent fréquemment des séquences codant pour des protéines (147). L'intégration de ces séquences dans le génome pose un dilemme cornélien pour les cellules hôtes. En effet, l'insertion de TEs dans le génome représente une opportunité évolutive exceptionnelle pour l'hôte, mais menace par le fait même l'intégrité du génome. L'organisme hôte doit donc maintenir un fragile équilibre entre la répression de leur expression et l'exaptation des fonctions des TE procurant un avantage évolutif.

1.2.1 Classification et caractéristiques

Les TEs représentent une grande diversité de séquences classées en fonction de leurs structures et de leurs mécanismes de transposition (**Figure 1.9**) (148). Les TEs sont généralement séparés en 2 catégories principales : les éléments de classe 1, aussi connus sous le nom de rétroéléments endogènes (ERE), et les éléments de classe 2, ou transposons à ADN. Les EREs sont eux-mêmes subdivisés en trois classes : les rétrovirus endogènes (ERV), qui possèdent de longues séquences terminales répétées (LTR), ainsi que les éléments nucléaires dispersés longs et courts (LINE et

SINE, respectivement) qui ne possèdent pas de LTRs (148). Les TEs sont finalement regroupés selon leur homologie en sous-familles ayant de quelques dizaines à des centaines de milliers de copies de leurs séquences dans le génome (**Figure 1.9**) (149). En parallèle, les TEs peuvent être qualifiés d'autonomes s'ils encodent toute la machinerie nécessaire à leur transposition, ou de non-autonomes s'ils doivent emprunter la machinerie de TEs autonomes pour se mobiliser (148).

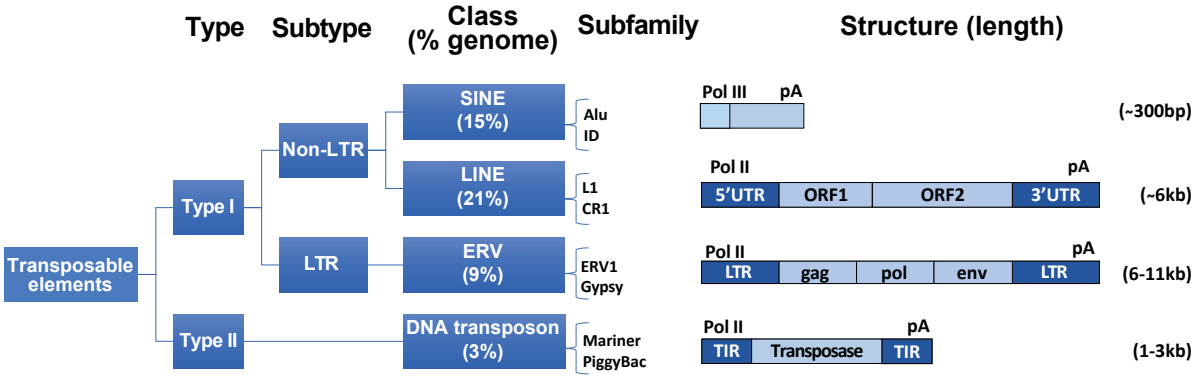


Figure 1.9 – Classification des TEs.

À gauche: classification des TEs en types, sous-types, classes et sous-familles. Le pourcentage du génome occupé par chaque classe de TEs est indiqué. À droite: Structure typique et longueur de chaque classe de TEs. Le type d'ARN polymérase responsable de la transcription de chaque type de TE et la présence d'un signal de polyadénylation sont indiqués. Reproduit avec permission d'*Elsevier* : Trends in Cancer (Ishak CA, Classon M & De Carvalho DD) ©2018 (150).

1.2.1.1 Transposons à ADN

Les transposons à ADN sont des séquences autonomes représentant 3% du génome humain qui mobilisent leurs séquences sans passer par un intermédiaire à ARN (151). Des analyses phylogénétiques ont montré que les séquences de plusieurs transposons à ADN sont associées à celles de bactéries (152-155). Deux mécanismes distincts peuvent être utilisés par les transposons à ADN pour mobiliser leur séquence (**Figure 1.10**): i) en excisant la séquence du transposon via des cassures double-brin au site donneur (156), ou ii) en clivant uniquement le brin sens du transposon, puis en synthétisant le brin antisens pour générer un intermédiaire double-brin qui

sera inséré au site receveur (157, 158). Les transposons à ADN sont donc des séquences d'une longueur de 1 000 à 3 000 nucléotides contenant la séquence d'une seule protéine, typiquement une transposase ou une hélicase essentielle à leur transposition, flanqués de séquences terminales répétées inversées (TIR) et dont l'expression est initiée par l'ARN polymérase II (**Figure 1.9**) (150, 154).

1.2.1.2 Éléments à LTRs

Les éléments à LTRs sont des rétrotransposons autonomes représentant 8-9% du génome humain (151). Les éléments à LTRs proviennent de rétrovirus ayant infecté la lignée germinale, permettant leur transmission verticale aux générations subséquentes (159, 160). Comme leur nom le suggère, les éléments à LTRs utilisent un mécanisme de transposition basé sur la rétrotranscription (161). Lors de leur réplication, l'expression des éléments à LTRs mène à la formation de capsules pseudo-virales dans lesquelles survient la rétrotranscription de l'élément à LTRs (162). L'ADN complémentaire ainsi formé peut ensuite être inséré dans le génome de la cellule hôte par le biais d'une intégrase (**Figure 1.10**) (163). La séquence des éléments à LTRs, typiquement d'une longueur de 6 000 à 11 000 nucléotides, contient trois gènes (*gag*, *pol* et *env*) transcrits par l'ARN polymérase II et flanqués de deux séquences LTRs (**Figure 1.9**) (164). Les gènes *gag* et *pol* codent pour des polyprotéines clivées de façon post-traductionnelle par une protéase encodée par *pol*. Le gène *gag* contient toute l'information nécessaire à la formation de la capsule pseudo-virale, alors que le gène *pol* encode une protéase, une transcriptase inverse, la ribonucléase H, et une intégrase (165, 166). Le gène *env* code quant à lui pour les protéines permettant la fusion de la capsule pseudo-virale avec la membrane plasmique, permettant la transmission horizontale à d'autres cellules (167, 168). La majorité des éléments à LTRs ont toutefois perdu leur gène *env* en raison de troncations et de mutations dans sa séquence, et ne peuvent donc se dupliquer dans le génome que de façon intracellulaire (169). Dans les génomes humain et murin, les éléments à LTRs sont enrichis dans les régions de la chromatine riches en AT et associées à des marques d'histones répressives (170).

1.2.1.3 LINE

Les LINEs sont des rétrotransposons autonomes représentant 21% du génome humain (151). Le mécanisme de transposition des LINEs s'appelle la rétrotranscription médiée par la cible : une endonucluéase encodée par le LINE cause un bris d'ADN simple-brin à sa séquence cible, puis le fragment d'ADN simple-brin ainsi formé est utilisé comme amorce pour la rétrotranscription (**Figure 1.10**) (171, 172). La séquence LINE, d'une longueur d'environ 6 000 nucléotides, contient typiquement deux gènes (ORF1 et ORF2) transcrits par l'ARN polymérase II (**Figure 1.9**) (173). La protéine générée par ORF1 est une chaperonne liant et stabilisant les transcrits dérivés du LINE (174), alors que la protéine codée par ORF2 possède les activités endonucléase et transcriptase inverse (175, 176). En raison de la séquence cible reconnue par la protéine ORF2 (5'-TT/AAAA-3'), les LINE sont enrichis dans les régions génomiques riches en AT (177). Chez l'humain, les éléments de la sous-famille L1 sont les seuls LINEs à avoir conservé la capacité de transposer de façon autonome, bien que seule une infime fraction de ces séquences soient actives (environ 100 des 500 000 séquences L1 dans le génome humain) (178, 179).

1.2.1.4 SINE

Les SINEs sont des rétrotransposons non-autonomes représentant environ 15% du génome humain (151). Ils empruntent donc la machinerie de transposition des LINEs pour se dupliquer. Bien qu'ils utilisent la même machinerie de réplication, les séquences des SINEs ne sont pas enrichies dans les mêmes régions du génome que les LINEs et sont plutôt retrouvées dans des régions riches en GC (180). Les SINEs étant des séquences dérivées des ARN de transfert ou ribosomaux, leurs séquences sont courtes (environ 600 nucléotides) et hétérogènes, et leur transcription est initiée par l'ARN polymérase III (181-183). Une seule sous-famille SINE, la sous-famille Alu, est toujours active dans le génome humain en détournant la machinerie des éléments L1 (184).

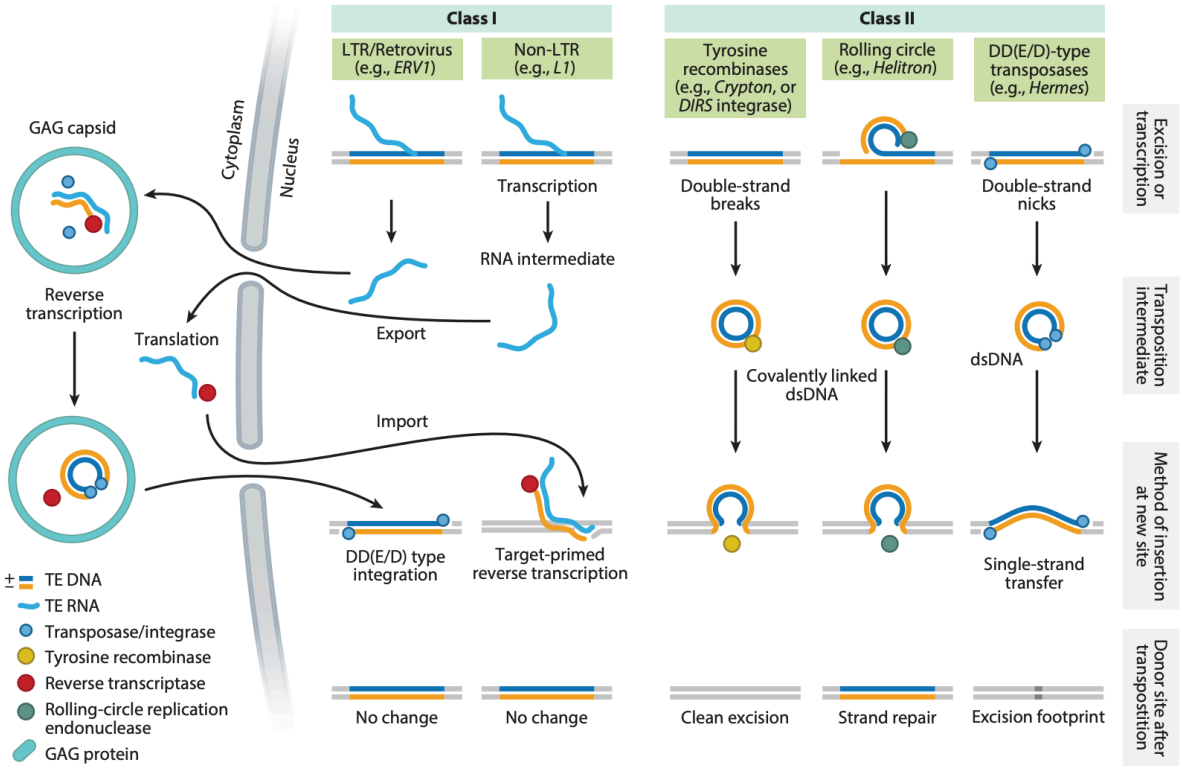


Figure 1.10 - Mécanismes de transposition des TEs.

Résumé schématisé des mécanismes de transposition des éléments transposables de classe I (à gauche) ou de classe II (à droite). Les éléments à LTRs produisent une capsule pseudo-virale, dans laquelle survient la rétrotranscription de leurs ARN messagers (ARNm). Les éléments sans LTR (LINEs et SINEs) utilisent plutôt un mécanisme de rétrotranscription médiée par la cible. Les éléments de classe II, ou transposons à ADN, peuvent utiliser plusieurs mécanismes de transposition. Les deux principaux sont i) via des cassures double-brin (*Tyrosine recombinases* et *DD(E/D)-transposases*) ou ii) en générant une cassure simple-brin et en synthétisant le brin complémentaire pour produire un intermédiaire d'ADN double-brin (*Rolling circle*). Reproduit avec permission d'*Annual Reviews* : Annual Review of Genetics (Wells JN & Feschotte C) ©2020 (148).

1.2.2 Impact des éléments transposables sur l'évolution du génome

L'intégration de TEs dans le génome représente à la fois une source de séquences régulatrices et de protéines ayant des fonctions variées pour la cellule hôte, et une menace pour la stabilité chromosomique (185, 186). En effet, les éléments transposables s'insèrent dans le génome sans considération pour les impacts qu'ils ont sur la cellule hôte et peuvent ainsi perturber la séquence de gènes (187-189). Les éléments transposables sont ainsi un couteau à double tranchant pour leurs cellules hôtes, et sont donc sujets à une pression évolutive forte pour les fixer lorsqu'ils sont bénéfiques ou les éliminer lorsqu'ils sont délétères pour l'hôte.

1.2.2.1 Transposition des TEs dans le génome

Après leur intégration dans le génome, les TEs entrent typiquement dans une phase d'amplification, durant laquelle le nombre de copies de leurs séquences augmente rapidement (190-192). Ces insertions dans le génome peuvent toutefois causer des évènements de délétion, d'amplification et de réorganisation chromosomique associés au cancer et à l'infertilité (185, 193, 194). Ainsi, la transposition des TEs dans le génome est soumise à une forte pression de purification pour éliminer les évènements d'insertion ayant des impacts négatifs sur l'hôte (195-197). Cette sélection de purification couplée à la dérive génétique entraine donc la dégénérescence progressive de la majorité des TEs (198-200). Bien que la quasi-totalité des TEs ait perdu la capacité de transposer leurs séquences, certains TEs ou certaines régions de leurs séquences peuvent être préservés et fixés dans la population s'ils procurent un avantage évolutif à leur hôte (201, 202). Des études phylogénétiques ont montré que les TEs fixés sont peu conservés entre espèces, suggérant que la fixation des TEs répond à des besoins précis de l'hôte, notamment l'adaptation des espèces à leur environnement (203-206). Ce phénomène, nommé exaptation ou domestication, permet à l'organisme hôte de détourner une propriété ou une fonction d'un TE à son avantage.

1.2.2.2 Exaptation des TEs

L'exaptation des TEs est une source importante de séquences régulatrices, d'exons, de gènes et d'ARN non-codants pour l'organisme hôte (**Figure 1.11**) (207). Ces fonctions domestiquées des TEs peuvent ainsi avoir de nombreux impacts sur le développement et le fonctionnement de

l'organisme hôte. Par soucis de clarté, les fonctions exaptées des TEs dans le développement et la fonction du système immunitaire seront abordés dans la section 1.3, « Interactions entre les TEs et le système immunitaire ».

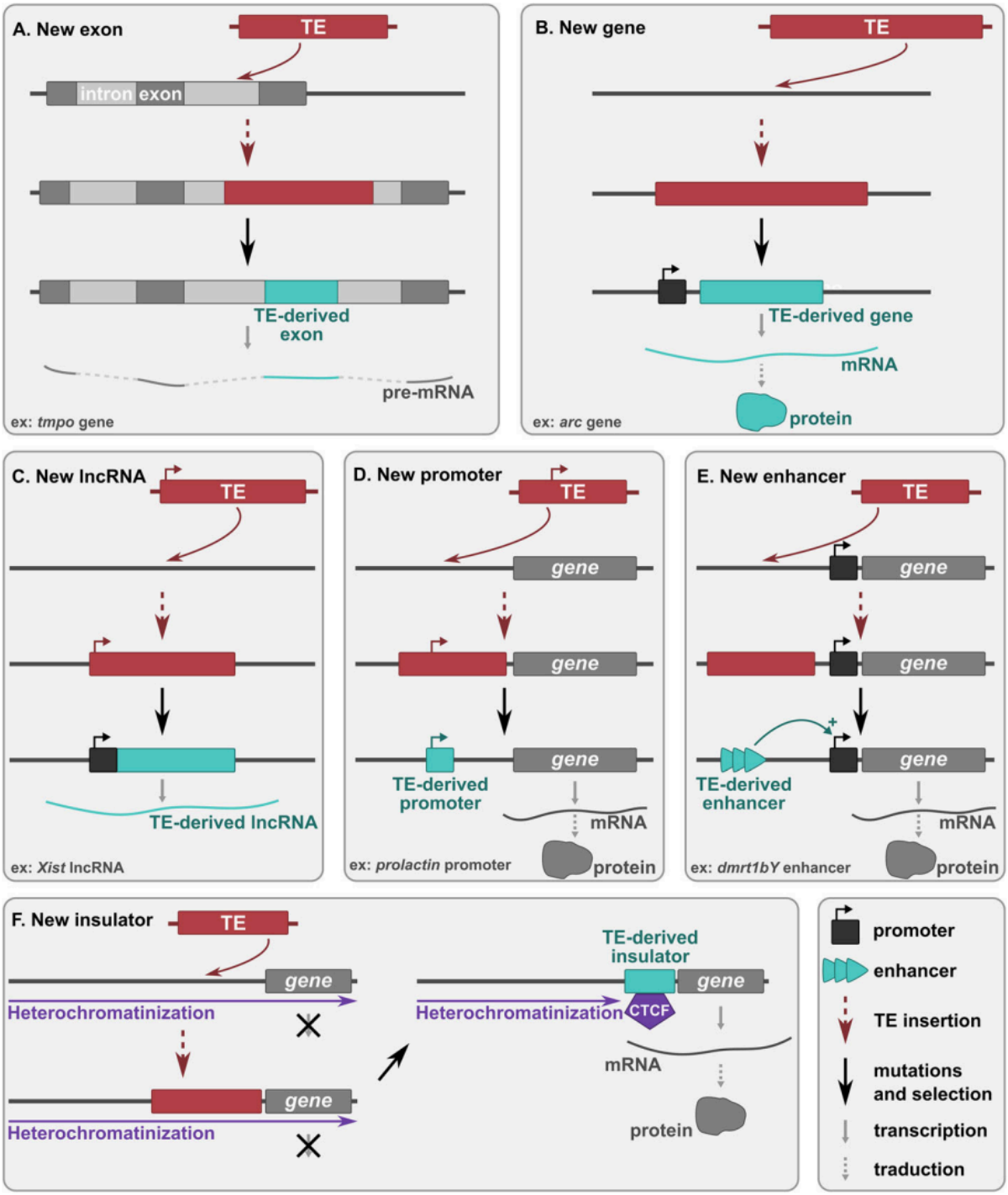


Figure 1.11 - Processus cellulaires obtenus grâce à l'exaptation des TEs par leurs cellules hôtes.

(A) L'insertion d'un TE, ou d'une partie d'un TE, dans la séquence d'un gène peut mener à la formation d'un nouvel exon. **(B)** L'insertion de la séquence d'un TE, ou d'une partie du TE, peut mener à la formation d'un nouveau gène dont l'expression est régulée par une séquence régulatrice de l'hôte ou du TE lui-même. **(C)** L'insertion d'un TE peut mener à la formation d'un nouveau lncRNA, ou de nouveaux domaines de lncRNA préexistants. **(D, E)** L'insertion de TEs en amont de la séquence d'un gène ou d'un lncRNA peut mener à la formation de nouveaux promoteurs (D) ou activateurs (E). **(F)** L'insertion de TEs dans le génome peut modifier la structure de la chromatine en fournissant des sites de liaison à la protéine CTCF, et ainsi permettre l'expression des gènes en relâchant la chromatine. Reproduit avec la permission de *Springer Nature* : Mobile DNA (Etchegaray E, Naville M, Volff JN & Haftek-Terreau Z) ©2021 (207).

Il a été montré que les séquences des TEs fournissent des sites de liaisons à l'ADN à de nombreuses protéines (**Figure 1.11D,E**). De fait, des analyses de données d'immunoprécipitation de la chromatine de 7 facteurs de transcription (ESR1, TP53, MYC, RELA, POU5F1, SOX2, CTCF) ont montré que 5 d'entre eux (ESR1, TP53, POU5F1, SOX2, CTCF) ont la capacité de lier les séquences des TEs et que les séquences de leurs motifs de liaison situés dans des TEs ont été conservés au cours de l'évolution (208). Des études subséquentes ont confirmé ces résultats, démontrant que les TEs contribuent à 12-18% des boucles d'ADN en fournissant des sites de liaison à CTCF, et que la délétion des TEs contenant ces sites de liaisons de CTCF avait un impact majeur sur la structure tridimensionnelle de la chromatine (**Figure 1.11F**) (209, 210). De plus, les TEs peuvent agir comme promoteurs ou activateurs (*enhancers*) pour une grande diversité de facteurs de transcription de façon tissu-spécifique, tant chez l'humain que chez la souris (211-218). De façon intéressante, il a été montré que la contribution des TEs aux promoteurs pouvait fournir des sites d'initiation de la transcription (TSS) de façon tissu-spécifique (219), incluant des cadres de lecture ouverts en amont (uORF, de l'anglais *upstream open reading frame*) qui sont des régulateurs connus de la traduction des ARNm (220). Ainsi, l'insertion de TEs dans les génomes des mammifères a favorisé l'évolution de réseaux de régulation géniques en fournissant des sites de liaisons à des facteurs de transcription ainsi qu'à des régulateurs de la structure de la chromatine. Finalement, bien que les évènements de transposition des TEs soient généralement associés à l'instabilité génomique, il a été montré que la transposition d'éléments L1 dans les

neurones contribue à la plasticité synaptique, une caractéristique importante pour le fonctionnement du cerveau (221, 222).

Les TEs ont aussi des fonctions diverses au niveau des ARN. Premièrement, il a été montré que l'insertion de TEs dans la séquence de gènes pouvait modifier l'épissage des ARNm. Par exemple, l'insertion d'un Alu dans un intron du gène *TBXT* cause l'exclusion de l'exon 6 du gène en favorisant la formation de boucle avec un autre TE, ce qui aurait causé la perte de la queue chez les hominoïdes (223). De plus, les ARNm encodés par les TEs peuvent former des complexes ribonucléoprotéiques (224), comme les ARNm dérivés d'éléments L1 interagissant avec la nucléoline pour stimuler le renouvellement des cellules souches embryonnaires et leur synthèse d'ARN ribosomaux (225). Un autre exemple est l'interaction de la protéine Spen avec les ARNm d'éléments à LTRs menant à l'inactivation du chromosome X par le long ARN non-codant (lncRNA) Xist (226). Il a aussi été démontré que les TEs contribuent de façon importante aux séquences des lncRNAs (**Figure 1.11C**) et peuvent modifier leur localisation intracellulaire et leur fonction (227-230). Finalement, il a été observé que les ARN interagissant avec Piwi (piRNA) ainsi que près de 20% des micro ARN (miRNA) peuvent être dérivés des séquences des TEs (231-239).

Enfin, il existe plusieurs exemples de gènes encodés par des TEs ayant été domestiqués par leurs cellules hôtes, soit en générant de nouveaux gènes ou sous-forme d'exons s'insérant dans la séquence de gènes préexistants (**Figure 1.11A,B**). Un exemple frappant est que plusieurs gènes dérivés de TEs sont essentiels à la formation du placenta : les gènes *peg10/11* dérivés de gènes *gag* (240, 241), ou encore les *syncytines 1/2,* dérivées de gènes *env* d'éléments à LTRs qui médient les fusions de membranes (167, 242). Un autre exemple est CENP-B, une protéine dérivée d'un transposon à ADN impliquée dans la progression de la fourche de réplication de l'ADN (243, 244). Les TEs peuvent aussi générer de nouveaux exons et modifier la fonction de gènes préexistants. Une étude récente a montré que l'incorporation de transposases encodées par des TEs dans la séquence de facteurs de transcription est un phénomène fréquent chez les vertébrés ayant impacté l'évolution de nombreux facteurs de transcription (245). Un autre exemple est TMPO, dont l'épissage alternatif génère un isoforme contenant un domaine dérivé d'un rétrotransposon

et interagissant avec la lamina nucléaire (246). En somme, la littérature regorge d'exemples de fonctions exaptées des TEs ayant impactées le développement et l'homéostasie des cellules des vertébrés, et ce même s'ils ont perdu leur capacité à produire des protéines.

1.2.3 Régulation de l'expression

Puisque la transposition des TEs peut être délétère au fonctionnement de la cellule hôte, plusieurs mécanismes de régulation transcriptionnelle et post-transcriptionnelle ont été développés pour réprimer leur expression et leur activité. Comme plusieurs TEs partagent des similarités avec des pathogènes, plusieurs mécanismes de défense immunitaire peuvent aussi reconnaitre les ARN et protéines des TEs pour les dégrader et bloquer leur réplication. S'en suit donc une course à l'armement entre les TEs et leurs cellules hôtes, où les cellules hôtes tentent de réprimer l'expression des TEs, et les TEs essaient d'échapper à cette régulation. Les sections subséquentes aborderont plus en détails les différents mécanismes régulant l'expression des TEs dans les cellules des mammifères.

1.2.3.1 Régulation épigénétique

La régulation épigénétique est un mécanisme central à la répression de l'expression des TEs. Cette régulation s'exerce autant au niveau de la méthylation de l'ADN que des modifications des histones. Ces mécanismes agissent de façon non-redondante en ciblant des répertoires distincts de TEs. En effet, la méthylation de l'ADN régule l'expression de TEs plus jeunes, alors que modifications des histones répriment l'expression des TEs plus anciens (247). Il a été montré que la méthylation des cytosines des dinucléotides CpG par les ADN méthyltransférases (DNMT) est essentielle au développement (**Figure 1.12**), puisque chez la souris la perte d'activité de DNMT1 cause une expression aberrante des TEs létale lors du développement embryonnaire (248). De façon similaire, l'inhibition de l'activité des DNMT par la 5-azadeoxycytidine dans des fibroblastes embryonnaires humains entraine une surexpression d'éléments L1 (249). Toujours chez l'humain, la délétion de DNMT3L mène à une expression anormale d'éléments à LTRs qui cause un arrêt de la méiose chez les spermatocytes (250). Finalement, il a été montré que la méthylation m^6A des ARN par les protéines méthyltransférases METTL3-METTL4 diminue le temps de demi-vie des ARN générés par les TEs dans les cellules souches embryonnaires (251).

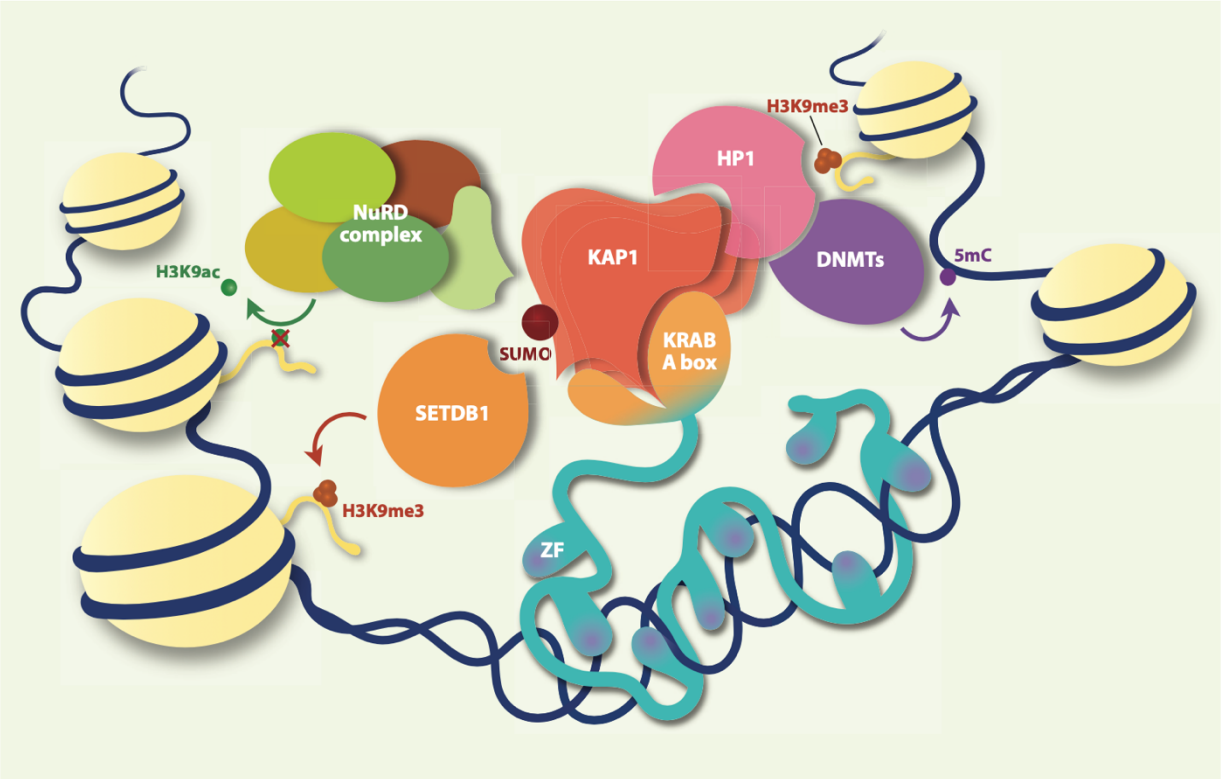


Figure 1.12 – Régulation épigénétique des TEs par les protéines KZFPs.

Les protéines KZFPs interagissent avec les séquences des TEs via leur domaine en doigt de zinc (ZF) et recrutent le corépresseur KAP1. KAP1 permet ensuite l'interaction avec les DNMTs, HP1, la méthyltransférase de l'histone H3K9 SETDB1, le complexe de remodelage de la chromatine et de désacétylation des histones NuRD. Reproduit avec la permission de *Annual Reviews* : Annual Reviews of Genetics (Bruno M, Mahgoub M & Macfarlan TS) ©2019 (252).

Les TEs subissent aussi une régulation épigénétique au niveau des histones. La marque d'histone la plus fréquemment retrouvés aux séquences des TEs est H3K9me3, qui peut être déposée par l'histone méthyltransférase SETDB1 ou encore la protéine d'hétérochromatine HP1 (**Figure 1.12**) (253-257). L'expression des TEs peut aussi être réprimée par le retrait de marques d'histones associées à la chromatine active; LSD1 et le complexe NuRD retirent respectivement les marques H3K4me2 et H3K9ac de la séquence des TEs pour réprimer leur expression (**Figure 1.12**) (258, 259). Étant donné la grande diversité de TEs présents dans les génomes eucaryotes, les mécanismes épigénétiques régulant leur expression peuvent différer d'un TE à l'autre (260, 261)

et d'autres marques d'histone ont été associées avec la répression des TEs : H4K20me3, H3K27me3, H4R3me2, ainsi que la biotinylation et la sumoylation des histones H2A, H3 et H4 (262-266).

Au centre de la régulation épigénétique des TEs se trouve une famille d'environ 400 facteurs de transcription : les protéines en doigt de zinc de la boîte associée Krüppel (KZFP). Les KZFPs lient la séquence des TEs grâce à leur domaine à doigt de zinc (267), et permettent le recrutement des DNMT, HP1, SETDB1 et du complexe NuRD pour médier la régulation épigénétique des TEs (**Figure 1.12**) (252). De nombreuses évidences suggèrent que les KZFPs évoluent en réponse aux nouvelles insertions de TEs dans le génome. Premièrement, il existe une forte corrélation entre le nombre de KZFPs et le nombre de TEs présents dans le génome des mammifères (268). De plus, il a été montré que des mutations dans le domaine de liaison à l'ADN des KZFPs sont soumises à la sélection naturelle positive, suggérant que l'émergence de KZFPs ayant des spécificités diverses a été importante pour l'évolution des primates (269). Ainsi, il existe de grandes divergences entre les KZFPs présentes dans les génomes des mammifères, et seulement ~20% des KZFPs sont conservées entre l'humain et la souris (270). Des analyses phylogénétiques plus poussées ont montré que l'insertion de nouveaux TEs dans les génomes eucaryotes menait à l'évolution de nouvelles KZFPs permettant leur répression (271). Finalement, il a été observé que l'extrémité 5' des éléments L1 récents évoluait pour perdre les sites de liaison des KZFPs régulant leur expression et échapper à leur régulation épigénétique (271). Ces évidences montrent que la régulation épigénétique des TEs est dynamique et que les cellules hôtes doivent constamment s'adapter aux nouvelles insertions et à l'évolution de TEs dans leur génome.

1.2.3.2 Mécanismes post-transcriptionnels

Plusieurs mécanismes post-transcriptionnels répriment aussi l'activité de TEs ayant échappé à la régulation épigénétique ou des nouvelles insertions de TEs n'étant pas encore régulées au niveau épigénétique. Plusieurs de ces mécanismes post-transcriptionnels font partie de la défense cellulaire antivirale et ciblent différentes étapes du processus de réplication des TEs.

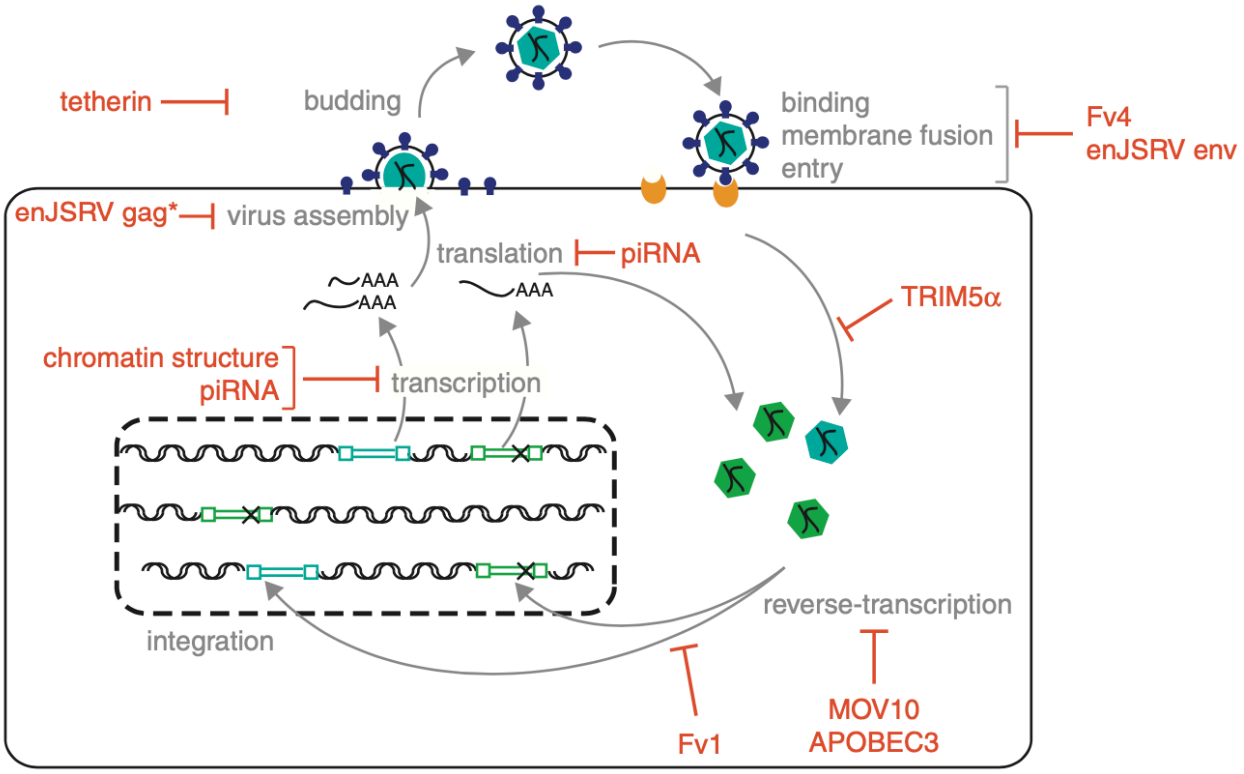


Figure 1.13 – Facteurs de restriction agissant aux différents stades de replication des TEs.

Schématisation du processus de réplication des éléments à LTR. L'action des différents facteurs de restriction pour bloquer la réplication des TEs est indiquée en rouge. Reproduit avec la permission d'*Elsevier* : Current Opinion in Virology (Dewannieux M & Heidmann T) ©2013 (272).

Les ARNm encodés par les TEs peuvent être ciblés par l'interférence à ARN (**Figure 1.13**), qui induit la méthylation de l'ADN et l'incorporation de marques d'histones répressives à la séquence des TEs ou encore stimule la dégradation des ARNm encodés par les TEs (273). Dans les cellules de mammifères, les piRNAs et les miRNAs semblent être les principaux mécanismes d'interférence à ARN ciblant les TEs, bien qu'il ait été montré que les courts ARN interférents (siRNA) régulent l'expression des TEs dans les ovocytes chez la souris (274, 275) et chez l'humain (276). Les piRNAs régulent l'expression des TEs dans la lignée germinale (277-280), alors que les miRNAs répriment quant à eux l'expression des TEs dans les cellules somatiques (281-283).

De nombreux mécanismes de défense antivirale contribuent aussi à l'inhibition de la réplication des TEs. De fait, plusieurs études ont montré que divers facteurs de restriction bloquant les différents stades de la réplication des rétrovirus (**Figure 1.13**), tels TRIM5α, SAMHD1, l'hélicase à ARN MOV10 et la protéine en doigt de zinc ZAP contribuent aussi à la répression de la réplication des TEs (284-288). Finalement, les protéines de la famille APOBEC, qui stimulent la désamination des cytosines en uraciles, modifient les séquences des protéines encodées par les TEs, ce qui empêche leurs réplications subséquentes (**Figure 1.13**) (289-292).

1.2.3.3 Expression au cours du développement

L'expression des TEs est typiquement réprimée de façon épigénétique pour prévenir des évènements de réplication qui menaceraient l'intégrité génomique. Cette régulation épigénétique est toutefois perturbée durant le développement par d'importantes vagues de réorganisation de la chromatine (293). La chromatine est relâchée lors de ces phases de réorganisation, permettant l'expression des TEs. Deux types cellulaires sont particulièrement affectés par ce phénomène : les cellules souches embryonnaires (ESC) et les cellules de la lignée germinale (294, 295). L'expression des TEs varie en effet rapidement lors des premiers stades de développement embryonnaire (**Figure 1.14**) (296). L'expression des TEs par les ESCs est essentielle au maintien de la pluripotence en fournissant des sites de liaison à divers facteurs de transcription (297, 298). De plus, environ 40% des lncRNA exprimés par les ESCs sont dérivés de TEs (299), dont plusieurs sont essentiels au développement. Par exemple, linc-RoR est un lncRNA presque entièrement composé de TEs de 7 sous-familles différentes qui maintient la pluripotence des ESCs en agissant comme une éponge à miRNA et en empêchant leur liaison à leurs gènes cibles (300, 301). L'activité des TEs dans la lignée germinale représente quant à elle une importante source de variation génétique. De fait, l'activité des TEs cause de 12-15% des mutations spontanées de la lignée germinale de la souris (302). Bien que la transposition des TEs soit plus rare dans la lignée germinale humaine (302), ces données montrent que les TEs contribuent à l'évolution des génomes de leurs organismes hôtes et peuvent causer des maladies génétiques lorsqu'ils perturbent la séquence ou l'expression de gènes (303).

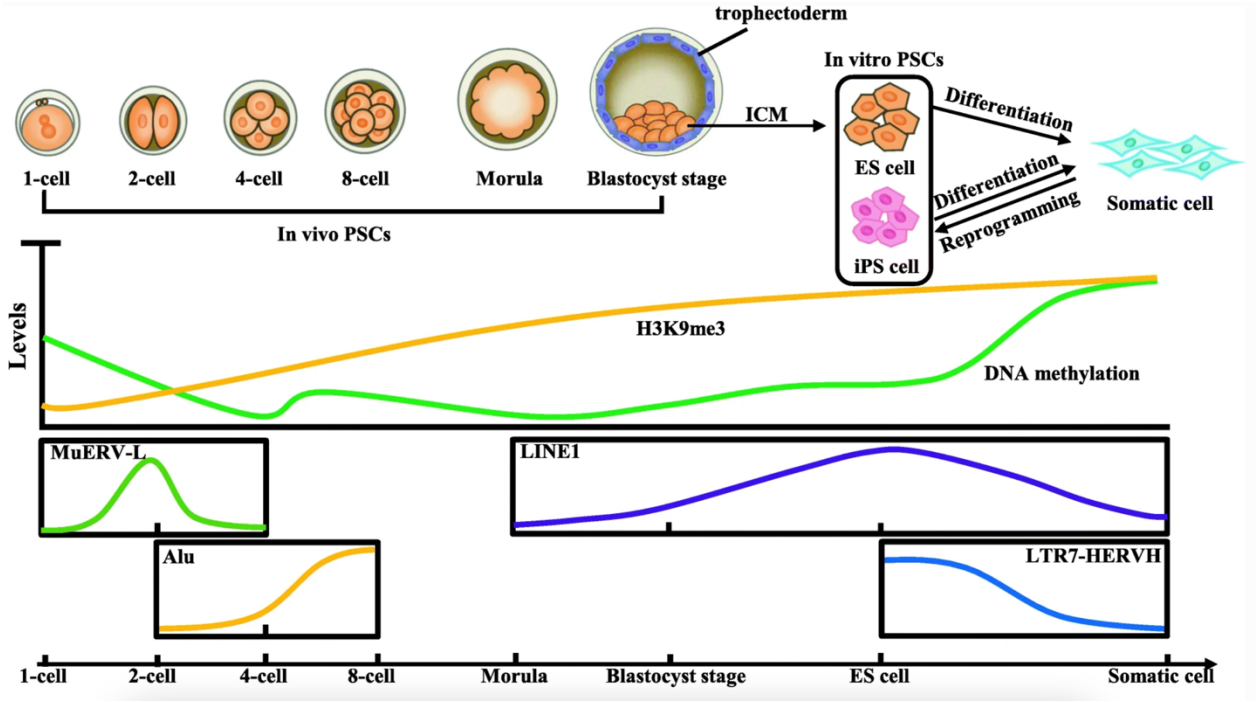


Figure 1.14 – Expression dynamique des TEs durant l'embryogenèse.

Niveaux de méthylation de l'ADN et de H3K9me3 ainsi que d'expression de différents TEs lors du développement embryonnaire et de la différenciation en cellules somatiques. Reproduit avec la permission de *Springer Nature* : Cell Regeneration (Wang J, Huang J & Shi G) ©2020 (304).

En dehors des ESCs, il est généralement considéré que l'expression des TEs est réprimée dans les cellules somatiques (**Figure 1.14**) (305). Une exception potentielle à ce modèle est le cerveau où l'expression et la transposition d'éléments L1 ont été observées (222, 306-310), bien que la fréquence de ces événements de transposition soit débattue (221, 311, 312). Certaines évidences suggèrent toutefois que certains TEs pourraient échapper à la répression exercée par leur hôte et être exprimés dans les cellules somatiques. En effet, il a été montré que différentes sous-familles de TEs sont enrichies aux TSS de façon tissu-spécifique (219). Il a aussi été observé que certains TEs sont enrichis dans des régions de chromatine active de façon tissu-spécifique, suggérant que leur expression est associée à des fonctions biologiques précises (214). Puisqu'un portrait global de l'expression des TEs dans l'ensemble des tissus humains n'a jamais été brossé, il est toutefois difficile d'évaluer l'expression et la fonction des TEs dans les cellules somatiques humaines.

1.2.3.4 Dérégulation de l'expression des TEs dans les cancers

Les modifications épigénétiques survenant lors de la tumorigenèse perturbent elles aussi la répression des TEs, menant à leur surexpression dans les cellules cancéreuses (313). De fait, l'expression aberrante des TEs a été observée dans plusieurs types de tumeurs (314-322) et est corrélée négativement avec le niveau de méthylation de l'ADN (323). L'expression aberrante des TEs peut aussi contribuer à la tumorigenèse i) en créant des réorganisations chromosomiques lors d'évènements de rétrotransposition (324-327), ou ii) en stimulant l'expression d'oncogènes via la formation de transcrits chimériques (328, 329). L'expression de protéines membranaires encodées par les gènes *env* des éléments à LTRs a aussi été associée à une plus grande prolifération des cellules cancéreuses et à la croissance de la tumeur (330), alors que la protéine Np9 encodée par les éléments HERV-K contrôle la migration des cellules cancéreuses (331).

De façon intéressante, l'expression aberrante des TEs dans les cellules cancéreuses n'est pas uniquement avantageuse pour la tumeur, puisqu'elle peut aussi induire des réponses immunitaires antitumorales. En effet, la nature hautement répétitive des TEs peut mener à la formation d'ARN double-brins (dsRNA) (332-334). Ces dsRNAs sont ensuite reconnus par des récepteurs de l'immunité innée comme RIG-I et MDA5, un état nommé « mimétisme viral » (*viral mimicry*) associé à l'activation de l'immunité innée et à la sécrétion d'IFN (335, 336). L'expression aberrante des TEs dans les cellules cancéreuses est aussi associée à une plus grande infiltration de la tumeur par les lymphocytes T CD8, une meilleure survie des patients et une meilleure réponse des patients aux immunothérapies par inhibiteur de points de contrôle (*immune checkpoint inhibitors*) (314, 319, 323, 337). Ces résultats suggèrent que les TEs surexprimés par les cellules cancéreuses génèrent des antigènes présentés par le CMH-I qui sont reconnus comme du non-soi par les lymphocytes T CD8.

1.3 Interactions entre les TEs et le système immunitaire

Tel que mentionné précédemment, les TEs ont un impact important sur l'évolution du génome de leur hôte. Les interactions complexes que forment les TEs avec le système immunitaire de leur hôte en sont un exemple flagrant. En effet, les TEs ont plusieurs contributions essentielles au développement et à la fonction du système immunitaire des gnathostomes. De par leur nature,

l'expression des TEs peut toutefois mener à des conflits avec le système immunitaire. En effet, puisque plusieurs TEs tirent leur origine de pathogènes, leur expression peut induire des réponses immunitaires. Ces conflits entre les TEs et le système immunitaire peuvent être lourds de conséquences pour leur hôte, mais peuvent aussi être exploités à des fins thérapeutiques.

1.3.1 Contribution des TEs au développement et à la fonction immunitaires

Plusieurs des fonctions exaptées des TEs décrites précédemment ont aussi un impact sur le développement et la fonction des cellules immunitaires. Premièrement, les TEs régulent l'expression de gènes impliqués dans la différenciation et l'activation des cellules immunitaires en fournissant des sites de liaison à divers facteurs de transcription. En effet, il a été montré que les TEs fournissent des sites de liaison à STAT1 et à IRF1 lors de la réponse aux interférons (338), lors de l'activation des lymphocytes T (218, 339), ou encore à NF-κB et AP-1 lors de l'activation des macrophages (340). Des données transcriptomiques montrent aussi que l'expression des TEs change drastiquement lors de l'activation des lymphocytes B (320), mais des analyses plus poussées seraient nécessaires pour déterminer si ces changements d'expression des TEs causent l'activation des lymphocytes B ou en résultent. Finalement, il a été observé que les TEs peuvent agir comme séquences cis-régulatrices lors de la différenciation des lymphocytes T Th2 : l'ajout de marques d'histones répressives à la séquence de TEs réprime l'expression de gènes en aval et permet la différenciation de lymphocytes T CD4 naïfs en lymphocytes T Th2 (341). Ainsi, les TEs impactent le développement et l'activation de plusieurs populations de cellules immunitaires en régulant l'expression de gènes.

Une des particularités des vertébrés est le développement du système immunitaire adaptatif, qui procure une plus grande spécificité ainsi qu'une mémoire immunitaire. Il a été observé que les gènes *RAG1/2*, qui sont essentiels au développement des lymphocytes T et B en médiant la recombinaison V(D)J, sont dérivés de transposons à ADN (342, 343). De façon intéressante, des gènes homologues à *RAG1/2* ont été détectés chez l'oursin, suggérant que ces gènes étaient présents dans le génome bien avant d'acquérir une fonction dans le développement de l'immunité adaptative (344). Les TEs auraient donc joué un rôle central dans l'émergence de l'immunité adaptative en permettant la diversification des récepteurs antigéniques.

Finalement, en plus de toutes ces contributions au développement et à la fonction des cellules immunitaires, il a été montré que les TEs peuvent directement défendre leurs cellules hôtes contre les infections virales. En effet, différentes protéines encodées par les gènes *env* d'éléments à LTRs interagissent avec les récepteurs des virus pour bloquer leur internalisation (**Figure 1.13**) : *Fv4* et *Rmcf2* chez la souris (345, 346), *chf* chez la poule (347), enFeLV chez le chat (348) et *SUPYN* chez l'humain (349). Ces résultats montrent donc que, suite à leur insertion dans le génome, les TEs peuvent contribuer à protéger leur hôte contre l'intégration des séquences de nouveaux pathogènes.

1.3.3 Reconnaissance des TEs par le système immunitaire et thérapies ciblant les TEs

Puisque les transposons à ADN et les éléments à LTRs tirent respectivement leur origine de bactéries et de virus, leurs séquences contiennent des similarités avec ces pathogènes et peuvent donc entrainer des réponses immunitaires. Tel que mentionné plus tôt, l'expression aberrante des TEs peut mener à la formation de dsRNAs, qui sont des structures normalement associées aux infections virales, ce qui cause l'activation des récepteurs MDA5 et RIG-I. La transposition des TEs peut également mener à la formation d'hybrides ARN:ADN ou d'ADN cytosolique, des structures causant l'activation de la voie cGAS-STING et la sécrétion d'interféron (350, 351). L'activation de ces voies de l'immunité innée entraine la sécrétion de différentes cytokines qui stimulent un microenvironnement pro-inflammatoire, ce qui augmente l'infiltration lymphocytaire de la tumeur. Finalement, plusieurs études montrent que la surexpression des TEs dans les cellules cancéreuses mène à la présentation de MAPs dans différents types de cancers (329, 352-355). Ainsi, l'expression anormale des TEs dans les tumeurs peut être reconnue par les système immunitaires inné et adaptatif et causer des réponses antitumorales. Ces exemples d'activation du système immunitaire contre les TEs ont mené à l'élaboration de plusieurs thérapies anticancéreuses exploitant l'expression aberrante des TEs dans les tumeurs.

De fait, l'état de mimétisme viral peut être provoqué dans les tumeurs en perturbant la régulation de l'expression des TEs. Ainsi, il a été montré que les agents hypométhylants tels l'azacytidine et la décitabine (323, 336, 356) ou encore les inhibiteurs des histones méthyltransférases (357, 358)

causent une surexpression des TEs. De même, l'inhibition des enzymes éditant les ARNs TEs telles ADAR1 permet la reconnaissance des dsRNAs générés par les TEs par MDA5 et cause un état de mimétisme viral (359, 360). De façon intéressante, cet état de mimétisme viral crée un environnement pro-inflammatoire qui contribue au recrutement et à l'activation des cellules immunitaires adaptatives dans les tumeurs (323, 359).

Finalement, plusieurs études ont montré que l'expression aberrante des TEs dans les tumeurs mène à la présentation d'antigènes dérivés des TEs qui peuvent activer des réponses des lymphocytes B et T. De fait, les éléments à LTRs ayant conservé leurs séquences codant pour le gène *env* peuvent générer des protéines exprimées à la membrane plasmique (316, 361-363). Il a été montré que la liaison d'anticorps à ces protéines TEs membranaires permet le recrutement de cellules NK et est associée à une meilleure réponse aux immunothérapies par inhibiteur de points de contrôle (364). L'expression aberrante des TEs mène aussi à la présentation de MAPs dans un grand nombre de cancers (323, 329, 352-355, 365-367). Les études ayant testé l'immunogénicité des MAPs dérivés des TEs ont toutefois donné des résultats contradictoires : dans certaines études, les MAPs TEs induisent des réponses des lymphocytes T CD8 (352, 355, 367), alors que dans d'autres études les MAPs TEs semblent tolérés par les lymphocytes T CD8 (317, 354).

Ces données suggèrent que les lymphocytes T apprennent à tolérer les séquences des TEs lors de leur développement dans le thymus. Le manque de données sur l'expression et la fonction des TEs dans le thymus rend cependant difficile d'évaluer leur rôle dans l'établissement de la tolérance au soi des lymphocytes T. Chez l'humain, il a été démontré que l'élément à LTRs ERVPb1 était exprimé plus fortement dans le thymus que dans les autres tissus somatiques étudiés et que sa séquence était conservée entre les génomes des primates, suggérant un rôle dans le développement ou la fonction du thymus (368). Puisque l'expression d'ERVPb1 a été mesurée dans le thymus entier, il est toutefois difficile d'évaluer dans quelles cellules ce TE est exprimé et quelle y serait sa fonction. Une autre étude réalisée chez la souris suggère qu'un antigène encodé par le gène *env* d'un TE exprimé par les mTECs induirait la sélection négative de lymphocytes T CD4, mais cette étude est basée sur l'expression du TE au niveau transcriptomique et la présentation de l'antigène par le CMH-I n'a pas été validée (369). Une meilleure compréhension

de la contribution des TEs à l'établissement de la tolérance au soi des lymphocytes T serait donc nécessaire pour faciliter l'identification d'antigènes dérivés des TEs pouvant être ciblés dans des contextes d'immunothérapie du cancer.

1.3.4 Objectifs de la thèse

La littérature regorge d'exemples de fonctions exaptées des TEs dans les cellules de leurs hôtes. La très grande majorité des études se concentrant toutefois sur les fonctions des TEs lors des stades précoces du développement, l'expression et le rôle des TEs dans les tissus somatiques demeurent nébuleux. Dans un contexte ou un grand nombre d'équipes de recherche tentent d'identifier des MAPs TEs à la surface de cellules cancéreuses pour la mise en place d'immunothérapies, évaluer l'expression des TEs dans les tissus somatiques humains permettrait d'évaluer la toxicité de telles thérapies pour le patient. De plus, bien que des données suggèrent que les TEs sont impliqués dans l'éducation des lymphocytes T dans le thymus chez la souris, la présentation de MAPs TEs par les cellules thymiques doit être démontrée. Puisque les lymphocytes T apprennent à tolérer les antigènes présentés dans le thymus, il serait inutile de tenter de cibler les MAPs TEs retrouvés dans le thymus dans des contextes d'immunothérapies du cancer.

Cette thèse vise donc à répondre à trois questions principales : i) l'expression des TEs est-elle répandue dans les tissus somatiques humains, ii) cette expression au niveau transcriptomique mène-t-elle à la présentation de MAPs à la surface de cellules non-cancéreuses, et iii) les lymphocytes T apprennent-ils à tolérer les TEs lors de leur développement dans le thymus? Nos travaux de recherche avaient donc pour objectif de mieux comprendre l'expression et la fonction des TEs dans les tissus somatiques et le thymus.

Chapitre 2 : Projet 1

2.1 Article #1: Widespread and tissue-specific expression of endogenous retroelements in human somatic tissues.

2.1.1 Résumé en français

Titre en français : Expression répandue et tissu-spécifique des rétroéléments endogènes dans les tissus somatiques humains.

Les rétroéléments endogènes (ERE) représentent environ 42% du génome humain et ont été associés à des pathologies fréquentes telles les maladies autoimmunes et le cancer. Le dogme central stipule que les EREs sont exprimés dans les cellules souches embryonnaires (ESC) et dans la lignée germinale, mais sont réprimés dans les cellules souches différenciées. Malgré des évidences que les EREs peuvent être exprimés aux niveaux des ARN et des protéines dans certains contextes, une analyse exhaustive de l'expression des EREs dans les tissus humains est nécessaire.

À l'aide de données transcriptomiques (RNA-seq), nous avons analysé l'expression des EREs dans un panel de 32 tissus et types cellulaires somatiques humains incluant les cellules épithéliales de la médulla (mTEC). Un indice de tissu-spécificité a été mesuré afin d'identifier les sous-familles EREs exprimées de façon tissue-spécifique. Nous avons aussi analysé le transcriptome de mTECs provenant de souris sauvages ou pour lesquelles AIRE est déplété. Finalement, nous avons développé une approche protéogénomique combinant le RNA-seq et la spectrométrie de masse (MS) pour déterminer si les EREs sont traduits et génèrent des MAPs dans des lignées B-LCL de 16 individus.

Nos travaux montrent que tous les tissus et types cellulaires étudiés expriment les EREs, bien que la magnitude de leur expression varie grandement d'un tissu à l'autre. L'expression des EREs était particulièrement élevée dans deux tissus n'exprimant par le CMH I (ESCs et les testicules), et dans un type cellulaire exprimant très fortement le CMH I (mTECs). L'utilisation de notre modèle murin a permis de démontrer que l'expression forte des EREs dans les mTECs est indépendante de AIRE. L'analyse en MS des lignées de B-LCL a identifié 103 ereMAPs non redondants. Ces ereMAPs dérivaient préférentiellement de la traduction sense d'EREs introniques. Notamment, une analyse détaillée de la composition en acides aminés des ereMAPs a révélé qu'ils possèdent une homologie de séquence avec des MAPs viraux.

Cette étude montre que l'expression des EREs dans les tissus somatiques est plus répandue et hétérogène qu'anticipée. L'expression forte et diversifiée des EREs dans les mTECs, couplée à leur capacité de générer des MAPs, suggère que les EREs pourraient jouer un rôle important dans l'établissement de la tolérance au soi. Les propriétés similaires aux MAPs viraux des ereMAPs pourrait expliquer la grande immunogénicité des ereMAPs non-exprimés par les mTECs.

2.1.2 Contribution des auteurs

Jean-David Larouche : Conception du projet. Écriture des scripts, réalisation des analyses bio-informatiques et analyse des résultats pour les figures 2.1 à 2.6 et les figures supplémentaires 2.1 à 2.7. Écriture de la première version du manuscrit.

Assya Trofimov : Contribution à l'écriture des scripts et à la conception des analyses bio-informatiques (figure 2.2 et figure supplémentaire 2.2)

Leslie Hesnard : Contribution à la figure 2.3A. Isolation des mTECs de thymus humains (figures 2.1 et 2.2, figures supplémentaires 2.1 et 2.2). Réalisation des expériences d'ELISPOT de la figure supplémentaire 2.7, B et C.

Gregory Ehx : Contribution à la figure 2.6A et à la figure supplémentaire 2.7A.

Qingchuan Zhao : Contribution à l'analyse de la méthylation de l'ADN (figure supplémentaire 2.5C).

Krystel Vincent : Conception du projet, analyse des résultats et écriture de la première version du manuscrit.

Chantal Durette : Recherches des bases de données de MS dans Peaks (figure 2.4).

Patrick Gendron : Téléchargement des données transcriptomiques de GTEx et de TCGA.

Jean-Philippe Laverdure : Génération du protéome canonique personnalisé (figure 2.4). Contribution aux figures supplémentaires 2.4 et 2.5A.

Éric Bonneil : Validation manuelle des spectres MS/MS des ereMAPs (figure 2.4).

Caroline Côté : Isolation des mTECs de thymus humains (figures 2.1 et 2.2, figures supplémentaires 2.1 et 2.2).

Sébastien Lemieux : Analyse des résultats.

Pierre Thibault : Analyse des résultats.

Claude Perreault : Conception du projet, analyse des résultats et écriture de la première version du manuscrit.

2.1.3 Version originale publiée dans Genome Medicine

Larouche JD, *et al*. (2020). "Widespread and tissue-specific expression of endogenous retroelements in human somatic tissues". Genome Medicine; 12(1):40.

Jean-David Larouche[1,2], Assya Trofimov[1,3], Leslie Hesnard[1,2], Gregory Ehx[1,2], Qingchuan Zhao[1,2], Krystel Vincent[1,2], Chantal Durette[1], Patrick Gendron[1], Jean-Philippe Laverdure[1], Éric Bonneil[1], Caroline Côté[1], Sébastien Lemieux[1,4], Pierre Thibault[1,5*] and Claude Perreault[1,2,6*].

1. Institute of Research in Immunology and Cancer, Université de Montréal, Montréal, QC, Canada.
2. Department of Medicine, Université de Montréal, Montréal, QC, Canada.
3. Department of Computer Science and Operations Research, Université de Montréal, Montréal, QC, Canada.
4. Department of Biochemistry and Molecular Medicine, Université de Montréal, Montréal, QC, Canada.
5. Department of Chemistry, Université de Montréal, Montréal, QC, Canada.
6. Division of Hematology-Oncology, Hôpital Maisonneuve-Rosemont, Montréal, QC, Canada.

*Correspondence:

Pierre Thibault
IRIC - Université de Montréal,
P.O. Box 6128, Downtown Station
QC, Canada, H3C 3J7
pierre.thibault@umontreal.ca

Claude Perreault
IRIC - Université de Montréal,
P.O. Box 6128, Downtown Station
QC, Canada, H3C 3J7
claude.perreault@umontreal.ca

2.1.3.1 Abstract

Background: Endogenous retroelements (ERE) constitute about 42% of the human genome and have been implicated in common human diseases such as autoimmunity and cancer. The dominant paradigm holds that EREs are expressed in embryonic stem cells (ESC) and germline cells but are repressed in differentiated somatic cells. Despite evidence that some EREs can be

expressed at the RNA and protein levels in specific contexts, a systems-level evaluation of their expression in human tissues is lacking.

Methods: Using RNA-sequencing data, we analyzed ERE expression in 32 human tissues and cell types, including medullary thymic epithelial cells (mTECs). A tissue-specificity index was computed to identify tissue-restricted ERE families. We also analyzed the transcriptome of mTECs in wild-type and Autoimmune regulator (AIRE)-deficient mice. Finally, we developed a proteogenomic workflow combining RNA-sequencing and mass spectrometry (MS) in order to evaluate whether EREs might be translated and generate MHC I-associated peptides (MAP) in B-lymphoblastoid cell lines (B-LCL) from 16 individuals.

Results: We report that all human tissues express EREs but the breadth and magnitude of ERE expression are very heterogeneous from one tissue to another. ERE expression was particularly high in two MHC-I-deficient tissues (ESCs and testis) and one MHC-I-expressing tissue, mTECs. In mutant mice, we report that the exceptional expression of EREs in mTECs was AIRE-independent. MS analyses identified 103 non-redundant ERE-derived MAPs (ereMAPs) in B-LCLs. These ereMAPs preferentially derived from sense translation of intronic EREs. Notably, detailed analyses of their amino acid composition revealed that ERE-derived MAPs presented homology to viral MAPs.

Conclusions: This study shows that ERE expression in somatic tissues is more pervasive and heterogeneous than anticipated. The high and diversified expression of EREs in mTECs and their ability to generate MAPs suggest that EREs may play an important role in the establishment of

self-tolerance. The viral-like properties of ERE-derived MAPs suggest that those not expressed in mTECs can be highly immunogenic.

Keywords: Endogenous retroelements, immunopeptidome, major histocompatibility complex, medullary thymic epithelial cells, somatic tissues, systems biology, transcriptome.

2.1.3.2 Background

Endogenous retroelements (ERE) are remnants of transposable elements that successfully integrated our germline DNA millions of years ago (198, 272). After initial integration in the genome, EREs further increased their copy number via several successive waves of retrotransposition (370, 371). Now, most ERE sequences contain mutated or truncated open reading frames and have lost their capacity to transpose in the genome (198). Phylogenic analyses have allowed the classification of EREs in families based on sequence homology (372, 373). Most EREs are categorized in three groups, which altogether comprise ~42% of the human genome: the long-terminal repeats (LTR) as well as the long and short interspersed nuclear elements (LINE and SINE) (147, 151, 374).

Hosts repress ERE expression in order to protect their genomic integrity from deleterious insertions of EREs in open reading frames (375, 376). Indeed, a strict epigenetic regulation of ERE sequences is applied at both the DNA and histone levels (377). Growing evidence suggests that KRAB zinc finger proteins (KZFPs) are involved in an evolutionary arms race to repress the expression of novel ERE integrations (271). KZFPs recruit numerous restriction factors to silence ERE sequences: the histone methyltransferase SETDB1, DNA methyltransferase proteins, the

nucleosome remodeling and deacetylase complex NuRD and the heterochromatin protein HP1 (252). KZFP-independent mechanisms, such as the HUSH complex (378) and the histone demethylase LSD1 (258), also apply non-redundant epigenetic silencing on ERE sequences. Nevertheless, some “domesticated” EREs contribute at many levels to human development and survival. Specifically, ERE sequences are key components of several promoters and enhancers of genes implicated in interferon responses, DNA damage response in the male germline and maintenance of stem cell pluripotency (212, 338, 379). Additionally, a LINE-derived transcript is essential to embryonic stem cells (ESCs) self-renewal via activation of rRNA synthesis (225). Finally, syncytins are ERE-derived proteins that mediate cell-cell fusion to allow formation of the placental syncytium (167, 380).

The dominant paradigm holds that EREs are expressed in ESCs as well as in germline cells, but are repressed in other differentiated cells outside specific contexts in which they have relevant functions (377). However, studies on ERE expression have been limited to subsets of ERE families in one or few tissues. Additionally, to our knowledge, no study has addressed ERE expression in the thymus where central T-cell immune tolerance is established. Hence, we have no clue as to the ability of EREs to induce T-cell tolerance. In the present report we demonstrate that ERE expression is widespread in human tissues, but with tissue-specific profiles. In addition, our mass spectrometry (MS) analyses revealed that the three main groups of EREs generate MHC I-associated peptides (MAPs) retaining similarities with viral peptides. Finally, we found that mTECs express top levels of EREs, in a fashion that is independent of the Autoimmune regulator (AIRE), which could mediate self-tolerance to the antigens deriving from them.

2.1.3.3 Methods

Transcriptomic data manifest

RNA-seq data of 30 non-redundant human tissues were downloaded from the Genotype-Tissue Expression (GTEx) on the dbGaP portal (accession number phs000424.v8.p2.c1) (381). When possible, 50 samples were randomly selected per tissue, otherwise all available samples were analyzed. Transcriptomic data of ESCs were downloaded from the sequence read archive from Lister *et al* (382). RNA-seq data of purified hematopoietic cells were obtained from the Gene Expression Omnibus (GEO) (projects PRJNA384650 and PRJNA225999). Six human mTEC samples were analyzed: four from (352) and two additional samples processed with the same protocol with minor modifications: i) after transfer to our laboratory, thymic samples were frozen in cryovials containing a cryoprotective medium composed of 5% DMSO and 95% Dextran-40 solution (5% concentration), ii) $CD45^-$ cells were magnetically enriched with the CD45 Microbeads human kit from Miltenyi Biotec (no. 130-045-801) prior to mTEC sorting, iii) cDNA libraries were prepared with the KAPA mRNAseq stranded kit (KAPA, Cat no. KK8421), and iv) sequencing generated around $400x10^6$ reads per sample. For the complete list of human samples analyzed, see Table S1 of Additional File 2. Mature murine mTECs ($mTEC^{hi}$) data were obtained from St-Pierre *et al* (383) on GEO (accession GSE65617).

Expression of transcripts derived from EREs and canonical genes

RNA-seq reads of human samples were trimmed with Trimmomatic *0.35* (384) to remove adapters and low quality sequences. Expression levels of transcripts and EREs were quantified in

transcripts per million (TPM) with kallisto *0.43.1* (385) with indexes composed of i) Ensembl 88 (GRCh38.88) transcripts and human ERE sequences from RepeatMasker or ii) Mouse mm10 (GRCm38) transcripts and murine ERE sequences from RepeatMasker for human and murine samples, respectively. TPM values of transcripts and ERE sequences were summed in genes and ERE families based on Ensembl and RepeatMasker annotations, respectively, using the aggregate function in R.

ERE expression profiling in human tissues

Expression levels of ERE families were computed for each tissue by calculating the median expression across all samples for a given tissue. The numbers of standard deviations from the mean (row Z-score) of ERE families for each tissue were determined using the scale function in R. The Euclidean distance was then calculated between all tissues based on the row Z-scores of ERE families, followed by an unsupervised hierarchical clustering. The pvClust package in R (386) was used to assess the statistical significance of the clustering using a bootstrap procedure (1000 iterations). Finally, standard deviations of expression of each ERE family between samples of a given tissue were computed.

Quintile ranking of ERE expression in somatic tissues

Median expression of ERE families were calculated among all samples of a given tissue. Tissues were then ranked based on their expression level of each ERE family individually and assigned to

quintiles of 6, 6, 8, 6 and 6 tissues, respectively. Finally, tissues were sorted based on the number of times they were assigned to the fifth quintile.

Identification and characterization of tissue-restricted EREs (TREs)

The τ-index of tissue specificity was calculated as per Yanai *et al* (387). Briefly, the τ-index is defined as:

$$\tau = \frac{\Sigma_{i=1}^{N}(1 - x_i)}{N - 1}$$

where x_i is the level of expression of a gene or ERE family in tissue i normalized to its maximal expression level among all tissues and N is the number of tissues. Genes and ERE families with $\tau \geq 0.8$ were considered as tissue-restricted. To determine in which tissue(s) a tissue-restricted gene or ERE family was overexpressed, a binary pattern was computed as reported by Yanai *et al* (387). Briefly, tissues were sorted based on their expression level for each tissue-restricted gene (TRG) or ERE family (TRE). The distance between neighboring tissues was calculated, and the maximal distance or 'gap' was used as threshold for the binary pattern. Tissues with an expression level above the gap were considered as overexpressing the TRG or TRE while other tissues were considered as underexpressing them, and were given a value of 1 or 0, respectively. ERE groups were determined for all identified TREs, and the proportions of LINE, LTR and SINE elements in TREs were compared to their representation among ERE families. A chi-squared test was performed to assess enrichment of discrete ERE groups among TREs. Using the above described binary pattern, the number of overexpressing tissues was determined for each TRG or TRE.

Impact of AIRE on ERE expression in mTECs

Lists of AIRE-dependent, AIRE-independent and constitutively expressed genes were generated as per St-Pierre *et al* (383). Expression levels of these three sets of genes as well as ERE families were compared between wild-type (n=3) and AIRE knock-out (n=3) murine mTEC^{hi} using Wilcoxon tests. Expression levels of each individual ERE family were also compared between wild-type and AIRE knock-out mice using Wilcoxon tests.

MS analyses

Immunopeptidomic data of a cohort of 16 B-lymphoblastoid cell lines (B-LCL) samples from Pearson *et al* (52) were downloaded from the Pride Archive (Project PXD004023). For the detailed protocol of mild acid elution and peptide processing, see Granados *et al* (388). Peptides were identified using Peaks X (Bioinformatics Solution Inc.) and peptide sequences were searched against the personalized proteome of each sample. For peptide identification, tolerance was set at 5 ppm and 0.02 Da for precursor and fragment ions, respectively. Occurrence of oxidation (M) and deamination (NQ) were considered as post-translational modifications.

Identification of ereMAPs

For individual B-LCL samples, RNA-seq reads were aligned to the Ensembl 88 human reference genome (GRCh38.88) using STAR (389) with default parameters. Using the intersect mode of the

BEDTools suite (390), reads entirely mapping in RepeatMasker and Ensembl annotations were separated in ERE and canonical datasets respectively, and any read seen in the canonical dataset was discarded from the ERE dataset. Unmapped reads, secondary alignments and low quality reads were then removed from the ERE dataset using Samtools view (391) with the following parameters: -f “163”, “147”, “99” or “83” and -F “3852”. In order to keep a manageable database size, ambiguous nucleotides were trimmed from reads of the ERE dataset, followed by translation in all possible reading frames. Finally, the resulting ERE amino acid sequences were spliced to remove sequences following stop codons. Only sequences of at least 8 amino acids were kept and given a unique ID to generate a theoretical ERE proteome. In parallel, a canonical personalized proteome containing the polymorphisms of the donor was generated as per (352) for each sample. Briefly, single-nucleotide variants were detected using freebayes version 1.0.2 (392), and variants with a minimal alternate count of 5 were inserted in transcript sequences using pyGeno (393). Expression levels of transcripts were quantified with kallisto using GRCh38.88 transcripts (downloaded from Ensembl) as index, and only transcripts with a TPM>0 were translated into a canonical proteome, which was concatenated with the ERE proteome to generate a Personalized Proteome unique to each sample. To validate our proteogenomic workflow, we also analyzed matched transcriptomic and immunopeptidomic data of an ovarian cancer cell line (OVCAR-3) treated with interferon-ɣ (IFNɣ) to increase MHC-I expression.

Peptide annotation and validation

Following peptide identification, a list of unique peptides was extracted for each sample and a false discovery rate (FDR) of 5% was applied on the peptide scores. Binding affinities to the sample's HLA alleles were predicted with NetMHC4.0 (394) or with NetMHCpan-4.0 (395) when an HLA allele was not included in NetMHC4.0, and only 8 to 11-amino-acid-long peptides with a percentile rank ≤ 2% were included for further annotation. For each peptide, a binary code was generated based on the presence or absence of its amino acid sequence in the ERE and canonical proteomes and an ERE status of "Yes", "Maybe" or "No" was given to the peptide accordingly. Peptides that were seen only in the ERE proteome or the canonical proteome were classified as "Yes" and "No" respectively. To determine if candidates with a "Maybe" status were ereMAP candidates, we retrieved all their possible nucleotide coding sequences from the sample's reads and split them in a set of 24-nucleotide-long subsequences (k-mers). These k-mers were then queried in 24-nucleotide-long k-mer databases generated from our ERE and canonical reads datasets using Jellyfish version 2.2.3 (396) (with the -C argument to consider the read's sequence and its reverse complement). Only peptides encoded by more than one read were kept for further validation to reduce risks of sequencing errors. If at least one of the MAP-coding sequences (MCS) was only seen in the canonical read dataset, the peptide was discarded. "Maybe" peptides were considered as ereMAP candidates if the minimal occurrence of their most abundant MCS was at least 10 times higher in the ERE k-mer database than in the canonical k-mer database. Because leucine and isoleucine variants are not distinguishable by standard MS approaches, all possible I/L variants for each ereMAPs candidates were searched in the personalized proteome. If one of the I/L variants had a higher expression in the personalized proteome, the ereMAP candidate was discarded. The genomic region generating each ereMAP candidate was determined by mapping

the reads coding for the peptide on the GRCh38.88 assembly of the reference genome with the BLAT algorithm of the UCSC Genome Browser. If a clear genomic region could not be found, the peptide was discarded. Genomic regions coding for ereMAPs candidates were then inspected in IGV (397) to see if the MCS contained known germline polymorphisms (using dbSNP v.149), and candidates were kept or discarded based on their orientation in ERE and annotated sequences. Briefly, any ereMAP candidate whose MCS mapped in the sense of a gene coding sequence was discarded, whereas candidates whose coding sequences mapped in intergenic regions were considered as ereMAPs no matter their orientation. Candidates were also discarded if they fulfilled these two conditions: i) their MCS mapped in the sense of an intron and in antisense of the ERE, and ii) if their MCS did not map in other ERE sequences (for the complete decision tree, see Figure S3). Finally, MS/MS spectra of the ereMAPs candidates were manually validated to ensure the quality of the identification. Peptides that passed all these validation steps were then considered as ereMAPs.

Characterization of ereMAPs

During manual validation in IGV, characteristics regarding the family and group of the ERE generating the peptides, the type of genomic region encoding the peptide (coding sequence, intronic or intergenic) and the orientation of the peptide sequences (sense or antisense) were retrieved for individual ereMAPs. When a peptide was identified in multiple samples and had different characteristics depending upon the sample, all possibilities were kept, otherwise they were aggregated to reduce redundancy. The expression levels of ERE families that were source or

non-source of ereMAPs were averaged among B-LCL samples, and their distributions were compared with a Mann-Whitney test. We next compared the proportions of the three main groups of EREs (LINE, LTR and SINE) in the genome, transcriptome and immunopeptidome. Representation of EREs in the transcriptome was assessed in our B-LCL samples: the expression levels of LINE, LTR and SINE elements were summed in each sample and divided by the expression level of all EREs. We then averaged these transcriptomic proportions across all B-LCL samples. We used immunopeptidomic proportions of LINE, LTR and SINE elements from the ereMAPs identified in this work, whereas the genomics proportions were taken from Treangen *et al* (374). A chi-squared test was performed to compare the proportions of ERE groups at the genomic, transcriptomic and immunopeptidomic levels. The proportions of ERE sequences located in intergenic and intronic regions as well as in coding sequences were determined by intersecting the genomic localization of ERE sequences with the localization of introns and exons from the UCSC Table Browser (files downloaded on August 21, 2019). A chi-squared test was used to determine the enrichment of a certain genomic region for ereMAPs generation. Last, Kendall tau correlation between the number of ereMAPs generated by each ERE family and the number of copies of the family's sequence in the human genome (determined from RepeatMasker annotations) was computed with a confidence level of 95%.

Expression profiling of ereMAPs' coding sequences

To evaluate the expression of the ereMAP-coding sequences in peripheral tissues, we downloaded RNA-seq data of 30 tissues from the GTEx consortium (phs000424.v7.p2). For the

complete protocol of this analysis, see Laumont *et al* (352). Briefly, we generated 24-nucleotide-long k-mer databases for each sample, in which we queried each ereMAP-coding sequence's 24-nucleotide-long k-mer set. For each ereMAP, the minimal occurrence in the k-mer set was used as the number of reads coding for the peptide in a given sample ($r_{overlap}$). The number of reads coding for a peptide was normalized between RNA-seq experiments by dividing $r_{overlap}$ by the total number of reads of the sample and multiplying this number by 10^8 to obtain the number of reads detected per hundred million reads sequenced (rphm). We then averaged the log-transformed rphm values ($log_{10}(rphm + 1)$) for each tissue, and an average expression superior to 10 rphm in a tissue was considered as significant. This analysis was also performed on 12 TCGA cohorts (50 randomly selected samples per cohort) to assess the expression level of the ereMAPs identified on B-LCLs in the following cancer types: urothelial bladder carcinoma, breast invasive carcinoma, colon adenocarcinoma, head-neck squamous cell carcinoma, kidney renal clear cell carcinoma, liver hepatocellular carcinoma, lung adenocarcinoma, lung squamous cell carcinoma, ovarian cancer, pancreatic adenocarcinoma, prostate adenocarcinoma and skin cutaneous melanoma. Last, methylation data (HM27 array for ovarian cancer, HM450 for other cancer types) matched with the RNA-seq samples used to profile ereMAPs' expression in TCGA cohorts were downloaded when available. Only probes located in a window of 5000 nucleotides from the ereMAPs' genomic locations were used for this analysis. We then computed the Pearson correlation between the ereMAP's RNA expression (in rphm) and the methylation level of the genomic region coding for the peptide.

Amino acid composition of ereMAPs

In addition to the list of ereMAPs identified on our B-LCL samples, two linear and MHC I-restricted epitopes' sequences datasets were downloaded from the Immune Epitope Database: a first dataset of 36 472 MAPs from any virus infecting human cells and a second one of 282 069 human canonical MAPs (downloaded on August 7, 2019). Lists of 8 to 11-amino-acid-long MAPs were extracted from these two datasets. Usage frequency of each amino acid was calculated by dividing their occurrences by the total number of amino acids in the ERE, viral and human canonical MAPs datasets. In parallel, datasets were separated in subsets of 8, 9, 10 and 11-amino-acid-long MAPs, and frequencies of amino acids were computed for each peptide position of each subset of MAPs. The 11-amino-acid-long MAP subset was discarded because of an insufficient number of ereMAPs (n = 2).

Viral homology

To assess the similarity between ereMAPs and viral peptides, we used the same datasets of viral and human canonical MAPs from the Immune Epitope Database used for the amino acid composition analysis (see section "Amino acid composition of ereMAPs" of the Methods). We aligned ereMAP sequences to this database of viral peptides using version 2.2.28 of the Protein Basic Local Alignment Tool (BLASTp) (398) in the blastp-short mode with the following arguments: -word_size 2, -gapopen 5, -gapextend 2, -matrix PAM30, and -evalue 10 000 000. As a control, human canonical MAPs were aligned to the viral peptides dataset with BLASTp. For the viral homology analysis, we compared the 103 ERE MAPs to 10,000 groups of 103 randomly sampled

canonical MAPs. We calculated the percentage of identity (%I) of ereMAPs and canonical MAPs with viral peptides as:

$$\%_I = \frac{M_{max} \times L_a}{L_p} \times 100\%$$

where M_{max} is the maximal percentage of identical matches with the viral MAPs database, L_a is the length of the alignment and L_p is the length of the ereMAP or the canonical MAP. The average percentage of identity of ereMAPs and each subgroup of the bootstrap distribution was computed, and the p-value was determined as the number of times that the percentage of identity of the bootstrap distribution was higher than the percentage of identity of ereMAPs divided by the number of bootstrap iterations (10,000) as per Granados *et al* (53).

ereMAPs' immunogenicity prediction

We used the Repitope algorithm (399) with default settings to predict ereMAPs' immunogenicity to determine their potential to activate CD8 T cells. As negative control, immunogenicity scores of thymic MAPs identified by *Adamopoulou et al* (400) were also computed using Repitope, and the distributions of thymic MAPs' and ereMAPs' immunogenicity scores were compared with a Mann-Whitney test.

Generation of monocyte-derived dendritic cells

Monocyte-derived dendritic cells were generated from frozen PBMCs, as previously described (401, 402). Briefly, DCs were prepared from the adherent PBMC fraction by culture for 8 days in X-vivo 15 medium (Lonza Bioscience) complemented with 5% human serum (Sigma-Aldrich), Sodium pyruvate (1 mM), IL-4 (100 ng/mL, Peprotech) and GM-CSF (100 ng/mL, Peprotech). After 7 days of culture, DCs were matured overnight with IFNγ (1000 IU/mL, Gibco) and LPS (100 ng/mL, Sigma Aldrich). DCs were loaded with 2 µg/mL of peptide during 2h after maturation process and were then irradiated (40 Gy) before they were used as APCs in T-DC culture. As control, the experiment was performed for the MelanA peptide when the number of T cells was sufficient.

In vitro peptide-specific T cell expansion

Peptide-specific CD8+ T cells were expanded as previously described, with some minor modifications (402, 403). Briefly, thawed PBMCs were first $CD8^+$ T-cell enriched using the Human $CD8^+$ T cell isolation kit (Miltenyi Biotech) and co-incubated with autologous peptide-pulsed DCs at an APC:T cell ratio of 1:10. Expanding T cells were cultured for four weeks (with pulsed-DC stimulation every 7 days) in Advanced RPMI medium (Gibco) supplemented with 8% human serum (Sigma-Aldrich), L-glutamine (Gibco) and cytokines. For the first coculture week, IL-12 (10 ng/mL) and IL-21 (30 ng/mL) were added to the medium. Two days after, IL-2 (100 UI/mL) was also added to the cytokine mix. The second week, IL-2 (100 UI/mL), Il-7 (10 ng/mL), IL-15 (5 ng/mL) and Il-21 (30 ng/ml) were added to the medium. For the two last weeks of coculture, IL-2 (100 UI/mL), Il-7 (10 ng/mL) and IL-15 (5 ng/mL) were used. Medium supplemented with the appropriate cytokine mix was added in the cocultures every two days. At the end of the fourth

week of coculture, cells were harvested in order to perform an ELISPOT assay. If the number of specific T cells was not sufficient at the end of the fourth week of coculture, cocultures were maintained for an additional week (week 5).

IFNγ ELISPOT assay

ELISpot Human IFNγ (R&D Systems, USA) kit was used according to the manufacturer's recommendations to perform the experiment. Harvested CD8+ T cells were then plated and incubated at 37°C for 24 hours in the presence of irradiated peptide-pulsed PBMCs (40 Gy) that were used as stimulator cells. As negative control sorted CD8 T cells were incubated with irradiated nonpulsed PBMCs. Spots were revealed as mentioned in the manufacturer protocol and were counted using an ImmunoSpot S5 UV Analyzer (Cellular Technology Ltd, Shaker Heights, OH). IFNγ production was expressed as the number of peptide-specific spot-forming cells (SFC) per 10^6 CD8+ T cells after subtracting the spot counts from negative control wells.

2.1.3.4 Results

Expression of ERE transcripts in normal human tissues and cells

To assess ERE expression in healthy human tissues, we quantified the expression levels of the 809 ERE families contained in the RepeatMasker annotations in 1371 samples from 30 different healthy human tissues and 2 cell types (mTECs and ESCs). For brevity, mTECs and ESCs will be referred to as tissues in the rest of the manuscript. We calculated the median expression of each ERE family among samples of a given tissue (Table S2) and then computed the row Z-score across

tissues. Unsupervised hierarchical clustering identified a significant cluster of three cell types with high ERE expression: ESCs, testis and mTECs (Figure 2.1). Remaining tissues could then be visually separated in two groups of low and intermediate ERE expression (Figure 2.1). High ERE expression (cluster 1) in ESCs and testis was expected. The salient finding was the high ERE expression in mTECs which, to the best of our knowledge, has never been reported before. Comparison with hematopoietic cell types at several differentiation stages confirmed the high ERE expression in mTECs and ESCs (Supplementary Figure 2.1A). Computing the standard deviation of ERE expression among individual samples for each tissue also revealed that most ERE families displayed low interindividual variability (Supplementary Figure 2.1B). Finally, while quintile ranking analysis showed that ERE expression was generally concordant between ERE families in each tissue analyzed, almost all tissues expressed some ERE families at high level (Supplementary Figure 2.2), suggesting that some tissue-specific factors regulate ERE expression in human tissues.

Most human tissues show a tissue-specific ERE expression.

To ascertain if expression of discrete ERE families was restricted to specific tissues, we computed the τ-index of tissue-specificity as defined by Yanai *et al* (387). Briefly, the τ-index compares the expression of a gene in a set of tissues and has a value ≤0.4 for housekeeping genes and ≥0.8 for tissue-restricted genes (404). We identified a total of 124 ERE families with a tissue-restricted expression. As control, we computed the τ-index for annotated genes and known tissue-restricted genes (TRGs), such as *INS*, *CRP* and *CHRNA1*. The majority (108/124) of the tissue-restricted ERE families (TREs) were identified in ESCs, testis and mTECs, revealing that in addition to their high

expression of EREs, these tissues express a broader repertoire of EREs than other tissues (Figure 2.1, Figure 2.2A). Nonetheless, tissue-restricted expression of EREs is a widespread phenomenon across human tissues because we identified TREs in 17 out of the 32 human tissues analyzed. For a given tissue, the number of TREs is positively associated with the number of TRGs (Figure 2.2A) suggesting some commonality between expression regulation of TRGs and TREs. We also identified a significant enrichment of LTRs in TREs (86.29%) relative to their proportion among all ERE families (71.45%), revealing an increased tissue specificity of LTR sequences compared to LINEs and SINEs (Figure 2.2B). Finally, TREs' expression was typically restricted to fewer tissues than TRGs, with 89.5% of TREs (111/124) being tissue-specific (Figure 2.2C, Table S3). Altogether, these results show that ERE expression in healthy human tissues is widespread but not homogeneous. Indeed, 124 ERE families, most of which are LTR elements with low copy numbers, showed tissue-specific expression.

Impact of the *AIRE* gene on ERE expression in mTECs

Out of the three tissues with high ERE expression (Figure 2.1), two express no or barely detectable MHC-I molecules (testis and ESCs, respectively), whereas mTECs express standard levels of MHC-I (38, 405, 406). Promiscuous expression of genomic sequences is a quintessential feature of mTECs that is driven in part by the *AIRE* gene and also by other genes whose identity is still debated (407). Since the role of mTECs is to induce tolerance to the MAPs that they display, EREs expressed in mTECs could be tolerogenic. However, T cell-mediated responses towards EREs were previously observed, suggesting that the establishment of central tolerance towards EREs in the

thymus is incomplete (369, 408). Therefore, we next investigated the contribution of the AIRE transcription factor to ERE expression in mTECs. To do so, we quantified the expression of ERE families as well as canonical genes in mTECs extracted from wild-type and AIRE knock-out mice. Canonical genes were sorted in three categories based on St-Pierre *et al* (383) : i) constitutively expressed genes, ii) AIRE-independent TRGs and iii) AIRE-dependent TRGs. As expected, expression of AIRE-dependent TRGs significantly decreased in the absence of AIRE, whereas constitutively expressed genes and AIRE-independent TRGs were minimally affected by AIRE depletion (Figure 2.3A). Strikingly, global ERE expression was independent of AIRE since it was unchanged in AIRE knock-out relative to wild-type mice (Figure 2.3A). Furthermore, computing Mann-Whitney tests for each ERE family revealed that the absence of AIRE did not affect the expression of any ERE family (Figure 2.3B). Hence, expression of all ERE families was independent of AIRE in mTECs.

Translation of ERE transcripts by healthy cells

We next sought to determine whether some ERE transcripts are translated in healthy cells. However, the identification of EREs by MS can be challenging due to their inherently low abundance in the corresponding proteome and the lack of appropriate protein databases for large-scale searches. We therefore decided to investigate the contribution of EREs to the immunopeptidome, which is mainly composed of peptides derived from rapidly degraded proteins (409, 410). To do so, we reanalyzed previously reported transcriptomic and immunopeptidomic data from 16 B-lymphoblastoid cell lines (B-LCL) (Table S4) (52). As

conventional approaches do not include ERE sequences, we developed a proteogenomic workflow combining RNA-sequencing and MS to enable ereMAP identification (Figure 2.4A, Supplementary Figure 2.3). Briefly, we generated for each B-LCL a personalized proteome that contained only the sample's expressed sequences as well as its polymorphisms. Canonical and ERE RNA sequences were translated *in silico* and concatenated to generate a personalized proteome that was used to identify MAPs in MS analyses (Figure 2.4A). For each MAP identified, we retrieved the peptide's coding sequence and proceeded to its annotation. Two categories of peptides were kept as ereMAP candidates to be further manually validated: i) peptides that were only seen in the ERE proteome, and ii) peptides seen in both the ERE and canonical proteomes ("Maybe" candidates) and for which the occurrence of the coding sequences was at least 10-fold higher in ERE reads compared to canonical reads.

Our proteogenomic approach enabled the identification of 129 ereMAPs in the 16 B-LCL samples analyzed, revealing that ERE sequences are translated in non-neoplastic cells (Figure 2.4B). Of those, 103 were non-redundant, confirming that ereMAPs can be shared by multiple individuals (Table S5). Of course, the extent of interindividual sharing would be considerably greater in cohorts of HLA-matched individuals since various HLA allotypes present different sets of MAPs (409). Profiling of the ereMAPs' RNA expression in healthy human tissues showed that 26% (27/103) of ereMAPs' coding sequences were expressed at high levels by multiple tissues (Supplementary Figure 2.4). Hence, since highly expressed transcripts are preferential sources of MAPs (52), ereMAPs derived from abundant transcripts could be presented on the surface of a wide range of tissues (Supplementary Figure 2.4). We also observed that ereMAPs were

generated by the three main groups of ERE sequences (SINE, LINE, LTR), confirming that they all have the potential to be translated in healthy cells (Figure 2.4C). As EREs are frequently dysregulated in cancer cells, we quantified the RNA expression of the ereMAPs identified on B-LCLs in 12 cohorts from TCGA (Supplementary Figure 2.5A). Strikingly, the majority of ereMAPs (94/103, 91.3%) were expressed at similar levels by healthy and cancer cells (Figure 2.5B), and ereMAPs' RNA expression in cancer cells did not correlate with DNA methylation levels (Supplementary Figure 2.5C). Additionally, applying our proteogenomic workflow on an ovarian cancer cell line (OVCAR-3) enabled the identification of 5 ereMAPs, including one peptide (TPRHIIVRF) also presented by B-LCL samples (Table S6). Together, these proteogenomic analyses show that several EREs are translated and generate ereMAPs in B-LCLs, and suggest that this is also the case in a wide range of human tissues and pathologies.

High expression of intronic regions is the main source of ereMAPs

We next investigated the mechanisms leading to presentation of ereMAPs on the cell surface. First, we noted that ereMAPs preferentially derived from highly expressed ERE transcripts (Figure 2.5A). For the majority of ereMAPs, this transcription was in the same sense as the ERE sequence in the genome, but ~30% of ereMAPs (34/103) resulted from antisense transcription (Figure 2.5B), which is common for EREs (335, 336, 411). Even though ereMAPs were generated by the three main groups of EREs (Figure 2.4C), the relative frequency of LTR translation was higher than that of LINEs and SINEs (Figure 2.5C). Indeed, the representation of LTRs in the immunopeptidome was superior to the space they occupy in the genome or their abundance in the transcriptome

(Figure 2.5C). Additionally, intronic EREs were a preferential source of ereMAPs: while 51% of EREs were intronic, ~79% of ereMAPs derived from intronic EREs (Figure 2.5D). Finally, we noted that some ERE families generated several distinct ereMAPs (Table S5). This can be explained in part by variations in the genomic space occupied by the various ERE families. Indeed, we observed a moderate, yet significant, correlation between the number of genomic copies and the number of ereMAPs (Figure 2.5E). Altogether, these results demonstrate that i) ereMAPs are generated by both sense and antisense transcripts that are preferentially located in introns and expressed at high levels, and ii) generation of ereMAPs is enhanced when a family belongs to the LTR group occupying a large genomic space.

ereMAPs have a viral-like amino acid composition

We next asked to what extent ereMAPs and their coding transcripts might retain some traces of their phylogeny ("viral features"). We found conspicuous differences between amino acid frequencies in ereMAPs relative to both viral MAPs and canonical human MAPs listed in the Immune Epitope Database (Figure 2.6A). Indeed, ereMAPs showed lower abundance of multiple amino acids (aspartic and glutamic acids, phenylalanine, methionine, asparagine, and tryptophan) and higher frequencies of leucine (L) and proline (P) residues. ereMAPs had therefore a less balanced (i.e., more skewed) amino acid composition. Furthermore, analysis of amino acid usage at individual MAP positions revealed that, relative to human MAPs, some residues were specifically enriched in ERE and viral MAPs, such as arginine (R) in P5 of 8 amino acid-long MAPs (Figure S6). We therefore aligned ereMAPs sequences to the viral MAPs dataset using BLAST and

calculated the average percentage of identity between ereMAPs and viral MAPs. We then compared this result with a bootstrap distribution (10,000 iterations) of randomly selected canonical MAPs that were also aligned to the viral MAPs dataset (Figure 2.6B). This analysis revealed that ereMAPs had a significantly higher percentage of identity with viral MAPs than all 10,000 randomly selected sets of canonical MAPs. Finally, we investigated if the viral features of ereMAPs might confer them the ability to activate CD8 T cells. First, immunogenicity prediction using the Repitope algorithm showed that ereMAPs have significantly higher immunogenicity scores than canonical MAPs presented in the thymus (Supplementary Figure 2.7A). Additionally, IFNɣ ELISpot assays demonstrated that two cancer-specific ereMAPs (*i.e.* not expressed by mTECs), identified by *Laumont et al* (352) on B-ALL samples, have the ability to activate CD8 T cells (Supplementary Figure 2.7B, C). Hence, ereMAPs clearly retain features that reflect their viral origin, conferring them the ability to elicit CD8 T cell responses when their expression is repressed in mTECs.

2.1.3.5 Discussion

Hundreds of scientific articles have alluded to the potential implication of EREs in various human diseases, particularly cancer and autoimmunity (198, 320, 337, 412-415). We therefore felt compelled to draw the global landscape of ERE expression in human somatic cells. One salient point emerging from this atlas is that ERE expression in somatic tissues is more pervasive and heterogeneous than anticipated. All tissues express EREs but the breadth and magnitude of ERE expression are very heterogeneous from one tissue to another. Thus, we identified 124 ERE families expressed in a tissue-restricted fashion, most of which were LTR elements. LTRs can act

as promoters and enhancers to stimulate gene expression (212, 338), and some LTR families are tissue-specifically enriched in intronic enhancer regions containing transcription factor binding sites (214). Our work therefore suggests that EREs, and more particularly LTRs, may regulate gene expression in a wide range of somatic tissues. In future experiments, single cell analyses might unveil a further level of heterogeneity that we could not capture by global tissue expression profiling. It was previously reported that EREs were expressed at high levels in two MHC I-deficient cell types: ESCs and testis (416, 417). That similar levels of expression were found in mTECs for three major groups of EREs (LINE, SINE and LTR) (Fig. 1) is remarkable and raises fundamental questions as to the mechanism and role of ERE expression in mTECs. The key role of mTECs is to induce central immune tolerance to a vast repertoire of self-peptides displayed by somatic tissues (407, 418). Given the large-scale expression of EREs in peripheral tissues highlighted in the present report, we speculate that it may be important for gnathostomes to be tolerant to a wide array of ERE-derived antigens. As a corollary, when EREs are overexpressed, for instance in cancer cells (319, 323), only those that are not expressed in mTECs may be immunogenic. Induction of tolerance to the multitude of self-peptides depends on the unique ability of mTECs to promiscuously express thousands of otherwise tissue-specific genes (21, 119). Promiscuous gene expression in mTECs is driven in part by *AIRE* and in part by other genes whose identity is unresolved, which may include *FEZF2* as well as genes involved in DNA methylation, histone modification and RNA splicing (121, 129, 383, 407, 419). Our data clearly show that the overexpression of numerous ERE families in mTECs is entirely AIRE-independent (Fig. 3). This observation underscores the relevance of further studies on the mechanisms of AIRE-independent promiscuous gene expression in mTECs.

A notable finding was that our MS analyses identified ereMAPs derived from LINEs (n = 47), SINEs (n = 29) and LTRs (n= 27). This means that these EREs are translated and produce peptides that are adequately processed for presentation by MHC-I molecules. Our analyses suggest a higher propensity of LTRs towards peptide generation. As SINEs do not contain protein-coding sequences, they were expected to generate fewer peptides. However, the reason why LTRs would be more efficiently translated than LINEs remains elusive but might include codon usage and sequence conservation. A few ereMAPs have previously been identified in cancer cells (323, 352, 414). The presence of ereMAPs on normal cells means that the mere identification of ereMAPs on cancer cells could not be sufficient to infer that these MAPs are cancer-specific nor immunogenic. Nevertheless, we have previously shown in mice that some ereMAPs are truly cancer-specific, immunogenic and can elicit protective anti-tumor responses (352). Furthermore, compelling evidence has been reported that some LTRs can generate immunogenic ereMAPs in clear cell renal cell carcinoma in humans (337). These studies coupled to our findings that ereMAPs i) retain viral-like features (Fig. 6) and ii) can be recognized by CD8 T cells (Fig. S7B and C) suggest that ereMAPs may represent particularly attractive targets for the development of cancer vaccines. In line with this, we must also emphasize that the number of translated EREs is certainly superior to the number of ereMAPs identified in our study: i) collectively our 16 B-LCLs expressed 39 MHC-I allotypes out of the thousands that can be found in human populations (Table S5), and ii) like canonical proteins (52), some translated EREs may not generate MAPs.

We anticipate that the biogenesis of ereMAPs in normal and neoplastic cells will be a fertile field of investigation. First, several observations suggest that the landscape of ereMAPs is highly diversified: i) the MAP repertoire is shaped by several cell type-specific variations in gene expression (47), and ii) ERE transcription is highly heterogeneous among various cell types (Fig. 1) and can be drastically affected by neoplastic transformation (420). The processing of ereMAPs is also intriguing. Indeed, following their integration in human genomes, EREs have undergone several rounds of mutation and truncation and very few have previously been shown to be translated (198, 242). Because ERE sequences are degenerate, they are not expected to yield stable polypeptides. However, MAPs preferentially derive from rapidly degraded unstable peptides, commonly referred to as defective ribosomal products (410). We therefore hypothesize that for most EREs, translation may yield ereMAPs but not stable long-lived proteins. In other words, the products of ERE translation may be detectable only in the immunopeptidome and not in the proteome.

2.1.3.6 Conclusions

In summary, transcriptomic analysis demonstrated that ERE expression is heterogeneous in healthy human tissues, with a higher expression in mTECs, ESCs and testis than in other tissues. mTECs are the sole normal human cells that express high levels of both EREs and MHC-I molecules. In mutant mice, we report that the exceptional expression of EREs in mTECs is AIRE-independent. We also identified ERE families expressed in a tissue-restricted manner, revealing that most healthy human tissues have a unique ERE signature. MS analyses of 16 B-LCL samples enabled the identification of 103 non-redundant ereMAPs, showing that EREs contribute to the immunopeptidome of healthy cells. Interestingly, sharing of ereMAPs by multiple B-LCL samples

was observed, and ereMAPs' coding sequences are expressed at similar levels in other somatic tissues, suggesting that ereMAPs could also be presented by other cell types. Finally, we found that ereMAPs bear strong homology to viral MAPs and therefore have the potential to be particularly immunogenic. We hope that this work will serve as a reference in further studies on EREs in various physiological and pathological conditions.

2.1.3.7 Abreviations

AIRE : Autoimmune regulator; B-LCL: B-lymphoblastoid cell line; ELISpot: Enzyme-Linked Immunospot ; ERE: Endogenous Retroelements; ereMAP: ERE-derived MAP; ESC: Embryonic stem cells; FDR: False discovery rate; GTEx: Genotype-Tissue Expression project; IFNɣ: Interferon-ɣ; LINE: Long interspersed nuclear element; LTR: Long terminal repeat; MCS : MAP-coding sequence; MAP: MHC I-associated peptide; mTEC: medullary thymic epithelial cells; MS: Mass spectrometry; SINE: Short interspersed nuclear element; TPM: transcripts per million; TRE: Tissue-restricted ERE; TRG: Tissue-restricted gene; WT: Wild-type; KZFP: KRAB Zinc Finger Protein

2.1.3.8 Declarations

Ethics approval and consent to participate

The study of MHC-associated peptides on human lymphoid cells was approved by the Comité d'Éthique de la Recherche de l'Hôpital Maisonneuve-Rosemont (Permit Number CÉR 2018-1396).

Consent for publication

Not applicable.

Availability of data and material

mTECs' RNA sequencing datasets generated during this study are available on GEO as GSE127826 (BioProject accession number: PRJNA525591). Transcriptomic data of four additional mTEC samples, previously reported by Laumont et al. (352), are publicly available on GEO (BioProject accession number: PRJNA525590). ESCs' transcriptomic data from Lister et al. (382) are available on the short read archive (Accessions: SRR488684 and SRR488685). RNA-seq data of purified hematopoietic cells were obtained from the Gene Expression Omnibus (GEO) (projects PRJNA384650 and PRJNA225999) (421, 422). RNA sequencing data of WT and AIRE-deficient mice were reported by St-Pierre et al. (383). Transcriptomic and immunopeptidomic data of B-LCL samples from Pearson et al. (52) were downloaded from GEO (BioProject accession number: PRJNA286122) and the PRIDE Archive (Project PXD004023), respectively. Transcriptomic and immunopeptidomic data of the OVCAR-3 cell line are available on GEO as GSE147570 (BioProject accession number: PRJNA615537) and the PRIDE Archive database (Project PXD018124), respectively.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by grants from the Canadian Institutes of Health Research (FDN 148400) and the Canadian Cancer society (#705604).

Authors' contributions

JDL, KV and CP designed the study. LH and CC digested the thymic samples, isolated the mTECs and did the RNA extraction. JDL, AT, GE, PG and JPL contributed to the bioinformatic analyses. CD and EB did the PEAKS database searches and the MS/MS spectra validation. JDL and CP wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge Annie Gosselin and Gaël Dulude for cell sorting. We thank Raphaëlle Lambert and Jennifer Huber for performing the RNA sequencing. We also thank Céline M Laumont and Qingchuan Zhao for their conceptual input, as well as Marie-Pierre Hardy and all other members of our laboratory for their suggestions. We thank the Leucegene group for sharing transcriptomic data for hematopoietic cells. Finally, we thank the Genotype-Tissue Expression (GTEx) Project for providing RNA-seq data from human tissues used in this study. The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

2.1.3.9 Figures

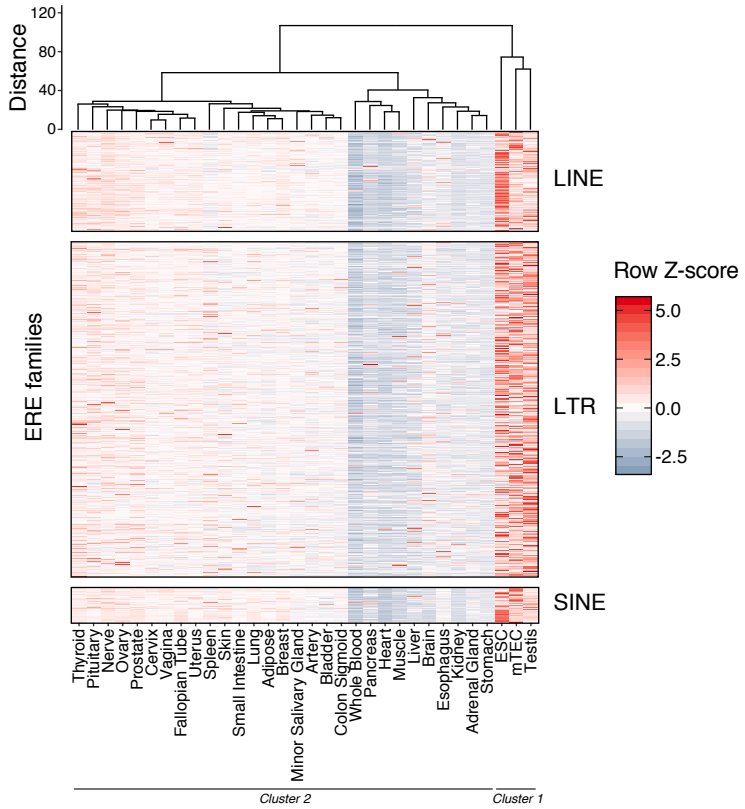


Figure 2.1 - Expression profiling of endogenous retroelements in 30 healthy human tissues and 2 cell types.

Hierarchical clustering of tissues based on the expression levels of the 809 ERE families sorted in LINE, LTR and SINE elements. For each tissue, mean expression of ERE families was computed among available samples. Row Z-scores were then determined for each ERE family across tissues. Significant clusters identified by pvClust are indicated.

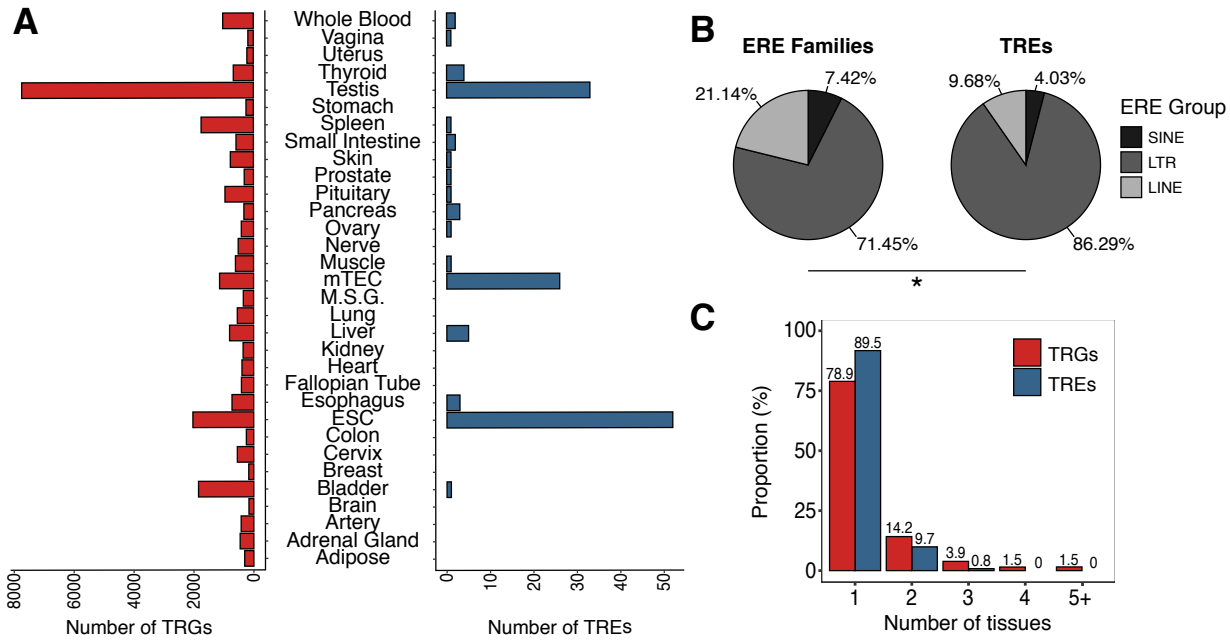


Figure 2.2 - Tissue specificity of ERE expression in healthy human tissues.

Tissue-specificity indexes were computed for ERE families as well as annotated genes. (A) Barplots showing the number of TRGs and TREs for each of the 32 healthy human tissues analyzed. (B) Pie charts depicting the proportions of the 809 ERE families (left panel) or TREs (right panel) belonging to the LINE, LTR and SINE groups (Chi-squared test, *$P \leq 0.05$). (C) Histogram showing the number of tissues in which each identified TRGs and TREs are overexpressed.

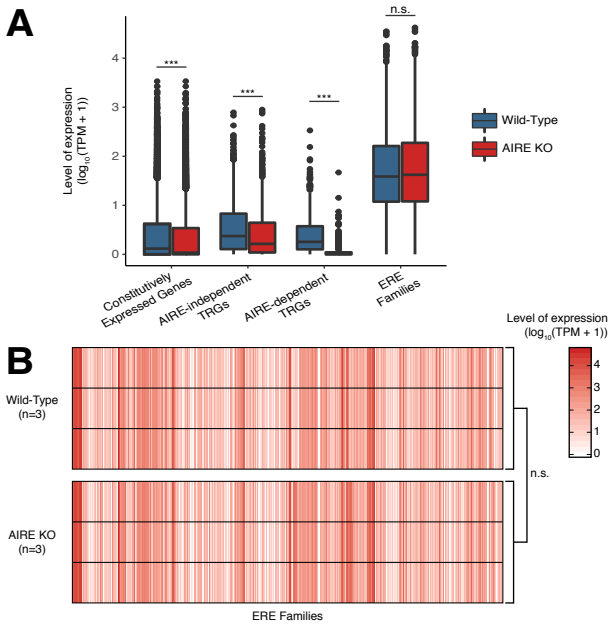


Figure 2.3 - ERE expression is independent of AIRE in mouse mTECs.

(A) Boxplot showing the expression levels of constitutively expressed genes, AIRE-dependent TRGs, AIRE-independent TRGs (lists of genes based on St-Pierre *et al* (383)) as well as ERE families in wild-type (n=3) and AIRE knock-out (n=3) mice. (B) Heatmap depicting the expression levels of ERE families in each replicate of wild-type and AIRE knock-out murine mTECs. A Mann-Whitney test was used for statistical analysis in both panels, n.s. not significant ($P>0.05$), ***$P\leq0.001$.

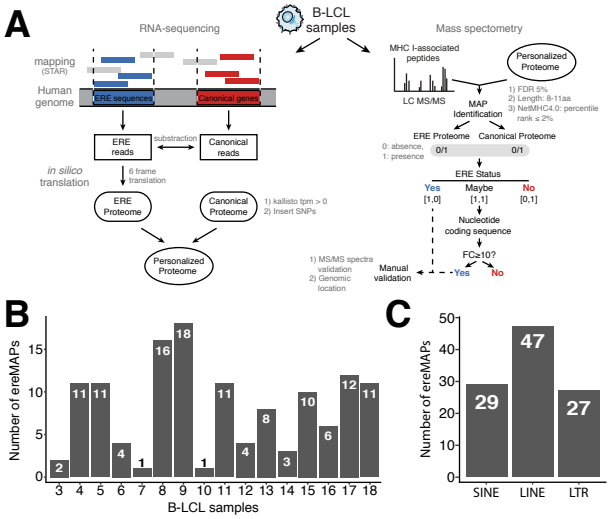


Figure 2.4 - ERE sequences are translated and contribute to the immunopeptidome of B-LCLs.

(A) Schematic depicting how the personalized proteome of each B-LCL sample was generated. The personalized proteome was generated by combining the ERE and the canonical proteomes and then used to identify MAPs by MS. MAPs were annotated to keep only ereMAPs. (B, C) Barplots showing the number of ereMAPs identified in B-LCL samples separated by (B) individual samples analyzed and (C) according to the three main groups of EREs.

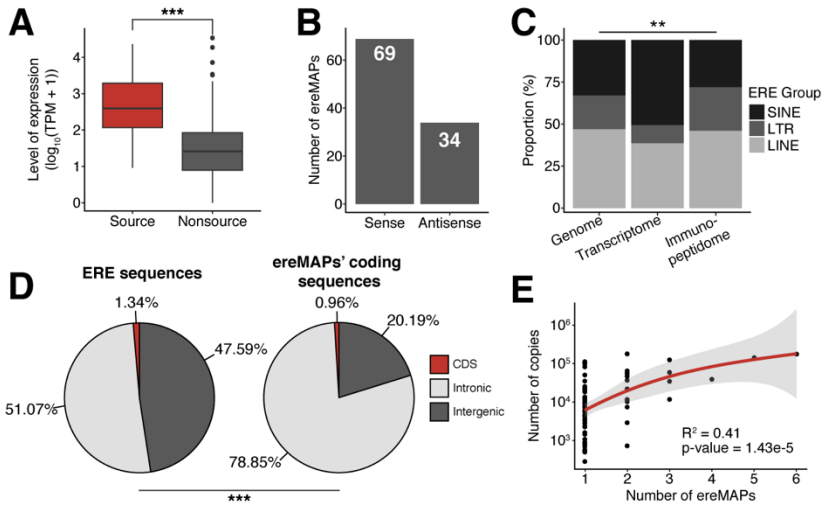


Figure 2.5 - Sense transcription of intronic EREs is the main source of ereMAPs.

(A) Boxplot showing the mean expression levels (log_{10}(TPM + 1)) of ERE families that are source or non-source of ereMAPs in B-LCLs (Mann-Whitney test, ****P*≤0.001). (B) Barplot showing the number of ereMAPs generated by sense or antisense transcription of ERE sequences. (C) Stacked barplot depicting the proportions of LINE, LTR and SINE groups in the genome, transcriptome and immunopeptidome. Statistical significance was computed with a chi-squared test (***P*≤0.01). (D) Pie charts depicting the percentages of all ERE sequences (left) and of ereMAPs-coding sequences (right) that are localized in intergenic regions, introns or coding sequences (Chi-squared test, ****P*≤0.001). (E) Scatterplot showing the Kendall tau correlation between the number of

ereMAPs generated by each ERE family and the number of copies of the ERE family's sequence in the human genome based on RepeatMasker annotations.

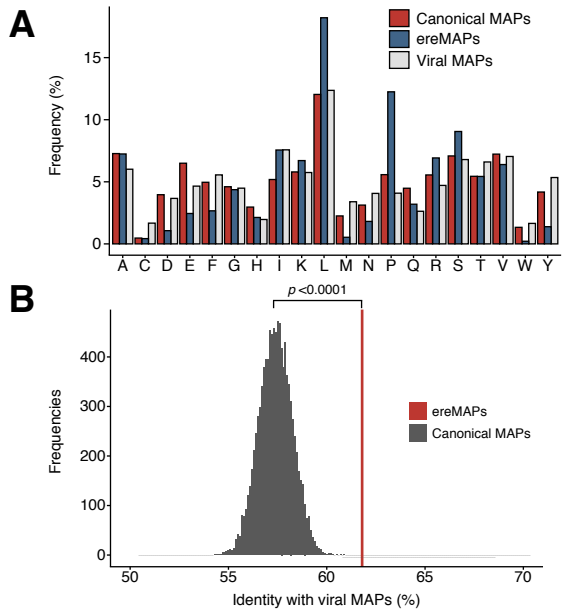
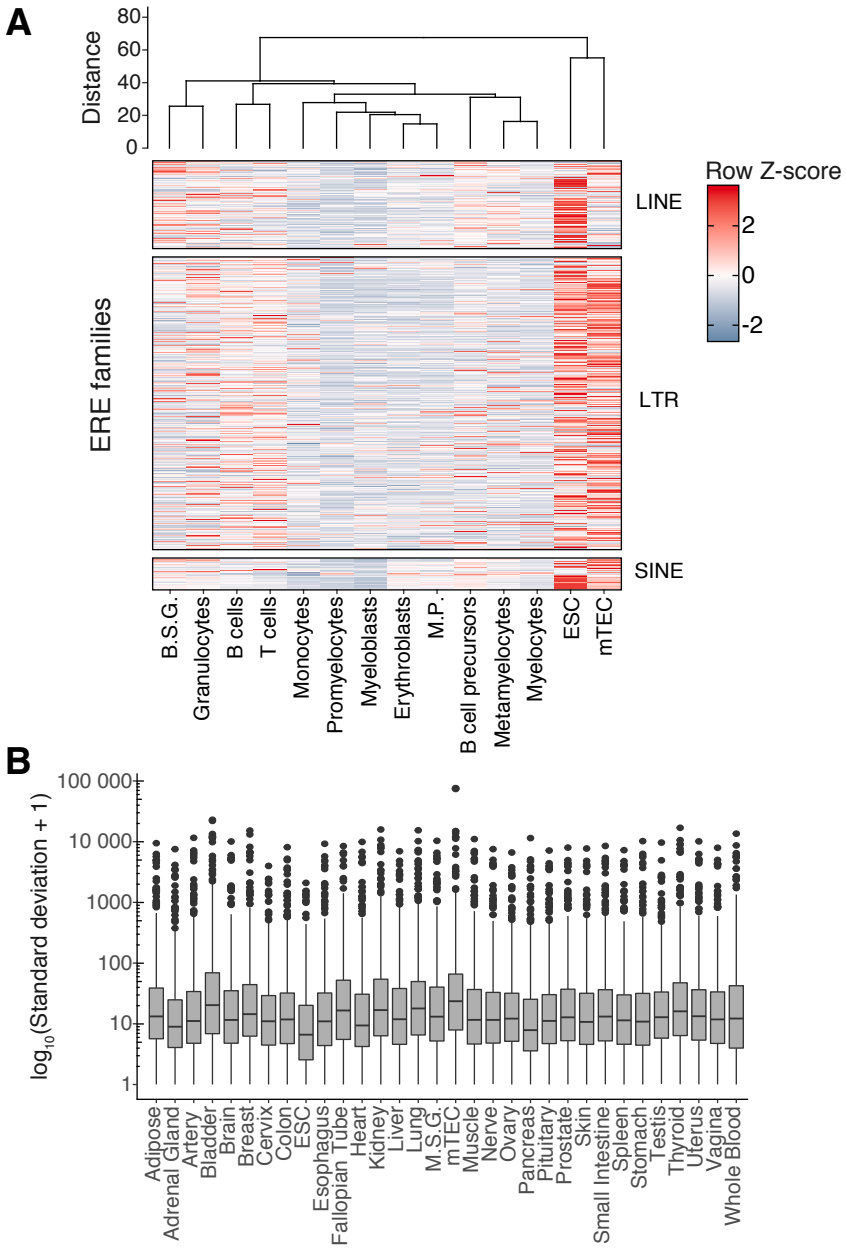


Figure 2.6 - Endogenous retroelements retained sequence homology with viruses.

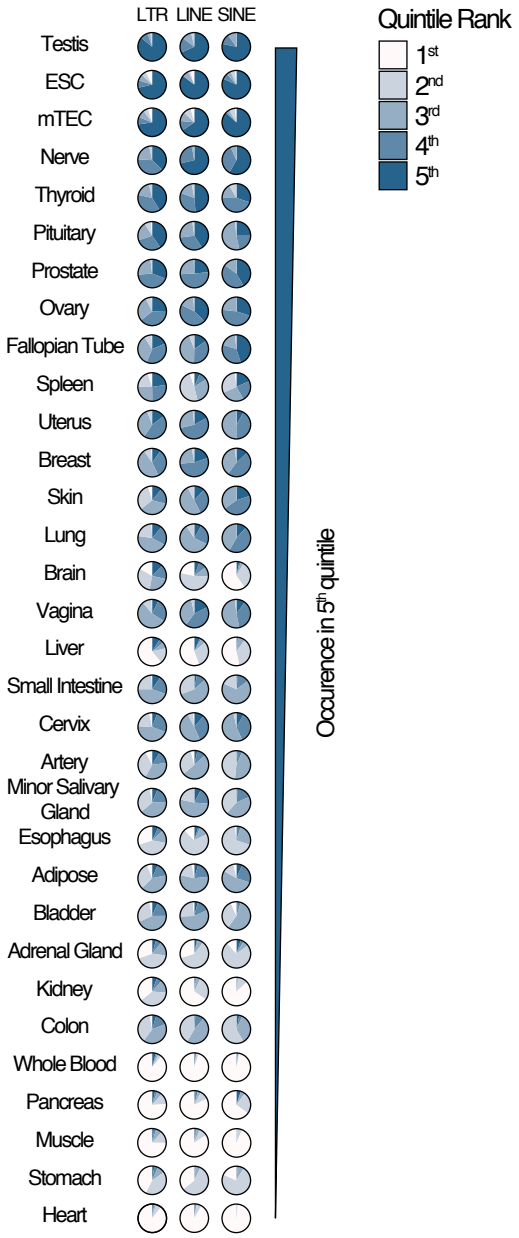
(A) Barplot showing the frequencies of each amino acid in ereMAPs, viral MAPs and human canonical MAPs. Abbreviations for amino acids: Y, Tyrosine; W, Tryptophan; V, Valine; T, Threonine; S, Serine; R, Arginine; Q, Glutamine; P, Proline; N, Asparagine; M, Methionine; L, Leucine; K, Lysine; I, Isoleucine; H, Histidine; G, Glycine; F, Phenylalanine; E, Glutamic Acid; D, Aspartic Acid; C, Cysteine; A, Alanine. (B) Human canonical MAPs and ereMAPs were aligned to a database of viral peptides using BLAST, and the percentage of identity of their sequences with viral peptides was computed. The red line represents the average percentage of identity of ereMAPs with viral MAPs. A bootstrap procedure was used to calculate the percentage of identity of 10,000 sets of 103 randomly selected human canonical MAPs with viral MAPs. P-value was

calculated as the number of times the bootstrap distribution had a higher percentage of identity with viral MAPs than ereMAPs (P<0.0001).

2.1.2.10 Supplementary figures

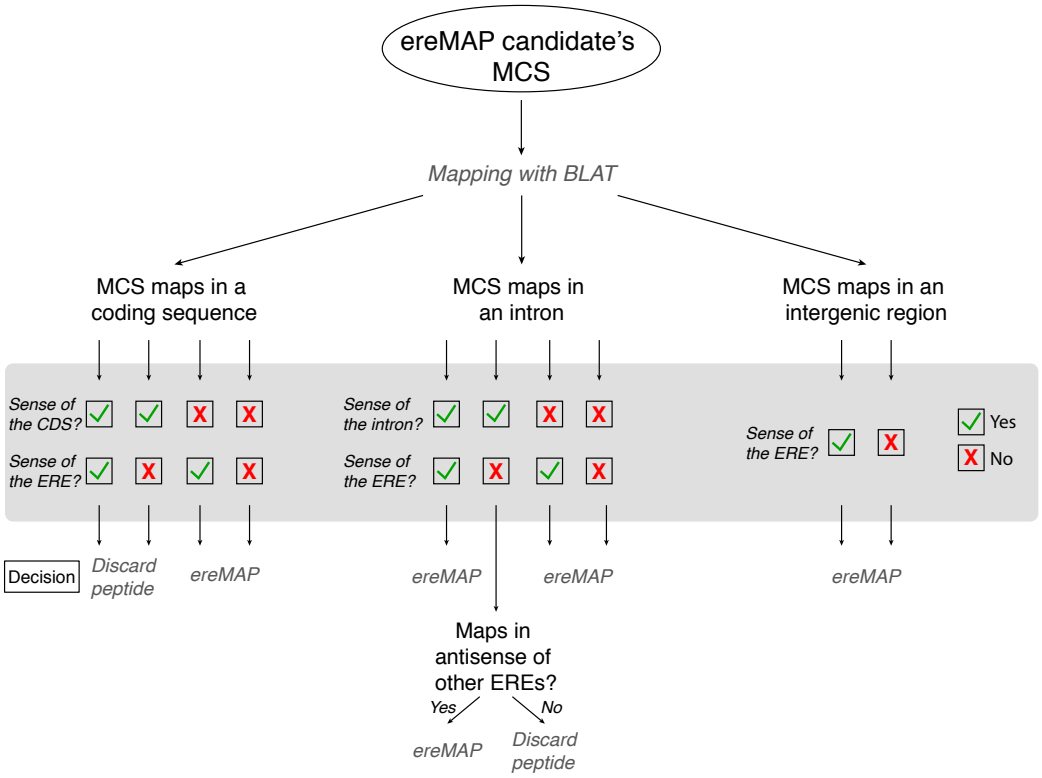


Supplementary Figure 2.1 - Comparison of ERE expression between mTECs and other cell types. (A) Hierarchical clustering of mTECs and multiple hematopoietic cell types based on the expression levels of the 809 ERE families sorted in LINE, LTR and SINE. For each cell type, the mean expression of ERE families was computed among available samples. Row Z-scores were then determined for each ERE family across cell types. (B) Low interindividual variation in ERE families' expression. Boxplot depicting the log-transformed value of the standard deviation of the expression of each ERE family between samples for the 32 healthy human tissues analyzed. Abbreviations: B.S.G.: Band segmented granulocytes; M.P.: Myelomonocytic progenitors; M.S.G.: Minor salivary gland.



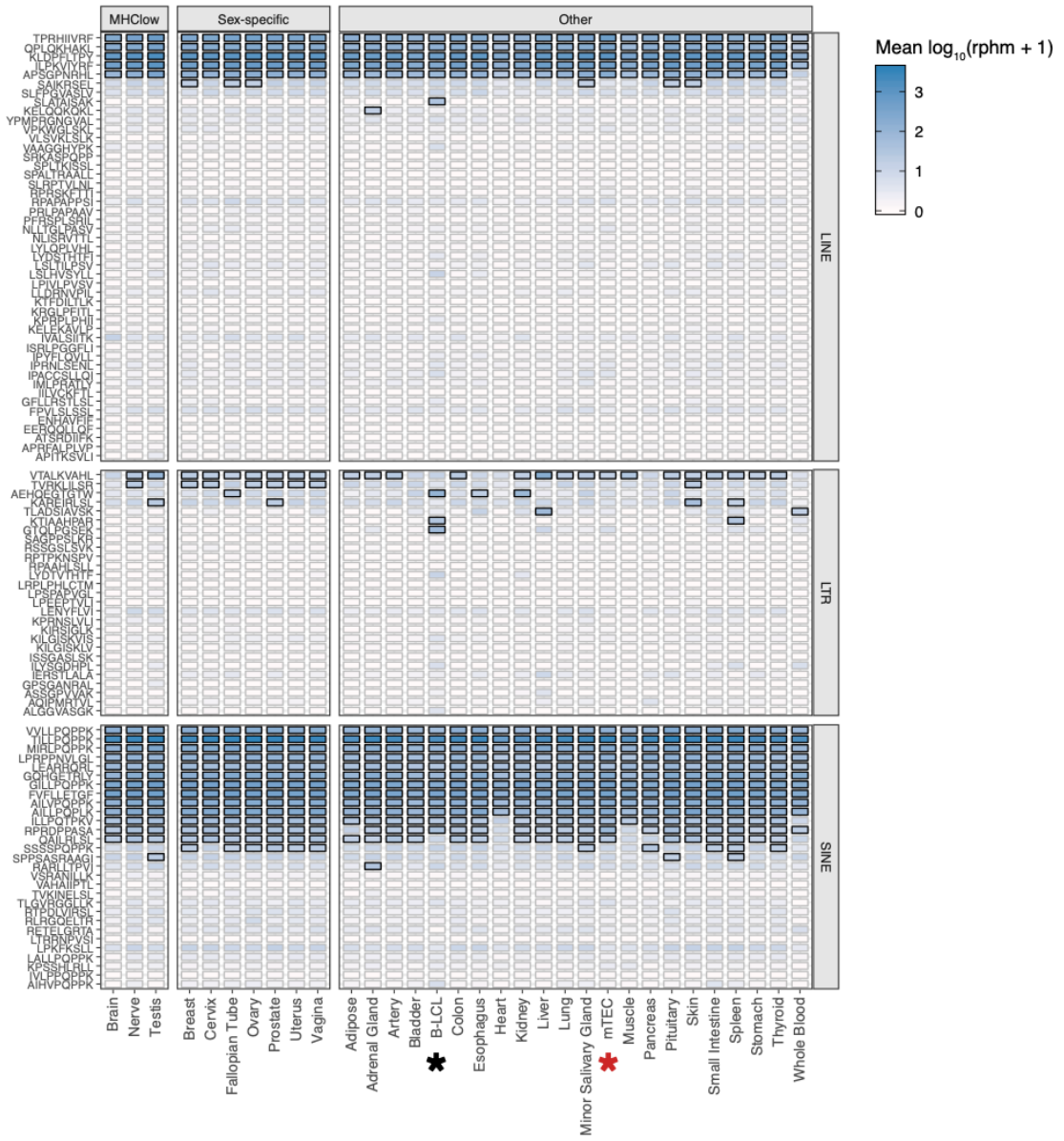
Supplementary Figure 2.2 - Quintile ranking of ERE families in healthy human tissues.

Pie charts represent the percentage of LTR, LINE and SINE families that were assigned to each quintile for the 32 healthy human tissues analyzed. Tissues were sorted based on the number of ERE families that were assigned to the 5th quintile.

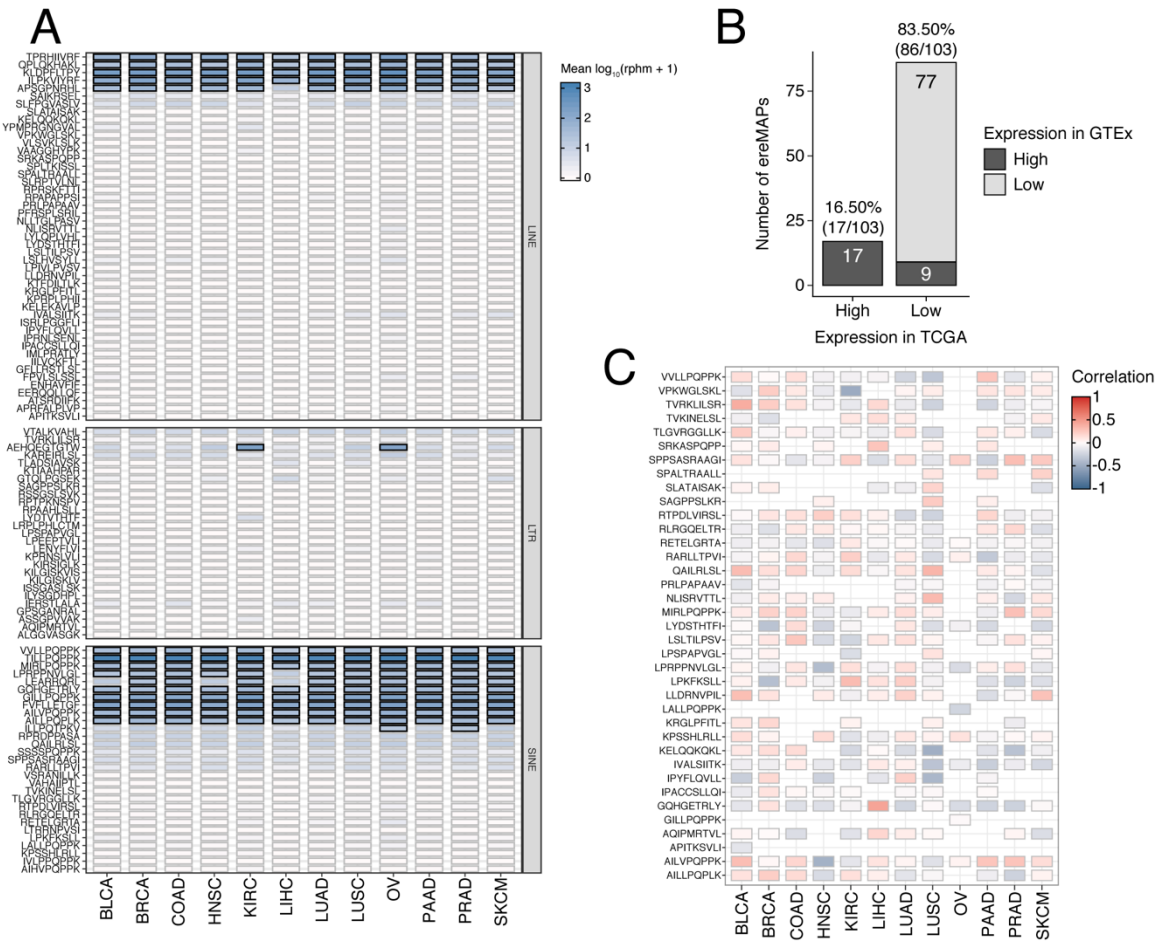


Supplementary Figure 2.3 - Manual validation of ereMAPs' nucleotide coding sequence in the human genome.

Flowchart depicting the decision tree for each ereMAP candidate during manual validation in IGV. After mapping of the peptide's coding sequence on the human genome with BLAT, candidates were considered as ereMAPs or discarded based on the orientation of the peptide's coding sequence towards the ERE sequence and other genomic regions (CDS, introns).



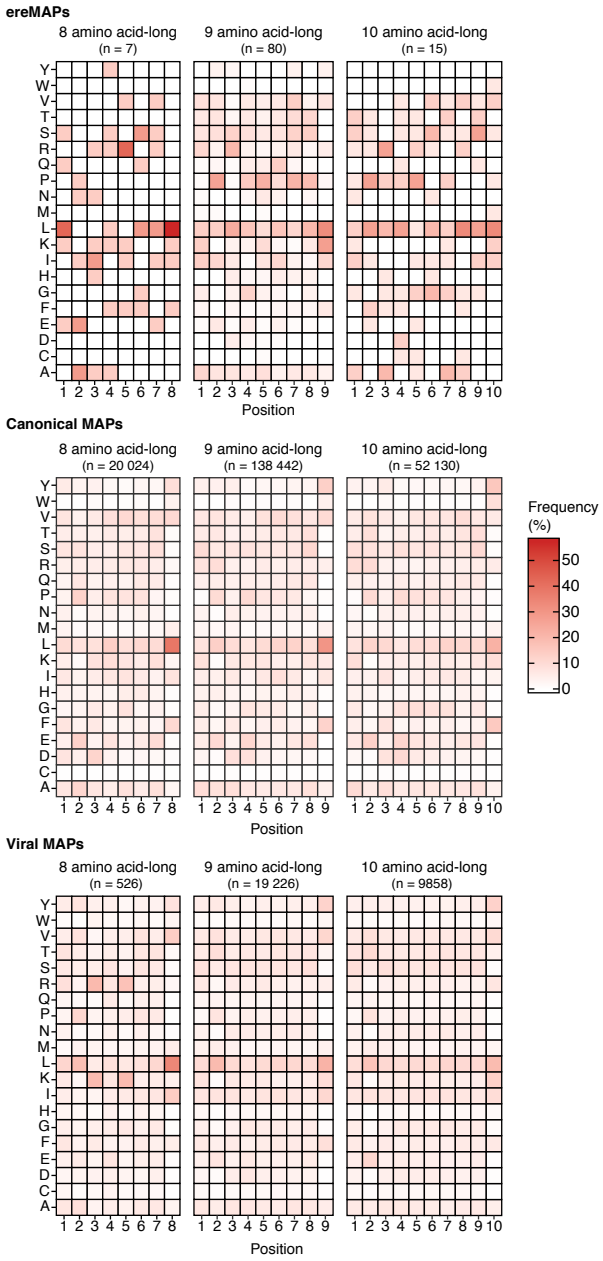
Supplementary Figure 2.4 - Expression of ereMAPs' coding sequence in healthy human tissues. Heatmap showing the average expression, in reads per hundred million reads sequenced (rphm), of ereMAPs' coding sequences in 32 human healthy tissues/cell types (see Table S1). Peptides were sorted based on the group of the ERE sequence generating the peptide (LINE, LTR or SINE). Positive tissues (rphm > 10) are shown with bold squares. B-LCL and mTECs are indicated with black and red stars, respectively.



Supplementary Figure 2.5 - Expression profiling of B-LCL ereMAPs in cancer.

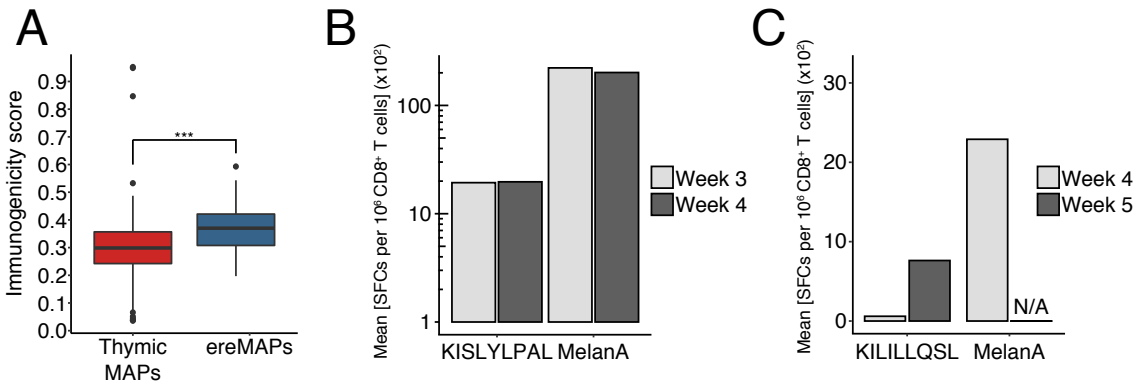
(A) Heatmap showing the average expression, in reads per hundred million reads sequenced (rphm), of B-LCL ereMAPs' coding sequences in 12 cohorts of cancers from TCGA. Peptides were sorted based on the group of the ERE sequence generating the peptide (LINE, LTR or SINE). Positive tissues (rphm > 10) are shown with bold squares. (B) Barplot depicting the number of B-LCL ereMAPs with high (above threshold in ≥2 cancer types) or low expression in TCGA cohorts. Shades of grey show the expression of ereMAPs in healthy tissues from GTEx (high if expression is above threshold in ≥2 tissues, otherwise expression is defined as low). (C) Heatmap showing the Pearson correlation between ereMAPs' RNA expression and DNA methylation level. Abbreviations for TCGA cohorts: BLCA, urothelial bladder carcinoma; BRCA, breast invasive

carcinoma; COAD, colon adenocarcinoma; HNSC, head-neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; LIHC, liver hepatocellular carcinoma; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; OV, ovarian cancer; PAAD, pancreatic adenocarcinoma; PRAD, prostate adenocarcinoma; SKCM, skin cutaneous melanoma.



Supplementary Figure 2.6 - Comparison of amino acid usage of ERE-derived, viral, and human MAPs.

Heatmaps showing amino acid frequencies at all positions of 8, 9 and 10 amino acid-long peptides for ereMAPs (top), canonical human (middle) and viral (bottom) MAPs. Abbreviations for amino acids: Y, Tyrosine; W, Tryptophan; V, Valine; T, Threonine; S, Serine; R, Arginine; Q, Glutamine; P, Proline; N, Asparagine; M, Methionine; L, Leucine; K, Lysine; I, Isoleucine; H, Histidine; G, Glycine; F, Phenylalanine; E, Glutamic Acid; D, Aspartic Acid; C, Cysteine; A, Alanine.



Supplementary Figure 2.7 - Assessment of ERE-derived MAPs' immunogenicity.

(A) Boxplot showing the immunogenicity scores of thymic MAPs and ereMAPs predicted by the Repitope algorithm. Statistical significance was computed with a Mann-Whitney test (***$P \leq 0.001$). (B , C) Barplots showing the numbers of spot-forming cells (SFCs) per 10^6 $CD8^+$ T cells measured by IFNɣ ELISpot assay for two cancer-specific ereMAPs, (B) KISLYLPAL and (C) KILILLQSL, and MelanA as positive control. N/A indicates that the experiment could not be performed due to a limited number of T cells.

Chapitre 3 : Projet 2

3.1 Article #2: Transposable elements regulate thymus development and function.

3.1.1 Résumé en français

Titre en français : Les éléments transposables régulent le développement et la fonction thymiques.

Les éléments transposables (TE) sont des séquences répétitives représentant environ 45% des génomes humain et murin qui sont exprimés fortement par les cellules épithéliales thymiques de la médulla (mTEC). Dans cette étude, nous avons réalisé des analyses multi-omiques des TEs dans des cellules thymiques humaines et murines pour déterminer leur rôle précis dans le développement des lymphocytes T. Nos données montrent que l'expression des TEs est forte dans le thymus humain, et qu'elle varie grandement i) entre les populations thymiques et ii) au cours du développement. Les TEs fournissent des sites de liaison à un grand nombre de facteurs de transcription dans tous les types cellulaires du thymus humain. Deux types cellulaires expriment des répertoires de TEs particulièrement larges : les mTECs et les cellules dendritiques plasmacytoïdes (pDC). Dans les mTECs, les TEs interagissent avec des facteurs de trancription essentiels au développement et à la fonction des mTECs (ex : PAX1 et RELB) et génèrent des peptides associés au CMH-I impliqués dans l'éducation des lymphocytes T. De façon marquée, AIRE, FEZF2 et CHD4 régulent des répertoires non-redondants de TEs dans les mTECs murines. Chez l'humain, les pDCs thymiques ont une expression homogène et forte des TEs qui mène à la formation de dsRNA qui induisent la signalisation de RIG-I et MDA5, ce qui explique la sécrétion

constitutive d'IFN α/β des pDCs thymiques. Cette étude illustre la diversité des interactions entre les TEs et le système immunitaire adaptatif. Les TEs sont des parasites génétiques, et les deux types cellulaires exprimant le plus fortement les TEs (mTECs et pDCs) sont essentiels à l'établissement de la tolérance au soi des lymphocytes T. Nous proposons donc que l'expression des TEs dans les cellules du thymus est critique pour prévenir des réactions auto-immunes chez les vertébrés.

3.1.2 Contribution des auteurs

Jean-David Larouche : Conception du projet. Écriture des scripts, réalisation des analyses bio-informatiques, réalisation des expériences et analyse des résultats pour les figures 3.1 à 3.6 et les figures supplémentaires 3.1 à 3.10. Écriture de la première version du manuscrit.

Céline Laumont : Conception des analyses des figures 3.3a et b. Contribution aux scripts de traitement et de contrôle de la qualité des données de scRNA-seq.

Assya Trofimov : Contribution aux analyses des figures 3.1c et d.

Krystel Vincent : Conception du projet, analyse des résultats et écriture de la première version du manuscrit.

Leslie Hesnard : Contribution à l'isolation des pDCs thymiques (figure 3.4c).

Sylvie Brochu : Isolation des mTECs et cTECs de thymus de souris K5D1 (figure 3.6).

Caroline Côté : Isolation des mTECs et cTECs de thymus de souris K5D1 (figure 3.6).

Juliette Humeau : Contribution à la quantification des images de microscopie confocale (figure 3.4d).

Éric Bonneil : Recherches des bases de données de MS dans Peaks (figure 3.6).

Joël Lanoix : Immunoprécipitation des complexes CMH-I-peptides (figure 3.6).

Chantal Durette : Validation manuelle des spectres MS/MS des ereMAPs (figure 3.6).

Patrick Gendron : Contribution aux analyses de la figure 3.2f.

Jean-Philippe Laverdure : Génération du protéome canonique personnalisé (figure 3.6).

Ellen Richie : Génération des souris K5D1 (figure 3.6).

Sébastien Lemieux : Analyse des résultats. Contribution à la conception de la figure 3.2b.

Pierre Thibault : Analyse des résultats.

Claude Perreault : Conception du projet, analyse des résultats et écriture de la première version du manuscrit.

3.1.3 Version révisée soumise à eLife

"Transposable elements regulate thymus development and function"

Jean-David Larouche[1,2], Céline M. Laumont[3,4], Assya Trofimov[1,5,6,7], Krystel Vincent[1], Leslie Hesnard[1], Sylvie Brochu[1], Caroline Côté[1], Juliette Humeau[1], Éric Bonneil[1], Joël Lanoix[1], Chantal Durette[1], Patrick Gendron[1], Jean-Philippe Laverdure[1], Ellen R. Richie[8], Sébastien Lemieux[1,9], Pierre Thibault[1,10], Claude Perreault[1,2]*.

[1] Institute for Research in Immunology and Cancer, Université de Montréal; Montréal, Canada.

[2] Department of Medicine, Université de Montréal; Montréal, Canada.

[3] Deeley Research Centre, BC Cancer; Victoria, Canada.

[4] Department of Medical Genetics, University of British Columbia; Vancouver, Canada.

[5] Department of Computer Science and Operations Research, Université de Montréal; Montréal, Canada.

[6] Department of Physics, University of Washington; Seattle, USA.

[7] Fred Hutchinson Cancer Center; Seattle, USA.

[8] Department of Epigenetics and Molecular Carcinogenesis, University of Texas M.D. Anderson Cancer Center; Houston, USA.

[9] Department of Biochemistry and Molecular Medicine, Université de Montréal; Montréal, Canada.

[10] Department of Chemistry, Université de Montréal; Montréal, Canada.

*Corresponding author. Email: claude.perreault@umontreal.ca

3.1.3.1 Abstract

Transposable elements (TE) are repetitive sequences representing ~45% of the human and mouse genomes and are highly expressed by medullary thymic epithelial cells (mTEC). In this study, we investigated the role of TEs on T-cell development in the thymus. We performed multi-omic analyses of TEs in human and mouse thymic cells to elucidate their role in T cell development. We report that TE expression in the human thymus is high and shows extensive age- and cell lineage-related variations. TE expression correlates with multiple transcription factors in all cell types of the human thymus. Two cell types express particularly broad TE repertoires: mTECs and plasmacytoid dendritic cells (pDC). In mTECs, transcriptomic data suggest that TEs interact with transcription factors essential for mTEC development and function (e.g., PAX1 and REL), and immunopeptidomic data showed that TEs generate MHC-I-associated peptides implicated in

thymocyte education. Notably, AIRE, FEZF2, and CHD4 regulate small yet non-redundant sets of TEs in murine mTECs. Human thymic pDCs homogenously express large numbers of TEs that likely form dsRNA, which can activate innate immune receptors, potentially explaining why thymic pDCs constitutively secrete IFN α/β. This study highlights the diversity of interactions between TEs and the adaptive immune system. TEs are genetic parasites, and the two thymic cell types most affected by TEs (mTEcs and pDCs) are essential to establishing central T-cell tolerance. Therefore, we propose that orchestrating TE expression in thymic cells is critical to prevent autoimmunity in vertebrates.

3.1.3.2 Introduction

Self/non-self discrimination is a fundamental requirement of life (423). In jawed vertebrates, the thymus is the only site where T lymphocytes can be properly educated to distinguish self from non-self (424, 425). This is vividly illustrated by Oncostatin M-transgenic mice, where T-cell production occurs exclusively in the lymph nodes (426). These mice harbor normal numbers of T-cell receptors (TCR) αβ T cells but present severe autoimmunity and cannot fight infections (427). Intrathymic generation of a functional T-cell repertoire depends on choreographed interactions between the TCRs of thymocytes and peptides presented by major histocompatibility complex (MHC) molecules on various antigen-presenting cells (APC) (428). Positive selection depends on self-antigens presented by cortical thymic epithelial cells (cTEC) and ensures that TCRs recognize antigens in the context of the host's MHC molecules (429, 430). The establishment of central tolerance depends on two main classes of APCs located in the thymic medulla: dendritic cells (DC) and medullary TEC (mTEC) (431-433). Two other APC types have a more limited contribution to central tolerance: thymic fibroblasts and B cells (94, 434). High avidity interactions between

thymic APCs and autoreactive thymocytes lead to thymocyte deletion (negative selection) or generation of regulatory T cells (Treg) (435).

The main drivers of central tolerance, mTECs and DCs, display considerable phenotypic and functional heterogeneity. Indeed, recent single-cell RNA-seq (scRNA-seq) studies have identified several subpopulations of mTECs: immature mTEC(I) that stimulate thymocyte migration to the medulla via chemokine secretion (436), mTEC(II) that express high levels of MHC and are essential to tolerance induction, fully differentiated corneocyte-like mTEC(III) that foster a pro-inflammatory microenvironment (437), and finally mimetic mTECs that express peripheral tissue antigens (100). Three different proteins whose loss of function leads to severe autoimmunity, AIRE, FEZF2, and CHD4, have been shown to drive the expression of non-redundant sets of peripheral tissue antigens in mTECs (129, 130, 438). DCs, on the other hand, are separated into three main populations. Conventional DC 1 and 2 (cDC1 and cDC2) have an unmatched ability to present both endogenous antigens and exogenous antigens acquired via cross-presentation or cross-dressing (439). Plasmacytoid DC (pDC) are less effective APCs than cDCs, their primary role being to produce interferon alpha (IFNα) (439). Notably, thymic pDCs originate from intrathymic IRF8^{hi} precursors, and, in contrast to extrathymic pDCs, they constitutively secrete high amounts of IFNα (105, 106, 113). This constitutive IFNα secretion by thymic pDCs regulates the late stages of thymocyte development by promoting the generation of Tregs and innate CD8 T cells (114-118).

Transposable elements (TE) are repetitive sequences representing ~45% of the human and mouse genomes (374, 377). Most TEs can be grouped into three categories: the long and short interspersed

nuclear elements (LINE and SINE, respectively) and the long terminal repeats (LTR). These broad categories are subdivided into over 800 subfamilies based on sequence homology (147). TE expression is typically repressed in host cells to prevent deleterious integrations of TE sequences in protein-coding genes (375). Unexpectedly, TEs were recently found to be expressed at higher levels in human mTECs than in any other MHC-expressing tissues and organs (i.e., excluding the testis) (440, 441), suggesting a role for TEs in thymopoiesis. Since some TEs are translated and generate MHC I-associated peptides (MAP) (440), they might induce TE-specific central tolerance (442). Additionally, TEs provide binding sites to transcription factors (TF) and stimulate cytokine secretion via the formation of double-stranded RNA (dsRNA) (338, 340, 341, 350, 443). Hence, TEs could have pleiotropic effects on thymopoiesis. To evaluate the role of TEs in thymopoiesis, we adopted a multipronged strategy beginning with scRNA-seq of human thymi and culminating in MS analyses of the MAP repertoire of mouse mTECs.

3.1.3.3 Results

LINE, LTR, and SINE expression shows extensive variations during ontogeny of the human thymus

We first profiled TE expression in various thymic cell populations during development. To do so, we quantified the expression of 809 TE subfamilies (classified according to the RepeatMasker annotations) in the scRNA-seq dataset of human thymi created by *Park et al.* (102). Cells were clustered in 19 populations representing the main constituents of the thymic hematolymphoid and stromal compartments (Figure 3.1a and Supplementary Figure 3.1). The expression of TE subfamilies was quantified at all developmental stages available, ranging from 7 post-conception weeks (pcw) to 40 years of age (Supplementary file 1 – Table 1). Unsupervised hierarchical clustering revealed three clusters of TE subfamilies based on their pattern of expression during

thymic development (Figure 3.1b, upper panel): i) maximal expression at early embryonic stages persisting, albeit at lower levels, throughout ontogeny (cluster 1), ii) an expression specific to a given timepoint (cluster 2), or iii) a high expression at early embryonic stages that decreases rapidly at later timepoints (cluster 3). LINE and SINE subfamilies were enriched in cluster 1, whereas LTR subfamilies were significantly enriched in clusters 2 and 3 (Figure 3.1b, lower panel). Expression of individual LINE and SINE subfamilies was highly shared among different cell types (Figure 3.1d). In contrast, the LTR subfamilies' expression pattern was shared by fewer cell subsets and adopted a quasi-random distribution (Figure 3.1d). The pattern of expression assigned to TE subfamilies (Figure 3.1c, innermost track) was not affected by the proportion of cells of different developmental stages (embryonic or postnatal) (Figure 3.1c, outermost track, and Supplementary Figure 3.2). This suggests that our observations do not result from a bias in the composition of the dataset. To gain further insights into the expression of TE subfamilies, we studied two biological processes known to regulate TE expression in other contexts: cell proliferation and expression of KRAB zinc-finger proteins (KZFP) (271, 444). Cell cycling scores negatively correlated with TE expression in various thymic cell subsets, particularly for LINE and SINE subfamilies shared among cell types (Supplementary Figure 3.3 and Supplementary file 1 – Table 2), whereas analysis of KZFP expression identified ZNF10 as a probable repressor of L1 subfamilies in Th17 and NK cells (Supplementary Figure 3.4 and Supplementary file 1 – Table 3). Thus, we conclude that the expression of the three main classes of TEs shows major divergences as a function of age and thymic cell types.

TEs form interactions with transcription factors regulating thymic development and function

TEs provide binding sites to TFs (211, 298, 338), and T-cell development is driven by the coordinated timing of multiple changes in transcriptional regulators (445). We, therefore, investigated interactions between TE subfamilies and TFs during the development of the human thymus. Two criteria defined an interaction: i) a significant and positive correlation between the expression of a TF and a TE subfamily in a given cell population, and ii) the presence of the TF binding motif in the loci of the TE subfamily (Figure 3.2a). Additionally, we validated the correlations we obtained using a bootstrap procedure to ascertain their reproducibility (see *Material and Methods* for details). This procedure removed weakly correlated TF-TE pairs (Figure 3.2b). TF-TE interactions were observed in all thymic cell populations (Figure 3.2c, d, Supplementary Figure 3.5, and Supplementary file 1 – Table 4). Numerous TF-TE interactions were conserved between hematolymphoid and stromal cell subsets (Figure 3.2e). However, the number of interactions and the complexity of the interaction networks were much higher in mTECs than in other cell populations (Figure 3.2c, d, Supplementary Figure 3.5, and Supplementary Figure 3.6).

Several TFs instrumental in thymus development and thymopoiesis interacted with TE subfamilies (Supplementary Figure 3.6, and Supplementary file 1 – Table 4). These TFs include the *NFKB1* and *REL* subunits of the NF-κB complex and *PAX1* in mTECs (446-448) and *JUND* in thymocytes (449). In DCs, the most notable TF-TE interactions involved interferon regulatory factors (IRF), which regulate the late stages of T-cell maturation, and *TCF4,* which is essential for pDC development (114, 450). This observation is consistent with evidence that TEs have shaped the evolution of IFN signaling networks (338). Finally, we found significant interactions between *CTCF* and TE subfamilies in mTECs and endothelial cells, suggesting that the binding of *CTCF* to TE sequences affects the tridimensional structure of the chromatin in the thymic stroma (209).

Interestingly, LINE and SINE subfamilies that occupy more genomic space interacted with higher numbers of transcription factors (Supplementary Figure 3.7).

Using data from the ENCODE consortium for hematopoietic cells (451, 452), we looked at the histone marks at the TE loci identified as TF interactors by our analyses (i.e., correlated with TF expression and containing the TF binding motif). The objective was to determine if they could act as promoters or enhancers (Figure 3.2a and Supplementary file 1 – Table 5). We found several TE promoter and enhancer candidates in all eight hematopoietic cell types analyzed, with a striking overrepresentation of LINE and SINE compared to LTR sequences (Figure 3.2f and Supplementary file 1 – Table 6). Finally, we analyzed publicly available ChIP-seq data of ETS1, an important TF for NK cell development (453), to confirm its ability to bind TE sequences. Indeed, 19% of ETS1 peaks overlap with TE sequences (Figure 3.2g). Notably, ETS1 peaks overlapped with TE sequences (Figure 3.2h, in red) in the promoter regions of PRF1 and KLRD1, two genes critical for NK cells' effector functions (454, 455). Hence, our data suggest that TEs affect thymic development and function by providing binding sites to multiple TFs.

TEs are highly and differentially expressed in human thymic APC subsets

We next sought to determine whether the high expression of TEs reported in mTECs (32, 33) was limited to this cell subset or was found in other thymic cell types. Since several thymic stromal cells reach maturity after birth (101), we selected postnatal samples for the following analyses. We computed two distinct Shannon entropy indices: one for the global diversity of TEs expressed by all cells of a given population, and another for the median value of TE diversity expressed by

individual cells of a population (Figure 3.3a). Then, we computed a linear model to represent the diversity of TEs expressed by a cell population based on the diversity of TEs expressed by individual cells (Figure 3.3a, blue curve). Two salient findings emerged from this analysis. First, the diversity of TEs expressed in the T-cell lineage decreases during differentiation according to the following hierarchy: DN thymocytes > DP thymocytes > SP thymocytes (Figure 3.3a, Supplementary Figure 3.8). Second, among the populations of thymic APCs implicated in positive and negative selection (Figure 3.3a, orange dots), cTECs, mTECs, and DCs expressed broader repertoires of TEs than B cells and fibroblasts. While cTECs and DCs expressed highly diverse TE repertoires at both the population and individual cell levels, the breadth of TE expression in mTECs was found only at the population level (Figure 3.3a). Accordingly, intercellular heterogeneity (i.e., deviation from the linear model) was higher for mTECs than other cell populations (Figure 3.3b).

We next focused on thymic APCs expressing the broadest TE repertoires: cTECs, mTECs, and DCs (Figure 3.3a). To this end, we annotated these APC subpopulations based on previously published lists of marker genes (Figure 3.3c and Supplementary Figure 3.9) (102, 103). We performed differential expression analyses to determine whether some TE subfamilies were overexpressed in specific APC subsets. pDCs and mTEC(II) overexpressed a broader TE repertoire than other APCs: 32.01% of subfamilies were overexpressed in pDCs and 10.88% in mTEC(II) (Figure 3.3d and Supplementary file 1 – Table 7). The nature of the overexpressed TEs differed between pDCs and other thymic APC subsets. Indeed, pDCs overexpressed LTRs, LINEs, and SINEs, including several Alu and L1 subfamilies (Figure 3.3d and Supplementary file 1 – Table 7). In contrast, other thymic APCs predominantly overexpressed LTRs.

TE expression showed wildly divergent levels of intercellular heterogeneity in APC subsets. Indeed, whereas most TE subfamilies were expressed by <25% of cells of the mTEC(II) population, an important proportion of TEs were expressed by >75% of pDCs (Figure 3.3e). To evaluate this question further, we compared TE expression between metacells of thymic APCs; metacells are small clusters of cells with highly similar transcription profiles. This analysis revealed that overexpression of TE subfamilies was shared between pDC metacells but not mTEC(II) metacells, reinforcing the idea that TE expression adopts a mosaic pattern in the mTEC(II) population (Supplementary Figure 3.10). We conclude that cTECs, mTECs, and DCs express broad TE repertoires. However, two subpopulations of thymic APCs clearly stand out. pDCs express an extremely diversified repertoire of LTRs, SINEs, and LINEs, showing limited intercellular heterogeneity, whereas the mTEC(II) population shows a highly heterogeneous overexpression of LTR subfamilies.

TE expression in human pDCs is associated with dsRNA structures

The high expression of a broad repertoire of TE sequences in thymic pDCs was unexpected (Figure 3.3d). LINE and SINE subfamilies, in particular, were highly and homogeneously expressed by thymic pDCs (Figure 3.4a). Constitutive IFNα secretion is a feature of thymic pDCs not found in extrathymic pDCs. We, therefore, hypothesized that this constitutive IFNα secretion by thymic pDCs might be mechanistically linked to their TE expression profile. We first assessed whether thymic and extrathymic pDCs have similar TE expression profiles by reanalyzing scRNA-seq data from human spleens published by *Madissoon et al.* (456) (Supplementary Figure 3.11a, b). This revealed that extrathymic pDCs express TE sequences at similar or lower levels than other splenic cells (Supplementary Figure 3.11c, d). We then used pseudobulk RNA-seq methods to perform a

differential expression analysis of TE subfamilies between thymic and splenic pDCs. This analysis confirmed that TE expression was globally higher in thymic than in extrathymic pDCs (Figure 3.4b). Since TE overexpression can lead to the formation of dsRNA (350, 443), we investigated if such structures were found in thymic pDCs. pDCs were magnetically enriched from primary human thymi following labeling with anti-CD303 antibody (a marker of pDCs). Then, pDC-enriched thymic cells were stained with an antibody against CD123 (another marker of pDCs) and the J2 antibody that stains dsRNA. The intensity of the J2 signal was more than 10-fold higher in $CD123^+$ relative to $CD123^-$ cells (Figure 3.4c, d). We conclude that thymic pDCs contain large amounts of dsRNAs. To evaluate if these dsRNAs arise from TE sequences, we analyzed in thymic APC subsets the proportion of the transcriptome assigned to two groups of genomic sequences known as important sources of dsRNAs: TEs and mitochondrial genes (457). Strikingly, whereas the percentage of reads from mitochondrial genes was typically lower in pDCs than in other thymic APCs, the proportion of the transcriptome originating from TEs was higher in pDCs (~22%) by several orders of magnitude (Supplementary Figure 3.12). Finally, we performed gene set enrichment analyses to ascertain if the high expression of TEs by thymic pDCs was associated with specific gene signatures. These analyses highlighted signatures of antigen presentation, immune response, and interferon signaling in thymic pDCs (Figure 3.4e and Supplementary file 1 – Table 8). Notably, thymic pDCs harbored moderate yet significant enrichment of gene signatures of RIG-I and MDA5-mediated IFN α/β signaling compared to all other thymic APCs (Figure 3.4e and Supplementary file 1 – Table 8). Altogether, these data support a model in which the high and ubiquitous expression of TEs in thymic pDCs would lead to the formation of dsRNAs triggering innate immune sensors, which might explain their constitutive secretion of IFN α/β.

AIRE, CHD4, and FEZF2 regulate distinct sets of TE sequences in murine mTECs

The essential role of mTECs in central tolerance hinges on their ability to ectopically express tissue-restricted genes, whose expression is otherwise limited to specific epithelial lineage (119, 140). This promiscuous gene expression is driven by AIRE, CHD4, and FEZF2 (129, 130, 438). We, therefore, investigated the contribution of these three genes to the expression of TE subfamilies in the mTEC(II) population (Figure 3.3d). First, we validated that mTEC(II) express *AIRE, CHD4,* and *FEZF2* in the human scRNA-seq dataset (Figure 3.5a). Next, we analyzed published murine mTEC RNA-seq data to assess the regulation of TE sequences by AIRE, CHD4, and FEZF2. Differential expression analyses between knock-out (KO) and wild-type (WT) mice showed that these three factors regulate TE sequences, but the magnitude and directionality of this regulation differed (Figure 3.5b and Supplementary file 1 – Table 9). Indeed, while CHD4 had the biggest impact on TE expression by inducing 433 TE loci and repressing 463, FEZF2's impact was minimal, with 97 TE loci induced and 60 repressed (Figure 3.5b). Besides, AIRE mainly acted as a repressor of TE sequences, with 326 loci repressed and 171 induced (Figure 3.5b). Interestingly, there was minimal overlap between the TE sequences regulated by AIRE, CHD4, and FEZF2, indicating that they have non-redundant roles in TE regulation (Figure 3.5c). Additionally, AIRE, CHD4, and FEZF2 preferentially targeted LTR and LINE elements, with significant enrichment of specific subfamilies such as MTA_Mm-int and RLTR4_Mm that are induced by Aire and Fezf2, respectively (Figure 3.5d and Supplementary Figure 3.13a). While AIRE and CHD4 preferentially targeted evolutionary young TE sequences, the age of the TE sequence did not seem to affect the regulation by FEZF2 (Supplementary Figure 3.13b). We also noticed that the distance between regulated TE loci was smaller than the distributions of randomly selected TEs (Supplementary Figure 3.13c). This

suggests that AIRE, CHD4, and FEZF2 nonrandomly affect the expression of TE sequences located in specific genomic regions. We observed no significant differences in the genomic localization of TE loci targeted by AIRE, CHD4, and FEZF2 relative to the genomic localization of all TE sequences in the murine genome: most TE loci were located in intronic and intergenic regions (Supplementary Figure 3.13d). Enrichment for intronic TEs could not be ascribed to induction of global intron retention: the intron retention ratio was similar for TEs regulated or not by AIRE, CHD4, and FEZF2 (Supplementary Figure 3.13e). ChIP-seq-based analysis of permissive histone marks showed that TE loci induced by AIRE, CHD4, and FEZF2 were all marked by H3K4me3 (Figure 3.5e). As a proof of concept, we validated that 31.42% of AIRE peaks overlap with TE sequences by reanalyzing ChIP-seq data, confirming AIRE's potential to bind TE sequences (Figure 3.5f). Hence, AIRE, CHD4, and FEZF2 regulate the expression of small yet non-redundant repertoires of TE sequences associated with permissive histone marks.

TEs are translated and presented by MHC class I molecules in murine TECs

Several TEs are translated and generate MAPs (440). Hence, the expression of TEs in cTECs and even more in mTECs raises a fundamental question: do these TEs generate MAPs that would shape the T cell repertoire? Mass spectrometry (MS) is the only method that can faithfully identify MAPs (458-460). Despite its quintessential role in central tolerance, the MAP repertoire of mTECs has never been studied by MS because of the impossibility of obtaining sufficient mTECs for MS analyses: mTECs represent ≤ 1% of thymic cells, and they do not proliferate *in vitro*. To get enough cTECs and mTECs for MS analyses, we used transgenic mice that express cyclin D1 under the control of the keratin 5 promoter (K5D1 mice). These mice develop dramatic thymic hyperplasia,

but their thymus is morphologically and functionally normal (461-463). Primary cTECs and mTECs (2 replicates of 70 x 10^6 cells from 121 and 90 mice, respectively) were isolated from the thymi of K5D1 mice as described (464). Following cell lysis and MHC I immunoprecipitation, MAPs were analyzed by liquid chromatography MS/MS (Figure 3.6a). To identify TE-coded MAPs, we generated a TE proteome by in silico translation of TE transcripts expressed by mTECs or cTECs, and this TE proteome was concatenated with the canonical proteome. MS analyses enabled the identification of a total of 1636 and 1714 MAPs in mTECs and cTECs, respectively. From these, we identified 4 TE-derived MAPs in mTECs and 2 in cTECs, demonstrating that TEs can be translated and presented by MHC I in the thymic cortex and medulla (Figure 3.6b and Supplementary file 1 – Table 10). These MAPs were coded by the three major groups of TE: LINEs (n=1), LTRs (n=1), and SINEs (n=4). Next, we evaluated whether the low number of TE MAPs identified could result from mass spectrometry detection limits (465, 466). We measured the level and frequency of TE expression in two subsets of cTECs (Figure 3.6c, left) or mTECs (Figure 3.6c, right) using scRNA-seq data from *Baran-Gale et al.* (467). TE subfamilies generating MAPs in cTECs or mTECs are highlighted in red in their respective plots. Strikingly, TECs highly and ubiquitously expressed the MAP-generating TE subfamilies. These results suggest that the contribution of TEs to the MAP repertoire of cTECs and mTECs might be significantly underestimated by the limits of detection of MS. This is particularly true for mTECs because they express high levels of TEs (Figure 3.3d), but their TE profile displays considerable intercellular heterogeneity (Figure 3.3e and Supplementary Figure 3.10). Nonetheless, our data provide direct evidence that TEs can generate MAPs presented by cTECs and mTECs, which can contribute to thymocyte education.

3.1.3.4 Discussion

TEs are germline-integrated parasitic DNA elements that comprise about half of mammalian genomes. Over evolutionary timescales, TE sequences have been co-opted for host regulatory functions. Mechanistically, TEs encode proteins and noncoding RNAs that regulate gene expression at multiple levels (147, 468). Regulation of IFN signaling and triggering innate sensors are the best-characterized roles of TEs in the mammalian immune system (442). TEs are immunogenic and can elicit adaptive immune responses implicated in autoimmune diseases (440, 442, 469, 470). Pervasive TE expression in various somatic organs means that co-evolution with their host must depend on establishing immune tolerance, a concept supported by the highly diversified TE repertoire expressed in mTECs (440). This observation provided the impetus to perform multi-omic studies of TE expression in the thymus. At the whole organ level, we found that TE expression showed extensive age- and cell lineage-related variations and was negatively correlated with cell proliferation and expression of KZFPs. The negative correlation between TE expression and cell cycle scores in the thymus is coherent with recent data showing that transcriptional activity of L1s is increased in senescent cells (413). A potential rationale for this could be to prevent deleterious transposition events during DNA replication and cell division. On the other hand, the contribution of KZFPs to TE regulation in the thymus is likely underestimated due to their typically low expression (471) and scRNA-seq detection limit. Additionally, TEs interact with multiple TFs in all thymic cell subsets. This is particularly true for the LINE and SINE subfamilies that occupy larger genomic spaces. Notably, TEs appear to play particularly important roles in two cell types located in the thymic medulla: mTECs and pDCs.

As mTECs are the APC population crucial to central tolerance induction, their high and diverse TE expression is poised to impact the T cell repertoire's formation profoundly. The extent and complexity of TF-TE interactions were higher in mTECs than in all other thymic cell subsets. These interactions included *PAX1* and subunits of the NF-κB complex (e.g., *RELB*). PAX1 is essential for the development of TEC progenitors (448), and *RELB* is for the development and differentiation of mTECs (472). RelB-deficient mice have reduced thymic cellularity, markedly fewer mTECs, lack *Aire* expression, and suffer from autoimmunity (447, 473). Under the influence of *Aire, Fezf2,* and *Chd4,* mTECs collectively express almost the entire exome (119, 140). However, the expression of all genes in each mTEC would cause proteotoxic stress (140). Hence, promiscuous expression of tissue-restricted genes in mTECs adopts a mosaic pattern: individual tissue-restricted genes are expressed in a small fraction of mTECs (21, 100). The present work shows that mTECs also express an extensive repertoire of TEs in a mosaic pattern (i.e., with considerable intercellular heterogeneity). *Aire*, *Fezf2*, and *Chd4* regulate non-redundant sets of TEs and preferentially induce TE sequences associated with permissive histone marks. The immunopeptidome of thymic stromal cells is responsible for thymocyte education and represents one of the most fundamental "known unknowns" in immunology. Inferences on the immunopeptidome of thymic stromal cells are based on transcriptomic data. However, i) TCRs interact with MAPs, not transcripts, and ii) the MAP repertoire cannot be inferred from the transcriptome (47, 458, 474). Using K5D1 mice presenting prominent thymic hyperplasia, we conducted MS searches of TE MAPs, identifying 4 TE MAPs in mTECs and 2 in cTECs. These results demonstrate that cTECs and mTECs present TE MAPs and suggest they present different TE MAPs. However, the correlation between transcriptomic and immunopeptidomic data suggests that TECs can present many more TE MAPs. Their profiling will require MS analyses of enormous numbers of TECs or the development of more sensitive MS techniques. As TE MAPs have been detected in normal and neoplastic extrathymic cells (352, 365,

367, 440), the presentation of TEs by mTECs is likely essential to central tolerance. In line with vibrant plaidoyers for a collaborative Human Immunopeptidome Project (459, 475), our work suggests that immunopeptidomic studies should not be limited to protein-coding genes (2% of the genome) but also encompass non-coding sequences such as TEs.

The second population of cells exhibiting high TE expression, pDCs, are mainly seen as producers of IFN α/β and potentially as APCs (439). Thymic and extrathymic pDCs are ontogenically and functionally different. They develop independently from each other from different precursor cells (105, 106, 476). IFN α/β secretion is inducible in extrathymic pDCs but constitutive in thymic pDCs (113, 439). In line with the location of pDCs in the thymic medulla, their constitutive IFN α/β secretion is instrumental in the terminal differentiation of thymocytes and the generation of Tregs and innate CD8 T cells (114-118). We report here that high TE expression is also a feature of thymic, but not extrathymic, pDCs. Thus, the present study provides a rationale for the constitutive IFN α/β secretion by thymic pDCs: they homogeneously express large numbers of TEs (in particular LINEs and SINEs), leading to the formation of dsRNAs that trigger RIG-I and MDA5 signaling that causes the constitutive secretion of IFN α/β. As such, our data suggest that recognition of TE-derived dsRNAs by innate immune receptors promotes a pro-inflammatory environment favorable to the establishment of central tolerance in the thymic medulla.

At first sight, the pleiotropic effects of TEs on thymic function may look surprising. It should be reminded that the integration of genetic parasites such as TEs is a source of genetic conflicts with the host. Notably, the emergence of adaptive immunity gave rise to higher-order conflicts between TEs and their vertebrate hosts (442, 477). The crucial challenge for the immune system is

developing immune tolerance towards TEs to prevent autoimmune diseases that affect up to 10% of humans (478) without allowing selfish retrotransposition events that hinder genome integrity. The resolution of these conflicts has been proposed to be a determining factor in shaping the function of the immune system (477). Our data suggest that the thymus is the central battlefield for conflict resolution between TEs and T cells in vertebrates. Consistent with the implication of TEs in autoimmunity, more than 90% of putative causal variants associated with autoimmune diseases are in allegedly noncoding regions of the genome (478). In this context, our study illustrates the complexity of interactions between TEs and the vertebrate immune system and should provide impetus to explore them further in health and disease. We see two limitations to our study. First, as with all multi-omic systems immunology studies, our work provides a roadmap for many future mechanistic studies that could not be realized at this stage. Second, our immunopeptidomic analyses of TECs prove that TECs present TE MAPs but certainly underestimate the diversity of TE MAPs presented by cTECs and mTECs.

3.1.3.5 Methods

Experimental design

This study aimed to understand better the impacts of TE expression on thymus development and function. Thymic populations are complex and heterogeneous, so we opted for single-cell RNA-seq data to draw a comprehensive profile of TE expression in the thymus. To better understand the impact of AIRE, FEZF2, and CHD4 on TE expression in the mTEC(II) population, RNA-seq data from WT and KO murine mTEC, as well as ChIP-seq for different histone marks in murine mTECs, were reanalyzed to characterize the TE sequences regulated by these three proteins. Unless stated otherwise, studies were done in human cells. For MS analyses, two replicates of 70 million cells from K5D1 mice (461) were injected for both cTECs and mTECs. All experiments were in

accordance with the Canadian Council on Animal Care guidelines and approved by the *Comité de Déontologie de l'Expérimentation sur des Animaux* of Université de Montréal. Primary human thymi were obtained from 4-month-old to 12-year-old children undergoing cardiovascular surgeries at the CHU Sainte-Justine. This project was approved by the CHU Sainte-Justine Research Ethics Board (protocol and biobank #2126).

Transcriptomic data processing

Preprocessing of the scRNA-seq data was performed with kallisto (385), which uses an expectation-maximization algorithm to reassign multimapping reads based on the frequency of unique mappers at each sequence and bustools workflow. For human thymic data from *Park et al.* (102) and splenic data from *Madissoon et al.* (456), two different indexes were built for the pseudoalignment of reads with kallisto (version 0.46.0): one containing Ensembl 88 (GRCh38.88) transcripts used for the annotation of cell populations, and a second containing Ensembl 88 transcripts and human TE sequences (LINE, LTR, SINE) from RepeatMasker (479) which was used for all subsequent analyses of TE expression. For murine data from *Baran-Gale et al.* (467), cell-type annotations from the original publication were used, and an index containing mm10 transcripts and murine TE sequences from RepeatMasker was used to analyze TE expression. The cell barcodes were corrected, and the feature-barcode matrices were generated with the correct count functions of bustools (version 0.39.3) (480). For murine bulk RNA-seq data, an index composed of mm10 (GRCm38) transcripts and murine TE sequences from RepeatMasker was used for quantification with kallisto.

ChIP-seq data reanalysis

ChIP-seq data for i) ETS1 in human NK cells, ii) AIRE in murine mTECs, and iii) several histone marks of mTECs from WT mice were reanalyzed (see "**Availability of data and materials**" for

the complete list). ETS1 ChIP-seq reads were aligned to the reference *Homo sapiens* genome (GRCh38) using bowtie2 (version 2.3.5) (481) with the --very-sensitive parameter. Multimapping reads were removed using the samtools view function with the -q 10 parameter, and duplicate reads were removed using the samtools markdup function with the -r parameter (482). Peak calling was performed with macs2 with the -m 5 50 parameter (483). Peaks overlapping with the ENCODE blacklist regions (484) were removed with bedtools intersect (390) with default parameters. Overlap of ETS1 peaks with TE sequences was determined using bedtools intersect with default parameters. BigWig files were generated using the bamCoverage function of deeptools2 (485), and genomic tracks were visualized in the USCS Genome Browser (486). For the murine histone marks and AIRE data, reads were aligned to the reference *Mus musculus* genome (mm10) using bowtie2 with the --very-sensitive parameter. Multimapping reads were removed using the samtools view function with the -q 10 parameter, and duplicate reads were removed using the samtools markdup function with the -r parameter. For histone marks, read coverage at the sequence body and flanking regions (+/- 3000 base pairs) of TE loci induced by AIRE, FEZF2, and CHD4 was visualized using ngs.plot.r (version 2.63) (487). For AIRE, peaks overlapping with the ENCODE blacklist regions were removed with bedtools intersect, and overlap of peaks with TE sequences was determined using bedtools intersect with default parameters.

Cell population annotation

Feature-barcode matrices were imported in R with SingleCellExperiment (version 1.12.0) (488). As a quality control, cells with less than 2000 UMI detected, less than 500 genes detected, or more than 5% reads assigned to mitochondrial genes were considered low quality and removed from the dataset with scuttle (version 1.0.4) (489). Cells with more than 7000 genes detected were considered doublets and removed. Normalization of cell size factors was performed with scran

(version 1.18.7) (490), and log-normalization of read counts was done with scuttle with default parameters. Variable regions of TCR and IG genes, as well as ribosomal and cell cycle genes (based on *Park et al*. (102)), were removed, and highly variable features were selected based on a mean-variance trend based on a Poisson distribution of noise with scran. Adjustment of sequencing depths between batches and mutual nearest neighbors (MNN) correction were computed with batchelor (version 1.6.3) (491). Cell clustering was performed with scran using the Jaccard index for edge weighting and the Louvain method for community detection. Lists of marker genes for human thymic cell populations and TEC subsets were taken from *Park et al.* (102) and *Bautista et al.* (103), whereas marker genes of splenic populations were based on *Madissoon et al.* (456).

TE expression throughout thymic development

The expression of TE subfamilies was obtained by summing the read counts of loci based on the RepeatMasker annotations. For each TE subfamily in each cell population, expression levels amongst developmental stages were normalized by dividing them with the maximal expression value. Next, the Euclidean distance between each TE subfamily in each cell population (based on their normalized expression across developmental stages) was computed, followed by unsupervised hierarchical clustering. The tree was then manually cut into three clusters, and enrichment of LINE, LTR, and SINE elements in these three clusters was determined using Fisher's exact tests. The cluster assigned to each TE subfamily in each cell population was visualized in a circos plot using the circlize package (version 0.4.14) (492) in R, and the percentage of each cell population found in embryonic or postnatal samples. Finally, we computed the frequency that each TE family was assigned to the three clusters, and the maximal value was kept. As a control, a random distribution of the expression of 809 TE subfamilies in 18 cell populations was generated. A cluster (cluster 1, 2, or 3) was randomly attributed for each combination of TE subfamily and

cell type, and the maximal occurrence of a given cluster across cell types was then computed for each TE subfamily. Finally, the LINE, LTR, and SINE elements distributions were compared to the random distribution with Kolmogorov-Smirnov tests.

Regulation of TE expression by cell proliferation and KZFPs

Proliferation scores were generated for each dataset cell using the CellCycleScoring function of Seurat (version 4.1.0). As per *Cowan et al.* (493), we combined previously published lists of G2M and S phase marker genes (494) to compute the proliferation scores. For each thymic cell population, we calculated the Spearman correlation between proliferation scores and the expression of TE subfamilies. The Benjamini-Hochberg method was applied to correct for multiple comparisons. Correlations were considered positive if the correlation coefficient was ≥0.2 and the adjusted p-value≤0.05, and negative if the coefficient was ≤-0.2 and the adjusted p-value≤0.05. We also computed the median of all correlation coefficients for each cell population. We then assigned the class of each TE subfamily correlated with cell proliferation and compared this distribution to the distribution of classes of all TE subfamilies in the human genome. The percentage of overlap of the sets of TE subfamilies significantly correlated with cell proliferation was determined. A list of 401 human KZFPs was downloaded from *Imbeault et al.* (271). Spearman correlations between KZFP and TE expression were independently computed in each cell population with the same methodology as the cell proliferation analysis, and Benjamini-Hochberg correction for multiple comparisons was applied. The information on the enrichment of KZFPs within TE subfamilies was downloaded from *Imbeault et al.* (271). Sharing of KZFP-TE pairs between cell populations was represented using the circlize package.

Estimation of TE sequences' age

The sequence divergence (defined as the number of mismatches per thousand) was given by the milliDiv value in RepeatMasker. The milliDiv values of each TE locus were divided by the substitution rate of its host's genome ($2{,}2x10^{-9}$ mutation/year for *Homo sapiens* and $4.5x10^{-9}$ mutation/year for *Mus musculus* (151, 495)). Finally, the age of each TE subfamily was determined by averaging the age of all loci of the subfamily.

Interactions between TE subfamilies and transcription factors

We downloaded a list of 1638 transcription factors (TF) manually curated by *Lambert et al.* (496). For each cell population of the thymus, Spearman correlations were computed for each possible pair of TF and TE subfamily, and the Benjamini-Hochberg method was applied to correct the p-values for multiple comparisons. Correlations were considered significant if i) the correlation coefficient was ≥0.2, ii) the adjusted p-value was ≤0.05, and iii) the TF was expressed by ≥10% of the cells of the population. The correlations were validated using a bootstrap procedure (1000 iterations) to ensure their reproducibility. Briefly, we randomly selected n cells out of the n cells of a given population (while allowing cells to be selected multiple times). The empirical p-value was determined by dividing the number of iterations with a correlation coefficient <0.2 by the total number of iterations (1000). In parallel, the curated binding motifs of 945 TFs were downloaded from the JASPAR database. We then used the *Find Individual Motif Occurences* (FIMO) software (497) to identify the 100 000 genomic positions with the most significant matches for the TF binding motif. These lists of binding motif positions were then intersected with the positions of TE loci with the intersect function of BEDTools (version 2.29.2) (390), and the percentage of TE loci of each subfamily harboring TF binding motifs was determined. Thus, in a specific cell population of the thymus, a TF was considered as interacting with a TE subfamily if it satisfied two criteria: i) its expression was correlated with the one of the TE family (spearman coefficient ≥0.2, adjusted

p-value ≤0.05 and expression of TF in ≥10% of cells), and ii) at least one locus of the TE subfamily contained a binding motif of the TF. For each cell population, networks of interactions between TF and TE subfamilies were generated with the network package (version 1.17.1) (498) in R and represented with the ggnetwork package. For the sake of clarity, only the most significant interactions were illustrated for each cell type (i.e., correlation coefficient ≥0.3, TF binding sites in ≥1% of the loci of the TE subfamily, and TF expression in ≥10% of cells of the population). Sharing of TF-TE interactions between cell populations was represented with a chord diagram using the circlize package. For each TE subfamily, the number of interactions with TFs and the number of loci of the TE subfamily in the human genome were determined. Wilcoxon-Mann-Whitney tests were used to compare the number of interactions with TF of LTR, LINE, and SINE elements, whereas Kendall tau correlation was calculated between the number of interactions with TF and the number of loci of TE subfamilies.

Identification of TE promoter and enhancer candidates

From the previously identified list of TF-TE interactions, we isolated the specific loci containing TF binding sites from the subfamilies whose expression was positively correlated with the TF. To determine if these TE loci could act as promoters or enhancers, we used histone ChIP-seq data from the ENCODE consortium for H3K27ac, H3K4me1, and H3K4me3. BED files from the ENCODE consortium were downloaded for eight immune cell populations: B cells, CD4 Single Positive T cells (CD4 SP), CD8 Single Positive T cells (CD8 SP), dendritic cells (DC), monocytes and macrophages (Mono/Macro), NK cells, Th17, and Treg. TE loci colocalizing with peaks in histone ChIP-seq data were identified using the intersect function of BEDTools (version 2.29.2). To be considered enhancer candidates, TE loci had to colocalize with H3K27ac and H3K4me1 but not H3K4me3. To be considered as promoter candidates, TE loci had to colocalize with H3K27ac

and H3K4me3, but not H3K4me1, and be located at ≤ 1000 nucleotides from a transcription start site (TSS) annotated in the refTSS database (499).

Diversity of TE expression

The human thymic scRNA-seq dataset was subsampled to retain only postnatal cells, as it was shown by *Bornstein et al.* (56) that thymic APCs are mainly found in postnatal samples. The diversity of TE sequences expressed by thymic populations was assessed using Shannon entropy. Using the vegan package (version 2.5-7) (500) in R, two distinct Shannon entropy metrics were computed for each cell population. First, the Shannon entropy was computed based on the expression level (i.e., log(read count)) of TE subfamilies for each cell individually. The median entropy was calculated for each cell population. In parallel, the diversity of TE sequences expressed by an entire population was also assessed. For this purpose, a binary code was generated to represent the expression status of TE subfamilies in each cell (where 1 is expressed and 0 is not expressed). For each population separately, the binary codes of individual cells were summed to obtain the frequency of expression of each TE subfamily in the population, which was used to compute the Shannon entropy of TE sequences expressed by the population. A linear model was generated with the lm function of the stats package in R to summarize the data distribution. The deviation (Δy) from the observed population's TE diversity and the one expected by the linear model was computed for each cell population.

TE expression in thymic APC

A differential expression analysis of TE subfamilies between the subsets of thymic APC was performed with the FindAllMarkers function with default parameters of Seurat (501) with the MAST model. Finally, the heterogeneity of TE expression inside thymic APC subsets was evaluated with the MetaCell package (version 0.3.5) (502). The composition of the metacells was

validated based on manual annotation (see the "Single-cell RNA-seq preprocessing" section), and only metacells with >50% of cells belonging to the same subset of thymic APCs were kept. Differential expression of TE subfamilies between metacells was performed as described above, and the percentage of overlap between the sets of TEs overexpressed by the different metacells was computed.

Isolation of human thymic pDCs and immunostaining of dsRNAs

Primary human thymi were obtained from 4-month-old to 12-year-old children undergoing cardiovascular surgeries at the CHU Sainte-Justine. This project was approved by the CHU Sainte-Justine Research Ethics Board (protocol and biobank #2126). Thymi from 4-month-old to 12-year-old individuals were cryopreserved in liquid nitrogen in the following solution: 95% (PBS-5% Dextran 40 (Sigma-Aldrich)) – 5% DMSO (Fisher Scientific). Protocol for thymic pDCs isolation was based on *Stoeckle et al.* (503). Briefly, thymic samples were cut in ~2mm pieces, followed by three rounds of digestion (40 min, 180 RPM at 37°C) in RPMI 1640 (Gibco) supplemented with 2mg/mL of Collagenase A (Roche) and 0.1mg/mL of DNase I (Sigma-Aldrich). APCs were then enriched using Percoll (Sigma-Aldrich) density centrifugation (3500g, 35min at 4°C), followed by an FBS cushion density gradient (5mL of RPMI 1640 containing enriched APCs layered on 5mL of heat-inactivated FBS (Invitrogen, 12483020), 1000RPM for 10min at 4°C) to remove cell debris. Finally, thymic pDCs were magnetically enriched using the QuadroMACS Separator (Miltenyi). Cells were stained with a CD303 (BDCA-2) MicroBead Kit (Miltenyi), and labeled cells were loaded on LS columns (Miltenyi) for magnetic-activated cell sorting.

Purified thymic pDCs were pipetted on poly-L-lysine (Sigma-Aldrich, 1:10 in dH_2O) coated 15µ-Slide 8 well (ibid) and incubated for 2h at 37°C in RPMI 1640 supplemented with 10% BSA (Sigma-Aldrich). Cells were fixed using 1% [w/v] paraformaldehyde (PFA, Sigma-Aldrich) in PBS

1X (Sigma-Aldrich) for 30min at room temperature. Cells were permeabilized for 30min at room temperature with 0.1% [v/v] Triton X-100 (Sigma-Aldrich) in PBS 1X, followed by blocking using 5% [w/v] BSA (Sigma-Aldrich) in PBS 1X for 30min at room temperature. Immunostaining was performed in four steps to avoid unspecific binding of the secondary antibodies: i) incubation overnight at 4°C with the mouse monoclonal IgG2a J2 antibody anti-dsRNA (Jena Bioscience, cat. RNT-SCI-10010500, dilution 1:200), ii) incubation with the donkey anti-mouse IgG (H+L) antibody coupled to Alexa Fluor 555 (Invitrogen, cat. A-31570, dilution 1:500) for 30min at room temperature, iii) incubation with the mouse monoclonal IgG1 clone 6H6 anti-CD123 (eBioscience, cat. 14-1239-82, 1:100) for one hour at room temperature, and iv) incubation with the goat anti-mouse IgG1 polyclonal Alexa Fluor 488 antibody (Invitrogen, cat. A-21121, 1:1000) for 30min at room temperature. Finally, cells were stained with DAPI (Invitrogen, cat. D3571, 1:1000) for 5 minutes at room temperature. All antibodies and DAPI were diluted in a blocking solution. Image acquisition was made with an LSM 700 laser scanning confocal microscope (Zeiss) using a 40x oil objective (Zeiss, Plan-Neofluar N.A. 1.4) and the ZEN software. Using the whiteTopHat function of the EBImage package and the sigmoNormalize function of the MorphoR package in R, the background of the DAPI signal was removed. The nuclei were segmented on the resulting images as circular shapes based on the DAPI signal. The mean intensity of CD123 and J2 staining was determined for each cytoplasm, defined as 19nm rings around nuclei. Based on the distribution of the CD123 signal across cells, a threshold between $CD123^-$ and $CD123^+$ cells was set up for each replicate independently. J2 signal intensity was compared between $CD123^-$ and $CD123^+$ cells using the Wilcoxon Rank Sum test in R.

Gene set enrichment analysis

Gene set enrichment analyses were performed to determine which biological processes are enriched in mTEC(II) and pDCs. Differential gene expression analyses were performed between each possible pair of thymic APCs subsets using MAST with the FindMarkers function of Seurat. The gene set enrichment analysis was performed using the iDEA package (version 1.0.1) (504) in R. As per *Ma et al.* (504), the fold change and standard error of gene expression were used as input for iDEA, in addition to predefined lists of gene sets compiled in the iDEA package. Gene sets associated with antigen presentation, interferon signaling, and immune response were manually annotated. iDEA was launched with default parameters, except for the 500 iterations of the Markov chain Monte Carlo algorithm, and p-values were corrected with the Louis method. We also visualized the expression of *AIRE, FEZF2,* and *CHD4* in the TEC lineage to validate their expression in mTEC(II).

TE loci regulated by AIRE, FEZF2, and CHD4

A differential expression analysis of TE subfamilies between WT and *Aire*-, *Fezf2*-, or *Chd4*-KO mice was performed with the voom method of the limma package (version 3.46.0) (505, 506). Stringent criteria (i.e., an expression below 2 transcripts per million (TPM) in all samples) were applied to remove lowly expressed TEs. TE subfamilies with i) a fold change ≥2 and an adjusted p-value ≤0.05 or ii) a fold change ≤-2 and an adjusted p-value ≤0.05 were considered as induced and repressed, respectively. The percentage of overlap between the sets of TE loci induced or repressed by AIRE, FEZF2, and CHD4 was computed. The class and subfamily were assigned to each regulated TE locus, and the distributions of classes and subfamilies across all TE sequences of the murine genome were used as controls. Significant enrichment of classes or subfamilies was determined with Chi-squared tests, and a Bonferroni correction for multiple comparisons was performed to enrich subfamilies in induced or repressed TEs. The distance between TE loci induced

or repressed by AIRE, FEZF2, or CHD4 was defined as the minimal distance between the middle position of TE loci on the same chromosome. As a control, distributions of randomly selected TE loci whose expression is independent of AIRE, FEZF2, and CHD4 and equal size to the sets of regulated TEs were generated (for example, if 433 TE loci are induced by CHD4, 433 independent TE loci were randomly selected). Wilcoxon rank-sum tests were used to compare random and regulated distributions. Genomic positions of exons, introns 3' and 5' untranslated transcribed region (UTR) were downloaded from the UCSC Table Browser. The genomic localization of regulated TEs was determined using the intersect mode of the BEDTools suite version 2.29.2. TE loci not located in exons, introns, 3'UTR, or 5'UTR were considered intergenic. The percentage of regulated TE loci in each type of genomic region was determined and compared to the genomic localization of all TE loci in the murine genome with chi-squared tests. Finally, we estimated the frequency of intron retention events for introns containing TE loci regulated by AIRE, FEZF2, or CHD4 with S-IRFindeR (507). Sequencing reads were aligned to the reference *Mus musculus* genome (mm10) using STAR version 2.7.1a (389) with default parameters. Each intron's Stable Intron Retention ratio (SIRratio) was computed with the computeSIRratio function of S-IRFindeR. Introns containing TE loci induced by AIRE, FEZF2, or CHD4 were filtered using BEDTools intersect. Random distributions of equivalent sizes of introns containing TE sequences independent of AIRE, FEZF2, and CHD4 were generated as control. A SIRratio of 0.1 was used as a threshold of significant intron retention events.

Enzymatic digestion and isolation of murine TECs

Thymic stromal cell enrichment was performed as previously described (464, 508). Briefly, thymi from 16- to 22-week-old K5D1 mice were mechanically disrupted and enzymatically digested with papain (Worthington Biochemical Corporation), DNase I (Sigma-Aldrich), and collagenase IV

(Sigma-Aldrich) at 37°C. Next, the single-cell suspension obtained after enzymatic digestion was maintained at 4°C in FACS buffer (PBS, 0.5% [w/v] BSA, 2mM EDTA) and enriched in thymic epithelial cells using anti-EpCAM (CD326) or anti-CD45 microbeads (mouse, Miltenyi) and LS columns (Miltenyi). Then, the enriched epithelial cell suspension was stained for flow cytometry cell sorting with the following antibodies and dyes: anti-EpCAM-APC-Cy7 clone G8.8 (BioLegend, cat. 118218), anti-CD45-APC clone 30-F11 (BD Biosciences, cat. 559864), anti-UEA1-biotinylated (Vector Laboratories, cat. B-1065), anti-I-A/I-E-Alexa Fluor 700 clone M5/114.15.2 (BioLegend, cat. 107622), anti-Ly51-FITC clone 6C3 (BioLegend, cat. 553160), anti-streptavidin-PE-Cy7 (BD Biosciences, cat. 557598), and 7-AAD (BD Biosciences, cat. 559925). Cell sorting was performed using a BD FACSAria (BD Biosciences), and data were analyzed using the FACSDiva. TECs were defined as $EpCAM^{+}CD45^{-}$, while the cTEC and mTEC subsets were defined as $UEA1^{-}Ly51^{+}$ and $UEA1^{+}Ly51^{-}$ TEC, respectively.

RNA-Sequencing

Total RNA from 80 000 mTECs or cTECs was isolated using TRIzol and purified with an RNeasy micro kit (Qiagen). Total RNA was quantified using Qubit (Thermo Scientific), and RNA quality was assessed with the Agilent 2100 Bioanalyzer (Agilent Technologies). Transcriptome libraries were generated using a KAPA RNA HyperPrep kit (Roche) using a poly(A) selection (Thermo Scientific). Sequencing was performed on the Illumina NextSeq 500, obtaining ~200 million paired-end reads per sample.

Preparation of CNBR-activated Sepharose beads for MHC I immunoprecipitation

CNBR-activated Sepharose 4B beads (Sigma-Aldrich, cat. 17-0430-01) were incubated with 1 mM HCl at a ratio of 40 mg of beads per 13.5 ml of 1 mM HCl for 30 minutes with tumbling at room

temperature. Beads were spun at 215g for 1 minute at 4°C, and supernatants were discarded. 40 mg of beads were resuspended with 4 ml of coupling buffer (0.1M $NaHC0_3$/0.5M NaCl pH 8.3), spun at 215g for 1 minute at 4°C, and the supernatants were discarded. Mouse antibodies Pan-H2 (clone M1/42), H2-K^b (clone Y-3), and H2-D^b (clone 28-14-8S) were coupled to beads at a ratio of 1 mg of antibody to 40 mg of beads in coupling buffer for 120 minutes with tumbling at room temperature. Beads were spun at 215g for 1 minute at 4°C, and supernatants were discarded. 40 mg of beads were resuspended with 1 ml of blocking buffer (0.2M glycine), incubated for 30 minutes with tumbling at room temperature, and the supernatants were discarded. Beads were washed by centrifugation twice with PBS pH 7.2, resuspended at a concentration of 1 mg of antibody per ml of PBS pH 7.2, and stored at 4°C.

Immuno-isolation of MAPs

Frozen pellets of mTECs (90 mice, 191 million cells total) and cTECs (121 mice, 164 million cells total) were thawed, pooled, and resuspended with PBS pH 7.2 up to 4 ml and then solubilized by adding 4 mL of detergent buffer containing PBS pH 7.2, 1% (w/v) CHAPS (Sigma, cat. C9426-5G) supplemented with Protease inhibitor cocktail (Sigma, cat. P8340-5mL). Solubilized cells were incubated for 60 minutes with tumbling at 4°C and then spun at 16,600g for 20 minutes at 4°C. Supernatants were transferred into new tubes containing 1.5 mg of Pan-H2, 0.5 mg of H2-K^b, and 0.5 mg of H2-D^b antibodies covalently-cross-linked CNBR-Sepharose beads per sample and incubated with tumbling for 180 minutes at 4°C. Samples were transferred into BioRad Poly prep chromatography columns and eluted by gravity. Beads were first washed with 11.5 mL PBS, then with 11.5 mL of 0.1X PBS, and finally with 11.5 mL of water. MHC I complexes were eluted from the beads by acidic treatment using 1% trifluoroacetic acid (TFA). Acidic filtrates containing

peptides were separated from MHC I subunits (HLA molecules and β-2 microglobulin) using home-made stage tips packed with two 1 mm diameter octadecyl (C-18) solid phase extraction disks (EMPORE). Stage tips were pre-washed with methanol, then with 80% acetonitrile (ACN) in 0.1% TFA, and finally with 1% TFA. Samples were loaded onto the stage tips and washed with 1% TFA and 0.1% TFA. Peptides were eluted with 30% ACN in 0.1% TFA, dried using vacuum centrifugation, and then stored at -20 °C until MS analysis.

MS analyses

Peptides were loaded and separated on a home-made reversed-phase column (150-µm i.d. by 200 mm) with a 106-min gradient from 10 to 38% B (A: formic acid 0.1%, B: 80% CAN 0.1% formic acid) and a 600-nl/min flow rate on an Easy nLC-1200 connected to an Orbitrap Exploris 480 (Thermo Fisher Scientific). Each full MS spectrum acquired at a resolution of 240,000 was followed by tandem-MS (MS-MS) spectra acquisition on the most abundant multiply charged precursor ions for a maximum of 3s. Tandem-MS experiments were performed using higher energy collision-induced dissociation (HCD) at a collision energy of 34%. The generation of the personalized proteome containing TE sequences, as well as the identification of TE-derived MAPs, was performed as per *Larouche et al.* (440) with the following modifications: the mm10 murine reference genome was downloaded from the UCSC Genome Browser, the annotations for murine genes and TE sequences were downloaded from the UCSC Table Browser, and the Uniprot mouse database (16 977 entries) was used for the canonical proteome. MAPs were identified using PEAKS X Pro (Bioinformatics Solutions, Waterloo, ON). The level and frequency of expression of TE subfamilies generating MAPs or not were determined in thymic epithelial cells were determined by averaging the expression values across cells of a TEC subset and dividing the

number of cells with a positive (i.e.,> 0) expression of the TEs by the total number of cells of the TEC subset, respectively.

Availability of data and materials:

scRNA-seq data of human thymi and spleen were downloaded from ArrayExpress (accession number E-MTAB-8581) and the NCBI BIOPROJECT (accession code PRJEB31843), respectively. scRNA-seq data of murine thymi were downloaded from ArrayExpress (accession number E-MTAB-8560). RNA-seq data from WT, Aire-KO, Fezf2-KO, and Chd4-KO murine mTECs were downloaded from the Gene Expression Omnibus (GEO) under the accession code GSE144880. ChIP-seq data of ETS1 in human NK cells and AIRE in murine mTECs were downloaded from GEO (accession codes GSE124104 and GSE92654, respectively). ChIP-seq data for different histone marks in murine mTECs were also downloaded from GEO: H3K4me3 for mTECs (GSE53111); H3K4me1 and H3K27ac from MHCII^{hi} mTECs (GSE92597); H3K4me2 in mTEC-II (GSE103969); and H3K4ac and H3K9ac in mTECs (GSE114713). Transcriptomic and immunopeptidomic data of K5D1 mice mTECs and cTECs generated in this study are available on the Gene Expression Omnibus (GEO) under the accession GSE232011 and on the Proteomics Identification Database (PRIDE) under the accession PXD042241, respectively.

3.1.3.6 Declarations

Author contributions:

Conceptualization: JDL, CML, AT, KV, CP

Methodology: JDL, CML, AT, KV

Investigation: JDL, LH, CC, SB, JH, EB, JL, CD, PG, JPL

Visualization: JDL, CML, AT

Funding acquisition: JDL, CP

Project administration: JDL, KV, CP

Supervision: SL, PT, CP

Writing – original draft: JDL, KV, CP

Writing – review & editing: JDL, CML, AT, KV, LH, JH, SB, CC, EB, JL, CD, PG, ERR, BHN, SL, PT, CP

Acknowledgments:

The authors thank Christian Charbonneau and Raphaëlle Lambert from IRIC's bio-imaging and genomics platforms, respectively. We also thank Allan Sauvat for the help with the microscopy quantification. We thank Mathilde Soulez, Bernhard Lehnertz, Biljana Culjkovic, Brian Wilhelm, and Michaël Imbeault for insightful discussions. We are indebted to Kathie Béland and Elie Haddad, from the CHU Sainte-Justine Research Center, for providing the primary thymic samples.

3.1.3.7 Figures

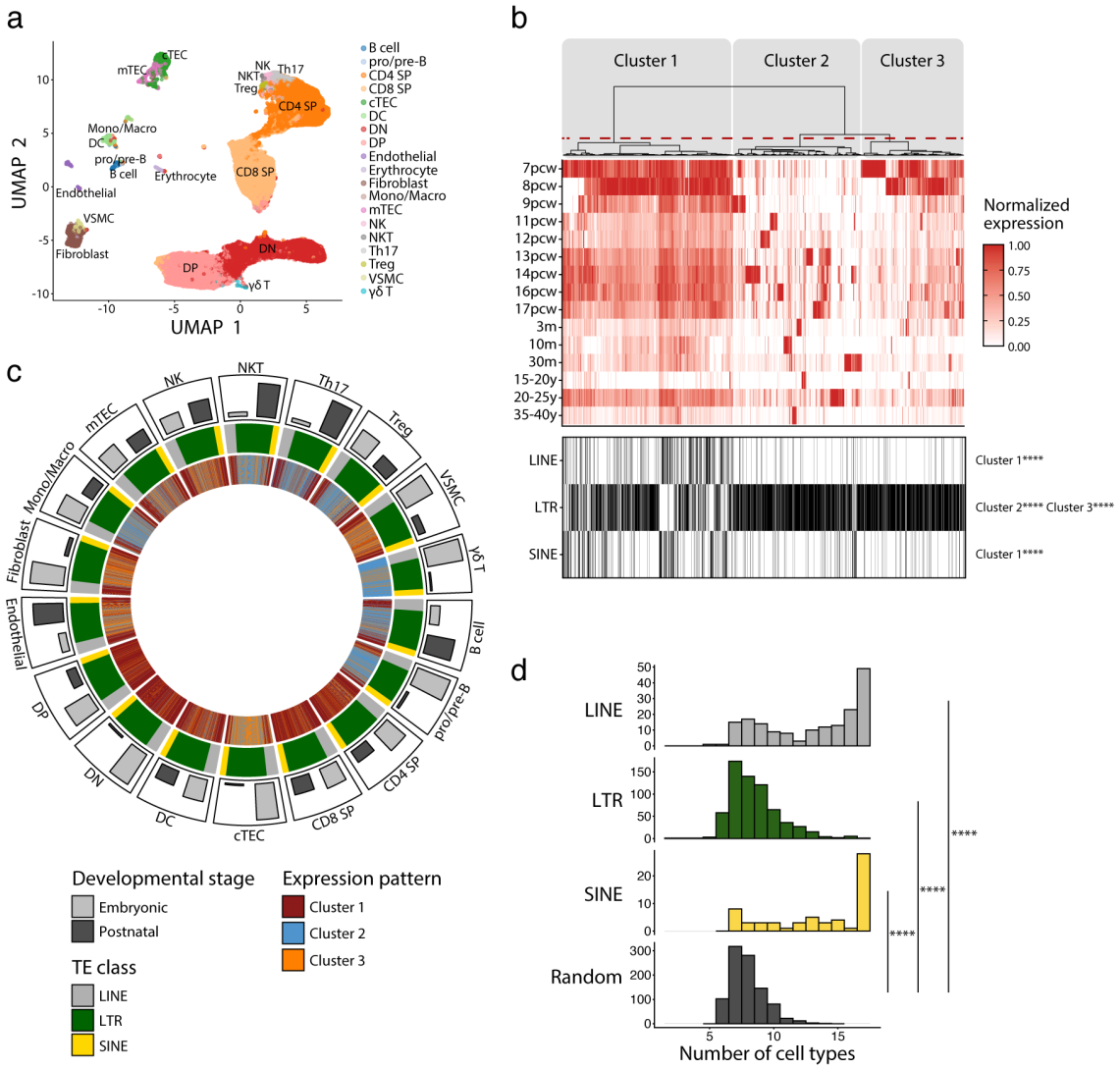
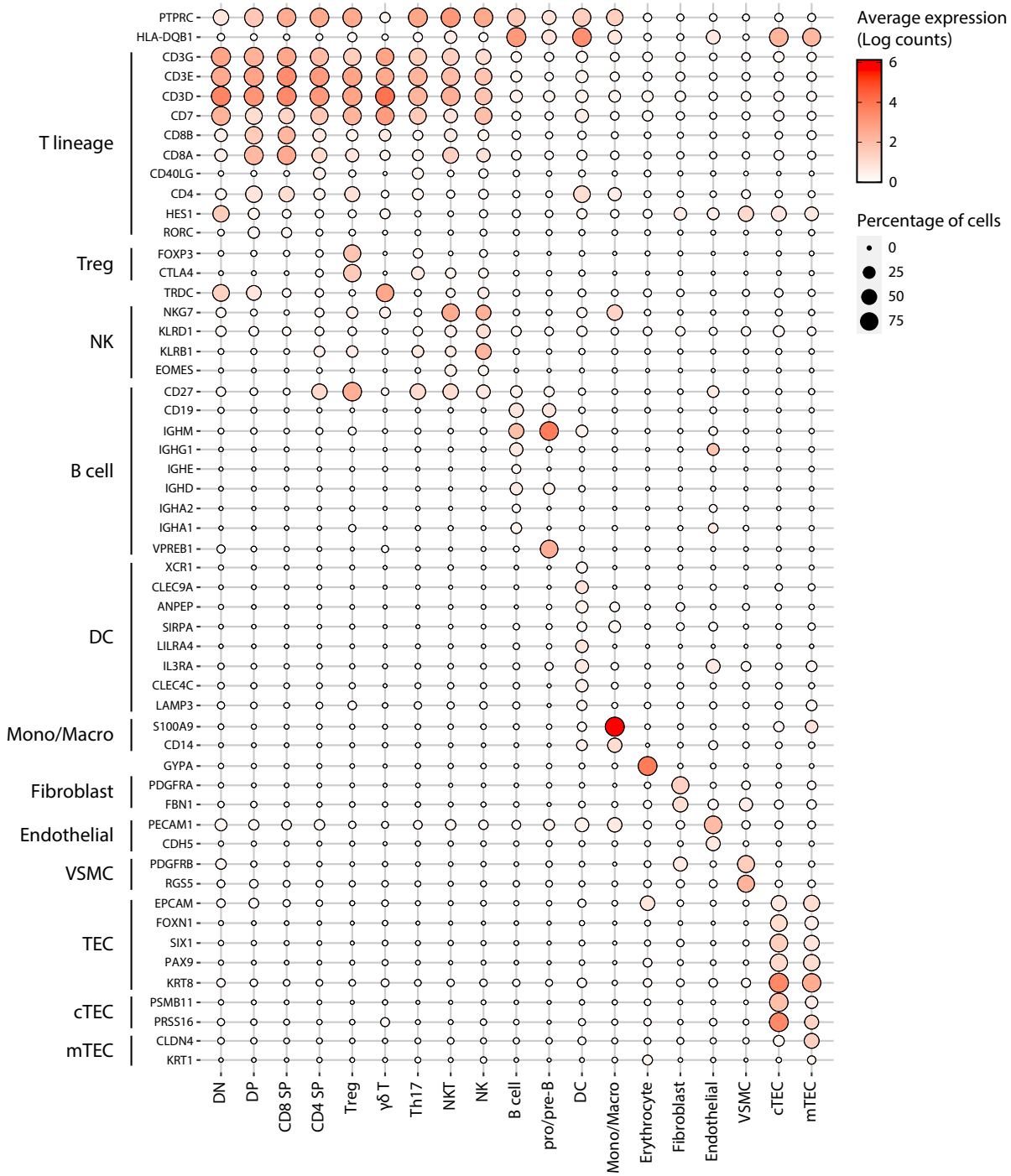


Figure 3.1 - LINEs, SINEs, and LTRs exhibit distinct expression profiles in human thymic cell populations.

(a) UMAP depicting the cell populations present in human thymi (CD4 SP, CD4 single positive thymocytes; CD8 SP, CD8 single positive thymocytes; cTEC, cortical thymic epithelial cells; DC, dendritic cells; DN, double negative thymocytes; DP, double positive thymocytes; Mono/Macro, monocytes and macrophages; mTEC, medullary thymic epithelial cells; NK, natural killer cells; NKT, natural killer T cells; pro/pre-B, pro-B and pre-B cells; Th17, T helper 17 cells; Treg,

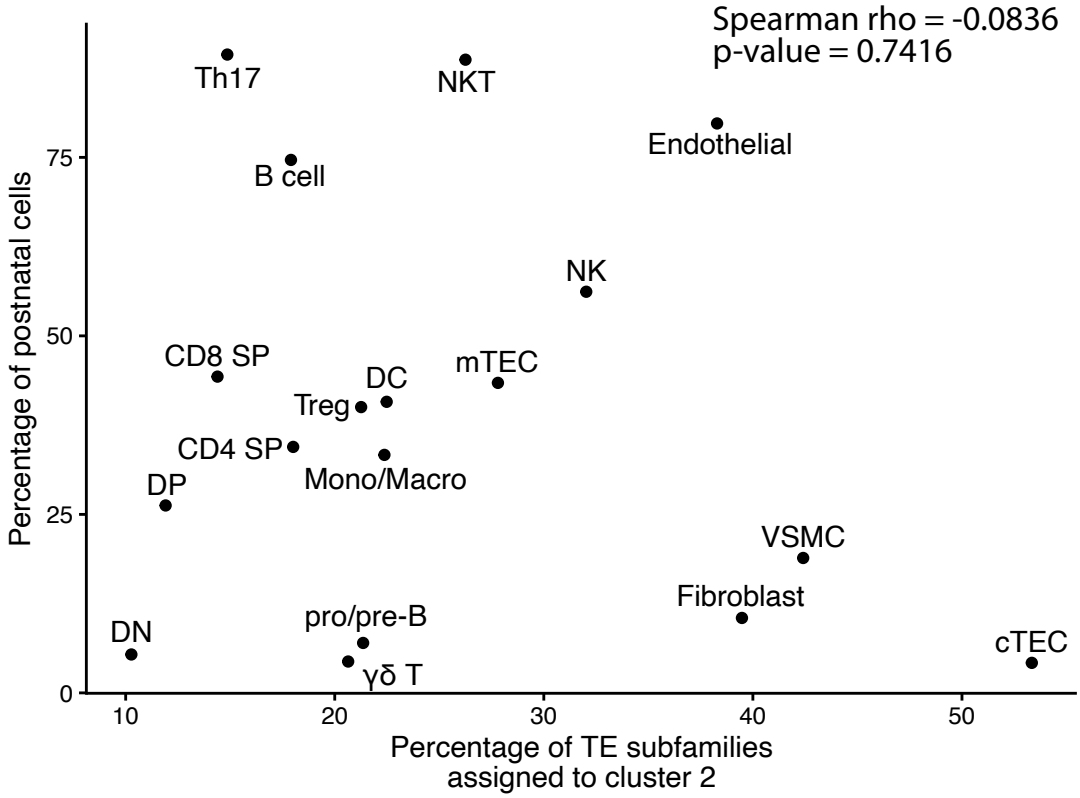
regulatory T cells; VSMC, vascular smooth muscle cell). Cells were clustered in 19 populations based on the expression of marker genes from *Park et al.* (40). **(b)** *Upper panel:* Heatmap of TE expression during thymic development, with each column representing the expression of one TE subfamily in one cell type. Unsupervised hierarchical clustering was performed, and the dendrogram was manually cut into 3 clusters (red dashed line). *Lower panel:* The class of TE subfamilies and significant enrichments in the 3 clusters (Fisher's exact tests; ****$p \leq 0.0001$). (pcw, post-conception week; m, month; y, year). **(c)** Circos plot showing the expression pattern of TE subfamilies across thymic cells. From outermost to innermost tracks: i) proportion of cells in embryonic and postnatal samples, ii) class of TE subfamilies, iii) expression pattern of TE subfamilies identified in (b). TE subfamilies are in the same order for all cell types. **(d)** Histograms showing the number of cell types sharing the same expression pattern for a given TE subfamily. LINE (n=171), LTR (n=577), and SINE (n=60) were compared to a randomly generated distribution (n=809) (Kolmogorov-Smirnov tests, ****$p \leq 0.0001$).



Supplementary Figure 3.1 - Annotation of human thymic cell populations.

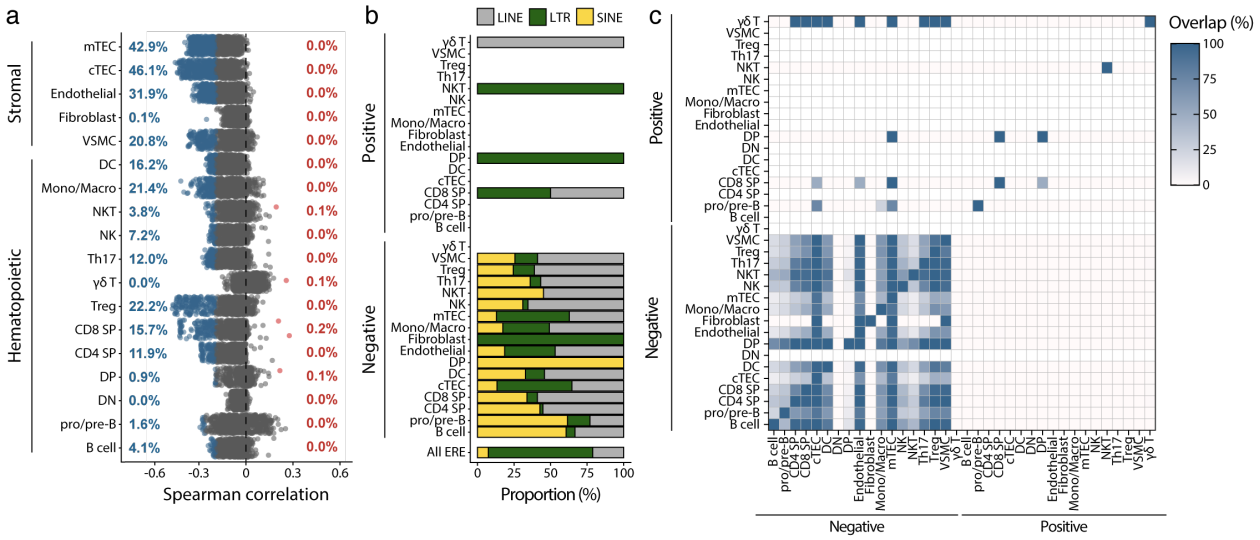
Dot plot depicting the expression of marker genes in the annotated cell types of the thymus. The average expression and percentage of cells expressing the gene are represented by the color and

size of the dot, respectively (DN, double negative thymocytes; DP, double positive thymocytes; CD8 SP, CD8 single positive thymocytes; CD4 SP, CD4 single positive thymocytes, Treg, regulatory T cells; NKT, natural killer T cells; NK, natural killer cells; DC, dendritic cells; Mono/Macro, monocytes and macrophages; VSMC, vascular smooth muscle cells; cTEC, cortical thymic epithelial cells; mTEC, medullary thymic epithelial cells).



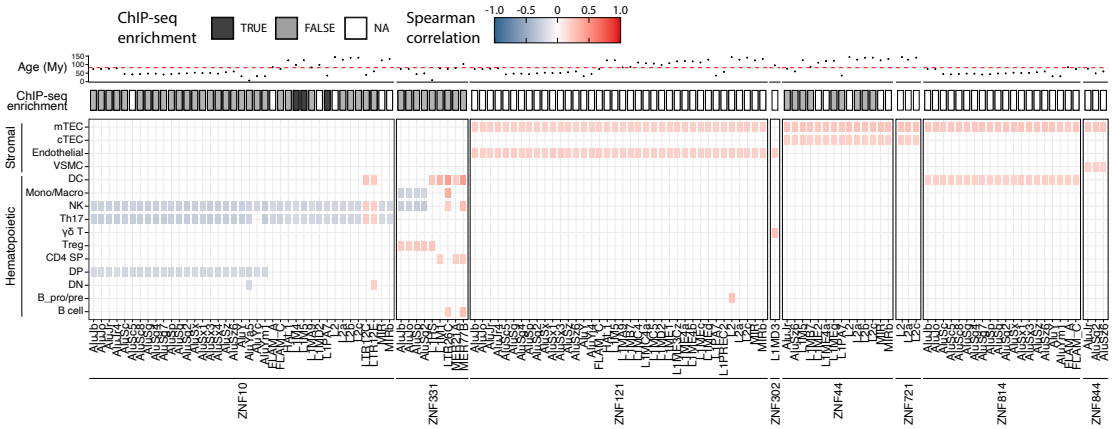
Supplementary Figure 3.2 - Assignment to cluster 2 is independent of the developmental stage of cells.

The graph depicts the correlation between the proportion of cells of a population originating from a postnatal sample and the proportion of TE subfamilies assigned to cluster 2 by the hierarchical clustering in Figure 1B.



Supplementary Figure 3.3 - TE expression is negatively correlated with cell proliferation.

(a) Spearman correlation between the expression of TE subfamilies and cell cycle scores. Positively (r≥0.2 and adj. p≤0.01) and negatively (r≤-0.2 and adj. p≤0.01) correlated subfamilies are red and blue, respectively. P-values were corrected for multiple comparisons with the Benjamini-Hochberg method. **(b)** Proportion of subfamilies positively or negatively correlated with cell proliferation belonging to each TE class. **(c)** Percentage of overlap of TE subfamilies positively or negatively correlated with cell proliferation between cell types.



Supplementary Figure 3.4 - KZFPs repress TE expression in the hematopoietic lineage of the thymus.

Lower panel: pairs of TE subfamilies and KZFPs significantly correlated in at least two cell types (significant correlation: r>0.2 and adj. p≤0.05, or r<-0.2 and adj. p≤0.05, p-values corrected for multiple comparisons with the Benjamini-Hochberg method). *Middle panel*: Enrichment of the KZFP in the sequence of the correlated TE subfamily in ChIP-seq data from *Imbeault et al* (271). *Upper panel*: Age of TE subfamilies in millions of years (My). The estimated time of divergence between primates and rodents (82 million years ago) is indicated by the dashed line.

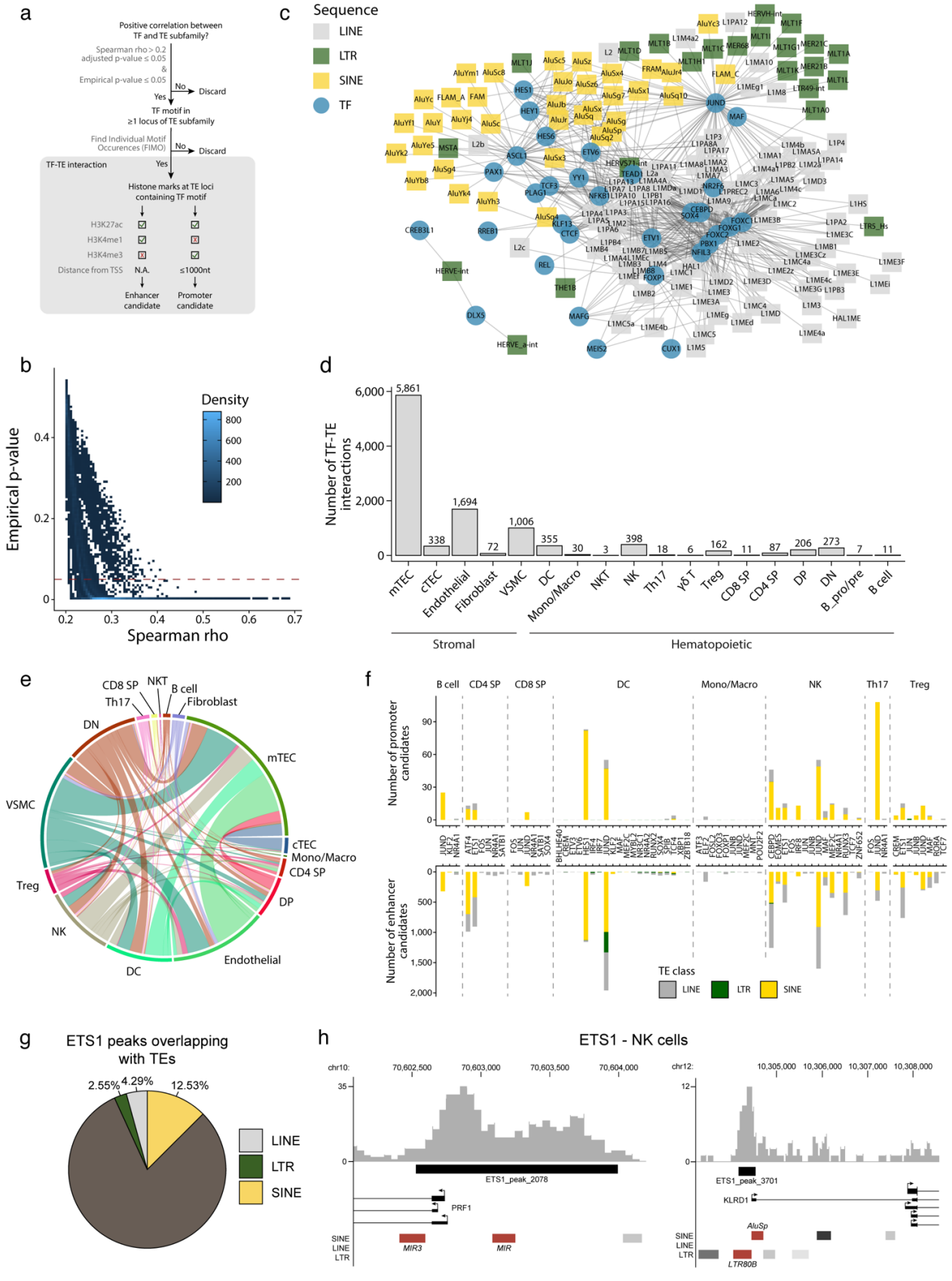
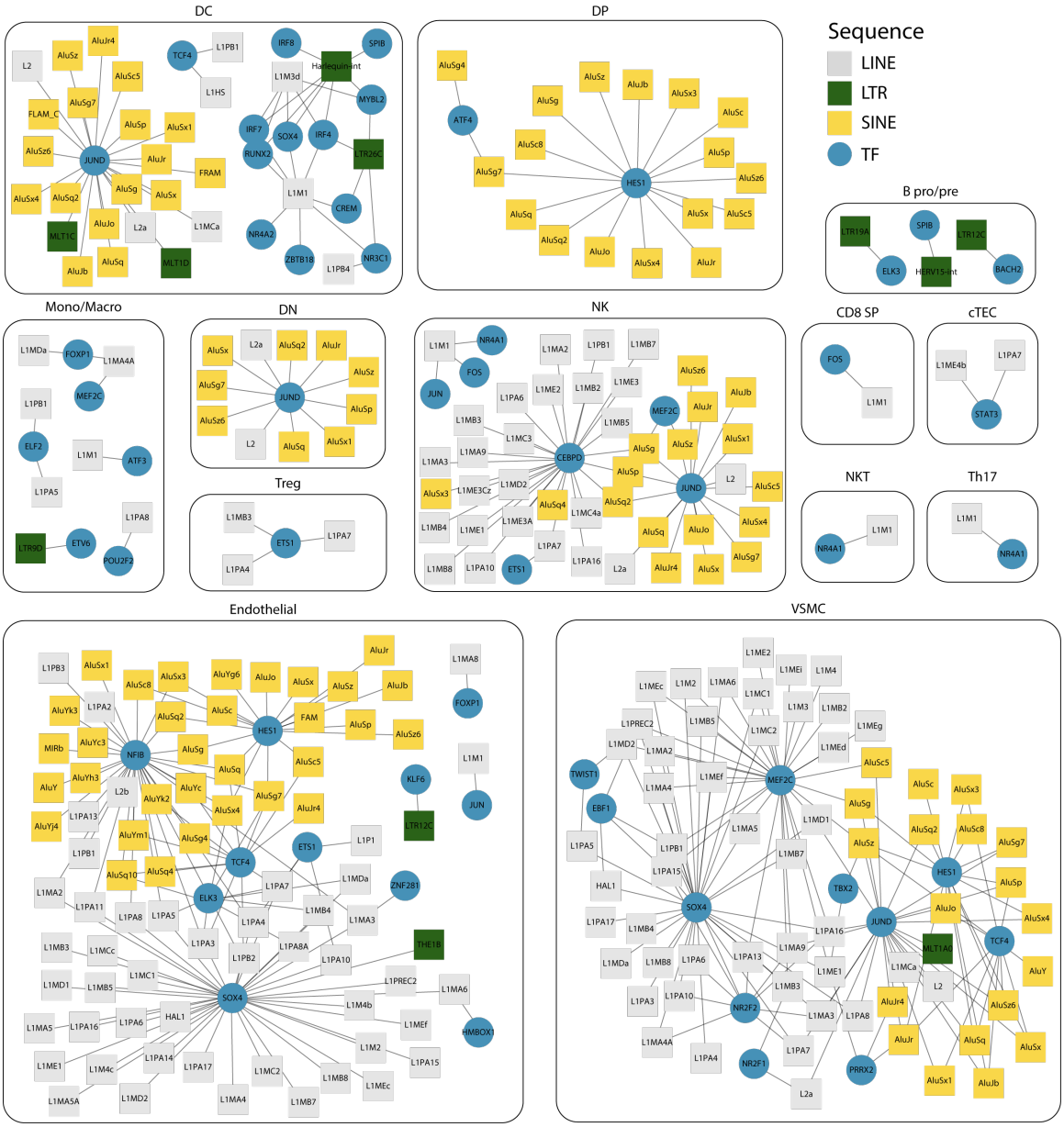

a
Positive correlation between TF and TE subfamily?
Spearman rho > 0.2
adjusted p-value ≤ 0.05
&
Empirical p-value ≤ 0.05
No
Discard
Yes
TF motif in ≥1 locus of TE subfamily
Find Individual Motif Occurences (FIMO)
No
Discard
Yes
TF-TE interaction
Histone marks at TE loci containing TF motif
H3K27ac
H3K4me1
H3K4me3
Distance from TSS
N.A.
≤1000nt
Enhancer candidate
Promoter candidate
c
Sequence
LINE
LTR
SINE
TF
b
Density
800
600
400
200
Empirical p-value
Spearman rho
d
Number of TF-TE interactions
5,861
338
1,694
72
1,006
355
30
3
398
18
6
162
11
87
206
273
7
11
mTEC
cTEC
Endothelial
Fibroblast
VSMC
DC
Mono/Macro
NKT
NK
Th17
γδ T
Treg
CD8 SP
CD4 SP
DP
DN
B_pro/pre
B cell
Stromal
Hematopoietic
e
CD8 SP
NKT
B cell
Th17
Fibroblast
DN
mTEC
VSMC
cTEC
Mono/Macro
CD4 SP
Treg
DP
NK
Endothelial
DC
f
B cell
CD4 SP
CD8 SP
DC
Mono/Macro
NK
Th17
Treg
Number of promoter candidates
Number of enhancer candidates
TE class
LINE
LTR
SINE
g
ETS1 peaks overlapping with TEs
2.55%
4.29%
12.53%
LINE
LTR
SINE
h
ETS1 - NK cells
chr10:
70,602,500
70,603,000
70,603,500
70,604,000
ETS1_peak_2078
PRF1
MIR3
MIR
chr12:
10,305,000
10,306,000
10,307,000
10,308,000
ETS1_peak_3701
KLRD1
AluSp
LTR80B
SINE
LINE
LTR

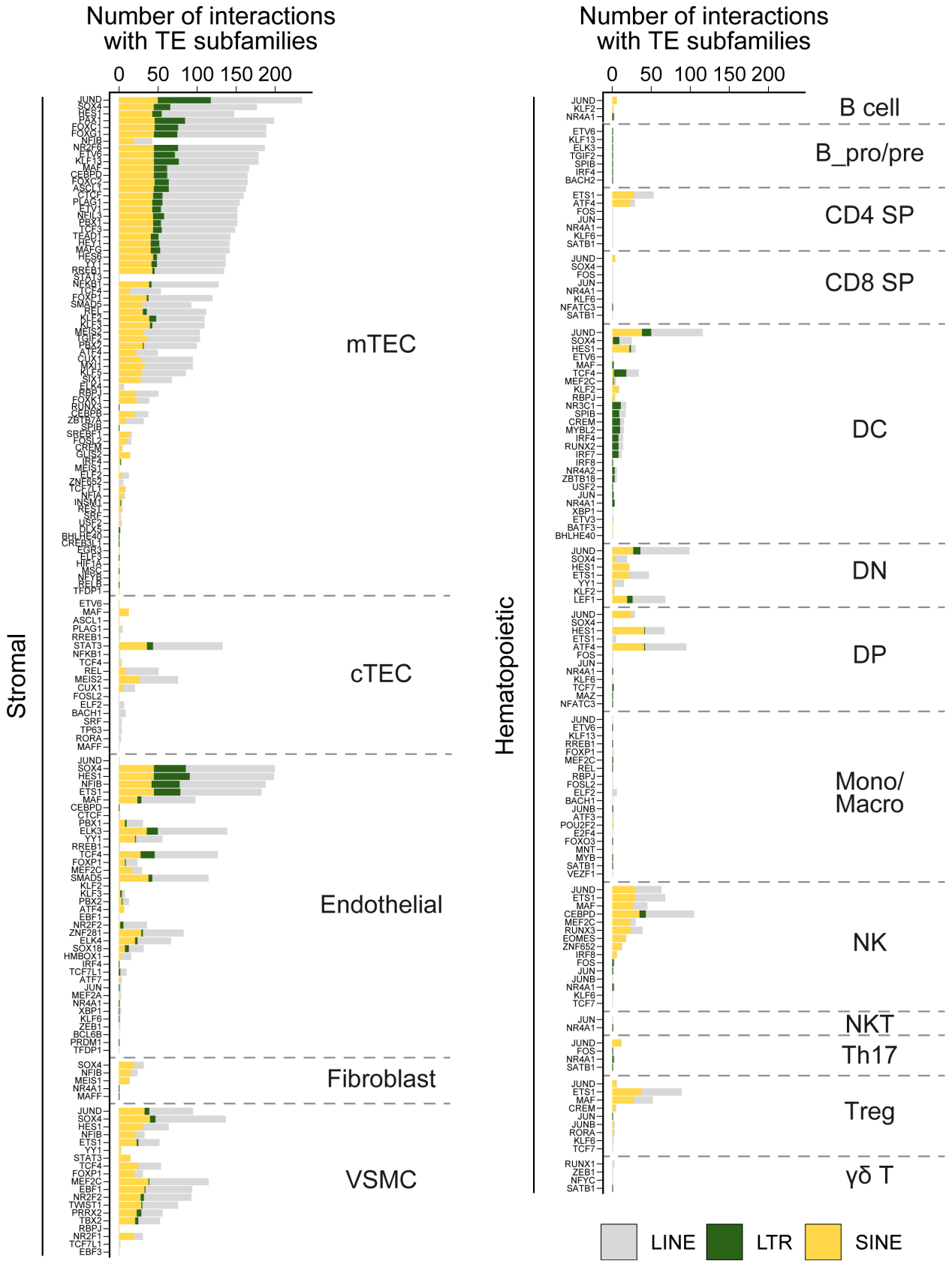
Figure 3.2 - TEs shape complex gene regulatory networks in thymic cells.

(a) The flowchart depicts the decision tree for each TE promoter or enhancer candidate. **(b)** Density heatmap representing the correlation coefficient and the empirical p-value determined by bootstrap for TF and TE pairs in each cell type of the dataset. The color code shows density (i.e., the occurrence of TF-TE pairs at a specific point). **(c)** Connectivity map of interactions between TEs and TFs in mTECs. For visualization purposes, only TF-TE pairs with high positive correlations (Spearman correlation coefficient ≥ 0.3 and p-value adjusted for multiple comparisons with the Benjamini-Hochberg procedure ≤ 0.05) and TF binding sites in $\geq 1\%$ of TE loci are shown. **(d)** Number of TF-TE interactions for each thymic cell population. **(e)** Sharing of TF-TE pairs between thymic cell types. **(f)** Number of promoter (*top*) or enhancer (*bottom*) TE candidates per transcription factor in hematopoietic cells of the thymus. **(g)** The proportion of statistically significant peaks overlapping with TE sequences in ETS1 ChIP-seq data from NK cells. **(h)** Genomic tracks depicting the colocalization of ETS1 occupancy (i.e., read coverage) and TE sequences (*in red)* in the upstream region of two genes in ETS1 ChIP-seq data from NK cells. Statistically significant ETS1 peaks are indicated by the black rectangles.



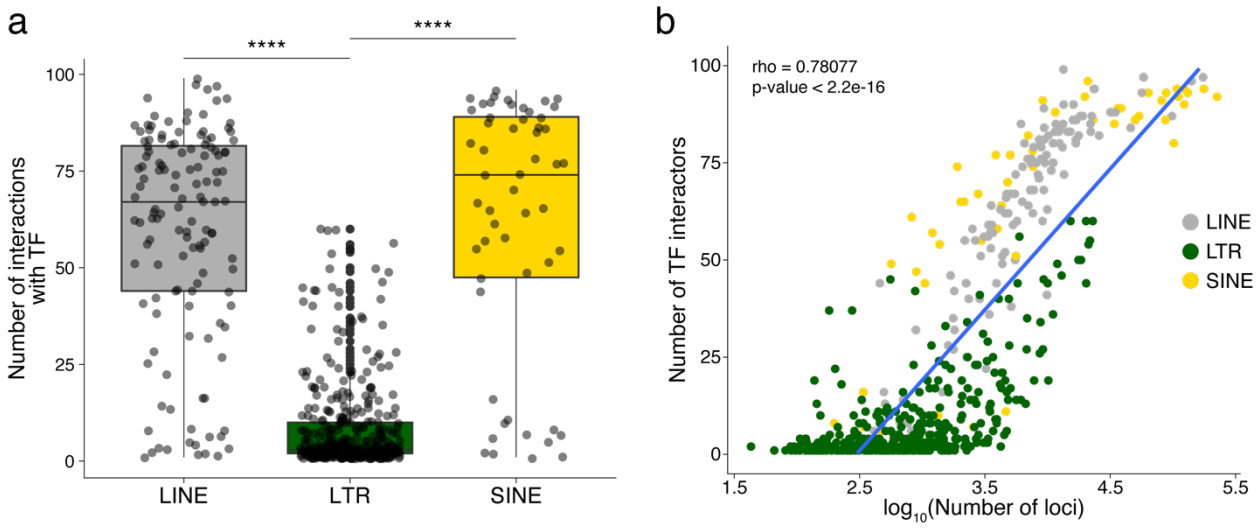
Supplementary Figure 3.5 - Interaction networks between transcription factors and TE subfamilies.

For each cell type, networks illustrate the interactions between TF and TE subfamilies. Pairs of TF and TE are connected by edges when i) their expressions are significantly correlated (Spearman correlation coefficient ≥ 0.2) and ii) the TF binding motifs are found in the loci of the TE subfamily. TE subfamilies are colored based on the class of TE subfamily (LINE, LTR, and SINE).


Number of interactions with TE subfamilies
0 50 100 150 200
Stromal
mTEC
cTEC
Endothelial
Fibroblast
VSMC
Hematopoietic
B cell
B_pro/pre
CD4 SP
CD8 SP
DC
DN
DP
Mono/ Macro
NK
NKT
Th17
Treg
γδ T
LINE
LTR
SINE

Supplementary Figure 3.6 - Frequency of interactions between transcription factors and TE subfamilies in thymic cells.

For each cell type of the stromal (*left*) or hematopoietic (*right*) compartments of the thymus, the graph shows the number of interactions between transcription factors and TE subfamilies of the LINE, LTR, or SINE groups.



Supplementary Figure 3.7 - TE subfamilies occupying larger genomic spaces interact more frequently with TF.

(a) Number of interactions formed with TFs for each TE subfamily of the LINE, LTR, and SINE classes (Wilcoxon-Mann-Whitney tests, ****p≤0.0001). **(b)** Scatterplot depicting the Kendall tau correlation between the number of interactions with TFs of a TE subfamily and the number of loci of that subfamily in the human genome. The color code indicates the class of TE subfamilies.

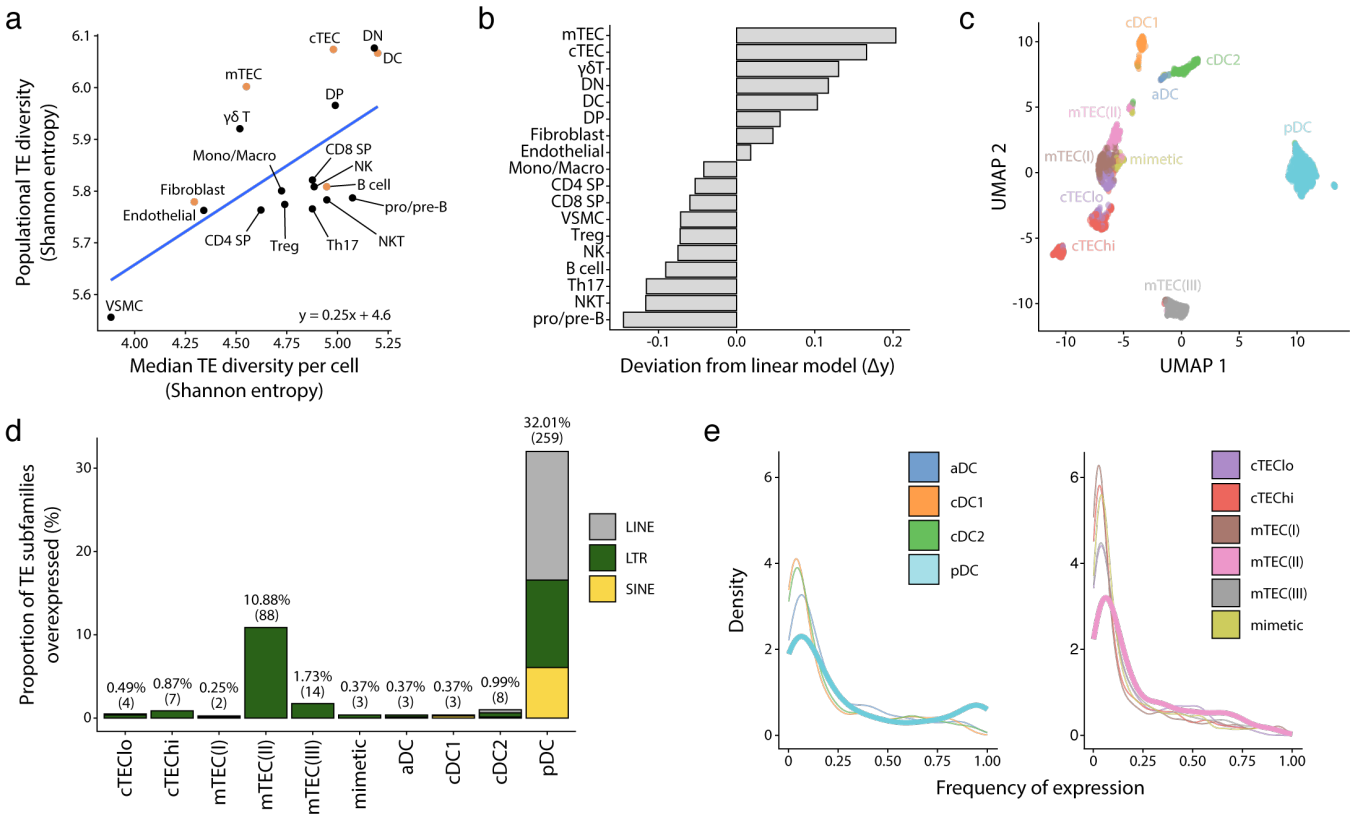
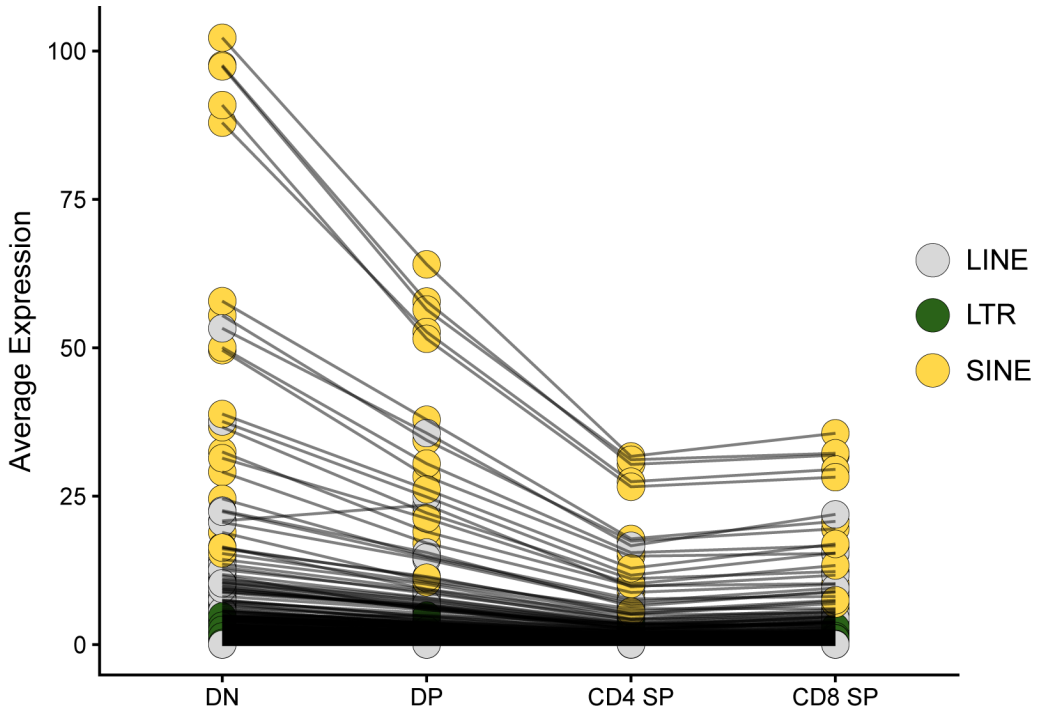
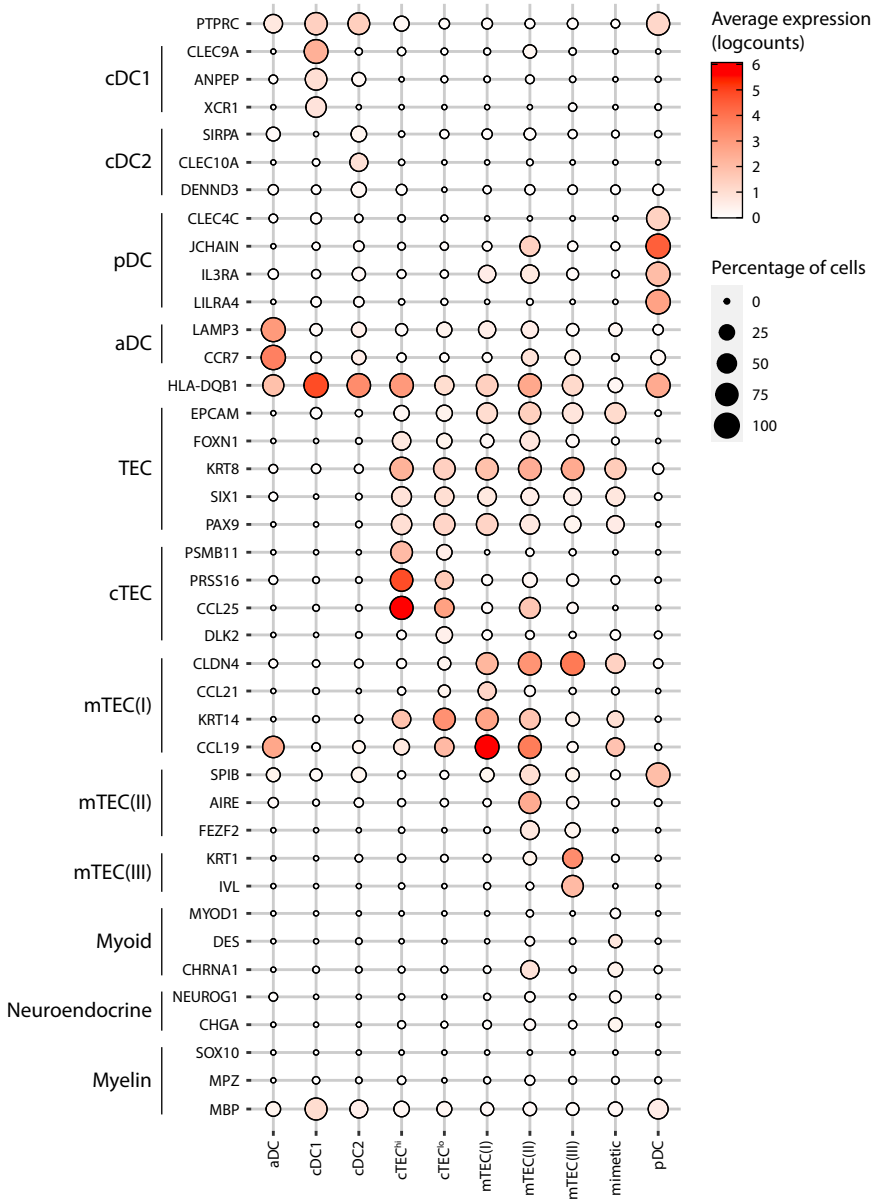


Figure 3.3 - Human pDCs and mTEC(II) express diverse and distinct repertoires of TE sequences. Diversity of TEs expressed by thymic populations measured by Shannon entropy. The x and y axes represent the median diversity of TEs expressed by individual cells in a population and the global diversity of TEs expressed by an entire population, respectively. The equation and blue curve represent a linear model summarizing the data. Thymic APC subsets are indicated in orange. **(b)** Difference between the observed diversity of TEs expressed by cell populations and the one expected by the linear model in (A). **(c)** UMAP showing the subsets of thymic APCs (aDC, activated DC; cDC1, conventional DC1; cDC2, conventional DC2; pDC, plasmacytoid DC). **(d)** Bar plot showing the number and class of differentially expressed TE subfamilies between APC subsets. **(e)** Frequency of expression of TE subfamilies by the different APC subsets. The distributions for pDCs and mTEC(II) are highlighted in bold.



Supplementary Figure 3.8 - TE expression decreases during thymocyte differentiation.

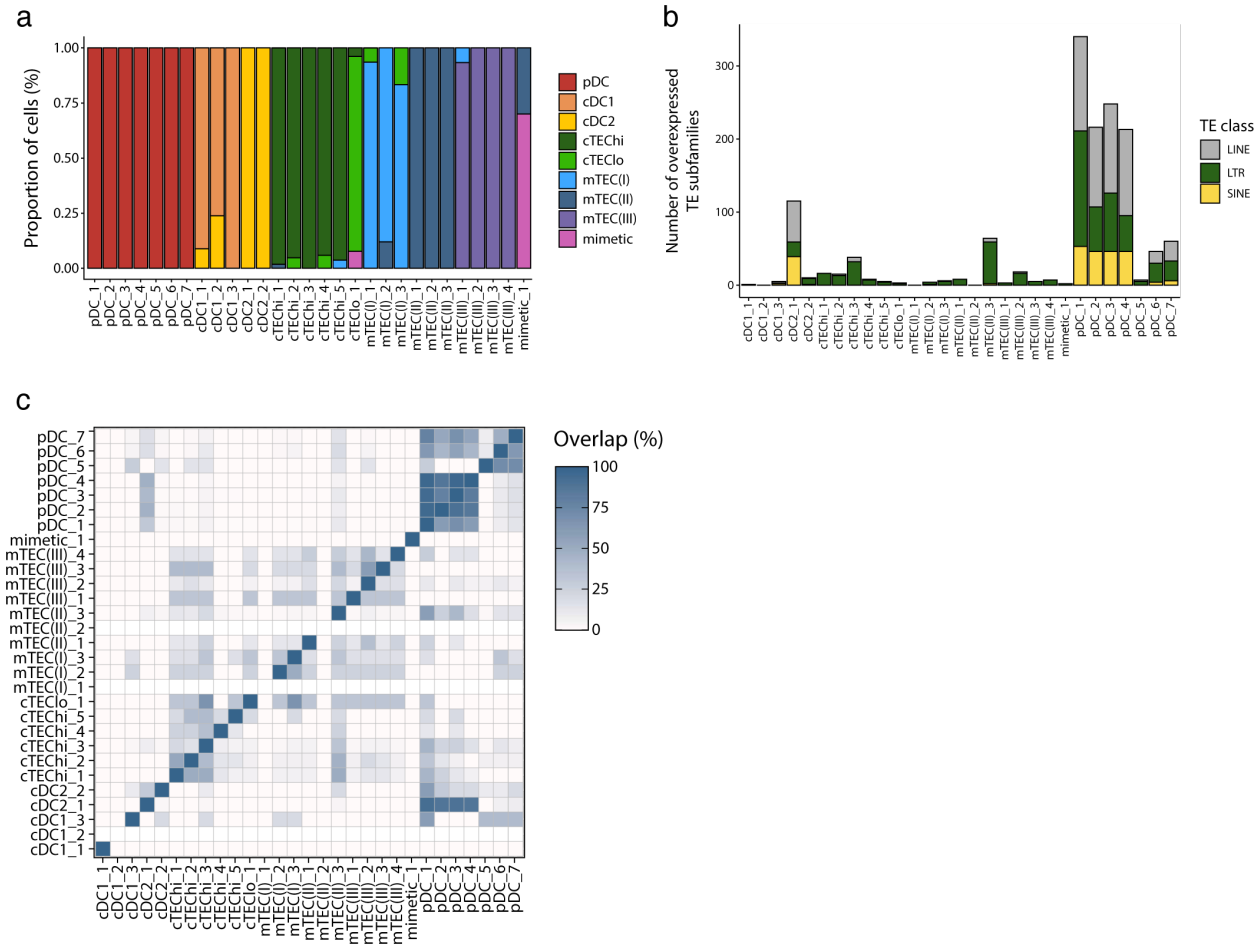
The average expression level of TE subfamilies across cells of the four main populations of thymocytes is shown: DN, DP, CD4 SP, and CD8 SP. Black lines between thymocyte subsets connect expression values for the same TE subfamily.



Supplementary Figure 3.9 - Annotation of human thymic antigen presenting cell subsets.

Dot plot depicting the expression of marker genes in the annotated cell types of the thymus. The average expression and percentage of cells expressing the gene are represented by the color and size of the dot, respectively. Myoid-, myeloid- and neuroendocrine-related genes are used as markers of mimetic mTEC. (aDC, activated dendritic cell; cDC1, conventional dendritic cell 1; cDC2,

conventional dendritic cell 2; cTEC, cortical thymic epithelial cell; mTEC, medullary thymic epithelial cell; pDC, plasmacytoid dendritic cell).



Supplementary Figure 3.10 - Differential TE expression in metacells of human thymic antigen presenting cells.

(a) Cellular composition of the metacells (x-axis) based on the manual annotation of the thymic cell populations (see Fig. S1). **(b)** Number of TE subfamilies overexpressed expressed between the metacells. TE subfamilies are colored based on class (LINE, LTR, and SINE). **(c)** Percentage of overlap of the TE subfamilies overexpressed by each metacell.

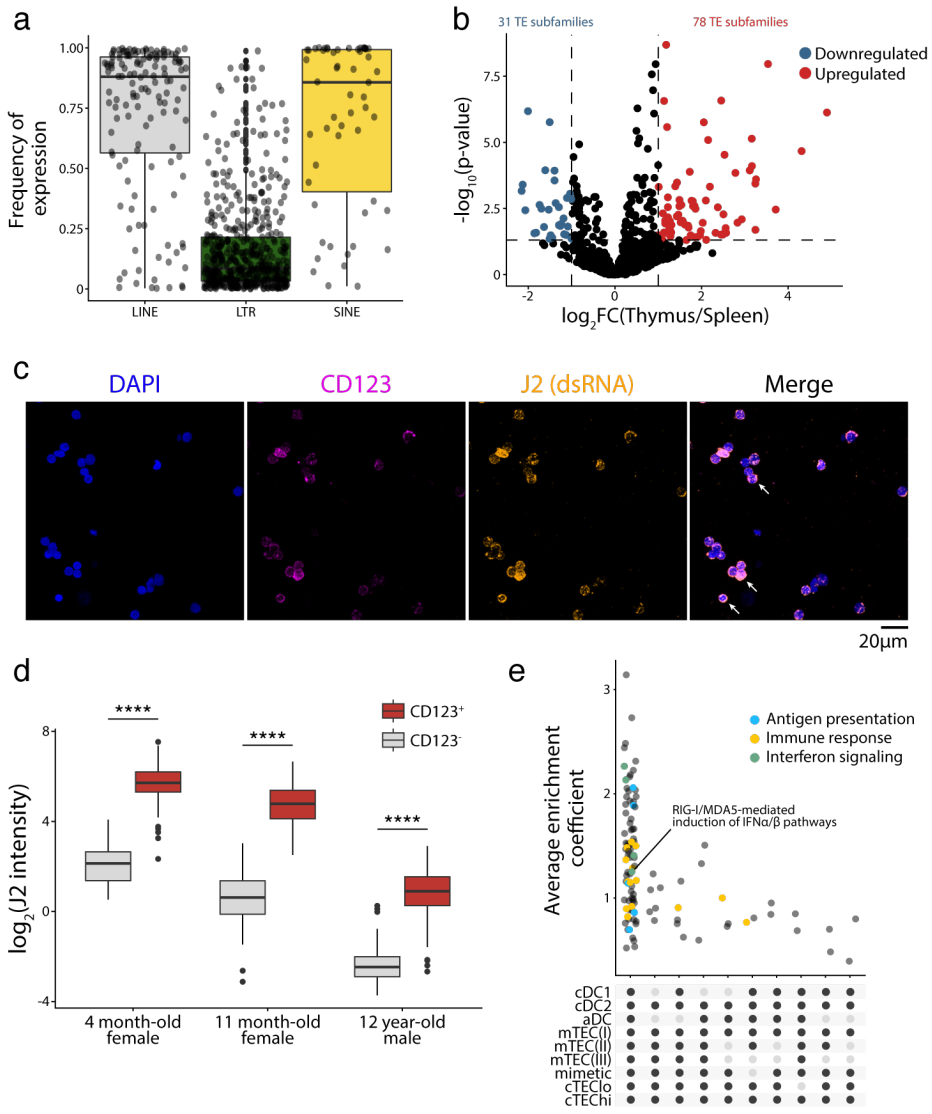
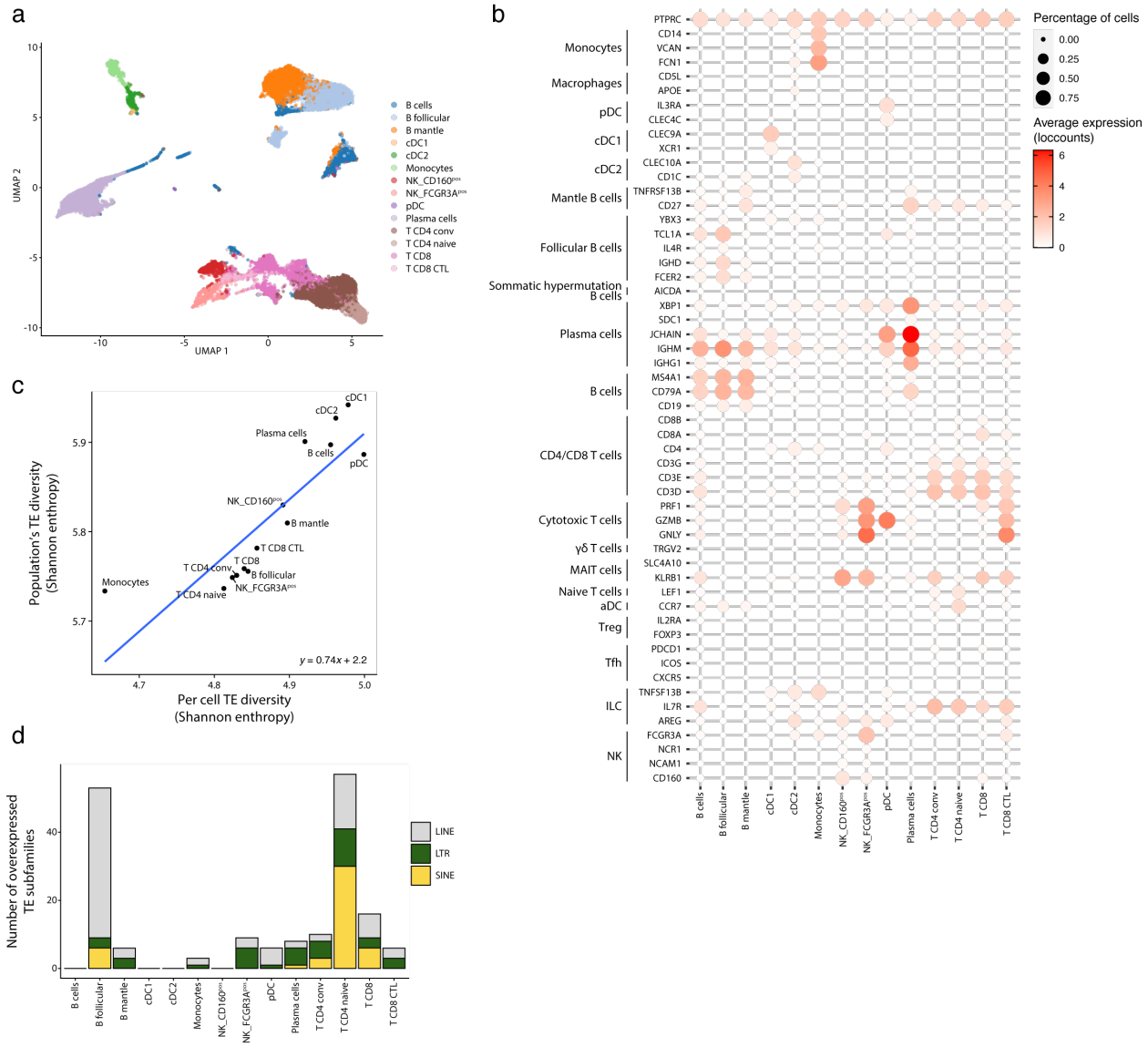


Figure 3.4 - TE expression in human pDCs is associated with dsRNA formation and type I IFN signaling.

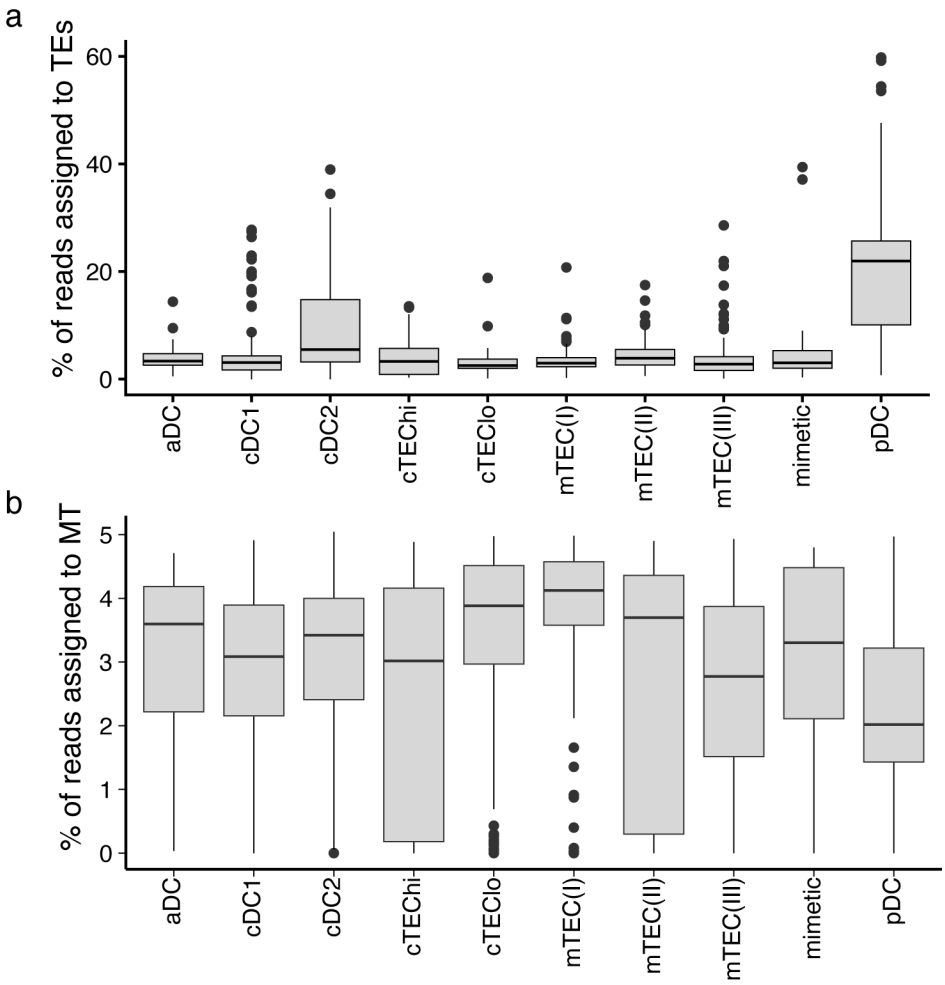
(a) Frequency of LINE, LTR, and SINE subfamilies expression in thymic pDCs. **(b)** Differential expression of TE subfamilies between splenic and thymic pDCs. TE subfamilies significantly upregulated or downregulated by thymic pDCs are indicated in red and blue, respectively (Upregulated, $\log_2$(Thymus/Spleen)≥1 and adj. p≤0.05; Downregulated, $\log_2$(Thymus/Spleen)≤-1 and adj. p≤0.05). **(c,d)** Immunostaining of dsRNAs in human thymic pDCs ($CD123^+$) using the J2 antibody (n=3). **(c)** One representative experiment. Three examples of CD123 and J2 colocalization

are shown with white arrows. **(d)** J2 staining intensity in $CD123^{+}$ and $CD123^{-}$ cells from three human thymi (Wilcoxon Rank Sum test, ****p-value≤0.0001). **(e)** UpSet plot showing gene sets enriched in pDCs compared to the other populations of thymic APCs. On the lower panel, black dots represent cell populations for which gene signatures are significantly depleted compared to pDCs. All comparisons where gene signatures were significantly enriched in pDCs are shown.



Supplementary Figure 3.11 - TE expression in human splenic pDCs.

(a) UMAP depicting the cell populations present in the human spleen. **(b)** Dot plot showing the expression of marker genes in the annotated cell types of the spleen. The average expression and percentage of cells expressing the gene are represented by the color and size of the dot, respectively. **(c)** Diversity of TE expressed by splenic populations measured by Shannon entropy. The x and y axes represent the median diversity of TE expressed by individual cells of a population and the global diversity of TE expressed by discrete populations, respectively. The equation and blue curve represent a linear model summarizing the data. **(d)** Bar plot showing the number (y-axis) and class (color) of differentially expressed TE subfamilies between splenic cell populations.



Supplementary Figure 3.12 - A higher proportion of reads originates from TEs in pDCs than in other thymic APCs.

Boxplots depicting the percentage of reads assigned to **(a)** TE sequences or **(b)** mitochondrial reads in the different subpopulations of thymic APCs.

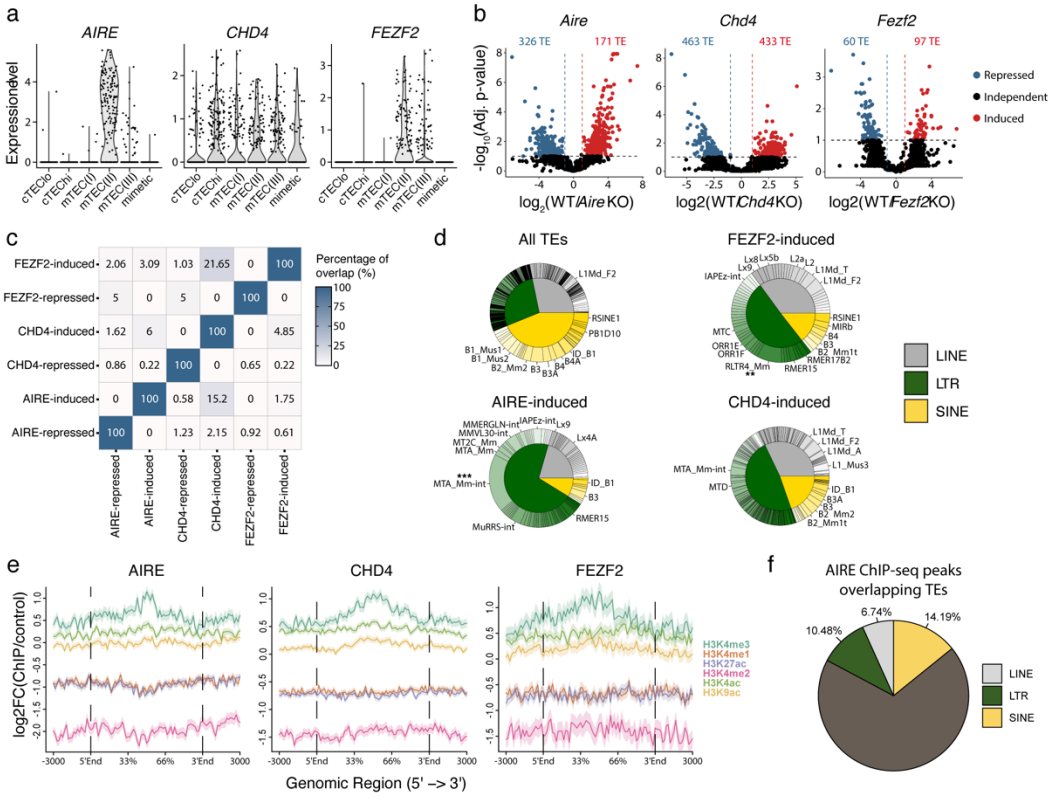
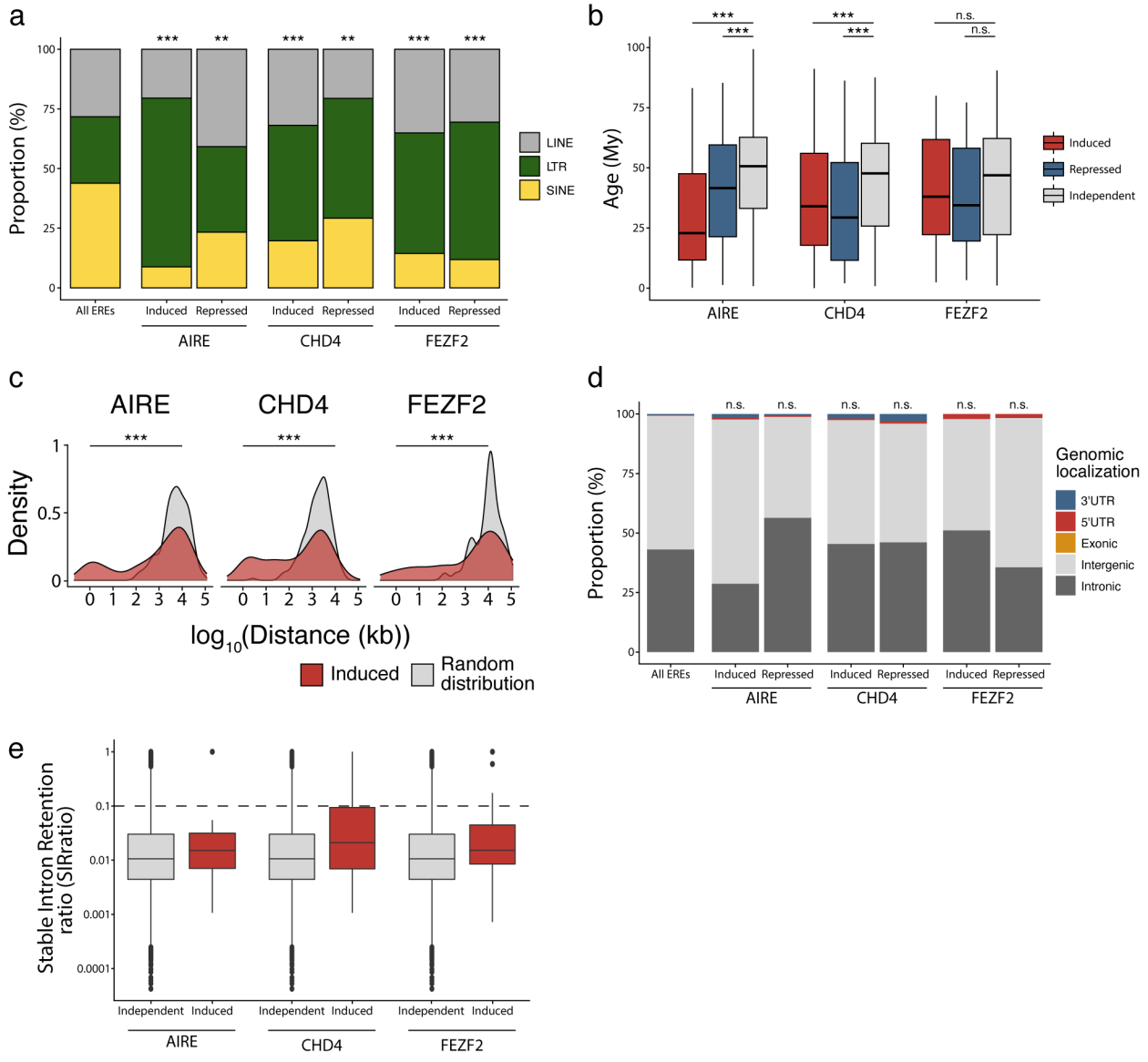


Figure 3.5 - *AIRE, FEZF2,* and *CHD4* regulate non-redundant sets of TEs in murine mTECs.

(a) Expression of *AIRE*, *CHD4*, and *FEZF2* in human TEC subsets. **(b)** Differential expression of TE loci between wild-type (WT) and *Aire*-, *Chd4*- or *Fezf2*-knockout (KO) mice (Induced, $\log_2$(WT/KO)≥2 and adj. p≤0.05; Repressed, $\log_2$(WT/KO)≤-2 and adj. p≤0.05). P-values were corrected for multiple comparisons with the Benjamini-Hochberg procedure. The numbers of induced (red) and repressed (blue) TE loci are indicated on the volcano plots. **(c)** Overlap of TE loci repressed or induced by AIRE, FEZF2, and CHD4. **(d)** Proportion of TE classes and

subfamilies in the TE loci regulated by AIRE, FEZF2, or CHD4, as well as all TE loci in the murine genome for comparison (Chi-squared tests with Bonferroni correction, **adj. p≤0.01, ***adj. p≤0.001). **(e)** Plots for the tag density of H3K4me3 and H3K4me2 on the sequence and flanking regions (3000 base pairs) of TE loci induced by AIRE, FEZF2, and CHD4. **(f)** Proportion of statistically significant peaks overlapping TE sequences in AIRE ChIP-seq data from murine mTECs.



Supplementary Figure 3.13 - Characterization of TE subfamilies regulated by AIRE, CHD4 and FEZF2 in murine mTECs.

(a) Class of TEs induced or repressed by AIRE, CHD4, and FEZF2. Distributions were compared to the proportion of LINEs, LTRs, and SINEs amongst all TE sequences of the murine genome with Chi-squared tests (**p≤0.01, ***p≤0.001). **(b)** Age of TEs induced, repressed, or independent of AIRE, CHD4, and FEZF2 (Wilcoxon-Mann-Whitney test, *p≤0.05, ***p≤0.001) (My, millions of years). **(c)** Distance between TE loci induced by AIRE, FEZF2, and CHD4, and random selections of TE loci (Wilcoxon rank-sum tests, ***p≤0.001). **(d)** Genomic localization of the TE loci induced or repressed by AIRE, CHD4, and FEZF2. **(e)** Intron retention ratio of intronic TE induced or independent of AIRE, CHD4, and FEZF2. The dashed line represents intron retention events occurring in at least 10% of transcripts.

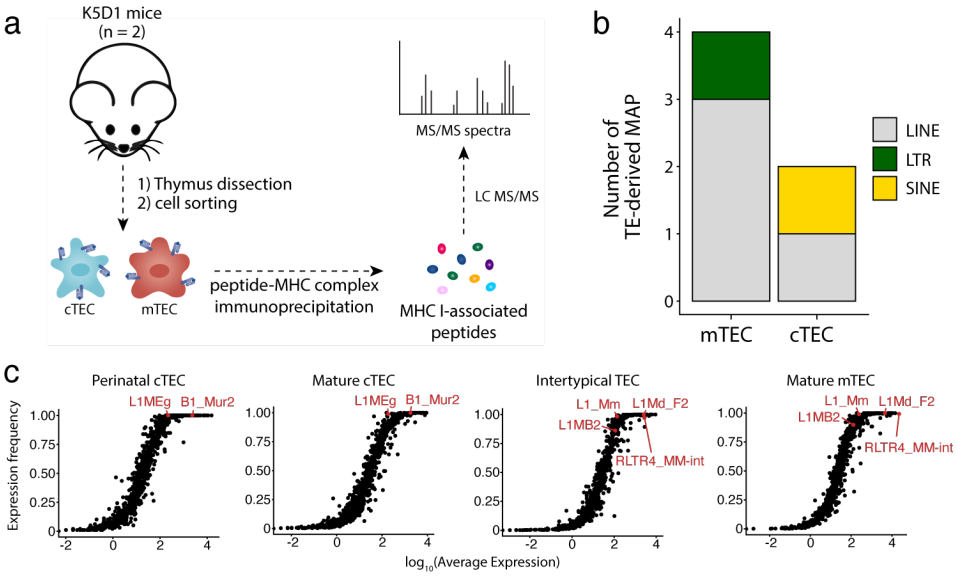


Figure 3.6 - Murine cTECs and mTECs present TE MAPs..

(a) mTECs and cTECs were isolated from the thymi of K5D1 mice (n=2). The peptide-MHC I complexes were immunoprecipitated independently for both populations, and MAPs were sequenced by MS analyses. **(b)** Number of LINE-, LTR-, and SINE-derived MAPs in mTECs and cTECs from K5D1 mice. **(c)** Distributions of TE subfamilies in murine TECs subsets based on expression level (*x-axis*) and frequency of expression (*y-axis*).

Chapitre 4 : Discussion

L'objectif principal de cette thèse était d'approfondir notre connaissance des interactions formées entre les TEs et le système immunitaire adaptatif. En effet, bien que leur rôle lors du développement ainsi que leur expression aberrante dans les maladies auto-immunes et les cancers soient abondamment décrits, leur expression et fonctions dans les tissus somatiques humains à l'état basal demeurent nébuleux. Nos travaux montrent que l'expression des TEs est répandue dans les tissus somatiques humains, mais que leur niveau d'expression varie grandement d'un tissu à l'autre et que plusieurs TEs présentent des profils d'expression tissu-spécifiques. De plus, l'expression des TEs peut mener à la présentation de MAPs dérivés des TEs par des cellules non-cancéreuses, démontrant que ces séquences sont impliquées dans l'immunosurveillance des lymphocytes T.

Parmi le panel de tissus et cellules somatiques étudiés, les mTECs présentaient une expression exceptionnellement haute d'un large répertoire de TEs. Ce résultat suggère que les lymphocytes T apprennent à tolérer les séquences des TEs lors de leur développement dans le thymus. Une analyse plus approfondie de l'expression des TEs dans le thymus a montré que ces séquences jouent trois rôles potentiels dans le développement des lymphocytes T : i) en fournissant des sites de liaison à des facteurs de transcription dans toutes les populations thymiques étudiées, ii) en favorisant la génération de Tregs via la sécrétion d'IFNα/β par les pDCs thymiques, et iii) en contribuant aux sélections positive et négative des thymocytes.

Bien que nos résultats suggèrent qu'il y ait induction de la tolérance immunitaire envers les TEs, cette tolérance semble partielle puisqu'il existe de nombreux exemples d'activation des lymphocytes T contre des TEs dans des pathologies. Cette discussion tentera donc de réconcilier ces deux observations en abordant pourquoi il est important qu'une forme de tolérance soit mise en place à l'endroit les TEs, et d'évaluer l'étendue des TEs tolérés par le système immunitaire.

4.1 Revisiter le dogme central de l'expression des TEs

4.1.1 L'expression des TEs dans les tissus somatiques humains est complexe.

Lorsque nous avons débuté les travaux de cette thèse de doctorat, relativement peu de choses étaient connues de l'expression des TEs dans les tissus somatiques. Le dogme central généralement accepté stipulait que l'expression des TEs, forte dans les cellules souches embryonnaires (212, 213, 225) et dans les cellules de la lignée germinale (295, 509), diminuait rapidement lors de la différenciation cellulaire, à l'exception de certains contextes précis comme lors du développement du cerveau (**Figure 4.1**) (221, 222, 231, 308, 510).

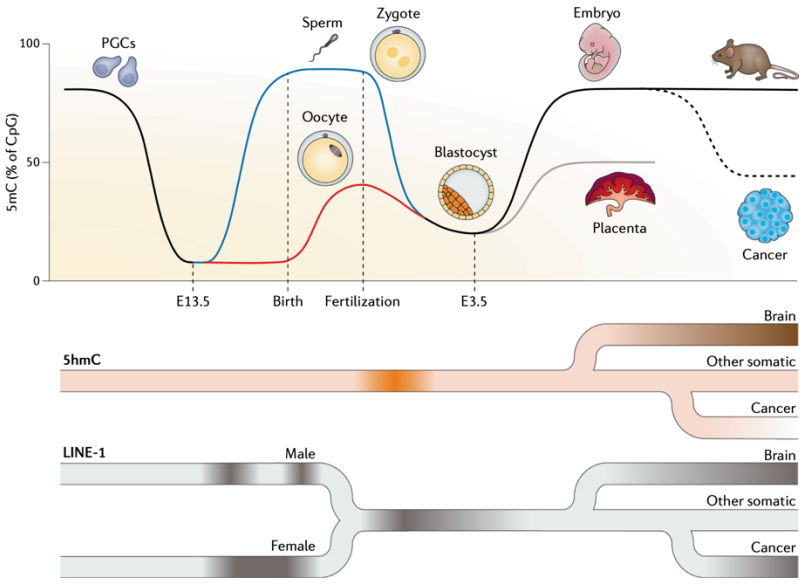


Figure 4.1 – Expression des éléments L1 au cours du développement.

Dynamique de l'expression des éléments L1 (en gris) et de la méthylation de l'ADN (5hmC) lors du développement de la souris. Reproduit avec la permission de *Springer Nature* : Nature Reviews Genetics (Deniz O, Frost JM & Branco MR) ©2019 (377).

Nos travaux viennent apporter d'importantes nuances à ce dogme. De fait, nos données montrent que l'expression des TEs est répandue dans les tissus somatiques humains, bien que leur niveau d'expression varie grandement d'un tissu à l'autre. De façon intéressante, nos résultats montrent aussi que 124/809 sous-familles TEs ont une expression tissu-spécifique, illustrant que la variabilité d'expression des TEs entre les tissus somatiques s'applique aussi au répertoire de séquences TEs exprimées. Puisque 17/32 tissus et types cellulaires que nous avons étudiés exprimaient des sous-familles TEs de façon spécifique, nos travaux montrent que la plupart des

tissus somatiques possèdent une signature d'expression des TEs et suggère que les TEs pourraient être impliqués dans le développement ou la fonction de plusieurs organes. Nos résultats pourraient toutefois grandement sous-estimer la complexité de l'expression des TEs dans les tissus périphériques puisque les données transcriptomiques utilisées avaient été générées à partir d'organes entiers. En effet, nos analyses réalisées sur des données de scRNA-seq ont montré qu'il existe une grande hétérogénéité dans l'expression des TEs entre les différentes populations cellulaires du thymus. Ces observations sont cohérentes avec une étude indépendante qui a confirmé que l'expression des TEs est répandue dans les tissus somatiques humains et présente des profils d'expression tissu-spécifiques, supportant nos conclusions (511). De plus, des études ont montré que les TEs régulent fréquemment l'expression génique en agissant comme des promoteurs ou *enhancers* (512) et que différentes sous-familles TEs sont enrichies dans des régions de chromatine active de façon tissu-spécifique (214). Nos travaux supportent cette observation en montrant que les TEs fournissent des promoteurs et des *enhancers* à un grand nombre de facteurs de transcription dans toutes les populations cellulaires du thymus, dont plusieurs facteurs de transcription importants du développement et de la fonction des cellules thymiques. Ainsi, l'expression des TEs dans les tissus somatiques humains est plus répandue et complexe qu'anticipée, puisque tous les tissus somatiques présentent un certain niveau d'expression des TEs et que plusieurs TEs sont exprimés de façon tissu-spécifique. Ceci suggère que les TEs pourraient être impliqués dans le développement et l'homéostasie des tissus somatiques.

Des études subséquentes seraient donc nécessaires pour clarifier les fonctions des TEs dans les tissus somatiques. Premièrement, il serait intéressant de réanalyser l'expression des TEs dans les tissus somatiques en utilisant la technologie de séquençage Nanopore (513), qui permet d'obtenir des reads de plusieurs milliers de paires de bases, ce qui faciliterait grandement l'alignement des *reads* sur les séquences des éléments transposables et permettrait d'analyser spécifiquement l'expression de chaque locus TE. De plus, il serait important d'utiliser des données de scRNA-seq pour déterminer avec précision dans quelles cellules d'un tissu les TEs sont exprimés. Lorsque des loci TE exprimés spécifiquement par certaines populations cellulaires auraient été identifiés, nous

pourrions utiliser le système CRISPR-Cas9 ainsi que des traitements avec des ARN interférants et des oligonucléotides antisens (ASO) (contenant les modifications chimiques phosphorothioate et 2'-O-Methoxyethyl sur la totalité de la séquence afin de stabiliser l'oligonucléotide et prévenir la liaison de la RNase H (514)) pour déterminer si ces TEs agissent au niveau de l'ADN, des ARNs ou des protéines. Ces expériences pourraient être réalisées sur des lignées cellulaires lorsque disponibles, ou sur des organoïdes qui permettraient aussi d'évaluer le rôle des TEs dans le développement de l'organe et sur l'ensemble des populations cellulaires du tissu.

4.1.2 Les TEs contribuent à l'immunopeptidome à l'état basal.

Une autre avancée importante de cette thèse est l'élaboration d'une approche protéogénomique permettant l'identification de MAPs dérivés des TEs, qui nous a permis de démontrer que l'expression des TEs mène à la présentation de peptides par le CMH-I à la surface de cellules non-cancéreuses. Puisque l'immunopeptidome donne une représentation des ARN exprimés à l'intérieur de la cellule (47), ceci suggère que la présentation de MAPs dérivés des TEs est fréquente dans les tissus somatiques humains. De plus, puisqu'il a été observé que les signatures géniques tissu-spécifiques sont aussi présentes au sein de l'immunopeptidome (460, 515), les sous-familles TEs tissu-spécifiques que nous avons identifiées devraient aussi générer des MAPs dans leurs tissus respectifs. Il est toutefois difficile d'estimer la proportion de l'immunopeptidome provenant des TEs. En effet, puisque les TEs représentent près de 50% du génome humain (374), les bases de données utilisées pour l'identification des MAPs séquencés en spectrométrie de masse étaient volumineuses. Or, il est préférable que les bases de données de spectrométries de masse soient aussi petites que possible pour augmenter la quantité des identifications (516, 517). Dans le futur, l'utilisation de technologies de séquençage Nanopore réduirait grandement la redondance des séquences contenues dans nos bases de données, ce qui permettrait d'augmenter le nombre de MAPs TEs identifiés par échantillon, et ainsi de mieux évaluer leur contribution totale à l'immunopeptidome. De plus, le séquençage Nanopore permettrait d'évaluer la présentation de MAPs produits par des transcripts chimériques entre des TEs et des gènes canoniques à l'état basal (329). Nos résultats approfondissent néanmoins notre compréhension de la composition de l'immunopeptidome à l'état basal, et suggèrent que des MAPs TEs pourraient être retrouvés à la surface de plusieurs organes périphériques.

4.1.3 L'expression des TEs est-elle soumise à une pression immunitaire?

La confirmation que les TEs contribuent à l'immunopeptidome des cellules non-cancéreuses (**Figure 2.3**) suggère que des MAPs dérivés des TEs sont présentés par les tissus somatiques humains. Une observation intéressante peut être faite à partir de notre analyse de l'expression des TEs dans les tissus somatiques humains (**Figure 2.1**): à l'exception des mTECs, l'expression des TEs est particulièrement élevée dans deux types cellulaires et tissus immunoprivilégiés, soit les ESCs et les testicules. En effet, les ESCs expriment faiblement le CMH ainsi que la machinerie de présentation antigénique (405, 518), alors que les spermatozoïdes n'expriment pas le CMH-I et sont protégés du système immunitaire par la barrière hémato-testiculaire (519). Plusieurs études ont aussi montré que des éléments L1 sont actifs dans le cerveau (221, 222, 311), un autre organe immunoprivilégié (520). Ceci suggère que l'expression des TEs doit être réprimée dans les cellules exposées à la surveillance des lymphocytes T CD8 pour éviter des maladies auto-immunes. De fait, la surexpression des TEs a fréquemment été associée aux maladies auto-immunes. Premièrement, l'expression aberrante des TEs a été associée à un état inflammatoire en stimulant la sécrétion d'interféron (413, 521-523). Des auto-anticorps spécifiques aux TEs ont aussi été détectés dans le sérum de patients atteints de troubles rhumatismaux auto-immuns (par exemple, le lupus ou le polyarthrite rhumatoïde) ou de diabète de type 1 (524). De plus, il a été observé que des protéines encodées par HERV-W et HERV-Fc1 (deux éléments à LTRs) étaient surexprimées dans des échantillons de cerveaux provenant de patients atteints de sclérose en plaques (525, 526). Ainsi, nos données suggèrent que l'expression des TEs dans les tissus somatiques est plus strictement régulée dans les tissus exposés à la surveillance immunitaire pour éviter des réactions auto-immunes sévères.

Dans le futur, il serait intéressant d'utiliser notre approche protéogénomique pour comparer la composition de l'immunopeptidome de tissus provenant d'individus sains ou atteints de maladies auto-immunes pour déterminer si certains MAPs dérivés des TEs sont associés à l'auto-immunité. Ce faisant, nous pourrions déterminer si les TEs tissus-spécifiques que nous avons identifiés dans nos analyses transcriptomiques pourraient être liés au développement de maladies auto-immunes spécifiques à certains tissus.

4.2 Tolérance à l'endroit des TEs

4.2.1 Mécanismes d'induction de la tolérance médiés par les TEs.

Nos travaux illustrent deux implications potentielles des TEs dans l'établissement des tolérances centrale et périphérique. Premièrement, nos analyses immunopeptidomiques ont montré que des MAPs dérivés des TEs sont présentés par les mTECs murines, suggérant une implication des TEs dans la sélection négative des lymphocytes T. L'identification d'un faible nombre de MAPs dérivés des TEs sur les mTECs murines (**Figure 3.6b**) peut paraitre surprenante compte tenu de l'expression forte des TEs que nous avons détectée dans les mTECs humaines (**Figure 2.1**). Toutefois, la contribution des TEs à la sélection négative est probablement plus grande que ce que nous avons observé pour deux raisons : i) la spectrométrie de masse identifie les MAPs les plus abondants, et ii) l'expression des TEs dans les mTECs est fréquemment faible et/ou restreinte à un faible pourcentage de cellules (**Figures 3.3e, 3.6c**). La fonction principale des mTECs étant d'établir la tolérance au soi via la présentation antigénique, nos données suggèrent néanmoins fortement une implication des TEs dans l'établissement de la tolérance centrale, bien que l'étendue de cette implication reste nébuleuse.

Dans le futur, il serait important de vérifier que la présentation de MAPs TEs par les mTECs mène à la sélection négative des lymphocytes T. Pour ce faire, nous pourrions quantifier l'abondance de TCRs spécifiques aux MAPs TEs que nous avons identifiés en réalisant des marquages par tétramères sur les populations de thymocytes DP et SP, c'est-à-dire les thymocytes pré et post sélection négative. Nous pourrions ainsi mesurer la délétion des TCRs spécifiques aux MAPs TEs lors de la sélection négative. De plus, puisqu'il existe des différences importantes entre l'humain et la souris quant à la nature et à l'activité des TEs, il serait intéressant de valider que les TEs sont aussi traduits et présentés par le CMH-I à la surface des mTECs humaines. Le nombre de mTECs pouvant être isolées d'un échantillon de thymus humain ne permettant pas pour l'instant de procéder à des analyses immunopeptidomiques en spectrométrie de masse, il serait possible de réaliser des expériences de séquençage des polysomes pour déterminer si les TEs sont activement traduits chez l'humain (527). Ces expériences ne confirmeraient pas la présentation de MAPs dérivés des TEs par les mTECs humaines, mais puisque la fonction principale des mTECs est

d'induire la tolérance au soi il est possible d'assumer que la traduction d'ARNs dans ces cellules mène à la présentation antigénique.

Deuxièmement, nos travaux suggèrent que l'expression des TEs dans le thymus contribue aussi à la mise en place de la tolérance périphérique en induisant la génération de Tregs. En effet, nos données supportent un modèle selon lequel la sécrétion constitutive d'IFNα/β par les pDCs thymiques, qui stimule la différenciation des thymocytes en Tregs, est causée par la détection de dsRNAs dérivés des TEs par RIG-I et MDA5 (**Figure 3.4**). En stimulant la production de cellules immunosuppressives, l'expression des TEs dans le thymus serait cruciale pour éviter l'activation de lymphocytes T ayant échappé à la sélection négative contre des antigènes du soi présentés dans les tissus périphériques. Dans le futur, il serait intéressant de réaliser certaines validations pour renforcer notre modèle. Premièrement, des expériences d'hybridation *in situ* nous permettraient de confirmer hors de tout doute que les dsRNAs détectés dans les pDCs thymiques sont générés par des TEs. Finalement, l'utilisation d'un modèle d'organoïdes thymiques humains récemment développé (528, 529) permettrait de dépléter RIG-I et MDA5 spécifiquement dans les pDCs pour valider leur implication dans la sécrétion d'IFNα/β. Nos travaux soulignent tout de même deux mécanismes distincts par lesquels les TEs contribuent à l'établissement de la tolérance au soi : i) en éliminant les thymocytes spécifiques aux MAPs TEs présentés dans la médulla thymique, et ii) en stimulant la production de cellules immunosupressives (Tregs).

4.2.2 Expression des TEs dans le contexte de la PGE

Une caractéristique unique des mTECs est la PGE, c'est-à-dire l'expression d'un répertoire de gènes exceptionnellement diversifié incluant des gènes tissu-spécifiques. Nos analyses transcriptomiques ont montré que l'expression des TEs est beaucoup plus forte dans les mTECs que dans les autres tissus somatiques humains (**Figure 2.1**), un résultat qui a par la suite été confirmé par une étude indépendante (441). De plus, les auteurs de cette publication avait observé que l'expression des TEs est plus élevée dans les mTECs matures (mTECs CMH-II^{hi}, soit des mTEC(II)) que dans les mTECs immatures (mTECs CMH-II^{lo}, c'est-à-dire majoritairement des mTEC(I)), un résultat cohérent avec nos analyses d'expression différentielle des TEs en scRNA-seq (**Figure 3.4d**).

Puisque la PGE et l'expression forte des TEs sont toutes deux spécifiques aux mTEC(II), nous nous sommes demandé si ces deux phénomènes sont liés. Nous avons donc évalué l'impact des trois facteurs de transcription essentiels à la PGE (AIRE, FEZF2 et CHD4) sur l'expression des TEs dans les mTECs à l'aide de données transcriptomiques de souris pour lesquelles AIRE, FEZF2 ou CHD4 avaient été déplétés. Nos analyses ont montré que ces trois facteurs de transcription régulent l'expression des groupes de TEs non-redondants (**Figure 3.5, b et c**). De façon intéressante, nos données montrent que l'impact de ces trois facteurs de transcription sur l'expression des TEs peut être classé selon la hiérarchie suivante : CHD4 > AIRE > FEZF2. Puisque CHD4 est un régulateur épigénétique et qu'AIRE interagit avec des marques d'histone répressives, nos travaux suggèrent que la PGE pourrait permettre de contourner la répression épigénétique des TEs dans les mTECs pour permettre l'établissement de la tolérance au soi à l'égard des TEs.

À prime abord ces résultats pourraient sembler contradictoires avec ceux de notre première étude, où nous avions observé qu'AIRE ne régulait pas l'expression des TEs dans les mTECs (**Figure 2.3**). Il est toutefois important de noter que ces premières analyses étaient réalisées sur les sous-familles TEs entières, ce qui pourrait avoir masqué le signal de la régulation de certains loci spécifiques par AIRE. Ainsi, nos travaux montrent que l'expression particulièrement forte des TEs dans les mTEC(II) est associée à la PGE, supportant notre modèle selon lequel les lymphocytes T pourraient apprendre à tolérer un grand nombre de TEs.

4.2.3 La tolérance à l'endroit des TEs est-elle partielle?

L'idée que les TEs soient impliqués dans l'établissement de la tolérance au soi peut paraitre contradictoire avec les nombreuses évidences que les TEs soient impliqués dans les maladies auto-immunes. Il est toutefois connu que la tolérance centrale n'est pas parfaite et que des lymphocytes T auto-réactifs, principalement ceux ayant une affinité intermédiaire pour le soi, peuvent survivre à la sélection négative (530, 531). Dans la périphérie, la tolérance à l'endroit des MAPs dérivés des TEs serait alors maintenue par les Tregs, et des changements dans l'abondance des MAPs dérivés des TEs lorsque leur expression est dérégulée pourraient être à l'origine de maladies auto-immunes (532, 533).

Un autre facteur qui pourrait expliquer que les MAPs TEs sont fréquemment associés à des réponses immunitaires est leur homologie avec des antigènes viraux. En effet, il a été montré que les antigènes tumoraux présentant des similarités de séquences avec des pathogènes, ou autrement dit avec le non-soi, avaient une plus grande immunogénicité (534, 535). Des lymphocytes T spécifiques aux antigènes viraux pourraient ainsi réagir à la présentation de MAPs TEs par cross-réactivité (536, 537). La nature même des TEs et leurs similitudes avec le non-soi pourraient donc prévenir l'établissement complet de la tolérance à leur égard, malgré la présentation de MAPs TEs par les mTECs.

4.3 Implications de nos découvertes pour la mise en place d'immunothérapies du cancer

De nombreuses études ont identifié les TEs comme des cibles prometteuses pour des immunothérapies du cancer (538-541). À ce jour, aucune donnée clinique n'est toutefois disponible pour confirmer le potentiel thérapeutique des TEs, bien qu'un essai clinique basé sur l'injection de lymphocytes T exprimant un TCR transgénique spécifique à un épitope du HERV-E à des patients atteints de cancer du rein soit présentement en cours (identifiant sur le site ClinicalTrials.gov : NCT03354390). De nombreuses équipes de recherche tentent actuellement d'identifier des MAPs TEs à la surface de cellules cancéreuses, ce qui permettrait de mettre en place des immunothérapies basées sur les lymphocytes T. Les travaux de cette thèse soulèvent deux préoccupations importantes que les équipes de recherche devraient avoir en tête lors de l'identification de MAPs TEs sur des échantillons cancéreux : s'assurer que les MAPs TEs identifiés i) sont bel et bien tumeur-spécifiques, et ii) ne sont pas exprimés dans le thymus pour assurer leur immunogénicité.

4.3.1 Identifier des cibles tumeur-spécifiques pour éviter de la toxicité pour le patient

Plusieurs études visant à identifier des cibles d'immunothérapies comparent l'expression des TEs entre la tumeur et le tissu normal associé, une stratégie permettant de déterminer quels TEs sont surexprimés par les cellules de la tumeur (314, 323, 337, 542). Bien que de telles méthodes

permettent d'évaluer l'impact des changements métaboliques et épigénétiques survenant lors de la tumorigenèse sur l'expression des TEs, elles sont sous-optimales pour identifier des cibles thérapeutiques. En effet, bien que l'expression des MAPs TEs soient plus élevées dans la tumeur que dans le tissu normal associé, rien ne garantit qu'elle soit tumeur-spécifique. Ce type d'approches mène régulièrement à l'identification d'antigènes associés aux tumeurs (TAA), soit des antigènes surexprimés par la tumeur mais tout de même exprimés par les cellules normales. Or, les TAAs provenant de gènes canoniques ayant été testés en essais cliniques ont démontré une faible immunogénicité (543, 544); les TAAs générés par des TEs identifiés par de telles approches entraineraient donc probablement de faibles réponses anti-tumorales.

Une solution pour favoriser l'identification d'antigènes tumeur-spécifiques (TSA) est de quantifier l'expression des TEs générant ces antigènes dans l'ensemble des tissus somatiques humains. L'augmentation exponentielle de la quantité de données transcriptomiques ainsi que l'avènement de bases de données transcriptomiques telles GTEx (*Genotype-Tissue expression*) facilite grandement ce type de validation. Puisque nos travaux montrent que plusieurs sous-familles TEs sont exprimées de façon tissu-spécifiques, quantifier l'expression des transcrits codant pour les MAPs TEs dans l'ensemble des tissus somatiques humains permettrait de valider que les antigènes TEs identifiés sur des tumeurs sont des TSAs. L'avantage de cibler des TSAs dans le contexte d'immunothérapies est de limiter la toxicité pour le patient, qui peut entrainer des effets secondaires importants ou même la mort lorsque l'antigène ciblé est exprimé par d'autres cellules du corps (545-548). Finalement, les TSAs ont une plus faible probabilité d'être tolérés par le système immunitaire, puisque par définition ils ne sont pas exprimés par les cellules normales et ne font donc pas parti du soi immunitaire (549). Ainsi, assurer que les antigènes TEs identifiés sur des tumeurs sont véritablement tumeur-spécifiques améliorerait la spécificité et l'efficacité des immunothérapies ciblant ces antigènes.

4.3.2 Vérifier l'expression de la cible dans le thymus pour assurer son immunogénicité

Une autre découverte majeure de nos travaux de recherche est que les TEs sont exprimés à des niveaux élevés par les mTECs humaines, et que leur expression peut mener à la présentation d'antigènes TEs par le CMH-I chez la souris. Ces résultats suggèrent que les lymphocytes T pourraient apprendre à tolérer un vaste répertoire de séquences TEs lors de leur développement dans le thymus, tant chez la souris que chez l'humain. Ces données ont des implications importantes pour le développement d'immunothérapies ciblant des TEs, puisque les TEs générant des MAPs présentés par les mTECs appartiennent au soi immunitaire. L'expression forte d'un large répertoire de TEs par les mTECs humaines implique qu'un nombre important d'antigènes TEs seraient tolérés par les lymphocytes T. Ceci pourrait expliquer pourquoi plusieurs MAPs TEs identifiés sur des tumeurs ont démontré une faible immunogénicité (317, 420), mais les auteurs de ces études n'ont pas quantifié l'expression des transcrits codant pour leurs antigènes dans les mTECs.

Nos travaux soulignent l'importance d'inclure une étape de quantification de l'expression des transcrits codant pour les antigènes TEs dans les mTECs lors de l'identification de cibles d'immunothérapies. En effet, nos analyses ont montré que deux antigènes TEs identifiés sur des échantillons de leucémies, qui ne sont ni exprimés par les mTECs ni par les tissus somatiques humains, induisent des réponses des lymphocytes T CD8, confirmant leur immunogénicité (**Supplementary Figure 2.7,B et C**). Ces résultats sont cohérents avec ceux de deux autres études ayant démontré que des antigènes TEs tumeur-spécifiques, et n'étant pas exprimés par les mTECs, sont immunogènes (352, 367). Quantifier l'expression des antigènes TEs dans les mTECs réduirait le nombre de cibles identifiées pour des immunothérapies du cancer, mais augmenterait les probabilités que ces cibles soient immunogènes.

4.4 Conclusion

Nos travaux de recherche soulignent la complexité des interactions entre les TEs et notre système immunitaire et illustrent l'importance d'établir la tolérance des lymphocytes T à l'endroit de ces séquences « non-codantes » du génome. De plus, cette thèse apporte des nuances importantes au dogme actuel stipulant que l'expression des TEs dans les tissus somatiques est typiquement réprimée dans les tissus somatiques.

De fait, nos analyses ont montré que l'expression des TEs est omniprésente dans les tissus somatiques humains, que plusieurs TEs sont exprimés de façon tissu-spécifique, et que cette expression peut mener à la présentation d'antigènes par le CMH I à la surface de cellules non-cancéreuses. Ces observations sont importantes, car elles suggèrent que des MAPs TEs pourraient être présentés par l'ensemble des tissus somatiques humains. Ces observations supportent un nouveau paradigme selon lequel l'établissement de la tolérance centrale à l'égard des TEs serait crucial pour éviter une auto-immunité généralisée.

Dans le même ordre d'idée, l'expression forte par les mTECs d'un large répertoire de TEs suggère l'implication de ces séquences lors de l'éducation des lymphocytes T dans le thymus, une hypothèse supportée par l'identification de MAPs TEs présentés par les mTECs murines. Nos analyses ont aussi montré que les TEs ont des effets pléiotropiques sur le développement des cellules immunitaires du thymus en contribuant à la régulation génique, en stimulant un microenvironnement pro-inflammatoire supportant la maturation des thymocytes, et en participant aux sélections positive et négative.

À prime abord, de telles fonctions des TEs dans le développement et la fonction du système immunitaire peuvent paraitre contradictoires, puisque les TEs sont à l'origine des envahisseurs du génome de leur hôte. Favoriser ainsi la survie de l'hôte est toutefois une méthode ingénieuse pour les TEs d'assurer leur propre pérennité. De tels exemples d'exaptation des TEs représenteraient alors l'atteinte d'un armistice dans la course à l'armement entre le génome de l'hôte et ses parasites génétiques.

Références bibliographiques

1. Miller JF. Immunological function of the thymus. Lancet. 1961;2(7205):748-9.

2. Silverstein AM, Rose NR. On the mystique of the immunological self. Immunol Rev. 1997;159:197-206; discussion 7-18.

3. Guillet JG, Lai MZ, Briner TJ, Buus S, Sette A, Grey HM, et al. Immunological self, nonself discrimination. Science. 1987;235(4791):865-70.

4. Isaacson PG, Norton AJ, Addis BJ. The human thymus contains a novel population of B lymphocytes. Lancet. 1987;2(8574):1488-91.

5. Brocker T. Survival of mature CD4 T lymphocytes is dependent on major histocompatibility complex class II-expressing dendritic cells. J Exp Med. 1997;186(8):1223-32.

6. Vandenabeele S, Hochrein H, Mavaddat N, Winkel K, Shortman K. Human thymus contains 2 distinct dendritic cell populations. Blood. 2001;97(6):1733-41.

7. Hadeiba H, Lahl K, Edalati A, Oderup C, Habtezion A, Pachynski R, et al. Plasmacytoid dendritic cells transport peripheral antigens to the thymus to promote central tolerance. Immunity. 2012;36(3):438-50.

8. Surh CD, Sprent J. T-cell apoptosis detected in situ during positive and negative selection in the thymus. Nature. 1994;372(6501):100-3.

9. Zhou TA, Hsu HP, Tu YH, Cheng HK, Lin CY, Chen NJ, et al. Thymic macrophages consist of two populations with distinct localization and origin. Elife. 2022;11.

10. Gray DH, Tull D, Ueno T, Seach N, Classon BJ, Chidgey A, et al. A unique thymic fibroblast population revealed by the monoclonal antibody MTS-15. J Immunol. 2007;178(8):4956-65.

11. Muller SM, Terszowski G, Blum C, Haller C, Anquez V, Kuschert S, et al. Gene targeting of VEGF-A in thymus epithelium disrupts thymus blood vessel architecture. Proc Natl Acad Sci U S A. 2005;102(30):10587-92.

12. Raviola E, Karnovsky MJ. Evidence for a blood-thymus barrier using electron-opaque tracers. J Exp Med. 1972;136(3):466-98.

13. von Gaudecker B. Functional histology of the human thymus. Anat Embryol (Berl). 1991;183(1):1-15.

14. Porritt HE, Gordon K, Petrie HT. Kinetics of steady-state differentiation and mapping of intrathymic-signaling environments by stem cell transplantation in nonirradiated mice. J Exp Med. 2003;198(6):957-62.

15. Kurobe H, Liu C, Ueno T, Saito F, Ohigashi I, Seach N, et al. CCR7-dependent cortex-to-medulla migration of positively selected thymocytes is essential for establishing central tolerance. Immunity. 2006;24(2):165-77.

16. McCaughtry TM, Wilken MS, Hogquist KA. Thymic emigration revisited. J Exp Med. 2007;204(11):2513-20.

17. Wu L, Antica M, Johnson GR, Scollay R, Shortman K. Developmental potential of the earliest precursor cells from the adult mouse thymus. J Exp Med. 1991;174(6):1617-27.

18. Lind EF, Prockop SE, Porritt HE, Petrie HT. Mapping precursor movement through the postnatal thymus reveals specific microenvironments supporting defined stages of early lymphoid development. J Exp Med. 2001;194(2):127-34.

19. Dudley EC, Petrie HT, Shah LM, Owen MJ, Hayday AC. T cell receptor beta chain gene rearrangement and selection during thymocyte development in adult mice. Immunity. 1994;1(2):83-93.

20. Baldwin TA, Sandau MM, Jameson SC, Hogquist KA. The timing of TCR alpha expression critically influences T cell development and selection. J Exp Med. 2005;202(1):111-21.

21. Klein L, Kyewski B, Allen PM, Hogquist KA. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). Nat Rev Immunol. 2014;14(6):377-91.

22. Matsuyama M, Wiadrowski MN, Metcalf D. Autoradiographic analysis of lymphopoiesis and lymphocyte migration in mice bearing multiple thymus grafts. J Exp Med. 1966;123(3):559-76.

23. Scollay RG, Butcher EC, Weissman IL. Thymus cell migration. Quantitative aspects of cellular traffic from the thymus to the periphery in mice. Eur J Immunol. 1980;10(3):210-8.

24. Miller JF. The golden anniversary of the thymus. Nat Rev Immunol. 2011;11(7):489-95.

25. Bevan MJ. In a radiation chimaera, host H-2 antigens determine immune responsiveness of donor cytotoxic cells. Nature. 1977;269(5627):417-8.

26. Dausset J. [Iso-leuko-antibodies]. Acta Haematol. 1958;20(1-4):156-66.

27. Kissmeyer-Nielsen F, Svejgaard A, Hauge M. Genetics of the human HL-A transplantation system. Nature. 1968;219(5159):1116-9.

28. Thorsby E, Sandberg L, Lindholm A, Kissmeyer-Nielsen F. The HL-A system: evidence of a third sub-locus. Scand J Haematol. 1970;7(3):195-200.

29. Hirschberg H, Kaakinen A, Thorsby E. Presence of HLA-D determinants on human macrophages. Nature. 1976;263(5572):63-4.

30. Sachs DH. The Ia antigens. Contemp Top Mol Immunol. 1976;5:1-33.

31. Barker DJ, Maccari G, Georgiou X, Cooper MA, Flicek P, Robinson J, et al. The IPD-IMGT/HLA Database. Nucleic Acids Res. 2023;51(D1):D1053-D60.

32. Karnaukhov V, Paes W, Woodhouse IB, Partridge T, Nicastri A, Brackenridge S, et al. HLA variants have different preferences to present proteins with specific molecular functions which are complemented in frequent haplotypes. Front Immunol. 2022;13:1067463.

33. Kubo RT, Sette A, Grey HM, Appella E, Sakaguchi K, Zhu NZ, et al. Definition of specific peptide motifs for four major HLA-A alleles. J Immunol. 1994;152(8):3913-24.

34. Bassani-Sternberg M, Chong C, Guillaume P, Solleder M, Pak H, Gannon PO, et al. Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity. PLoS Comput Biol. 2017;13(8):e1005725.

35. Prugnolle F, Manica A, Charpentier M, Guegan JF, Guernier V, Balloux F. Pathogen-driven selection and worldwide HLA class I diversity. Curr Biol. 2005;15(11):1022-7.

36. Lilly F, Boyse EA, Old LJ. Genetic Basis of Susceptibility to Viral Leukaemogenesis. Lancet. 1964;2(7371):1207-9.

37. Benhammadi M, Mathe J, Dumont-Lagace M, Kobayashi KS, Gaboury L, Brochu S, et al. IFN-lambda Enhances Constitutive Expression of MHC Class I Molecules on Thymic Epithelial Cells. J Immunol. 2020.

38. Boegel S, Lower M, Bukur T, Sorn P, Castle JC, Sahin U. HLA and proteasome expression body map. BMC Med Genomics. 2018;11(1):36.

39. Pai RK, Askew D, Boom WH, Harding CV. Regulation of class II MHC expression in APCs: roles of types I, III, and IV class II transactivator. J Immunol. 2002;169(3):1326-33.

40. Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. Structure of the human class I histocompatibility antigen, HLA-A2. Nature. 1987;329(6139):506-12.

41. Brown JH, Jardetzky TS, Stern LJ, Gorga JC, Strominger JL, Wiley DC. Human class II MHC molecule HLA-DR1: X-ray structure determined from three crystal forms. Acta Crystallogr D Biol Crystallogr. 1995;51(Pt 6):946-61.

42. Tadros DM, Eggenschwiler S, Racle J, Gfeller D. The MHC Motif Atlas: a database of MHC binding specificities and ligands. Nucleic Acids Res. 2023;51(D1):D428-D37.

43. Rock KL, Gramm C, Rothstein L, Clark K, Stein R, Dick L, et al. Inhibitors of the proteasome block the degradation of most cell proteins and the generation of peptides presented on MHC class I molecules. Cell. 1994;78(5):761-71.

44. Caron E, Charbonneau R, Huppe G, Brochu S, Perreault C. The structure and location of SIMP/STT3B account for its prominent imprint on the MHC I immunopeptidome. Int Immunol. 2005;17(12):1583-96.

45. Chu E, Umetsu D, Lareau M, Schneeberger E, Geha RS. Analysis of antigen uptake and presentation by Epstein-Barr virus-transformed human lymphoblastoid B cells. Eur J Immunol. 1984;14(4):291-8.

46. Chesnut RW, Colon SM, Grey HM. Antigen presentation by normal B cells, B cell tumors, and macrophages: functional and biochemical comparison. J Immunol. 1982;128(4):1764-8.

47. Caron E, Vincent K, Fortier MH, Laverdure JP, Bramoulle A, Hardy MP, et al. The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. Mol Syst Biol. 2011;7:533.

48. Gfeller D, Liu Y, Racle J. Contemplating immunopeptidomes to better predict them. Semin Immunol. 2023;66:101708.

49. Granados DP, Tanguay PL, Hardy MP, Caron E, de Verteuil D, Meloche S, et al. ER stress affects processing of MHC class I-associated peptides. BMC Immunol. 2009;10:10.

50. Apcher S, Millot G, Daskalogianni C, Scherl A, Manoury B, Fahraeus R. Translation of pre-spliced RNAs in the nuclear compartment generates peptides for the MHC class I pathway. Proc Natl Acad Sci U S A. 2013;110(44):17951-6.

51. Bianchi F, Textor J, van den Bogaart G. Transmembrane Helices Are an Overlooked Source of Major Histocompatibility Complex Class I Epitopes. Front Immunol. 2017;8:1118.

52. Pearson H, Daouda T, Granados DP, Durette C, Bonneil E, Courcelles M, et al. MHC class I-associated peptides derive from selective regions of the human genome. J Clin Invest. 2016;126(12):4690-701.

53. Granados DP, Yahyaoui W, Laumont CM, Daouda T, Muratore-Schroeder TL, Cote C, et al. MHC I-associated peptides preferentially derive from transcripts bearing miRNA response elements. Blood. 2012;119(26):e181-91.

54. Laumont CM, Daouda T, Laverdure JP, Bonneil E, Caron-Lizotte O, Hardy MP, et al. Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. Nat Commun. 2016;7:10238.

55. Ouspenskaia T, Law T, Clauser KR, Klaeger S, Sarkizova S, Aguet F, et al. Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. Nat Biotechnol. 2022;40(2):209-17.

56. Apcher S, Daskalogianni C, Lejeune F, Manoury B, Imhoos G, Heslop L, et al. Major source of antigenic peptides for the MHC class I pathway is produced during the pioneer round of mRNA translation. Proc Natl Acad Sci U S A. 2011;108(28):11572-7.

57. Dolan BP, Li L, Takeda K, Bennink JR, Yewdell JW. Defective ribosomal products are the major source of antigenic peptides endogenously generated from influenza A virus neuraminidase. J Immunol. 2010;184(3):1419-24.

58. Cardinaud S, Starck SR, Chandra P, Shastri N. The synthesis of truncated polypeptides for immune surveillance and viral evasion. PLoS One. 2010;5(1):e8692.

59. Mackay LK, Long HM, Brooks JM, Taylor GS, Leung CS, Chen A, et al. T cell detection of a B-cell tropic virus infection: newly-synthesised versus mature viral proteins as antigen sources for CD4 and CD8 epitope display. PLoS Pathog. 2009;5(12):e1000699.

60. Spies T, DeMars R. Restored expression of major histocompatibility class I molecules by gene transfer of a putative peptide transporter. Nature. 1991;351(6324):323-4.

61. Ortmann B, Androlewicz MJ, Cresswell P. MHC class I/beta 2-microglobulin complexes associate with TAP transporters before peptide binding. Nature. 1994;368(6474):864-7.

62. Sadasivan B, Lehner PJ, Ortmann B, Spies T, Cresswell P. Roles for calreticulin and a novel glycoprotein, tapasin, in the interaction of MHC class I molecules with TAP. Immunity. 1996;5(2):103-14.

63. Jiang J, Taylor DK, Kim EJ, Boyd LF, Ahmad J, Mage MG, et al. Structural mechanism of tapasin-mediated MHC-I peptide loading in antigen presentation. Nat Commun. 2022;13(1):5470.

64. Hughes EA, Cresswell P. The thiol oxidoreductase ERp57 is a component of the MHC class I peptide-loading complex. Curr Biol. 1998;8(12):709-12.

65. Serwold T, Gonzalez F, Kim J, Jacob R, Shastri N. ERAAP customizes peptides for MHC class I molecules in the endoplasmic reticulum. Nature. 2002;419(6906):480-3.

66. Hughes EA, Hammond C, Cresswell P. Misfolded major histocompatibility complex class I heavy chains are translocated into the cytoplasm and degraded by the proteasome. Proc Natl Acad Sci U S A. 1997;94(5):1896-901.

67. Neefjes JJ, Ploegh HL. Allele and locus-specific differences in cell surface expression and the association of HLA class I heavy chain with beta 2-microglobulin: differential effects of inhibition of glycosylation on class I subunit association. Eur J Immunol. 1988;18(5):801-10.

68. Neefjes J, Jongsma ML, Paul P, Bakke O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol. 2011;11(12):823-36.

69. Cosgrove D, Chan SH, Waltzinger C, Benoist C, Mathis D. The thymic compartment responsible for positive selection of CD4+ T cells. Int Immunol. 1992;4(6):707-10.

70. Laufer TM, DeKoning J, Markowitz JS, Lo D, Glimcher LH. Unopposed positive selection and autoreactivity in mice expressing class II MHC only on thymic cortex. Nature. 1996;383(6595):81-5.

71. Capone M, Romagnoli P, Beermann F, MacDonald HR, van Meerwijk JP. Dissociation of thymic positive and negative selection in transgenic mice expressing major histocompatibility complex class I molecules exclusively on thymic cortical epithelial cells. Blood. 2001;97(5):1336-42.

72. Liu CP, Crawford F, Marrack P, Kappler J. T cell positive selection by a high density, low affinity ligand. Proc Natl Acad Sci U S A. 1998;95(8):4522-6.

73. Murata S, Sasaki K, Kishimoto T, Niwa S, Hayashi H, Takahama Y, et al. Regulation of CD8+ T cell development by thymus-specific proteasomes. Science. 2007;316(5829):1349-53.

74. Florea BI, Verdoes M, Li N, van der Linden WA, Geurink PP, van den Elst H, et al. Activity-based profiling reveals reactivity of the murine thymoproteasome-specific subunit beta5t. Chem Biol. 2010;17(8):795-801.

75. Nakagawa T, Roth W, Wong P, Nelson A, Farr A, Deussing J, et al. Cathepsin L: critical role in Ii degradation and CD4 T cell selection in the thymus. Science. 1998;280(5362):450-3.

76. Bowlus CL, Ahn J, Chu T, Gruen JR. Cloning of a novel MHC-encoded serine peptidase highly expressed by cortical epithelial cells of the thymus. Cell Immunol. 1999;196(2):80-6.

77. Apavaloaei A, Brochu S, Dong M, Rouette A, Hardy MP, Villafano G, et al. PSMB11 Orchestrates the Development of CD4 and CD8 Thymocytes via Regulation of Gene Expression in Cortical Thymic Epithelial Cells. J Immunol. 2019;202(3):966-78.

78. Honey K, Nakagawa T, Peters C, Rudensky A. Cathepsin L regulates CD4+ T cell selection independently of its effect on invariant chain: a role in the generation of positively selecting peptide ligands. J Exp Med. 2002;195(10):1349-58.

79. Viret C, Lamare C, Guiraud M, Fazilleau N, Bour A, Malissen B, et al. Thymus-specific serine protease contributes to the diversification of the functional endogenous CD4 T cell receptor repertoire. J Exp Med. 2011;208(1):3-11.

80. Marrack P, Ignatowicz L, Kappler JW, Boymel J, Freed JH. Comparison of peptides bound to spleen and thymus class II. J Exp Med. 1993;178(6):2173-83.

81. van Meerwijk JP, Marguerat S, Lees RK, Germain RN, Fowlkes BJ, MacDonald HR. Quantitative impact of thymic clonal deletion on the T cell repertoire. J Exp Med. 1997;185(3):377-83.

82. Kappler JW, Roehm N, Marrack P. T cell tolerance by clonal elimination in the thymus. Cell. 1987;49(2):273-80.

83. Murphy KM, Heimberger AB, Loh DY. Induction by antigen of intrathymic apoptosis of CD4+CD8+TCRlo thymocytes in vivo. Science. 1990;250(4988):1720-3.

84. Robey EA, Ramsdell F, Kioussis D, Sha W, Loh D, Axel R, et al. The level of CD8 expression can determine the outcome of thymic selection. Cell. 1992;69(7):1089-96.

85. Wack A, Ladyman HM, Williams O, Roderick K, Ritter MA, Kioussis D. Direct visualization of thymocyte apoptosis in neglect, acute and steady-state negative selection. Int Immunol. 1996;8(10):1537-48.

86. Bensinger SJ, Bandeira A, Jordan MS, Caton AJ, Laufer TM. Major histocompatibility complex class II-positive cortical epithelium mediates the selection of CD4(+)25(+) immunoregulatory T cells. J Exp Med. 2001;194(4):427-38.

87. Liston A, Nutsch KM, Farr AG, Lund JM, Rasmussen JP, Koni PA, et al. Differentiation of regulatory Foxp3+ T cells in the thymic cortex. Proc Natl Acad Sci U S A. 2008;105(33):11903-8.

88. Fontenot JD, Dooley JL, Farr AG, Rudensky AY. Developmental regulation of Foxp3 expression during ontogeny. J Exp Med. 2005;202(7):901-6.

89. Tai X, Indart A, Rojano M, Guo J, Apenes N, Kadakia T, et al. How autoreactive thymocytes differentiate into regulatory versus effector CD4(+) T cells after avoiding clonal deletion. Nat Immunol. 2023;24(4):637-51.

90. Bonomo A, Matzinger P. Thymus epithelium induces tissue-specific tolerance. J Exp Med. 1993;177(4):1153-64.

91. Klein L, Klugmann M, Nave KA, Tuohy VK, Kyewski B. Shaping of the autoreactive T-cell repertoire by a splice variant of self protein expressed in thymic epithelial cells. Nat Med. 2000;6(1):56-61.

92. Bonasio R, Scimone ML, Schaerli P, Grabie N, Lichtman AH, von Andrian UH. Clonal deletion of thymocytes by circulating dendritic cells homing to the thymus. Nat Immunol. 2006;7(10):1092-100.

93. Duncan SR, Capetanakis NG, Lawson BR, Theofilopoulos AN. Thymic dendritic cells traffic to thymi of allogeneic recipients and prolong graft survival. J Clin Invest. 2002;109(6):755-64.

94. Perera J, Zheng Z, Li S, Gudjonson H, Kalinina O, Benichou JIC, et al. Self-Antigen-Driven Thymic B Cell Class Switching Promotes T Cell Central Tolerance. Cell Rep. 2016;17(2):387-98.

95. Nitta T, Tsutsumi M, Nitta S, Muro R, Suzuki EC, Nakano K, et al. Fibroblasts as a source of self-antigens for central immune tolerance. Nat Immunol. 2020.

96. Kozai M, Kubo Y, Katakai T, Kondo H, Kiyonari H, Schaeuble K, et al. Essential role of CCL21 in establishment of central self-tolerance in T cells. J Exp Med. 2017;214(7):1925-35.

97. Gavanescu I, Kessler B, Ploegh H, Benoist C, Mathis D. Loss of Aire-dependent thymic expression of a peripheral tissue antigen renders it a target of autoimmunity. Proc Natl Acad Sci U S A. 2007;104(11):4583-7.

98. Yano M, Kuroda N, Han H, Meguro-Horike M, Nishikawa Y, Kiyonari H, et al. Aire controls the differentiation program of thymic epithelial cells in the medulla for the establishment of self-tolerance. J Exp Med. 2008;205(12):2827-38.

99. Wang X, Laan M, Bichele R, Kisand K, Scott HS, Peterson P. Post-Aire maturation of thymic medullary epithelial cells involves selective expression of keratinocyte-specific autoantigens. Front Immunol. 2012;3(March):19.

100.Michelson DA, Hase K, Kaisho T, Benoist C, Mathis D. Thymic epithelial cells co-opt lineage-defining transcription factors to eliminate autoreactive T cells. Cell. 2022.

101.Bornstein C, Nevo S, Giladi A, Kadouri N, Pouzolles M, Gerbe F, et al. Single-cell mapping of the thymic stroma identifies IL-25-producing tuft epithelial cells. Nature. 2018;559(7715):622-6.

102.Park JE, Botting RA, Dominguez Conde C, Popescu DM, Lavaert M, Kunz DJ, et al. A cell atlas of human thymic development defines T cell repertoire formation. Science. 2020;367(6480).

103.Bautista JL, Cramer NT, Miller CN, Chavez J, Berrios DI, Byrnes LE, et al. Single-cell transcriptional profiling of human thymic stroma uncovers novel cellular heterogeneity in the thymic medulla. Nat Commun. 2021;12(1):1096.

104.Michelson DA, Mathis D. Thymic mimetic cells: tolerogenic masqueraders. Trends Immunol. 2022;43(10):782-91.

105.Lavaert M, Liang KL, Vandamme N, Park JE, Roels J, Kowalczyk MS, et al. Integrated scRNA-Seq Identifies Human Postnatal Thymus Seeding Progenitors and Regulatory Dynamics of Differentiating Immature Thymocytes. Immunity. 2020;52(6):1088-104 e6.

106.Le J, Park JE, Ha VL, Luong A, Branciamore S, Rodin AS, et al. Single-Cell RNA-Seq Mapping of Human Thymopoiesis Reveals Lineage Specification Trajectories and a Commitment Spectrum in T Cell Development. Immunity. 2020;52(6):1105-18 e9.

107.Cosway EJ, Ohigashi I, Schauble K, Parnell SM, Jenkinson WE, Luther S, et al. Formation of the Intrathymic Dendritic Cell Pool Requires CCL21-Mediated Recruitment of CCR7(+) Progenitors to the Thymus. J Immunol. 2018;201(2):516-23.

108.Li J, Park J, Foss D, Goldschneider I. Thymus-homing peripheral dendritic cells constitute two of the three major subsets of dendritic cells in the steady-state thymus. J Exp Med. 2009;206(3):607-22.

109.Atibalentja DF, Murphy KM, Unanue ER. Functional redundancy between thymic CD8alpha+ and Sirpalpha+ conventional dendritic cells in presentation of blood-derived lysozyme by MHC class II proteins. J Immunol. 2011;186(3):1421-31.

110.Humblet C, Rudensky A, Kyewski B. Presentation and intercellular transfer of self antigen within the thymic microenvironment: expression of the E alpha peptide-I-Ab complex by isolated thymic stromal cells. Int Immunol. 1994;6(12):1949-58.

111.Gallegos AM, Bevan MJ. Central tolerance to tissue-specific antigens mediated by direct and indirect antigen presentation. J Exp Med. 2004;200(8):1039-49.

112.Koble C, Kyewski B. The thymic medulla: a unique microenvironment for intercellular self-antigen transfer. J Exp Med. 2009;206(7):1505-13.

113.Colantonio AD, Epeldegui M, Jesiak M, Jachimowski L, Blom B, Uittenbogaart CH. IFN-alpha is constitutively expressed in the human thymus, but not in peripheral lymphoid organs. PLoS One. 2011;6(8):e24252.

114.Xing Y, Wang X, Jameson SC, Hogquist KA. Late stages of T cell maturation in the thymus involve NF-kappaB and tonic type I interferon signaling. Nat Immunol. 2016;17(5):565-73.

115.Hanabuchi S, Ito T, Park WR, Watanabe N, Shaw JL, Roman E, et al. Thymic stromal lymphopoietin-activated plasmacytoid dendritic cells induce the generation of FOXP3+ regulatory T cells in human thymus. J Immunol. 2010;184(6):2999-3007.

116.Martin-Gayo E, Sierra-Filardi E, Corbi AL, Toribio ML. Plasmacytoid dendritic cells resident in human thymus drive natural Treg cell development. Blood. 2010;115(26):5366-75.

117.Martinet V, Tonon S, Torres D, Azouz A, Nguyen M, Kohler A, et al. Type I interferons regulate eomesodermin expression and the development of unconventional memory CD8(+) T cells. Nat Commun. 2015;6:7089.

118.Epeldegui M, Blom B, Uittenbogaart CH. BST2/Tetherin is constitutively expressed on human thymocytes with the phenotype and function of Treg cells. Eur J Immunol. 2015;45(3):728-37.

119.Sansom SN, Shikama-Dorn N, Zhanybekova S, Nusspaumer G, Macaulay IC, Deadman ME, et al. Population and single-cell genomics reveal the Aire dependency, relief from Polycomb silencing, and distribution of self-antigen expression in thymic epithelia. Genome Res. 2014;24(12):1918-31.

120.Jansen K, Shikama-Dorn N, Attar M, Maio S, Lopopolo M, Buck D, et al. RBFOX splicing factors contribute to a broad but selective recapitulation of peripheral tissue splicing patterns in the thymus. Genome Res. 2021;31(11):2022-34.
121.Danan-Gotthold M, Guyon C, Giraud M, Levanon EY, Abramson J. Extensive RNA editing and splicing increase immune self-representation diversity in medullary thymic epithelial cells. Genome Biol. 2016;17(1):219.
122.Derbinski J, Schulte A, Kyewski B, Klein L. Promiscuous gene expression in medullary thymic epithelial cells mirrors the peripheral self. Nat Immunol. 2001;2(11):1032-9.
123.Jolicoeur C, Hanahan D, Smith KM. T-cell tolerance toward a transgenic beta-cell antigen and transcription of endogenous pancreatic genes in thymus. Proc Natl Acad Sci U S A. 1994;91(14):6707-11.
124.Antonia SJ, Geiger T, Miller J, Flavell RA. Mechanisms of immune tolerance induction through the thymic expression of a peripheral tissue-specific protein. Int Immunol. 1995;7(5):715-25.
125.Husbands SD, Schonrich G, Arnold B, Chandler PR, Simpson E, Philpott KL, et al. Expression of major histocompatibility complex class I antigens at low levels in the thymus induces T cell tolerance via a non-deletional mechanism. Eur J Immunol. 1992;22(10):2655-61.
126.Liblau R, Tournier-Lasserve E, Maciazek J, Dumas G, Siffert O, Hashim G, et al. T cell response to myelin basic protein epitopes in multiple sclerosis patients and healthy subjects. Eur J Immunol. 1991;21(6):1391-5.
127.Ota K, Matsui M, Milford EL, Mackin GA, Weiner HL, Hafler DA. T-cell recognition of an immunodominant myelin basic protein epitope in multiple sclerosis. Nature. 1990;346(6280):183-7.
128.Anderson MS, Venanzi ES, Klein L, Chen Z, Berzins SP, Turley SJ, et al. Projection of an immunological self shadow within the thymus by the aire protein. Science. 2002;298(5597):1395-401.
129.Takaba H, Morishita Y, Tomofuji Y, Danks L, Nitta T, Komatsu N, et al. Fezf2 Orchestrates a Thymic Program of Self-Antigen Expression for Immune Tolerance. Cell. 2015;163(4):975-87.
130.Tomofuji Y, Takaba H, Suzuki HI, Benlaribi R, Martinez CDP, Abe Y, et al. Chd4 choreographs self-antigen expression for central immune tolerance. Nat Immunol. 2020;21(8):892-901.

131.Wells KL, Miller CN, Gschwind AR, Wei W, Phipps JD, Anderson MS, et al. Combined transient ablation and single-cell RNA-sequencing reveals the development of medullary thymic epithelial cells. Elife. 2020;9.

132.Eckler MJ, Chen B. Fez family transcription factors: controlling neurogenesis and cell fate in the developing mammalian nervous system. Bioessays. 2014;36(8):788-97.

133.Koh AS, Miller EL, Buenrostro JD, Moskowitz DM, Wang J, Greenleaf WJ, et al. Rapid chromatin repression by Aire provides precise control of immune tolerance. Nat Immunol. 2018;19(2):162-72.

134.Org T, Rebane A, Kisand K, Laan M, Haljasorg U, Andreson R, et al. AIRE activated tissue specific genes have histone modifications associated with inactive chromatin. Hum Mol Genet. 2009;18(24):4699-710.

135.Waterfield M, Khan IS, Cortez JT, Fan U, Metzger T, Greer A, et al. The transcriptional regulator Aire coopts the repressive ATF7ip-MBD1 complex for the induction of immunotolerance. Nat Immunol. 2014;15(3):258-65.

136.Handel AE, Shikama-Dorn N, Zhanybekova S, Maio S, Graedel AN, Zuklys S, et al. Comprehensively Profiling the Chromatin Architecture of Tissue Restricted Antigen Expression in Thymic Epithelial Cells Over Development. Front Immunol. 2018;9:2120.

137.Benlaribi R, Gou Q, Takaba H. Thymic self-antigen expression for immune tolerance and surveillance. Inflamm Regen. 2022;42(1):28.

138.Derbinski J, Gabler J, Brors B, Tierling S, Jonnakuty S, Hergenhahn M, et al. Promiscuous gene expression in thymic epithelial cells is regulated at multiple levels. J Exp Med. 2005;202(1):33-45.

139.Brennecke P, Reyes A, Pinto S, Rattay K, Nguyen M, Kuchler R, et al. Single-cell transcriptome analysis reveals coordinated ectopic gene-expression patterns in medullary thymic epithelial cells. Nat Immunol. 2015;16(9):933-41.

140.St-Pierre C, Morgand E, Benhammadi M, Rouette A, Hardy MP, Gaboury L, et al. Immunoproteasomes Control the Homeostasis of Medullary Thymic Epithelial Cells by Alleviating Proteotoxic Stress. Cell Rep. 2017;21(9):2558-70.

141.Kadouri N, Nevo S, Goldfarb Y, Abramson J. Thymic epithelial cell heterogeneity: TEC by TEC. Nat Rev Immunol. 2019.

142.Miller CN, Proekt I, von Moltke J, Wells KL, Rajpurkar AR, Wang H, et al. Thymic tuft cells promote an IL-4-enriched medulla and shape thymocyte development. Nature. 2018;559(7715):627-31.

143.Dhalla F, Baran-Gale J, Maio S, Chappell L, Hollander GA, Ponting CP. Biologically indeterminate yet ordered promiscuous gene expression in single medullary thymic epithelial cells. EMBO J. 2020;39(1):e101828.

144.Panneck AR, Rafiq A, Schutz B, Soultanova A, Deckmann K, Chubanov V, et al. Cholinergic epithelial cell with chemosensory traits in murine thymic medulla. Cell Tissue Res. 2014;358(3):737-48.

145.Platt RN, 2nd, Vandewege MW, Ray DA. Mammalian transposable elements and their impacts on genome evolution. Chromosome Res. 2018;26(1-2):25-43.

146.Osmanski AB, Paulat NS, Korstian J, Grimshaw JR, Halsey M, Sullivan KAM, et al. Insights into mammalian TE diversity through the curation of 248 genome assemblies. Science. 2023;380(6643):eabn1430.

147.Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biol. 2018;19(1):199.

148.Wells JN, Feschotte C. A Field Guide to Eukaryotic Transposable Elements. Annu Rev Genet. 2020;54:539-61.

149.Arkhipova IR. Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. Mob DNA. 2017;8:19.

150.Ishak CA, Classon M, De Carvalho DD. Deregulation of Retroelements as an Emerging Therapeutic Opportunity in Cancer. Trends Cancer. 2018;4(8):583-97.

151.Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.

152.Bao W, Jurka MG, Kapitonov VV, Jurka J. New superfamilies of eukaryotic DNA transposons and their internal divisions. Mol Biol Evol. 2009;26(5):983-93.

153.Sarkar A, Sim C, Hong YS, Hogan JR, Fraser MJ, Robertson HM, et al. Molecular evolutionary analysis of the widespread piggyBac transposon family and related "domesticated" sequences. Mol Genet Genomics. 2003;270(2):173-80.

154.Feschotte C. Merlin, a new superfamily of DNA transposons identified in diverse animal genomes and related to bacterial IS1016 insertion sequences. Mol Biol Evol. 2004;21(9):1769-80.
155.Kojima KK. Structural and sequence diversity of eukaryotic transposable elements. Genes Genet Syst. 2020;94(6):233-52.
156.Hickman AB, Dyda F. DNA Transposition at Work. Chem Rev. 2016;116(20):12758-84.
157.Grabundzija I, Messing SA, Thomas J, Cosby RL, Bilic I, Miskey C, et al. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. Nat Commun. 2016;7:10716.
158.Grabundzija I, Hickman AB, Dyda F. Helraiser intermediates provide insight into the mechanism of eukaryotic replicative transposition. Nat Commun. 2018;9(1):1278.
159.van der Laan LJ, Lockey C, Griffeth BC, Frasier FS, Wilson CA, Onions DE, et al. Infection by porcine endogenous retrovirus after islet xenotransplantation in SCID mice. Nature. 2000;407(6800):90-4.
160.Ishida Y, Zhao K, Greenwood AD, Roca AL. Proliferation of endogenous retroviruses in the early stages of a host germ line invasion. Mol Biol Evol. 2015;32(1):109-20.
161.Boeke JD, Garfinkel DJ, Styles CA, Fink GR. Ty elements transpose through an RNA intermediate. Cell. 1985;40(3):491-500.
162.Garfinkel DJ, Boeke JD, Fink GR. Ty element transposition: reverse transcriptase and virus-like particles. Cell. 1985;42(2):507-17.
163.Kirchner J, Sandmeyer SB. Ty3 integrase mutants defective in reverse transcription or 3'-end processing of extrachromosomal Ty3 DNA. Journal of virology. 1996;70(7):4737-47.
164.Lee YN, Bieniasz PD. Reconstitution of an infectious human endogenous retrovirus. PLoS Pathog. 2007;3(1):e10.
165.Kraus B, Boller K, Reuter A, Schnierle BS. Characterization of the human endogenous retrovirus K Gag protein: identification of protease cleavage sites. Retrovirology. 2011;8:21.
166.Baldwin ET, Gotte M, Tchesnokov EP, Arnold E, Hagel M, Nichols C, et al. Human endogenous retrovirus-K (HERV-K) reverse transcriptase (RT) structure and biochemistry reveals remarkable similarities to HIV-1 RT and opportunities for HERV-K-specific inhibition. Proc Natl Acad Sci U S A. 2022;119(27):e2200260119.

167.Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. Nature. 2000;403(6771):785-9.
168.Blond JL, Lavillette D, Cheynet V, Bouton O, Oriol G, Chapel-Fernandes S, et al. An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. J Virol. 2000;74(7):3321-9.
169.Magiorkinis G, Gifford RJ, Katzourakis A, De Ranter J, Belshaw R. Env-less endogenous retroviruses are genomic superspreaders. Proc Natl Acad Sci U S A. 2012;109(19):7385-90.
170.Campos-Sanchez R, Cremona MA, Pini A, Chiaromonte F, Makova KD. Integration and Fixation Preferences of Human and Mouse Endogenous Retroviruses Uncovered with Functional Data Analysis. PLoS Comput Biol. 2016;12(6):e1004956.
171.Luan DD, Korman MH, Jakubczak JL, Eickbush TH. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell. 1993;72(4):595-605.
172.Christensen SM, Bibillo A, Eickbush TH. Role of the Bombyx mori R2 element N-terminal domain in the target-primed reverse transcription (TPRT) reaction. Nucleic Acids Res. 2005;33(20):6461-8.
173.Scott AF, Schmeckpeper BJ, Abdelrazik M, Comey CT, O'Hara B, Rossiter JP, et al. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics. 1987;1(2):113-25.
174.Holmes SE, Singer MF, Swergold GD. Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. J Biol Chem. 1992;267(28):19765-8.
175.Feng Q, Moran JV, Kazazian HH, Boeke JD. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. Cell. 1996;87(5):905-16.
176.Mathias SL, Scott AF, Kazazian Jr HH, Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. Science. 1991;254(5039):1808-10.
177.Sultana T, van Essen D, Siol O, Bailly-Bechet M, Philippe C, Zine El Aabidine A, et al. The Landscape of L1 Retrotransposons in the Human Genome Is Shaped by Pre-insertion Sequence Biases and Post-insertion Selection. Mol Cell. 2019;74(3):555-70 e7.

178.Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, et al. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A. 2003;100(9):5280-5.

179.Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, et al. LINE-1 retrotransposition activity in human genomes. Cell. 2010;141(7):1159-70.

180.Korenberg JR, Rykowski MC. Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands. Cell. 1988;53(3):391-400.

181.Ohshima K, Okada N. Generality of the tRNA origin of short interspersed repetitive elements (SINEs). Characterization of three different tRNA-derived retroposons in the octopus. J Mol Biol. 1994;243(1):25-37.

182.Ullu E, Tschudi C. Alu sequences are processed 7SL RNA genes. Nature. 1984;312(5990):171-2.

183.Varshney D, Vavrova-Anderson J, Oler AJ, Cowling VH, Cairns BR, White RJ. SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. Nat Commun. 2015;6:6569.

184.Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet. 2003;35(1):41-8.

185.Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, et al. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. Int J Cancer. 2009;124(1):81-7.

186.Gasior SL, Wakeman TP, Xu B, Deininger PL. The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol. 2006;357(5):1383-93.

187.Teugels E, De Brakeleer S, Goelen G, Lissens W, Sermijn E, De Greve J. De novo Alu element insertions targeted to a sequence common to the BRCA1 and BRCA2 genes. Hum Mutat. 2005;26(3):284.

188.Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, et al. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res. 1992;52(3):643-5.

189.Rodriguez-Martin C, Cidre F, Fernandez-Teijeiro A, Gomez-Mariano G, de la Vega L, Ramos P, et al. Familial retinoblastoma due to intronic LINE-1 insertion causes aberrant and noncanonical mRNA splicing of the RB1 gene. J Hum Genet. 2016;61(5):463-6.
190.Goodchild NL, Wilkinson DA, Mager DL. Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements. Virology. 1993;196(2):778-88.
191.Cordonnier A, Casella JF, Heidmann T. Isolation of novel human endogenous retrovirus-like elements with foamy virus-related pol sequence. J Virol. 1995;69(9):5890-7.
192.Pace JK, 2nd, Feschotte C. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. Genome Res. 2007;17(4):422-32.
193.Hughes JF, Coffin JM. Evidence for genomic rearrangements mediated by human endogenous retroviruses during primate evolution. Nat Genet. 2001;29(4):487-9.
194.Sun C, Skaletsky H, Rozen S, Gromoll J, Nieschlag E, Oates R, et al. Deletion of azoospermia factor a (AZFa) region of human Y chromosome caused by recombination between HERV15 proviruses. Hum Mol Genet. 2000;9(15):2291-6.
195.Stritt C, Gordon SP, Wicker T, Vogel JP, Roulin AC. Recent Activity in Expanding Populations and Purifying Selection Have Shaped Transposable Element Landscapes across Natural Accessions of the Mediterranean Grass Brachypodium distachyon. Genome Biol Evol. 2018;10(1):304-18.
196.Ruggiero RP, Bourgeois Y, Boissinot S. LINE Insertion Polymorphisms are Abundant but at Low Frequencies across Populations of Anolis carolinensis. Front Genet. 2017;8:44.
197.Szitenberg A, Cha S, Opperman CH, Bird DM, Blaxter ML, Lunt DH. Genetic Drift, Not Life History or RNAi, Determine Long-Term Evolution of Transposable Elements. Genome Biol Evol. 2016;8(9):2964-78.
198.Kassiotis G, Stoye JP. Immune responses to endogenous retroelements: taking the bad with the good. Nat Rev Immunol. 2016;16(4):207-19.
199.Dazeniere J, Bousios A, Eyre-Walker A. Patterns of selection in the evolution of a transposable element. G3 (Bethesda). 2022;12(5).
200.Nellaker C, Keane TM, Yalcin B, Wong K, Agam A, Belgard TG, et al. The genomic landscape shaped by selection on transposable elements across 18 mouse strains. Genome Biol. 2012;13(6):R45.

201.Rishishwar L, Wang L, Wang J, Yi SV, Lachance J, Jordan IK. Evidence for positive selection on recent human transposable element insertions. Gene. 2018;675:69-79.
202.Sun X, Xiang Y, Dou N, Zhang H, Pei S, Franco AV, et al. The role of transposon inverted repeats in balancing drought tolerance and yield-related traits in maize. Nat Biotechnol. 2023;41(1):120-7.
203.Christmas MJ, Kaplow IM, Genereux DP, Dong MX, Hughes GM, Li X, et al. Evolutionary constraint and innovation across hundreds of placental mammals. Science. 2023;380(6643):eabn3943.
204.Van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, et al. The industrial melanism mutation in British peppered moths is a transposable element. Nature. 2016;534(7605):102-5.
205.Li ZW, Hou XH, Chen JF, Xu YC, Wu Q, Gonzalez J, et al. Transposable Elements Contribute to the Adaptation of Arabidopsis thaliana. Genome Biol Evol. 2018;10(8):2140-50.
206.Suh A, Smeds L, Ellegren H. Abundant recent activity of retrovirus-like retrotransposons within and among flycatcher species implies a rich source of structural variation in songbird genomes. Mol Ecol. 2018;27(1):99-111.
207.Etchegaray E, Naville M, Volff JN, Haftek-Terreau Z. Transposable element-derived sequences in vertebrate development. Mob DNA. 2021;12(1):1.
208.Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. 2008;18(11):1752-62.
209.Choudhary MN, Friedman RZ, Wang JT, Jang HS, Zhuo X, Wang T. Co-opted transposons help perpetuate conserved higher-order chromosomal structures. Genome Biol. 2020;21(1):16.
210.Choudhary MNK, Quaid K, Xing X, Schmidt H, Wang T. Widespread contribution of transposable elements to the rewiring of mammalian 3D genomes. Nat Commun. 2023;14(1):634.
211.Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, et al. Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res. 2014;24(12):1963-76.

212.Fort A, Hashimoto K, Yamada D, Salimullah M, Keya CA, Saxena A, et al. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. Nat Genet. 2014;46(6):558-66.

213.Todd CD, Deniz O, Taylor D, Branco MR. Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. Elife. 2019;8.

214.Trizzino M, Kapusta A, Brown CD. Transposable elements generate regulatory novelty in a tissue-specific fashion. BMC Genomics. 2018;19(1):468.

215.Ito J, Sugimoto R, Nakaoka H, Yamada S, Kimura T, Hayano T, et al. Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. PLoS Genet. 2017;13(7):e1006883.

216.Simo-Riudalbas L, Offner S, Planet E, Duc J, Abrami L, Dind S, et al. Transposon-activated POU5F1B promotes colorectal cancer growth and metastasis. Nat Commun. 2022;13(1):4913.

217.Pontis J, Pulver C, Playfoot CJ, Planet E, Grun D, Offner S, et al. Primate-specific transposable elements shape transcriptional networks during human development. Nat Commun. 2022;13(1):7178.

218.Ye M, Goudot C, Hoyler T, Lemoine B, Amigorena S, Zueva E. Specific subfamilies of transposable elements contribute to different domains of T lymphocyte enhancers. Proc Natl Acad Sci U S A. 2020;117(14):7905-16.

219.Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, et al. The regulated retrotransposon transcriptome of mammalian cells. Nat Genet. 2009;41(5):563-71.

220.Kitano S, Kurasawa H, Aizawa Y. Transposable elements shape the human proteome landscape via formation of cis-acting upstream open reading frames. Genes Cells. 2018;23(4):274-84.

221.Erwin JA, Paquola AC, Singer T, Gallina I, Novotny M, Quayle C, et al. L1-associated genomic regions are deleted in somatic cells of the healthy human brain. Nat Neurosci. 2016;19(12):1583-91.

222.Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV, Gage FH. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature. 2005;435(7044):903-10.

223.Xia B, Zhang W, Wudzinska A, Huang E, Brosh R, Pour M, et al. The genetic basis of tail-loss evolution in humans and apes. bioRxiv. 2021:2021.09.14.460388.
224.Kelley DR, Hendrickson DG, Tenen D, Rinn JL. Transposable elements modulate human RNA abundance and splicing via specific RNA-protein interactions. Genome Biol. 2014;15(12):537.
225.Percharde M, Lin CJ, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, et al. A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. Cell. 2018;174(2):391-405 e19.
226.Carter AC, Xu J, Nakamoto MY, Wei Y, Zarnegar BJ, Shi Q, et al. Spen links RNA-mediated endogenous retrovirus silencing and X chromosome inactivation. Elife. 2020;9.
227.Carlevaro-Fita J, Polidori T, Das M, Navarro C, Zoller TI, Johnson R. Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. Genome Res. 2019;29(2):208-22.
228.Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. PLoS Genet. 2013;9(4):e1003470.
229.Kannan S, Chernikova D, Rogozin IB, Poliakov E, Managadze D, Koonin EV, et al. Transposable Element Insertions in Long Intergenic Non-Coding RNA Genes. Front Bioeng Biotechnol. 2015;3:71.
230.Johnson WE. Origins and evolutionary consequences of ancient endogenous retroviruses. Nat Rev Microbiol. 2019;17(6):355-70.
231.Playfoot CJ, Sheppard S, Planet E, Trono D. Transposable elements contribute to the spatiotemporal microRNA landscape in human brain development. RNA. 2022;28(9):1157-71.
232.Qin S, Jin P, Zhou X, Chen L, Ma F. The Role of Transposable Elements in the Origin and Evolution of MicroRNAs in Human. PLoS One. 2015;10(6):e0131365.
233.Smalheiser NR, Torvik VI. Mammalian microRNAs derived from genomic repeats. Trends Genet. 2005;21(6):322-6.
234.Piriyapongsa J, Marino-Ramirez L, Jordan IK. Origin and evolution of human microRNAs from transposable elements. Genetics. 2007;176(2):1323-37.
235.Piriyapongsa J, Jordan IK. A family of human microRNA genes from miniature inverted-repeat transposable elements. PLoS One. 2007;2(2):e203.

236.Borchert GM, Holton NW, Williams JD, Hernan WL, Bishop IP, Dembosky JA, et al. Comprehensive analysis of microRNA genomic loci identifies pervasive repetitive-element origins. Mob Genet Elements. 2011;1(1):8-17.

237.Betel D, Sheridan R, Marks DS, Sander C. Computational analysis of mouse piRNA sequence and biogenesis. PLoS Comput Biol. 2007;3(11):e222.

238.Roovers EF, Rosenkranz D, Mahdipour M, Han CT, He N, Chuva de Sousa Lopes SM, et al. Piwi proteins and piRNAs in mammalian oocytes and early embryos. Cell Rep. 2015;10(12):2069-82.

239.Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science. 2007;316(5825):744-7.

240.Ono R, Nakamura K, Inoue K, Naruse M, Usami T, Wakisaka-Saito N, et al. Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. Nat Genet. 2006;38(1):101-6.

241.Sekita Y, Wagatsuma H, Nakamura K, Ono R, Kagami M, Wakisaka N, et al. Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. Nat Genet. 2008;40(2):243-8.

242.Bonnaud B, Bouton O, Oriol G, Cheynet V, Duret L, Mallet F. Evidence of selection on the domesticated ERVWE1 env retroviral element involved in placentation. Mol Biol Evol. 2004;21(10):1895-901.

243.Smit AF, Riggs AD. Tiggers and DNA transposon fossils in the human genome. Proc Natl Acad Sci U S A. 1996;93(4):1443-8.

244.Zaratiegui M, Vaughn MW, Irvine DV, Goto D, Watt S, Bahler J, et al. CENP-B preserves genome integrity at replication forks paused by retrotransposon LTR. Nature. 2011;469(7328):112-5.

245.Cosby RL, Judd J, Zhang R, Zhong A, Garry N, Pritham EJ, et al. Recurrent evolution of vertebrate transcription factors by transposase capture. Science. 2021;371(6531).

246.Abascal F, Tress ML, Valencia A. Alternative splicing and co-option of transposable elements: the case of TMPO/LAP2alpha and ZNF451 in mammals. Bioinformatics. 2015;31(14):2257-61.

247.Ohtani H, Liu M, Zhou W, Liang G, Jones PA. Switching roles for DNA and histone methylation depend on evolutionary ages of human endogenous retroviruses. Genome Res. 2018;28(8):1147-57.
248.Walsh CP, Chaillet JR, Bestor TH. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet. 1998;20(2):116-7.
249.Woodcock DM, Lawler CB, Linsenmeyer ME, Doherty JP, Warren WD. Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. J Biol Chem. 1997;272(12):7810-6.
250.Bourc'his D, Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. Nature. 2004;431(7004):96-9.
251.Chelmicki T, Roger E, Teissandier A, Dura M, Bonneville L, Rucli S, et al. m(6)A RNA methylation regulates the fate of endogenous retroviruses. Nature. 2021;591(7849):312-6.
252.Bruno M, Mahgoub M, Macfarlan TS. The Arms Race Between KRAB-Zinc Finger Proteins and Endogenous Retroelements and Its Impact on Mammals. Annu Rev Genet. 2019;53:393-416.
253.Fukuda K, Okuda A, Yusa K, Shinkai Y. A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells. Genome Res. 2018;28(6):846-58.
254.Kato M, Takemoto K, Shinkai Y. A somatic role for the histone methyltransferase Setdb1 in endogenous retrovirus silencing. Nat Commun. 2018;9(1):1683.
255.Garland W, Muller I, Wu M, Schmid M, Imamura K, Rib L, et al. Chromatin modifier HUSH co-operates with RNA decay factor NEXT to restrict transposable element expression. Mol Cell. 2022;82(9):1691-707 e8.
256.Rowe HM, Kapopoulou A, Corsinotti A, Fasching L, Macfarlan TS, Tarabay Y, et al. TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. Genome Res. 2013;23(3):452-61.
257.Maksakova IA, Thompson PJ, Goyal P, Jones SJ, Singh PB, Karimi MM, et al. Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. Epigenetics Chromatin. 2013;6(1):15.
258.Sheng W, LaFleur MW, Nguyen TH, Chen S, Chakravarthy A, Conway JR, et al. LSD1 Ablation Stimulates Anti-tumor Immunity and Enables Checkpoint Blockade. Cell. 2018;174(3):549-63 e19.

259.Montoya-Durango DE, Ramos KA, Bojang P, Ruiz L, Ramos IN, Ramos KS. LINE-1 silencing by retinoblastoma proteins is effected through the nucleosomal and remodeling deacetylase multiprotein complex. BMC Cancer. 2016;16:38.

260.Coluccio A, Ecco G, Duc J, Offner S, Turelli P, Trono D. Individual retrotransposon integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA methylation and TET-mediated hydroxymethylation in naive embryonic stem cells. Epigenetics Chromatin. 2018;11(1):7.

261.Day DS, Luquette LJ, Park PJ, Kharchenko PV. Estimating enrichment of repetitive elements from high-throughput sequence data. Genome Biol. 2010;11(6):R69.

262.Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448(7153):553-60.

263.Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K, Wutz A. Polycomb complexes act redundantly to repress genomic repeats and genes. Genes Dev. 2010;24(3):265-76.

264.Kim S, Gunesdogan U, Zylicz JJ, Hackett JA, Cougot D, Bao S, et al. PRMT5 protects genomic integrity during global DNA demethylation in primordial germ cells and preimplantation embryos. Mol Cell. 2014;56(4):564-79.

265.Chew YC, West JT, Kratzer SJ, Ilvarsonn AM, Eissenberg JC, Dave BJ, et al. Biotinylation of histones represses transposable elements in human and mouse cells and cell lines and in Drosophila melanogaster. J Nutr. 2008;138(12):2316-22.

266.Pestinger V, Wijeratne SS, Rodriguez-Melendez R, Zempleni J. Novel histone biotinylation marks are enriched in repeat regions and participate in repression of transcriptionally competent genes. J Nutr Biochem. 2011;22(4):328-33.

267.Pavletich NP, Pabo CO. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science. 1991;252(5007):809-17.

268.Thomas JH, Schneider S. Coevolution of retroelements and tandem zinc finger genes. Genome Res. 2011;21(11):1800-12.

269.Emerson RO, Thomas JH. Adaptive evolution in zinc finger transcription factors. PLoS Genet. 2009;5(1):e1000325.

270.Liu H, Chang LH, Sun Y, Lu X, Stubbs L. Deep vertebrate roots for mammalian zinc finger transcription factor subfamilies. Genome Biol Evol. 2014;6(3):510-25.

271.Imbeault M, Helleboid PY, Trono D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature. 2017;543(7646):550-4.

272.Dewannieux M, Heidmann T. Endogenous retroviruses: acquisition, amplification and taming of genome invaders. Curr Opin Virol. 2013;3(6):646-56.

273.Svoboda P, Stein P, Anger M, Bernstein E, Hannon GJ, Schultz RM. RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. Dev Biol. 2004;269(1):276-85.

274.Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, et al. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature. 2008;453(7194):534-8.

275.Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, et al. Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. Nature. 2008;453(7194):539-43.

276.Chen L, Dahlstrom JE, Lee SH, Rangasamy D. Naturally occurring endo-siRNA silences LINE-1 retrotransposons in human cells through DNA methylation. Epigenetics. 2012;7(7):758-71.

277.Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev. 2008;22(7):908-17.

278.Wenda JM, Homolka D, Yang Z, Spinelli P, Sachidanandam R, Pandey RR, et al. Distinct Roles of RNA Helicases MVH and TDRD9 in PIWI Slicing-Triggered Mammalian piRNA Biogenesis and Function. Dev Cell. 2017;41(6):623-37 e9.

279.De Fazio S, Bartonicek N, Di Giacomo M, Abreu-Goodger C, Sankar A, Funaya C, et al. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. Nature. 2011;480(7376):259-63.

280.Reuter M, Chuma S, Tanaka T, Franz T, Stark A, Pillai RS. Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. Nat Struct Mol Biol. 2009;16(6):639-46.

281.Heras SR, Macias S, Plass M, Fernandez N, Cano D, Eyras E, et al. The Microprocessor controls the activity of mammalian retrotransposons. Nat Struct Mol Biol. 2013;20(10):1173-81.
282.Choi YJ, Lin CP, Risso D, Chen S, Kim TA, Tan MH, et al. Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. Science. 2017;355(6325).
283.Hakim ST, Alsayari M, McLean DC, Saleem S, Addanki KC, Aggarwal M, et al. A large number of the human microRNAs target lentiviruses, retroviruses, and endogenous retroviruses. Biochem Biophys Res Commun. 2008;369(2):357-62.
284.Volkmann B, Wittmann S, Lagisquet J, Deutschmann J, Eissmann K, Ross JJ, et al. Human TRIM5alpha senses and restricts LINE-1 elements. Proc Natl Acad Sci U S A. 2020;117(30):17965-76.
285.Herrmann A, Wittmann S, Thomas D, Shepard CN, Kim B, Ferreiros N, et al. The SAMHD1-mediated block of LINE-1 retroelements is regulated by phosphorylation. Mob DNA. 2018;9:11.
286.Arjan-Odedra S, Swanson CM, Sherer NM, Wolinsky SM, Malim MH. Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses. Retrovirology. 2012;9:53.
287.Goodier JL, Cheung LE, Kazazian HH, Jr. MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. PLoS Genet. 2012;8(10):e1002941.
288.Goodier JL, Pereira GC, Cheung LE, Rose RJ, Kazazian HH, Jr. The Broad-Spectrum Antiviral Protein ZAP Restricts Human Retrotransposition. PLoS Genet. 2015;11(5):e1005252.
289.Niewiadomska AM, Tian C, Tan L, Wang T, Sarkis PT, Yu XF. Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association. J Virol. 2007;81(17):9577-83.
290.MacDuff DA, Demorest ZL, Harris RS. AID can restrict L1 retrotransposition suggesting a dual role in innate and adaptive immunity. Nucleic Acids Res. 2009;37(6):1854-67.
291.Ikeda T, Abd El Galil KH, Tokunaga K, Maeda K, Sata T, Sakaguchi N, et al. Intrinsic restriction activity by apolipoprotein B mRNA editing enzyme APOBEC1 against the mobility of autonomous retrotransposons. Nucleic Acids Res. 2011;39(13):5538-54.

292.Bogerd HP, Wiegand HL, Doehle BP, Lueders KK, Cullen BR. APOBEC3A and APOBEC3B are potent inhibitors of LTR-retrotransposon function in human cells. Nucleic Acids Res. 2006;34(1):89-95.

293.Cantone I, Fisher AG. Epigenetic programming and reprogramming during development. Nat Struct Mol Biol. 2013;20(3):282-9.

294.Fadloun A, Le Gras S, Jost B, Ziegler-Birling C, Takahashi H, Gorab E, et al. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. Nat Struct Mol Biol. 2013;20(3):332-8.

295.Molaro A, Falciatori I, Hodges E, Aravin AA, Marran K, Rafii S, et al. Two waves of de novo methylation during mouse germ cell development. Genes Dev. 2014;28(14):1544-9.

296.Goke J, Lu X, Chan YS, Ng HH, Ly LH, Sachs F, et al. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell Stem Cell. 2015;16(2):135-41.

297.Wang J, Xie G, Singh M, Ghanbarian AT, Rasko T, Szvetnik A, et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. Nature. 2014;516(7531):405-9.

298.Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet. 2010;42(7):631-4.

299.Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell. 2013;153(5):1134-48.

300.Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, et al. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. Dev Cell. 2013;25(1):69-80.

301.Kelley D, Rinn J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. 2012;13(11):R107.

302.Maksakova IA, Romanish MT, Gagnier L, Dunn CA, van de Lagemaat LN, Mager DL. Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. PLoS Genet. 2006;2(1):e2.

303.Hancks DC, Kazazian HH, Jr. Roles for retrotransposon insertions in human disease. Mob DNA. 2016;7:9.
304.Wang J, Huang J, Shi G. Retrotransposons in pluripotent stem cells. Cell Regen. 2020;9(1):4.
305.Yang N, Kazazian HH, Jr. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. Nat Struct Mol Biol. 2006;13(9):763-71.
306.Baillie JK, Barnett MW, Upton KR, Gerhardt DJ, Richmond TA, De Sapio F, et al. Somatic retrotransposition alters the genetic landscape of the human brain. Nature. 2011;479(7374):534-7.
307.Garza R, Atacho DAM, Adami A, Gerdes P, Vinod M, Hsieh P, et al. LINE-1 retrotransposons drive human neuronal transcriptome complexity and functional diversification. Sci Adv. 2023;9(44):eadh9543.
308.Playfoot CJ, Duc J, Sheppard S, Dind S, Coudray A, Planet E, et al. Transposable elements and their KZFP controllers are drivers of transcriptional innovation in the developing human brain. Genome Res. 2021;31(9):1531-45.
309.Macia A, Widmann TJ, Heras SR, Ayllon V, Sanchez L, Benkaddour-Boumzaouad M, et al. Engineered LINE-1 retrotransposition in nondividing human neurons. Genome Res. 2017;27(3):335-48.
310.Coufal NG, Garcia-Perez JL, Peng GE, Yeo GW, Mu Y, Lovci MT, et al. L1 retrotransposition in human neural progenitor cells. Nature. 2009;460(7259):1127-31.
311.Upton KR, Gerhardt DJ, Jesuadian JS, Richardson SR, Sanchez-Luque FJ, Bodea GO, et al. Ubiquitous L1 mosaicism in hippocampal neurons. Cell. 2015;161(2):228-39.
312.Evrony GD, Lee E, Park PJ, Walsh CA. Resolving rates of mutation in the brain using single-neuron genomics. Elife. 2016;5.
313.Kanholm T, Rentia U, Hadley M, Karlow JA, Cox OL, Diab N, et al. Oncogenic Transformation Drives DNA Methylation Loss and Transcriptional Activation of Transposable Element Loci. Cancer Res. 2023.
314.Rycaj K, Plummer JB, Yin B, Li M, Garza J, Radvanyi L, et al. Cytotoxicity of human endogenous retrovirus K-specific T cells toward autologous ovarian cancer cells. Clin Cancer Res. 2015;21(2):471-83.

315.Au L, Hatipoglu E, Robert de Massy M, Litchfield K, Beattie G, Rowan A, et al. Determinants of anti-PD-1 response and resistance in clear cell renal cell carcinoma. Cancer Cell. 2021;39(11):1497-518 e11.

316.Wang-Johanning F, Liu J, Rycaj K, Huang M, Tsai K, Rosen DG, et al. Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer. Int J Cancer. 2007;120(1):81-90.

317.Saini SK, Orskov AD, Bjerregaard AM, Unnikrishnan A, Holmberg-Thyden S, Borch A, et al. Human endogenous retroviruses form a reservoir of T cell targets in hematological cancers. Nat Commun. 2020;11(1):5660.

318.Wang-Johanning F, Rycaj K, Plummer JB, Li M, Yin B, Frerich K, et al. Immunotherapeutic potential of anti-human endogenous retrovirus-K envelope protein antibodies in targeting breast tumors. J Natl Cancer Inst. 2012;104(3):189-210.

319.Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell. 2015;160(1-2):48-61.

320.Attig J, Young GR, Stoye JP, Kassiotis G. Physiological and Pathological Transcriptional Activation of Endogenous Retroelements Assessed by RNA-Sequencing of B Lymphocytes. Front Microbiol. 2017;8:2489.

321.Larsen JM, Christensen IJ, Nielsen HJ, Hansen U, Bjerregaard B, Talts JF, et al. Syncytin immunoreactivity in colorectal cancer: potential prognostic impact. Cancer Lett. 2009;280(1):44-9.

322.Rodic N, Sharma R, Sharma R, Zampella J, Dai L, Taylor MS, et al. Long interspersed element-1 protein expression is a hallmark of many human cancers. Am J Pathol. 2014;184(5):1280-6.

323.Kong Y, Rose CM, Cass AA, Williams AG, Darwish M, Lianoglou S, et al. Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. Nat Commun. 2019;10(1):5228.

324.Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J, et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat Genet. 2020;52(3):306-19.

325.Solyom S, Ewing AD, Rahrmann EP, Doucet T, Nelson HH, Burns MB, et al. Extensive somatic L1 retrotransposition in colorectal tumors. Genome Res. 2012;22(12):2328-38.
326.Doucet-O'Hare TT, Rodic N, Sharma R, Darbari I, Abril G, Choi JA, et al. LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. Proc Natl Acad Sci U S A. 2015;112(35):E4894-900.
327.Helman E, Lawrence MS, Stewart C, Sougnez C, Getz G, Meyerson M. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. Genome Res. 2014;24(7):1053-63.
328.Jang HS, Shah NM, Du AY, Dailey ZZ, Pehrsson EC, Godoy PM, et al. Transposable elements drive widespread expression of oncogenes in human cancers. Nat Genet. 2019;51(4):611-7.
329.Attig J, Young GR, Hosie L, Perkins D, Encheva-Yokoya V, Stoye JP, et al. LTR retroelement expansion of the human cancer transcriptome and immunopeptidome revealed by de novo transcript assembly. Genome Res. 2019;29(10):1578-90.
330.Li M, Radvanyi L, Yin B, Li J, Chivukula R, Lin K, et al. Downregulation of Human Endogenous Retrovirus Type K (HERV-K) Viral env RNA in Pancreatic Cancer Cells Decreases Cell Proliferation and Tumor Growth. Clin Cancer Res. 2017;23(19):5892-911.
331.Chan SM, Sapir T, Park SS, Rual JF, Contreras-Galindo R, Reiner O, et al. The HERV-K accessory protein Np9 controls viability and migration of teratocarcinoma cells. PLoS One. 2019;14(2):e0212970.
332.Canadas I, Thummalapalli R, Kim JW, Kitajima S, Jenkins RW, Christensen CL, et al. Tumor innate immunity primed by specific interferon-stimulated endogenous retroviruses. Nat Med. 2018;24(8):1143-50.
333.Kawahara Y, Nishikura K. Extensive adenosine-to-inosine editing detected in Alu repeats of antisense RNAs reveals scarcity of sense-antisense duplex formation. FEBS Lett. 2006;580(9):2301-5.
334.Kim Y, Park J, Kim S, Kim M, Kang MG, Kwak C, et al. PKR Senses Nuclear and Mitochondrial Signals by Interacting with Endogenous Double-Stranded RNAs. Mol Cell. 2018;71(6):1051-63 e6.

335.Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, et al. DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. Cell. 2015;162(5):961-73.

336.Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, et al. Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. Cell. 2015;162(5):974-86.

337.Smith CC, Beckermann KE, Bortone DS, De Cubas AA, Bixby LM, Lee SJ, et al. Endogenous retroviral signatures predict immunotherapy response in clear cell renal cell carcinoma. J Clin Invest. 2018;128(11):4804-20.

338.Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. Science. 2016;351(6277):1083-7.

339.Cao Y, Chen G, Wu G, Zhang X, McDermott J, Chen X, et al. Widespread roles of enhancer-like transposable elements in cell identity and long-range genomic interactions. Genome Res. 2019;29(1):40-52.

340.Bogdan L, Barreiro L, Bourque G. Transposable elements have contributed human regulatory regions that are activated upon bacterial infection. Philos Trans R Soc Lond B Biol Sci. 2020;375(1795):20190332.

341.Adoue V, Binet B, Malbec A, Fourquet J, Romagnoli P, van Meerwijk JPM, et al. The Histone Methyltransferase SETDB1 Controls T Helper Cell Lineage Integrity by Repressing Endogenous Retroviruses. Immunity. 2019;50(3):629-44 e8.

342.Kapitonov VV, Jurka J. RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol. 2005;3(6):e181.

343.Huang S, Tao X, Yuan S, Zhang Y, Li P, Beilinson HA, et al. Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. Cell. 2016;166(1):102-14.

344.Fugmann SD, Messier C, Novack LA, Cameron RA, Rast JP. An ancient evolutionary origin of the Rag1/2 gene locus. Proc Natl Acad Sci U S A. 2006;103(10):3728-33.

345.Odaka T, Ikeda H, Akatsuka T. Restricted expression of endogenous N-tropic XC-positive leukemia virus in hybrids between G and AKR mice: an effect of the Fv-4r gene. Int J Cancer. 1980;25(6):757-62.

346.Wu T, Yan Y, Kozak CA. Rmcf2, a xenotropic provirus in the Asian mouse species Mus castaneus, blocks infection by polytropic mouse gammaretroviruses. J Virol. 2005;79(15):9677-84.

347.Robinson HL, Lamoreux WF. Expression of endogenous ALV antigens and susceptibility to subgroup E ALV in three strains of chickens (endogenous avian C-type virus). Virology. 1976;69(1):50-62.

348.McDougall AS, Terry A, Tzavaras T, Cheney C, Rojko J, Neil JC. Defective endogenous proviruses are expressed in feline lymphoid cells: evidence for a role in natural resistance to subgroup B feline leukemia viruses. J Virol. 1994;68(4):2151-60.

349.Frank JA, Singh M, Cullen HB, Kirou RA, Benkaddour-Boumzaouad M, Cortes JL, et al. Evolution and antiviral activity of a human protein of retroviral origin. Science. 2022;378(6618):422-8.

350.Lima-Junior DS, Krishnamurthy SR, Bouladoux N, Collins N, Han SJ, Chen EY, et al. Endogenous retroviruses promote homeostatic and inflammatory responses to the microbiota. Cell. 2021.

351.Zhao Y, Oreskovic E, Zhang Q, Lu Q, Gilman A, Lin YS, et al. Transposon-triggered innate immune response confers cancer resistance to the blind mole rat. Nat Immunol. 2021;22(10):1219-30.

352.Laumont CM, Vincent K, Hesnard L, Audemard E, Bonneil E, Laverdure JP, et al. Noncoding regions are the main source of targetable tumor-specific antigens. Sci Transl Med. 2018;10(470).

353.Ehx G, Larouche JD, Durette C, Laverdure JP, Hesnard L, Vincent K, et al. Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes. Immunity. 2021;54(4):737-52 e10.

354.Chong C, Muller M, Pak H, Harnett D, Huber F, Grun D, et al. Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. Nat Commun. 2020;11(1):1293.

355.Bonaventura P, Alcazer V, Mutez V, Tonon L, Martin J, Chuvin N, et al. Identification of shared tumor epitopes from endogenous retroviruses inducing high-avidity cytotoxic T cells for cancer immunotherapy. Sci Adv. 2022;8(4):eabj3671.

356.Liu M, Thomas SL, DeWitt AK, Zhou W, Madaj ZB, Ohtani H, et al. Dual Inhibition of DNA and Histone Methyltransferases Increases Viral Mimicry in Ovarian Cancer Cells. Cancer Res. 2018;78(20):5754-66.

357.Morel KL, Sheahan AV, Burkhart DL, Baca SC, Boufaied N, Liu Y, et al. EZH2 inhibition activates a dsRNA-STING-interferon stress axis that potentiates response to PD-1 checkpoint blockade in prostate cancer. Nat Cancer. 2021;2(4):444-56.

358.Deblois G, Tonekaboni SAM, Grillo G, Martinez C, Kao YI, Tai F, et al. Epigenetic Switch-Induced Viral Mimicry Evasion in Chemotherapy-Resistant Breast Cancer. Cancer Discov. 2020;10(9):1312-29.

359.Ishizuka JJ, Manguso RT, Cheruiyot CK, Bi K, Panda A, Iracheta-Vellve A, et al. Loss of ADAR1 in tumours overcomes resistance to immune checkpoint blockade. Nature. 2019;565(7737):43-8.

360.Mehdipour P, Marhon SA, Ettayebi I, Chakravarthy A, Hosseini A, Wang Y, et al. Publisher Correction: Epigenetic therapy induces transcription of inverted SINEs and ADAR1 dependency. Nature. 2021;591(7850):E20.

361.Goedert JJ, Sauter ME, Jacobson LP, Vessella RL, Hilgartner MW, Leitman SF, et al. High prevalence of antibodies against HERV-K10 in patients with testicular cancer but not with AIDS. Cancer Epidemiol Biomarkers Prev. 1999;8(4 Pt 1):293-6.

362.Buscher K, Trefzer U, Hofmann M, Sterry W, Kurth R, Denner J. Expression of human endogenous retrovirus K in melanomas and melanoma cell lines. Cancer Res. 2005;65(10):4172-80.

363.Humer J, Waltenberger A, Grassauer A, Kurz M, Valencak J, Rapberger R, et al. Identification of a melanoma marker derived from melanoma-associated endogenous retroviruses. Cancer Res. 2006;66(3):1658-63.

364.Ng KW, Boumelha J, Enfield KSS, Almagro J, Cha H, Pich O, et al. Antibodies against endogenous retroviruses promote lung cancer immunotherapy. Nature. 2023;616(7957):563-73.

365.Shah NM, Jang HJ, Liang Y, Maeng JH, Tzeng SC, Wu A, et al. Pan-cancer analysis identifies tumor-specific antigens derived from transposable elements. Nat Genet. 2023;55(4):631-9.

366.Merlotti A, Sadacca B, Arribas YA, Ngoma M, Burbage M, Goudot C, et al. Noncanonical splicing junctions between exons and transposable elements represent a source of immunogenic recurrent neo-antigens in patients with lung cancer. Sci Immunol. 2023;8(80):eabm6359.

367.Burbage M, Rocanin-Arjo A, Baudon B, Arribas YA, Merlotti A, Rookhuizen DC, et al. Epigenetically controlled tumor antigens derived from splice junctions between exons and transposable elements. Sci Immunol. 2023;8(80):eabm6360.

368.Aagaard L, Villesen P, Kjeldbjerg AL, Pedersen FS. The approximately 30-million-year-old ERVPb1 envelope gene is evolutionarily conserved among hominoids and Old World monkeys. Genomics. 2005;86(6):685-91.

369.Young GR, Ploquin MJ, Eksmond U, Wadwa M, Stoye JP, Kassiotis G. Negative selection by an endogenous retrovirus promotes a higher-avidity CD4+ T cell response to retroviral infection. PLoS Pathog. 2012;8(5):e1002709.

370.Sverdlov ED. Perpetually mobile footprints of ancient infections in human genome. FEBS Lett. 1998;428(1-2):1-6.

371.de Koning AP, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. PLoS Genet. 2011;7(12):e1002384.

372.Tristem M. Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. J Virol. 2000;74(8):3715-30.

373.Vargiu L, Rodriguez-Tome P, Sperber GO, Cadeddu M, Grandi N, Blikstad V, et al. Classification and characterization of human endogenous retroviruses; mosaic forms are common. Retrovirology. 2016;13:7.

374.Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet. 2011;13(1):36-46.

375.Argueso JL, Westmoreland J, Mieczkowski PA, Gawel M, Petes TD, Resnick MA. Double-strand breaks associated with repetitive DNA can reshape the genome. Proc Natl Acad Sci U S A. 2008;105(33):11845-50.

376.Mills RE, Bennett EA, Iskow RC, Devine SE. Which transposable elements are active in the human genome? Trends Genet. 2007;23(4):183-91.

377.Deniz O, Frost JM, Branco MR. Regulation of transposable elements by DNA modifications. Nat Rev Genet. 2019;20(7):417-31.

378.Robbez-Masson L, Tie CHC, Conde L, Tunbak H, Husovsky C, Tchasovnikarova IA, et al. The HUSH complex cooperates with TRIM28 to repress young retrotransposons and new genes. Genome Res. 2018;28(6):836-45.

379.Beyer U, Moll-Rocek J, Moll UM, Dobbelstein M. Endogenous retrovirus drives hitherto unknown proapoptotic p63 isoforms in the male germ line of humans and great apes. Proc Natl Acad Sci U S A. 2011;108(9):3624-9.

380.Blaise S, de Parseval N, Benit L, Heidmann T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. Proc Natl Acad Sci U S A. 2003;100(22):13013-8.

381.Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5.

382.Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009;462(7271):315-22.

383.St-Pierre C, Trofimov A, Brochu S, Lemieux S, Perreault C. Differential Features of AIRE-Induced and AIRE-Independent Promiscuous Gene Expression in Thymic Epithelial Cells. J Immunol. 2015;195(2):498-506.

384.Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20.

385.Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016;34(5):525-7.

386.Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. Bioinformatics. 2006;22(12):1540-2.

387.Yanai I, Benjamin H, Shmoish M, Chalifa-Caspi V, Shklar M, Ophir R, et al. Genome-wide midrange transcription profiles reveal expression level relationships in human tissue specification. Bioinformatics. 2005;21(5):650-9.

388.Granados DP, Sriranganadane D, Daouda T, Zieger A, Laumont CM, Caron-Lizotte O, et al. Impact of genomic polymorphisms on the repertoire of human MHC class I-associated peptides. Nat Commun. 2014;5:3600.

389.Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15-21.

390.Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-2.

391.Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25(16):2078-9.

392.Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv e-prints [Internet]. 2012 July 01, 2012. Available from: https://ui.adsabs.harvard.edu/abs/2012arXiv1207.3907G.

393.Daouda T, Perreault C, Lemieux S. pyGeno: A Python package for precision medicine and proteogenomics. F1000Res. 2016;5:381.

394.Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics. 2016;32(4):511-7.

395.Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. J Immunol. 2017;199(9):3360-8.

396.Marcais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6):764-70.

397.Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.

398.Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.

399.Ogishi M, Yotsuyanagi H. Quantitative Prediction of the Landscape of T Cell Epitope Immunogenicity in Sequence Space. Front Immunol. 2019;10:827.

400.Adamopoulou E, Tenzer S, Hillen N, Klug P, Rota IA, Tietz S, et al. Exploring the MHC-peptide matrix of central tolerance in the human thymus. Nat Commun. 2013;4:2039.

401.Bollard CM, Gottschalk S, Leen AM, Weiss H, Straathof KC, Carrum G, et al. Complete responses of relapsed lymphoma following genetic modification of tumor-antigen presenting cells and T-lymphocyte transfer. Blood. 2007;110(8):2838-45.

402.Wolfl M, Greenberg PD. Antigen-specific activation and cytokine-facilitated expansion of naive, human CD8+ T cells. Nat Protoc. 2014;9(4):950-66.

403.Janelle V, Carli C, Taillefer J, Orio J, Delisle JS. Defining novel parameters for the optimal priming and expansion of minor histocompatibility antigen-specific T cells in culture. J Transl Med. 2015;13:123.

404.Fergusson JR, Morgan MD, Bruchard M, Huitema L, Heesters BA, van Unen V, et al. Maturing Human CD127+ CCR7+ PDL1+ Dendritic Cells Express AIRE in the Absence of Tissue Restricted Antigens. Front Immunol. 2018;9:2902.

405.Drukker M, Katz G, Urbach A, Schuldiner M, Markel G, Itskovitz-Eldor J, et al. Characterization of the expression of MHC proteins in human embryonic stem cells. Proc Natl Acad Sci U S A. 2002;99(15):9864-9.

406.Klein L, Hinterberger M, Wirnsberger G, Kyewski B. Antigen presentation in the thymus for positive selection and central tolerance induction. Nat Rev Immunol. 2009;9(12):833-44.

407.Inglesfield S, Cosway EJ, Jenkinson WE, Anderson G. Rethinking Thymic Tolerance: Lessons from Mice. Trends Immunol. 2019;40(4):279-91.

408.Sacha JB, Kim IJ, Chen L, Ullah JH, Goodwin DA, Simmons HA, et al. Vaccination with cancer- and HIV infection-associated endogenous retrotransposable elements is safe and immunogenic. J Immunol. 2012;189(3):1467-79.

409.Granados DP, Laumont CM, Thibault P, Perreault C. The nature of self for T cells-a systems-level perspective. Curr Opin Immunol. 2015;34:1-8.

410.Yewdell JW, Dersh D, Fahraeus R. Peptide Channeling: The Key to MHC Class I Immunosurveillance? Trends Cell Biol. 2019;29(12):929-39.

411.Jung J, Lee S, Cho HS, Park K, Ryu JW, Jung M, et al. Bioinformatic analysis of regulation of natural antisense transcripts by transposable elements in human mRNA. Genomics. 2019;111(2):159-66.

412.Treger RS, Pope SD, Kong Y, Tokuyama M, Taura M, Iwasaki A. The Lupus Susceptibility Locus Sgp3 Encodes the Suppressor of Endogenous Retrovirus Expression SNERV. Immunity. 2019;50(2):334-47 e9.

413.De Cecco M, Ito T, Petrashen AP, Elias AE, Skvir NJ, Criscione SW, et al. L1 drives IFN in senescent cells and promotes age-associated inflammation. Nature. 2019;566(7742):73-8.

414.Attig J, Young GR, Hosie L, Perkins D, Encheva-Yokoya V, Stoye JP, et al. LTR retroelement expansion of the human cancer transcriptome and immunopeptidome revealed by de novo transcript assembly. Genome Res. 2019.

415.Smith CC, Selitsky SR, Chai S, Armistead PM, Vincent BG, Serody JS. Alternative tumour-specific antigens. Nat Rev Cancer. 2019.

416.Gainetdinov I, Skvortsova Y, Kondratieva S, Funikov S, Azhikina T. Two modes of targeting transposable elements by piRNA pathway in human testis. RNA. 2017;23(11):1614-25.

417.Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wossidlo M, Wesche DJ, et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. Nature. 2015;522(7555):221-5.

418.Abramson J, Anderson G. Thymic Epithelial Cells. Annu Rev Immunol. 2017;35:85-118.

419.Ucar O, Rattay K. Promiscuous Gene Expression in the Thymus: A Matter of Epigenetics, miRNA, and More? Front Immunol. 2015;6:93.

420.Chong C, Müller M, Pak H, Harnett D, Huber F, Grun D, et al. Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. bioRxiv. 2019.

421.Maiga A, Lemieux S, Pabst C, Lavallee VP, Bouvier M, Sauvageau G, et al. Transcriptome analysis of G protein-coupled receptors in distinct genetic subgroups of acute myeloid leukemia: identification of potential disease-specific targets. Blood Cancer J. 2016;6(6):e431.

422.Pabst C, Bergeron A, Lavallee VP, Yeh J, Gendron P, Norddahl GL, et al. GPR56 identifies primary human acute myeloid leukemia cells with high repopulating potential in vivo. Blood. 2016;127(16):2018-27.

423.Boehm T. Evolution of vertebrate immunity. Curr Biol. 2012;22(17):R722-32.

424.Boehm T, Swann JB. Origin and evolution of adaptive immunity. Annu Rev Anim Biosci. 2014;2:259-83.

425.Suo C, Dann E, Goh I, Jardine L, Kleshchevnikov V, Park JE, et al. Mapping the developing human immune system across organs. Science. 2022;376(6597):eabo0510.

426.Terra R, Louis I, Le Blanc R, Ouellet S, Zuniga-Pflucker JC, Perreault C. T-cell generation by lymph node resident progenitor cells. Blood. 2005;106(1):193-200.

427.Blais ME, Brochu S, Giroux M, Belanger MP, Dulude G, Sekaly RP, et al. Why T cells of thymic versus extrathymic origin are functionally different. J Immunol. 2008;180(4):2299-312.

428.Zuniga-Pflucker JC, Longo DL, Kruisbeek AM. Positive selection of CD4-CD8+ T cells in the thymus of normal mice. Nature. 1989;338(6210):76-8.

429.Breed ER, Lee ST, Hogquist KA. Directing T cell fate: How thymic antigen presenting cells coordinate thymocyte selection. Semin Cell Dev Biol. 2018;84:2-10.

430.Dervovic D, Zuniga-Pflucker JC. Positive selection of T cells, an in vitro view. Semin Immunol. 2010;22(5):276-86.

431.Lebel ME, Coutelier M, Galipeau M, Kleinman CL, Moon JJ, Melichar HJ. Differential expression of tissue-restricted antigens among mTEC is associated with distinct autoreactive T cell fates. Nat Commun. 2020;11(1):3734.

432.Srinivasan J, Lancaster JN, Singarapu N, Hale LP, Ehrlich LIR, Richie ER. Age-Related Changes in Thymic Central Tolerance. Front Immunol. 2021;12:676236.

433.Cheng M, Anderson MS. Thymic tolerance as a key brake on autoimmunity. Nat Immunol. 2018;19(7):659-64.

434.Nitta T, Ohigashi I, Nakagawa Y, Takahama Y. Cytokine crosstalk for thymic medulla formation. Curr Opin Immunol. 2011;23(2):190-7.

435.Malhotra D, Linehan JL, Dileepan T, Lee YJ, Purtha WE, Lu JV, et al. Tolerance is established in polyclonal CD4(+) T cells by distinct mechanisms, according to self-peptide expression patterns. Nat Immunol. 2016;17(2):187-95.

436.Lkhagvasuren E, Sakata M, Ohigashi I, Takahama Y. Lymphotoxin beta receptor regulates the development of CCL21-expressing subset of postnatal medullary thymic epithelial cells. J Immunol. 2013;190(10):5110-7.

437.Laan M, Salumets A, Klein A, Reintamm K, Bichele R, Peterson H, et al. Post-Aire Medullary Thymic Epithelial Cells and Hassall's Corpuscles as Inducers of Tonic Pro-Inflammatory Microenvironment. Front Immunol. 2021;12:635569.

438.Ramsey C, Winqvist O, Puhakka L, Halonen M, Moro A, Kampe O, et al. Aire deficient mice develop multiple features of APECED phenotype and show altered immune response. Hum Mol Genet. 2002;11(4):397-409.

439.Ginhoux F, Guilliams M, Merad M. Expanding dendritic cell nomenclature in the single-cell era. Nat Rev Immunol. 2022;22(2):67-8.

440.Larouche JD, Trofimov A, Hesnard L, Ehx G, Zhao Q, Vincent K, et al. Widespread and tissue-specific expression of endogenous retroelements in human somatic tissues. Genome Med. 2020;12(1):40.

441.Carter JA, Stromich L, Peacey M, Chapin SR, Velten L, Steinmetz LM, et al. Transcriptomic diversity in human medullary thymic epithelial cells. Nat Commun. 2022;13(1):4296.

442.Kassiotis G. The Immunological Conundrum of Endogenous Retroelements. Annu Rev Immunol. 2023;41:99-125.

443.Lefkopoulos S, Polyzou A, Derecka M, Bergo V, Clapes T, Cauchy P, et al. Repetitive Elements Trigger RIG-I-like Receptor Signaling that Regulates the Emergence of Hematopoietic Stem and Progenitor Cells. Immunity. 2020;53(5):934-51 e9.

444.Brocks D, Chomsky E, Mukamel Z, Lifshitz A, Tanay A. Single cell analysis reveals dynamics of transposable element transcription following epigenetic de-repression. bioRxiv. 2018:462853.

445.Hosokawa H, Rothenberg EV. How transcription factors drive choice of the T cell fate. Nat Rev Immunol. 2021;21(3):162-76.

446.Baik S, Sekai M, Hamazaki Y, Jenkinson WE, Anderson G. Relb acts downstream of medullary thymic epithelial stem cells and is essential for the emergence of RANK(+) medullary epithelial progenitors. Eur J Immunol. 2016;46(4):857-62.

447.Akiyama T, Shimo Y, Yanai H, Qin J, Ohshima D, Maruyama Y, et al. The tumor necrosis factor family receptors RANK and CD40 cooperatively establish the thymic medullary microenvironment and self-tolerance. Immunity. 2008;29(3):423-37.

448.Yamazaki Y, Urrutia R, Franco LM, Giliani S, Zhang K, Alazami AM, et al. PAX1 is essential for development and function of the human thymus. Sci Immunol. 2020;5(44).
449.Meixner A, Karreth F, Kenner L, Wagner EF. JunD regulates lymphocyte proliferation and T helper cell cytokine expression. EMBO J. 2004;23(6):1325-35.
450.Cisse B, Caton ML, Lehner M, Maeda T, Scheu S, Locksley R, et al. Transcription factor E2-2 is an essential and specific regulator of plasmacytoid dendritic cell development. Cell. 2008;135(1):37-48.
451.Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.
452.Luo Y, Hitz BC, Gabdank I, Hilton JA, Kagda MS, Lam B, et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res. 2020;48(D1):D882-D9.
453.Taveirne S, Wahlen S, Van Loocke W, Kiekens L, Persyn E, Van Ammel E, et al. The transcription factor ETS1 is an important regulator of human NK cell development and terminal differentiation. Blood. 2020;136(3):288-98.
454.Kim N, Kim M, Yun S, Doh J, Greenberg PD, Kim TD, et al. MicroRNA-150 regulates the cytotoxicity of natural killers by targeting perforin-1. J Allergy Clin Immunol. 2014;134(1):195-203.
455.Gunturi A, Berg RE, Forman J. The role of CD94/NKG2 in innate and adaptive immunity. Immunol Res. 2004;30(1):29-34.
456.Madissoon E, Wilbrey-Clark A, Miragaia RJ, Saeb-Parsy K, Mahbubani KT, Georgakopoulos N, et al. scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. Genome Biol. 2019;21(1):1.
457.Sadeq S, Al-Hashimi S, Cusack CM, Werner A. Endogenous Double-Stranded RNA. Noncoding RNA. 2021;7(1).
458.Shapiro IE, Bassani-Sternberg M. The impact of immunopeptidomics: From basic research to clinical implementation. Semin Immunol. 2023;66:101727.
459.Vizcaino JA, Kubiniok P, Kovalchik KA, Ma Q, Duquette JD, Mongrain I, et al. The Human Immunopeptidome Project: A Roadmap to Predict and Treat Immune Diseases. Mol Cell Proteomics. 2020;19(1):31-49.

460.Kubiniok P, Marcu A, Bichmann L, Kuchenbecker L, Schuster H, Hamelin DJ, et al. Understanding the constitutive presentation of MHC class I immunopeptidomes in primary tissues. iScience. 2022;25(2):103768.

461.Robles AI, Larcher F, Whalin RB, Murillas R, Richie E, Gimenez-Conti IB, et al. Expression of cyclin D1 in epithelial tissues of transgenic mice results in epidermal hyperproliferation and severe thymic hyperplasia. Proc Natl Acad Sci U S A. 1996;93(15):7634-8.

462.Klug DB, Crouch E, Carter C, Coghlan L, Conti CJ, Richie ER. Transgenic expression of cyclin D1 in thymic epithelial precursors promotes epithelial and T cell development. J Immunol. 2000;164(4):1881-8.

463.Ohigashi I, Tanaka Y, Kondo K, Fujimori S, Kondo H, Palin AC, et al. Trans-omics Impact of Thymoproteasome in Cortical Thymic Epithelial Cells. Cell Rep. 2019;29(9):2901-16 e6.

464.Dumont-Lagace M, Daouda T, Depoers L, Zumer J, Benslimane Y, Brochu S, et al. Qualitative Changes in Cortical Thymic Epithelial Cells Drive Postpartum Thymic Regeneration. Front Immunol. 2019;10:3118.

465.Ghosh M, Gauger M, Marcu A, Nelde A, Denk M, Schuster H, et al. Guidance Document: Validation of a High-Performance Liquid Chromatography-Tandem Mass Spectrometry Immunopeptidomics Assay for the Identification of HLA Class I Ligands Suitable for Pharmaceutical Therapies. Mol Cell Proteomics. 2020;19(3):432-43.

466.Nanaware PP, Jurewicz MM, Clement CC, Lu L, Santambrogio L, Stern LJ. Distinguishing Signal From Noise in Immunopeptidome Studies of Limiting-Abundance Biological Samples: Peptides Presented by I-A(b) in C57BL/6 Mouse Thymus. Front Immunol. 2021;12:658601.

467.Baran-Gale J, Morgan MD, Maio S, Dhalla F, Calvo-Asensio I, Deadman ME, et al. Ageing compromises mouse thymus function and remodels epithelial cell differentiation. Elife. 2020;9.

468.Frank JA, Feschotte C. Co-option of endogenous viral sequences for host cell function. Curr Opin Virol. 2017;25:81-9.

469.Groger V, Cynis H. Human Endogenous Retroviruses and Their Putative Role in the Development of Autoimmune Disorders Such as Multiple Sclerosis. Front Microbiol. 2018;9:265.

470.Volkman HE, Stetson DB. The enemy within: endogenous retroelements and autoimmune disease. Nat Immunol. 2014;15(5):415-22.

471.Huntley S, Baggott DM, Hamilton AT, Tran-Gyamfi M, Yang S, Kim J, et al. A comprehensive catalog of human KRAB-associated zinc finger genes: insights into the evolutionary history of a large family of transcriptional repressors. Genome Res. 2006;16(5):669-77.

472.Mouri Y, Nishijima H, Kawano H, Hirota F, Sakaguchi N, Morimoto J, et al. NF-kappaB-inducing kinase in thymic stroma establishes central tolerance by orchestrating cross-talk with not only thymocytes but also dendritic cells. J Immunol. 2014;193(9):4356-67.

473.O'Sullivan BJ, Yekollu S, Ruscher R, Mehdi AM, Maradana MR, Chidgey AP, et al. Autoimmune-Mediated Thymic Atrophy Is Accelerated but Reversible in RelB-Deficient Mice. Front Immunol. 2018;9:1092.

474.Admon A. The biogenesis of the immunopeptidome. Semin Immunol. 2023;67:101766.

475.Shao W, Pedrioli PGA, Wolski W, Scurtescu C, Schmid E, Vizcaino JA, et al. The SysteMHC Atlas project. Nucleic Acids Res. 2018;46(D1):D1237-D47.

476.Weijer K, Uittenbogaart CH, Voordouw A, Couwenberg F, Seppen J, Blom B, et al. Intrathymic and extrathymic development of human plasmacytoid dendritic cell precursors in vivo. Blood. 2002;99(8):2752-9.

477.Boehm T, Morimoto R, Trancoso I, Aleksandrova N. Genetic conflicts and the origin of self/nonself-discrimination in the vertebrate immune system. Trends Immunol. 2023;44(5):372-83.

478.Harroud A, Hafler DA. Common genetic factors among autoimmune diseases. Science. 2023;380(6644):485-90.

479.Smit A, Hubley, R & Green, P. RepeatMasker Open-4.0. 2013-2015.

480.Melsted P, Booeshaghi AS, Gao F, Beltrame E, Lu L, Hjorleifsson KE, et al. Modular and efficient pre-processing of single-cell RNA-seq. bioRxiv. 2019:673285.

481.Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9(4):357-9.

482.Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10(2).

483.Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137.

484.Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Sci Rep. 2019;9(1):9354.

485.Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44(W1):W160-5.

486.Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. Genome Res. 2002;12(6):996-1006.

487.Shen L, Shao N, Liu X, Nestler E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics. 2014;15:284.

488.Amezquita RA, Lun ATL, Becht E, Carey VJ, Carpp LN, Geistlinger L, et al. Orchestrating single-cell analysis with Bioconductor. Nat Methods. 2020;17(2):137-45.

489.McCarthy DJ, Campbell KR, Lun AT, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics. 2017;33(8):1179-86.

490.Lun AT, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Res. 2016;5:2122.

491.Haghverdi L, Lun ATL, Morgan MD, Marioni JC. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol. 2018;36(5):421-7.

492.Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. Bioinformatics. 2014;30(19):2811-2.

493.Cowan JE, Malin J, Zhao Y, Seedhom MO, Harly C, Ohigashi I, et al. Myc controls a distinct transcriptional program in fetal thymic epithelial cells that determines thymus growth. Nat Commun. 2019;10(1):5498.

494.Kowalczyk MS, Tirosh I, Heckl D, Rao TN, Dixit A, Haas BJ, et al. Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. Genome Res. 2015;25(12):1860-72.

495.Mouse Genome Sequencing C, Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, et al. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002;420(6915):520-62.

496.Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 2018;172(4):650-65.

497.Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. Bioinformatics. 2011;27(7):1017-8.

498.Butts CT. network: A Package for Managing Relational Data in R. Journal of Statistical Software. 2008;24(2):1 - 36.

499.Abugessaisa I, Noguchi S, Hasegawa A, Kondo A, Kawaji H, Carninci P, et al. refTSS: A Reference Data Set for Human and Mouse Transcription Start Sites. J Mol Biol. 2019;431(13):2407-22.

500.Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, et al. vegan: Community Ecology Package. 2020;https://CRAN.R-project.org/package=vegan.

501.Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015;33(5):495-502.

502.Baran Y, Bercovich A, Sebe-Pedros A, Lubling Y, Giladi A, Chomsky E, et al. MetaCell: analysis of single-cell RNA-seq data using K-nn graph partitions. Genome Biol. 2019;20(1):206.

503.Stoeckle C, Rota IA, Tolosa E, Haller C, Melms A, Adamopoulou E. Isolation of myeloid dendritic cells and epithelial cells from human thymus. J Vis Exp. 2013(79):e50951.

504.Ma Y, Sun S, Shang X, Keller ET, Chen M, Zhou X. Integrative differential expression and gene set enrichment analysis using summary statistics for scRNA-seq studies. Nat Commun. 2020;11(1):1585.

505.Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15(2):R29.

506.Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.

507.Broseus L, Ritchie W. S-IRFindeR: stable and accurate measurement of intron retention. bioRxiv. 2020:2020.06.25.164699.

508.Kim MJ, Miller CM, Shadrach JL, Wagers AJ, Serwold T. Young, proliferative thymic epithelial cells engraft and function in aging thymuses. J Immunol. 2015;194(10):4784-95.

509.Sakashita A, Maezawa S, Takahashi K, Alavattam KG, Yukawa M, Hu YC, et al. Endogenous retroviruses drive species-specific germline transcriptomes in mammals. Nat Struct Mol Biol. 2020;27(10):967-77.

510.Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV, McCormick J, et al. The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. Cell. 2018;172(1-2):275-88 e18.

511.Bogu GK, Reverter F, Marti-Renom MA, Snyder MP, Guigó R. Atlas of transcriptionally active transposable elements in human adult tissues. bioRxiv. 2019:714212.

512.Jordan IK, Rogozin IB, Glazko GV, Koonin EV. Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet. 2003;19(2):68-72.

513.Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. 2016;17(1):239.

514.Sheng L, Rigo F, Bennett CF, Krainer AR, Hua Y. Comparison of the efficacy of MOE and PMO modifications of systemic antisense oligonucleotides in a severe SMA mouse model. Nucleic Acids Res. 2020;48(6):2853-65.

515.Marcu A, Bichmann L, Kuchenbecker L, Kowalewski DJ, Freudenmann LK, Backert L, et al. HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. J Immunother Cancer. 2021;9(4).

516.Noble WS. Mass spectrometrists should search only for peptides they care about. Nat Methods. 2015;12(7):605-8.

517.Nesvizhskii AI. Proteogenomics: concepts, applications and computational strategies. Nat Methods. 2014;11(11):1114-25.

518.Suarez-Alvarez B, Rodriguez RM, Calvanese V, Blanco-Gelaz MA, Suhr ST, Ortega F, et al. Epigenetic mechanisms regulate MHC and antigen processing molecules in human embryonic and induced pluripotent stem cells. PLoS One. 2010;5(4):e10192.

519.Hotta C, Nagata T, Nakazawa M, Fujimaki H, Yoshinari M, Minami M. Impaired expression of MHC class I molecules on mouse testicular germ cells is mainly caused by the post-transcriptional mechanism. Immunogenetics. 2000;51(8-9):624-31.

520.Harris MG, Hulseberg P, Ling C, Karman J, Clarkson BD, Harding JS, et al. Immune privilege of the CNS is not the consequence of limited antigen sampling. Sci Rep. 2014;4:4422.
521.Stetson DB, Ko JS, Heidmann T, Medzhitov R. Trex1 prevents cell-intrinsic initiation of autoimmunity. Cell. 2008;134(4):587-98.
522.Gall A, Treuting P, Elkon KB, Loo YM, Gale M, Jr., Barber GN, et al. Autoimmunity initiates in nonhematopoietic cells and progresses via lymphocytes in an interferon-dependent autoimmune disease. Immunity. 2012;36(1):120-31.
523.Thomas CA, Tejwani L, Trujillo CA, Negraes PD, Herai RH, Mesci P, et al. Modeling of TREX1-Dependent Autoimmune Disease using Human Stem Cells Highlights L1 Accumulation as a Source of Neuroinflammation. Cell Stem Cell. 2017;21(3):319-31 e8.
524.Herve CA, Lugli EB, Brand A, Griffiths DJ, Venables PJ. Autoantibodies to human endogenous retrovirus-K are frequently detected in health and disease and react with multiple epitopes. Clin Exp Immunol. 2002;128(1):75-82.
525.Mameli G, Astone V, Arru G, Marconi S, Lovato L, Serra C, et al. Brains and peripheral blood mononuclear cells of multiple sclerosis (MS) patients hyperexpress MS-associated retrovirus/HERV-W endogenous retrovirus, but not Human herpesvirus 6. J Gen Virol. 2007;88(Pt 1):264-74.
526.Laska MJ, Brudek T, Nissen KK, Christensen T, Moller-Larsen A, Petersen T, et al. Expression of HERV-Fc1, a human endogenous retrovirus, is increased in patients with active multiple sclerosis. J Virol. 2012;86(7):3713-22.
527.Yoshikawa H, Sundaramoorthy R, Mariyappa D, Jiang H, Lamond AI. Efficient and Rapid Analysis of Polysomes and Ribosomal Subunits in Cells and Tissues Using Ribo Mega-SEC. Bio Protoc. 2021;11(15):e4106.
528.Zeleniak A, Wiegand C, Liu W, McCormick C, K R, Alavi A, et al. De novo construction of T cell compartment in humanized mice engrafted with iPSC-derived thymus organoids. Nat Methods. 2022;19(10):1306-19.
529.Ramos SA, Armitage LH, Morton JJ, Alzofon N, Handler D, Kelly G, et al. Generation of functional thymic organoids from human pluripotent stem cells. Stem Cell Reports. 2023;18(4):829-40.

530.Filion MC, Proulx C, Bradley AJ, Devine DV, Sekaly RP, Decary F, et al. Presence in peripheral blood of healthy individuals of autoreactive T cells to a membrane antigen present on bone marrow-derived cells. Blood. 1996;88(6):2144-50.

531.Danke NA, Koelle DM, Yee C, Beheray S, Kwok WW. Autoreactive T cells in healthy individuals. J Immunol. 2004;172(10):5967-72.

532.Tenzer S, Wee E, Burgevin A, Stewart-Jones G, Friis L, Lamberth K, et al. Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance. Nat Immunol. 2009;10(6):636-46.

533.Chapuis AG, Ragnarsson GB, Nguyen HN, Chaney CN, Pufnock JS, Schmitt TM, et al. Transferred WT1-reactive CD8+ T cells can mediate antileukemic activity and persist in post-transplant patients. Sci Transl Med. 2013;5(174):174ra27.

534.Balachandran VP, Luksza M, Zhao JN, Makarov V, Moral JA, Remark R, et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. Nature. 2017;551(7681):512-6.

535.Carrasco Pro S, Lindestam Arlehamn CS, Dhanda SK, Carpenter C, Lindvall M, Faruqi AA, et al. Microbiota epitope similarity either dampens or enhances the immunogenicity of disease-associated antigenic epitopes. PLoS One. 2018;13(5):e0196551.

536.Lee JK, Stewart-Jones G, Dong T, Harlos K, Di Gleria K, Dorrell L, et al. T cell cross-reactivity and conformational changes during TCR engagement. J Exp Med. 2004;200(11):1455-66.

537.Wooldridge L, Laugel B, Ekeruche J, Clement M, van den Berg HA, Price DA, et al. CD8 controls T cell cross-reactivity. J Immunol. 2010;185(8):4625-32.

538.Cherkasova E, Scrivani C, Doh S, Weisman Q, Takahashi Y, Harashima N, et al. Detection of an Immunogenic HERV-E Envelope with Selective Expression in Clear Cell Kidney Cancer. Cancer Res. 2016;76(8):2177-85.

539.Takahashi Y, Harashima N, Kajigaya S, Yokoyama H, Cherkasova E, McCoy JP, et al. Regression of human kidney cancer following allogeneic stem cell transplantation is associated with recognition of an HERV-E antigen by T cells. J Clin Invest. 2008;118(3):1099-109.

540.Mullins CS, Linnebacher M. Endogenous retrovirus sequences as a novel class of tumor-specific antigens: an example of HERV-H env encoding strong CTL epitopes. Cancer Immunol Immunother. 2012;61(7):1093-100.

541.Wang-Johanning F, Radvanyi L, Rycaj K, Plummer JB, Yan P, Sastry KJ, et al. Human endogenous retrovirus K triggers an antigen-specific immune response in breast cancer patients. Cancer Res. 2008;68(14):5869-77.

542.Kazachenka A, Young GR, Attig J, Kordella C, Lamprianidou E, Zoulia E, et al. Epigenetic therapy of myelodysplastic syndromes connects to cellular differentiation independently of endogenous retroelement derepression. Genome Med. 2019;11(1):86.

543.Haen SP, Loffler MW, Rammensee HG, Brossart P. Towards new horizons: characterization, classification and implications of the tumour antigenic repertoire. Nat Rev Clin Oncol. 2020;17(10):595-610.

544.Bezu L, Kepp O, Cerrato G, Pol J, Fucikova J, Spisek R, et al. Trial watch: Peptide-based vaccines in anticancer therapy. Oncoimmunology. 2018;7(12):e1511506.

545.Parkhurst MR, Yang JC, Langan RC, Dudley ME, Nathan DA, Feldman SA, et al. T cells targeting carcinoembryonic antigen can mediate regression of metastatic colorectal cancer but induce severe transient colitis. Mol Ther. 2011;19(3):620-6.

546.Morgan RA, Yang JC, Kitano M, Dudley ME, Laurencot CM, Rosenberg SA. Case report of a serious adverse event following the administration of T cells transduced with a chimeric antigen receptor recognizing ERBB2. Mol Ther. 2010;18(4):843-51.

547.Johnson LA, Morgan RA, Dudley ME, Cassard L, Yang JC, Hughes MS, et al. Gene therapy with human and mouse T-cell receptors mediates cancer regression and targets normal tissues expressing cognate antigen. Blood. 2009;114(3):535-46.

548.Morgan RA, Chinnasamy N, Abate-Daga D, Gros A, Robbins PF, Zheng Z, et al. Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy. J Immunother. 2013;36(2):133-51.

549.Apavaloaei A, Hardy MP, Thibault P, Perreault C. The Origin and Immune Recognition of Tumor-Specific Antigens. Cancers (Basel). 2020;12(9).