SCUOLA DI DOTTORATO
UNIVERSITÀ DEGLI STUDI DI MILANO-BICOCCA

Dipartimento di / Department of

School of Medicine and faculty of Science

Dottorato di Ricerca in / PhD program Translation and Molecular Medicine Ciclo / Cycle XXXIV

Epigenetic role of transposable elements in human T lymphocytes identity and plasticity

Cognome / Surname Vadalà Nome / Name Rebecca

Matricola / Registration number 767773

Tutore / Tutor: Prof. Beatrice Bodega

Coordinatore / Coordinator: Prof. Andrea Biondi

ANNO ACCADEMICO / ACADEMIC YEAR 2020/2021

Table of contents

Chapter 1. General introduction

1.1 Transposable elements orchestrate cell identity and transcriptional plasticity

Transposable elements (TEs) are DNA repetitive sequences with ability to move into the genome that represent the major components of eukaryotic genomes (de Koning et al., 2011; Percharde et al., 2020; Wells & Feschotte, 2020), they were considered for many decades as "junk DNA". Barbara McClintock has been the first scientist provides the initial insight that these elements are "*normal components of the chromosome responsible for controlling, differentially, the time and type of activity of individual genes*" (MCCLINTOCK, 1956). Indeed, in 1940 Barbara McClintock discovered that DNA elements were able to transpose throughout the maize genome, influencing gene expression (McCLINTOCK, 1950; MCCLINTOCK, 1956). Following these observations, Britten and Davidson hypothesized that TEs could generate *cis*-regulatory regions able to control gene expression (Britten & Davidson, 1971).

Although the scientific community took several years to take in consideration these studies, nowadays it is accepted that TEs contribute to genetic variability and regulate gene expression (Scacheri & Scacheri, 2015). Transposable elements are found in several type of organisms, from bacteria to humans (Kazazian, 2004). TEs shaped genomes

through evolution, they are one of the major contributors of genome complexity among different species (Deniz et al., 2019). Most of transposons are pieces of evidence of ancient transposition events that were maintained during evolution (Darby & Sabunciyan, 2014), indeed, several rounds of reverse transcription of TEs could have promoted the different size and complexity of human genome (Kazazian, 2004). LINE1 and Alu elements represent the major driving force of human genome evolution, not only as source of genetic variability, but also as contributor of genomic structure and transcriptional regulation of gene regulatory networks (Kazazian, 2004; Lu et al., 2021; Marasca et al., 2020a). Furthermore, TEs' homologous recombination could induce chromosome rearrangements, mutations, inversions, deletions and translocations, modifying DNA sequence (Belancio et al., 2008).

1.1.1 *Transposable elements are divided in DNA transposons and retrotransposons*

Approximately 2% of human genome accounts for protein coding sequences, more than the half (66%) is composed by repetitive elements (Lander et al., 2001). Among repetitive elements, TEs represent 40 - 45% of human genome (de Koning et al., 2011; Lander et al., 2001; Marian, 2014); TEs represent a class of interspersed repeats (Smit, 1999).

Considering their ability to transpose, TEs could be divided into two classes: DNA transposons and retrotransposons. DNA

transposons represent the 3% of the human genome and are divided into three major subclasses and ten different superfamilies (Cleacuteldric Feschotte & Pritham, 2007). These sequences are able to move throughout the genome by a “cut and paste” mechanism; indeed, they encode for a transposase enzyme that excises the repeat unit and permits its integration into a new position in the genome (Munoz-Lopez & Garcia-Perez, 2010). Although DNA transposons insertions could occur in several genomic positions, more often it occurs near the parental insertion, a mechanism defined “local hopping” (Kazazian, 2004).

Retrotransposons cover around 42-45% of the human genome and are transpose via a “copy and paste” mechanism that involve an RNA molecule as intermediate (Luan et al., 1993; Whitcomb & Hughes, 1992). Retrotransposons can be divided into two groups: long terminal repeat-containing elements (LTRs) and non-LTRs (figure 1) (Kazazian & Moran, 1998). LTRs account for 9% of the human genome and their genomic sequence is characterized by long terminal direct repeats that embed transcriptional regulatory elements. Endogenous retroviruses (ERVs) belong to LTR retrotransposons, they present transcriptional regulatory elements similar to those of viral genomes: *gag*, *pol* and *env* genes. *Gag* encodes for structural proteins for the virus core and *pol* encode respectively for reverse transcriptase that produce cDNA from RNA sequence, and integrase that permit their insertion in new genomic position (Malik et al., 2000; Padeken et al., 2015). The *env* gene codify

for a non-functional protein so ERVs have lost their ability to infect (Wilkinson & Lowenstein, 1994). There are 20 different families of ERVs and four superfamilies that are ERV- class I, ERV(K) class II, ERV(L) class III, and MalR. ERV-L is the oldest while MalR is the most represented across the genome with 240000 copies (Malik et al., 2000). ERVs are autonomous elements able to retrotranspose in several mammals but not in humans with the possible exception for HERV-K (Doolittle et al., 1989; Jansz & Faulkner, 2021).

Non-LTRs transposons account for 33% of human genome and are classified in Short Interspersed Nuclear Elements (SINEs) and Long Interspersed Nuclear Elements (LINEs) (Casa & Gabellini, 2012). SINEs account for 13% of the human genome and comprise sequences that are 100-400 base pairs long (bp). SINEs derive from tRNA and 7SL RNA; Alu, that represents the most abundant superfamily, derives from 7SL RNA (Deininger, 2011; Tucker & Glaunsinger, 2017). SINEs are non-autonomous elements indeed they take advantage of LINE machinery for reverse transcription and retrotransposition (Padeken et al., 2015).

LINEs represent the most abundant class covering 20% of the human genome (Casa & Gabellini, 2012; Viollet et al., 2014). They are constituted by different superfamilies: LINE1, LINE2 and LINE3, broadly distributed in both intragenic and intergenic genomic regions. Only evolutionary young LINE1 superfamilies

are still active in human. A full length LINE1 element sequence is around 6 kb of length and it is constituted by a 5' untranslated region (5'UTR) with an internal Polymerase II promoter, two open reading frames, ORF1 and ORF2, and a 3'UTR that presents a polyadenylation site (Beck et al., 2014). Once LINE1 elements are transcribed from an autonomous promoter, the RNA is exported in the cytoplasm where is translated in ORF1p and ORF2p; ORF1p and ORF2p interact with LINE1 RNA forming a ribonucleoprotein complex. ORF1p owns a nucleic acid chaperone activity, ORF2p retains an endonuclease and reverse transcriptase activity. The ribonucleoprotein complex, is then exported in the nucleus to be integrated in a new genomic location. The ORFp2 with its endonuclease activity generates a single stranded nick exposing 3'-OH that is used as a primer to start reverse transcription by its reverse transcriptase activity ; this process is called target primed reverse transcription (TPRT) (Beck et al., 2014; Feng et al., 1996; Viollet et al., 2014). During the TPRT process very often LINE1 elements result in truncated, non-functional sequence enriched at 3' end with short length, 900-1000bp (Beck et al., 2014; Viollet et al., 2014).

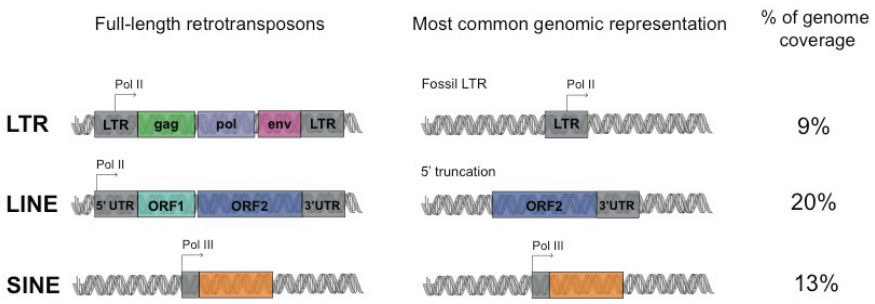


Figure 1. Schematic representation of retrotransposons sequences. Retrotransposons' major classes are: long terminal repeat-containing elements (LTRs), long intersperse nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). The full-length retrotransposons are on the left: in grey are the regulatory sequences, in color the protein coding sequence. In the middle are shown the most common genomic representation, in the right the percentage of genomic coverage (Marasca et al., 2020a).

1.1.2 *TEs expression is finely regulated*

The genetic information for the establishment of cellular identity and the regulation of processes are contained in chromosomes between protein coding genes and regulatory elements. Despite different cell types shared the same genome they have distinct morphological and structural features and respond differently to environmental, developmental and metabolic cues, due to a different epigenome (Bernstein et al., 2006). In 1940 Conrad Waddington introduce the word epigenetics that is defined as "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being" (Waddington, 2012). Nowadays we defined the epigenetics as the mechanisms that modulate gene expression without involving changes in DNA sequence (Peschansky & Wahlestedt, 2014). The major mechanisms known include DNA methylation, histone modifications, histone variants and

regulatory non-coding RNAs (Clark & Rager, 2020). TEs expression can be finely regulated through epigenetic mechanisms (Denli et al., 2015) as their transposition activity could generate detrimental insertion that can cause several diseases, as neurological and cancer diseases (Slotkin & Martienssen, 2007). Hence, the host genome developed many transcriptional and post- transcriptional mechanisms to control TEs expression and insertion (Payer & Burns, 2019). DNA methylation is a very well described mechanism of TEs silencing, in particular this mechanism allows the correct waves of expression of TEs in germ line development, embryogenesis (Bourc'his & Bestor, 2004). Indeed, different families of LTRs show different level of methylation during the processes of preimplantation during embryonic development (Wang et al., 2014), and LINE young elements have more methylation in human sperm but during fertilization become hypomethylated (Z. D. Smith et al., 2014). Several works, conduct in mouse models, have demonstrated that DNA methyltransferase 1 and 3 are required to the maintenance of retrotransposons' methylation (Yang & Wang, 2016). Indeed in male germ cells was demonstrated that the inactivation of DNA methyltransferase 3 cause a reactivation of TEs (in particular LINE1) producing a meiotic arrest and male sterility (Bourc'his & Bestor, 2004). While DNA methyltransferase 1 is required to maintain methylation of retrotransposons in somatic tissue of mouse embryos to carry out a correct development (Morgan et al., 1999).

Several pieces of evidence show that H3K9 methylation regulate TEs expression, indeed Martens et al, illustrated, in mouse embryonic stem cells, that mutations in genes responsible for repressive histone tail modification as *Suv39* lead TEs upregulation (Gendrel et al., 2002; Martens et al., 2005).Same results were demonstrated for LINE1: H3K9 methylation deposited by the H3K9 methyltransferase SETDB1, repress LINE1 expression a ; when LINE1 expression is de-regulated as occurs in several type of cancers as lung, colorectal, ovarian and pancreatic ones, the altered LINE1 retrotransposition co-participate to cancer progression (Liu et al., 2018; Payer & Burns, 2019).

Another level of regulation that occurs at post-transcriptional level, is the one mediated by non-coding RNAs. PIWI- interacting RNAs are one of the first defenses that the host developed to contrast TEs insertion (Molaro & Malik, 2016). As example, piRNA-PIWI complexes are able to drive, on retroelements insertion the deposition of repressive chromatin marks (Molaro & Malik, 2016).

1.1.3. *Transposable elements are novel regulatory players in the epigenetic regulation of cellular identity,*

Nowadays is accepted that TEs are strong source of regulatory sequences in eukaryotic genome, and that they have been co-opted to fine govern various biological mechanisms. Although at

the beginning the study of TEs' role in modulating gene expression were done following *de novo* TE insertion, recent works demonstrate that this type of activity spread through the genome and derived from TEs insertion occurred in the distant past that have been long fixed in the host genome (Chuong et al., 2016, 2017; Elbarbary et al., 2016; Cédric Feschotte, 2008). Genomically, TEs can create new or alternative promoters, Jordan et al., demonstrated that the 25% of examined human promoter contained TE-derived sequences (Jordan et al., 2003). TEs can disrupt *cis*-regulatory elements or generate new *cis*-regulatory elements as enhancers lineage-specific transcription factors (TFs) binding site, Sundaram et al showed that among TFs bindings sites analyzed TEs contributing to 5 to 40% of them (Sundaram et al., 2014). A seminal paper in the field, formally demonstrated by CAGE experiments (Cap Analysis Gene Expression) that 20% of transcripts expressed in mammalian tissues starts from a TEs; that TEs are expressed in tissue specific manner both in human and in mouse, (Faulkner et al., 2009; Jachowicz et al., 2017; Kapusta et al., 2013).

In line with these evidences, different TEs was found to be expressed in specific genes with definite spatiotemporal expression beating time of expression during the development (Lu et al., 2020). As an example, in placenta and oocytes there is a specific expression of LTRs, in general tissues of embryonic origin display an high percentage of transposable elements-

derived sequences in their transcriptomes (Rodriguez-Terrones et al., 2020).

Also, TEs could serve as motif or binding sites that participate in mRNA maturation regulating alternative splicing events generating different types of alternative splicing as intron retention or exon skipping (Faulkner et al., 2009; Ni et al., 2007). A TEs elements could also generates alternative polyadenylation sites or binding sites for RNA-binding protein (Attig et al., 2018; Perepelitsa-Belancio & Deininger, 2003; Roy-Engel et al., 2005; Thornburg et al., 2006).

Moreover TEs can generate and diversify non-coding regulatory RNA as long non-coding RNAs (lncRNAs) that in turn can regulate gene expression in *cis* and in *trans* (Kapusta et al., 2013; Kelley & Rinn, 2012). Furthermore, sequence derived from LTRs have been demonstrated to be involved the maintenance of pluripotency having a function of RNAs associated to enhancer regions (Fort et al., 2014). Also, ERV elements, and in particular MERVL are specifically expressed and regulate 2-cell stage in zygote development, regulating the expression of cell fate genes (Macfarlan et al., 2011).

Nowadays it is also known that TEs play a role in higher order chromatin folding (Hall et al., 2014). TEs could act as scaffold for 3D genome architecture, especially HERVs and SINEs elements are able to bind chromatin organizer (Ferrari et al., 2020; Lu et al., 2020, 2021; Schmidt et al., 2012; Y. Zhang et al., 2019).

In this context TEs could provide binding sites for cohesion and CCCTC-binding factor (CTCF), crucial proteins in higher order genome organization (Schmidt et al., 2012). In the last years transcribed ERV elements have been found to be involved in the formation of TAD boundaries in human pluripotent stem cells, indeed Zhang et al in 2019 demonstrated that the deletion of ERV elements promotes the deletion of corresponding TAD boundaries while their insertion generates new TAD boundaries (Y. Zhang et al., 2019). Importantly, RNA transcribed has been described to be associated with euchromatin regions, regulating chromatin accessibility (Hall., et al 2014).

All these evidences underline the importance of transposable elements as master regulators of human genome evolution and regulation.

1.1.4 LINE1 elements play a role in the maintenance of cell identity

LINE1 elements comprise almost the 17-18 % of the human genome (Lander et al., 2001). Nowadays there are several evidences that demonstrate the role of transcripts derived from LINE1 elements as epigenetic regulators able to modulate cell identity and plasticity, a function that is decoupled from their retrotransposition activity (Attig et al., 2018; Hall et al., 2014; Jachowicz et al., 2017; Lu et al., 2021; Percharde et al., 2018; Yuyang Lu et al., 2019). One of the first evidences showing a role of repeat-rich nuclear RNAs in the regulation of chromatin folding

was made by Hall et al, in 2014. They demonstrated that different repeated-rich RNAs, composed also by LINE1 derived transcripts, were found to be express at high level in interphase nuclei of primary cell lines and in cancer cells; demonstrating that euchromatin is embedded by LINE1 derived RNAs (Hall et al., 2014).

In 2017 Jachowicz et al., demonstrated that LINE1 expression, during early mouse embryos development, regulate chromatin accessibility; they demonstrated that modulating LINE1 elements expression through transcription-activator-like-effector (TALE) promotes a global chromatin remodeling. Indeed the silencing of LINE1 decreases chromatin accessibility while its prolonged activation avoids chromatin condensation; a phenomenon that occurs during developmental processes during the exit from 2-cells stage embryo (Jachowicz et al., 2017). This was one of the first evidence regarding the function of LINE1 independently from their retrotranpsosition activity.

Another very important step forward in understanding the role of LINE1 RNA in regulating gene expression and cell identity was made by Perchade et al., in 2018. For the first time, they revealed that in mouse embryonic stem cells a novel complex composed by LINE1 RNA in partnership with Nucleolin and Kap1 regulate gene expression in embryogenesis and stem cells differentiation (Percharde et al., 2018). In detail, one of the master regulator of embryonic stem cells development in mouse is Dux, whose

activation permits the passages into 2-cell embryo like cells in mouse, thanks also to the activation of ERV elements that promotes the expression of different transcripts specific to zygote genome activation. Dux and ERVs have to be repressed in blastocysts. At this stage LINE1 RNAs as nuclear RNA scaffold bind nucleolin and Kap1 and this complex are able to mediate Dux silencing with Kap1 and rRNA synthesis with nucleolin, promoting the exit from 2-cell stage in mouse embryonic stem cells (figure 2) (Percharde et al., 2018). Finally, LINE1 derived transcripts can be bound by RNA binding proteins (RBPs) participating to RNA processing machinery. MATR3 and PTBP1, for example, have different binding sites inside young evolutionary LINEs, they form ribonucleocomplexes able to act as splicing repressors promoting alternative splicing of tissue specific transcripts (Attig et al., 2018).

More evidences regarding the role of TEs and specifically LINE1 in organizing chromatin tridimensional space came out in the last few years. Lu et al., in 2020, assessed that, in embryonic stem cells, different classes of TEs, SINEs, LINEs and low complexity repeats are enriched in genes with different functions expressed in distinct developmental stages; SINEs localize in genes with housekeeping functions while LINE1 elements are in genes with for terminal differentiation in region depleted of SINE elements (Lu et al., 2020). SINEs and LINEs recruit different set of regulators in their associated genes promoting their segregation in active or inactive nuclear domain to regulate their activation or

silencing. They found also that LINE1 RNAs bind LINE1 DNA in embryonic stem cells, in repressive nuclear domain targeted with epigenetic repressor, in order to silence the expression of genes enriched in LINE1 elements guiding the temporal expression of genes during the development.

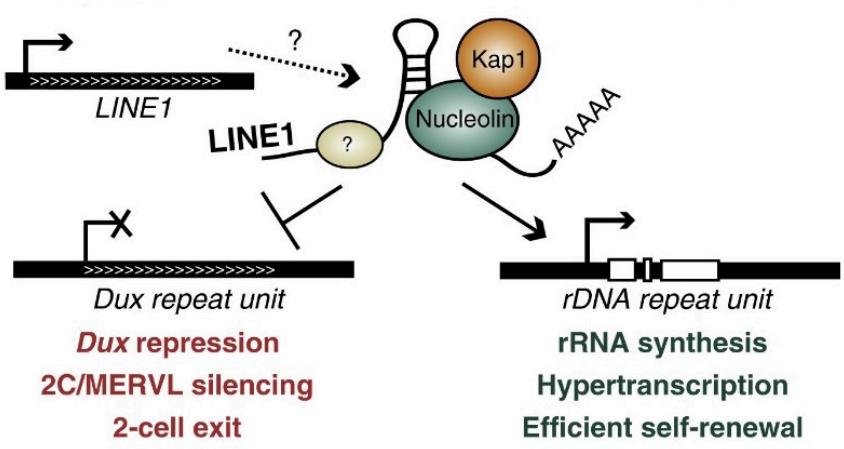


Figure 2. Schematic representation of LINE1 RNA-nucleolin-Kap1 complex in modulating gene expression in embryonic stem cells in mouse (Percharde et al., 2018).

Although recently several papers came out underlining the importance and the role of TEs in genome evolution and development there are still open questions regarding their possible roles and functions in shaping cell identity and plasticity especially in human cells where there are few evidences.

1.2. T cells are key players in adaptive immunity

Our organism establishes two specific strategy to protect itself against pathogens and external substances: innate and adaptive immune responses (Kubo, 2019).

Innate immune response is the first line of host defense and it is defined as nonspecific response, while adaptive immune response is characterized by a very specificity (Vivier & Malissen, 2005). Innate immunity arranges several mechanisms for a rapid elimination of pathogens and bacteria through an inflammatory response that is generic in respect to pathogens and tissues and does not provide memory (Caplan et al., 2017); whereas adaptive immunity is a specific second-line defense that arose days (or weeks) after the exposure to an infection or foreign antigens, it adapts an antigen-specific response and provides a great repertoire of receptors able to distinguish between self and nonself-antigens (Mirzaei, 2020). Immunological memory, the key characteristics of adaptive immunity, permits a rapid and strong response when the organism is attacked by the same pathogen (Starr et al., 2003).

Adaptive immunity is defined by the interplay between T and B lymphocytes and antigen-presenting cells (APCs). B cells derived from bone marrow and mediate humoral immunity against toxins and extracellular pathogens (Mirzaei, 2020). T cells derived from thymus and with their effector functions participate to the cell-mediate immunity against intracellular

pathogens (Jacqueline et al., 2016). Antigen-presenting cells (APCs) (as for examples dendritic cells and macrophages) are able to recognize and present the antigen to T cells (Jacqueline et al., 2016). To do this, APCs present pattern recognition receptors (PPRs), that can detect pathogens-associated molecular patterns (PAMPs) and damage-associated molecular patterns (DAMPs) expressed by host cells. After recognition of PAMPs or DAMPs, APCs internalized and processed their target since it is presented on the cell surface onto MHC class I or II receptors (Burgdorf et al., 2007). By this way, APCs, travelling throughout the lymphatic vessels, are able to present antigen to T cells activating the adaptive immune response. T cells with their different T helper (T_H) cell populations are one of the key players that orchestrate the adaptive immune response against non-self.

1.2.1 *$CD4^+$ T lymphocytes are plastic cells able to organize adaptive immune response*

T lymphocytes originate from bone marrow progenitors that migrate to the thymus where they undergo to thymopoiesis, the process of T cells generation, maturation and selection (Kumar et al., 2018; Schwarz & Bhandoola, 2006). T lymphocytes could be divided into two major groups based on the T-cell receptor (TCR) affinity to MHC molecules: $CD4^+$ T cells recognize MHC class II on APCs and $CD8^+$ T cells recognize MHC class I.

In the thymus, T lymphocytes progenitors lack both CD4 and CD8 molecules but during maturation, which involves TCR rearrangement, T cells, at first acquired both CD4 and CD8 (double positive cells), then double positive cells undergo to single positive selection generating $CD4^+$ T cells or $CD8^+$ T cells that ultimate their development when they are export in the periphery (Kumar et al., 2018; Schwarz & Bhandoola, 2006). In fact, Naïve T cells need different signals to be activated, to proliferate and further to differentiate in different helper (T_H) and regulatory subsets (figure 3). The first signal is constituted by the recognition of specific antigen through the binding of MHC class I or II, that is transduced by CD3, a receptor in complex with TCR (figure 3) (Abbas & Janeway, 2000); this signal could induce also a state named “clonal anergy” that represent the inability to respond to antigen or can fails to stimulate a response (Abbas & Janeway, 2000; Schwarz & Bhandoola, 2006). The signal 2, as shown in figure 3, is constituted by the binding of co-stimulatory molecules that amplify the first signal; APCs express several co-stimulatory molecules, as B7 proteins (CD80/B7.1 and CD86/B7.2) and CD40, that are recognize by co-receptor proteins, as CD28 and CD40L respectively (Kambayashi & Laufer, 2014). An appropriate signaling pathway through signal 1 and 2 guarantee a correct T cells activation instead when T cells received only signal 1 and not signal 2 apoptosis can be induced (Uzman et al., 2000); an incomplete activation as low-costimulatory signal or high co-inhibitory stimulation can cause T

cell anergy that induce an hyporesponsive state with low IL-2 production (Crespo et al., 2013); while a chronically stimulation through signal 1 could induce T cell exhaustion with a decrease of effector functions (Crespo et al., 2013) .

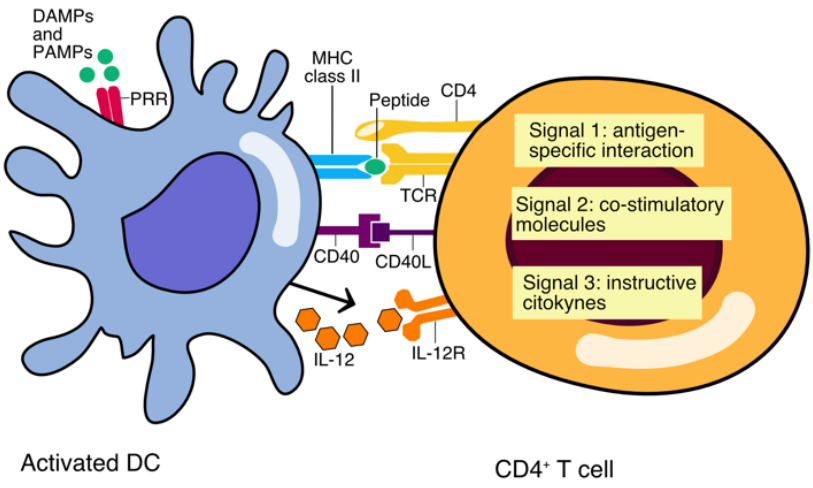


Figure 3. Schematic representation of the three signals that a T cells need to be activated, to proliferate and differentiate. Signal 1 is the antigen-specific recognition by MHC binding to TCR (for $CD4^+$ T cell). Signal 2: co-stimulatory molecules binding. Signal 3 is provided by instructive cytokines

The third signal, as highlight in figure 3, is provided by the environmental stimuli and more precisely by cytokines. This signal provides the polarization of Naïve T cells, in particular $CD4^+$ T cells, into a specific effector phenotype that suits with the

type of infection. Once T cells are activated and differentiated, effector T cells are able to secrete different cytokines to fight the infection accurately (Kambayashi & Laufer, 2014; Uzman et al., 2000). $CD8^+$ T cells, after the activation, are rapidly rendered functional against host cells; they express cytokines as perforins, that mediate the cytolysis of target cells, but also secrete tumor necrosis factor (TNF) and interferon- γ (IFN- γ) in order to build antimicrobial defense (Wong & Pamer, 2003). $CD4^+$ T cells are more plastic and can differentiate into several phenotypes, after their activation, in order to orchestrate immune protection throughout different mechanisms by cytokines and chemokines secretion: they promote B cells activation, by CD40-CD40L binding to stimulate antigen-specific antibodies production; they induce macrophages to increase their microbicide activity; they generate a positive feedback loop to amplify T cells response recruiting various innate immune cells (as neutrophils, eosinophils and basophils) to the site of infection and inflammation (Zhu & Paul, 2008). In detail, naïve $CD4^+$ T cells, based on specific environmental and cytokines milieu, could polarized into different helper (T_H) and regulatory subsets, delineated by cytokines production, according to the immune system needs (figure 4), T_H1 and T_H2 have been the first $CD4^+$ T effector subsets discovered. T_H1 cells are classified by the secretion of pro-inflammatory cytokines as IFN- γ, they are produced in response to intracellular pathogens and viral infection and they promote the stimulation of $CD8^+$ T cells

response, macrophage activation and a feedback loop that provide Naïve $CD4^+$ T cells into T_H1 polarization (Swain et al., 2012). T_H2 cells are characterized by the secretion of interleukin (IL) -4 and IL-13 (type II cytokines) that enhances humoral immunity and mediates inflammatory pathology associated with allergies (Stark et al., 2019). Naïve $CD4^+$ T cells polarized into T_H17 cells providing protection against bacteria and fungi and also preventing some autoimmune diseases. The master regulator of this T_H17 phenotype is orphan retinoic receptor γt (ROR-γt) and guarantees the secretion of IL-17 A and/or F (Weaver et al., 2006). $CD4^+$ T regulatory (Treg) cells are important to maintain peripheral tolerance as these cells secrete immunosuppressive cytokines, as IL-10 and transforming growth factor β (TGF-β), to inhibit T cell response against self-antigens, as such they are relevant in fighting autoimmune diseases (Vignali et al., 2008).

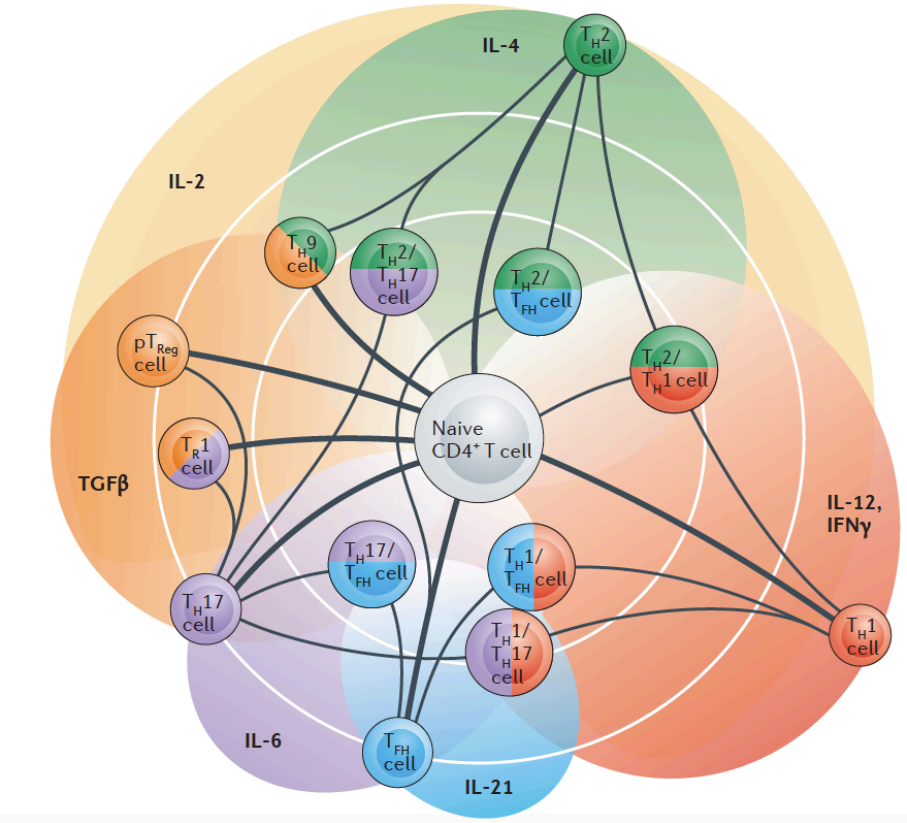


Figure 4. CD4⁺ T cells plasticity. Cytokines can polarize CD4⁺ T cells into several phenotypes with distinct functions but as shown by black lines different differentiation programs are interconnected with each other, underlined the plasticity of CD4⁺ T cells (Dupage & Bluestone, 2016).

During their lifetime, CD4⁺ T are not fully differentiated and are able to adapt rapidly and reversibly their fate and their function and to tune their response based on context and conditions, thanks to the integration of various stimuli that could act

simultaneously or not (Dupage & Bluestone, 2016; Natoli & Ostuni, 2019). As example, a repeated exposure of inflammatory agents, promotes a state of hypo-responsiveness to prevent the induction of inflammatory molecules (Seeley & Ghosh, 2017); even DNA methylation is useful to maintain a state of hypo-responsiveness during a chronic stimulation (Ghoneim et al., 2017). Moreover DNA methylation also define the adaptive state of T cells, in fact in naïve T cells the loci that encode for IL-4, IL-13 and IFN- γ are methylated and after T cells activation they undergo passive demethylation to promote faster reaction during stimulation (Monticelli, 2019; Wilson et al., 2009).

Extracellular cues as cytokines and TCR and co-stimulatory strength, cytosolic signaling, and gene regulation by transcription factors and DNA accessibility are some of the causes that promote T cell plasticity but an extensive investigation of these phenotype is still missing. Actually, there are several open questions as the unknown link between signaling and chromatin reprogramming in the establishment of a specific state in response to specific environmental cues, and/or the discriminating properties that promote the rapidly adaptation. Answering these questions could help us in finding novel strategies to adjust immune response in chronic infection, cancer and autoimmunity (Dupage & Bluestone, 2016; Natoli & Ostuni, 2019).

1.2.2 *TEs influence innate and adaptive immune response*

Several works have demonstrated that T cell identity and their immune response could be solicited by transposons. In fact, expressed TEs, though RNA and DNA sensing pathways, could be used as signaling molecules by the host sensing pathways of what to promote the expression of cytokines as IFN type I and INF-β (Kassiotis & Stoye, 2016). There is an extensive literature that describes how ERV elements, taking advantages of the machinery used by viral genomes, regulate both innate and adaptive immune response, as shown in figure 5 (Marasca et al., 2020).

In 2016, Choung et al., demonstrated that ERV regulate the expression of IFN type I response, acting as IFN-inducible enhancers. This peculiar activity is conserved among different mammalian species underlining an evolutive co-option of transposons to fine tune the expression of immune response genes (Chuong et al., 2016).

Within this frame, in 2015, Chiappinelli et al., found that dsRNAs transcribed by ERV sequences are able to bind MDA5 enhancing IFN-β secretion in order promote immune system response (Chiappinelli et al., 2015). ERV elements regulate also T_H1 differentiation, ERVs are collocated near genes involved in immune processes acting as T_H1 gene enhancer. In T_H2 cells, T_H1 gene expression profile was repressed by the histone methyltransferase SETDB1 that convey the deposition of

H3K9me3 at ERV elements located in proximity of T_H1 genes (Adoue et al., 2019).

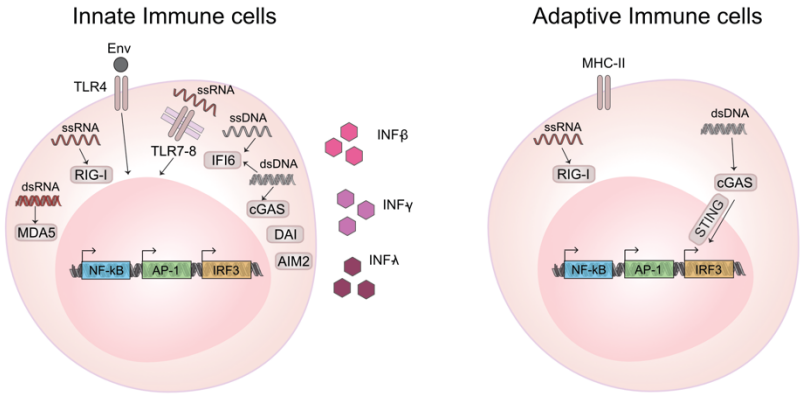


Figure 5. TEs promote innate and adaptive immune response. On the left are shown nucleic acid of TEs activate transcription factors that enhance IFNs production. On the right are described how TEs by RNA and DNA sensing pathway activate adaptive immune response in T lymphocytes (Marasca et al., 2020).

Other TEs are able to modulate immune response as ERVs do. LINE1 derived transcripts are AU-rich sequences that could be identified by RIG-I and MDA5 receptors promoting their activation and IFN-β production (Brisse & Ly, 2019; Zhao et al., 2018). Transposons can elicit adaptive immune response also by DNA-sensing pathways (Hornung et al., 2009; Takaoka et al., 2007; Unterholzner et al., 2010).

TEs are able to modulate adaptive immune response not only in T cells also in B cells. Zeng et al., demonstrated that in B cells

ERVs through their cytosolic RNAs activate RIG-I-MAVS pathway and through cDNAs activate cGAS-STING pathway. The activation of these pathway stimulate B cells activation and production of antigen-specific antibodies (Zeng et al., 2014).
All these evidences demonstrate that TEs play an important role in controlling and shaping innate and adaptive immune response, almost because their deregulation could be implicated in inflammation and autoimmune diseases.

1.3 T cells are key players in fighting cancer progression

1.3.1 *Tumor microenvironment is infiltrated by different T cell subsets*

Cancer is the second cause of dead worldwide; it is defined as “the pathology of the century” and “the modern disease par excellence”, suggesting its connotation as an endemic disease. In fact, World Health Organization (WHO) estimated that lead 10 million of dead per year (Falzone et al., 2018; Ferlay et al., 2020). One in six deaths is caused by one of the “big killers” that are lung, colorectal, prostate and breast cancer (Falzone et al., 2018).
In the last decade tumors have been identified as organs whose complexity exceed that of healthy tissues ref. Nowadays it is accepted that tumor microenvironment (TME) is a novel hallmark of cancer, in particular the composition of this environment is

crucial to define cancer phenotypes (Hanahan & Weinberg, 2011). TME, is a very heterogenous ecosystem that is composed by cancer cells, resident host cells, secreted factors, blood vessels and extracellular matrix (Hassan & Seno, 2020) (figure 6). Is worth to notice that TME composition changes in tumorigenesis and tumor cells promotes the molecular, physical and cellular rearrangement to sustain tumor progression (Truffi et al., 2020).

Cancer cells and components of TME set up a mutual relationship that promotes cancer cells survival, local invasion and metastatic propagation. Tumor cells can produce neoantigens that are detected by immune cells (Finn, 2017). Tumor antigens could be categorized into two different types based on their specificity: tumor specific antigens and tumor associated antigens. The former is tumor specific and could promote antitumor T cells response while the latter is not tumor specific and could lead to immunological tolerance (Finn, 2017; Gajewski et al., 2013).

TME is infiltrated by several immune cells, both from innate and adaptive immune system, that promotes both pro- and anti-tumorigenic functions.

Immune cells are one of the first components of TME and a persistent inflammation due to chronic infection, in several type of tumor underline the tumor formation. The presence of inflammatory cells is a positive signals, pointing out that the host is try to react and interfere against tumor progression (Zitvogel et

al., 2006). Immune cells are one of the critical elements of the TME, actually they can suppress or promote tumor growth (Anderson & Simon, 2020). As example, $CD8^+$ T cells with their cytotoxic actions are able to target and destroy cancer cells, moreover in concert with $CD4^+$ T cells, secrete IFN-γ in order to suppress angiogenesis and prime immune response against tumor generating an inflammatory environment (Maimela et al., 2019). On the other hand, also $CD4^+$ Treg cells belong to TME, usually suppressing inflammatory response with the production of inhibitory cytokines as TGF-β and IL-10. In TME context $CD4^+$ Treg cells repress anti-tumor response supporting tumor progression to maintain immunologiacal self-tolerance (Plitas & Rudensky, 2020).

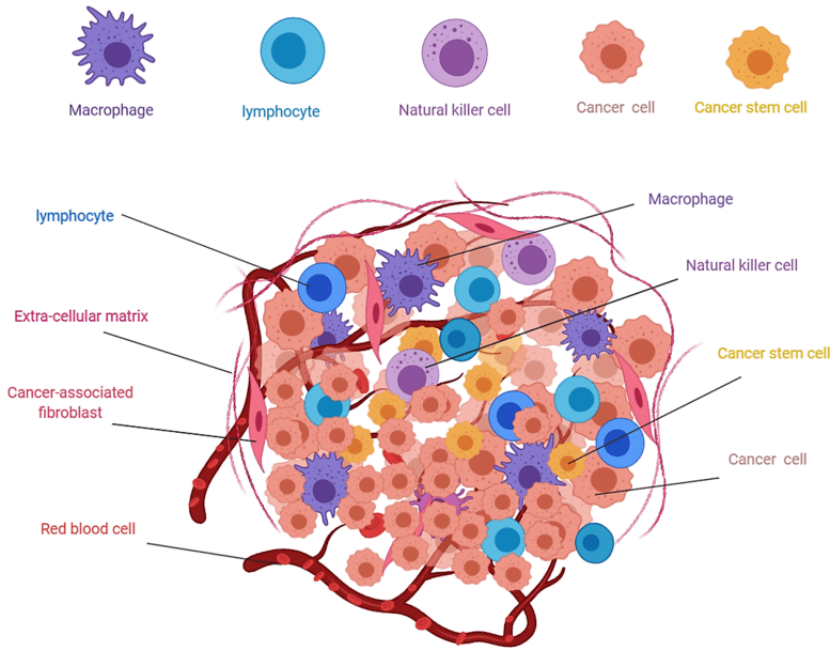


Figure 6. Representation of heterogeneity of tumor microenvironment. TME is a very heterogeneous ecosystem constituted by immune cells, cancer cells, vessels and extracellular matrix (Hassan & Seno, 2020).

Recent studies demonstrated that there is a resistance program expressed by cancer cells that permits T cells exclusion and immune evasion (Jerby-Arnon et al., 2018). Cancer cells are detected by APCs and presented to T cells but often the immune response to tumor antigen exposure is not sufficient to block tumor progression (Finn, 2017). Indeed, during tumor progression cancer cells became able to overcome immune recognition and escape immune response, by this way tumor growth is immoderate and the disease become clinically apparent (Schreiber et al., 2011).

Recent evidences suggest that type, density and location of immune cells in TME could have a prognostic and therapeutic value (Galon et al., 2006; Pagès et al., 2005).

Actually, tumor infiltrating lymphocytes (TILs) are not randomly distributed but are organized in specific regions within TME (Sasada & Suekane, 2011).

The reason behind the high or low levels of TILs is still under investigation but several groups are arguing, also at single-cell level, the characterization of TILs and the dynamics beyond these type of tumors, in order to enhance immunotherapy strategies (Duan et al., 2020; Jerby-Arnon et al., 2018; Zheng et

al., 2017). Indeed, single-cell RNA sequencing analysis permits not only to profile the gene expression pattern of every cells unveiling the heterogeneity of T cell population but also to determine dynamic relationship between T cells through transcriptome and TCR analysis.

Recently Tirosh et al., demonstrated by single-cell transcriptome analysis in melanoma patients that there are different microenvironments that are associated with distinct malignant cells profiles that influence the proportion of other cell type, as T cells (Tirosh et al., 2016). Study of TILs composition in melanoma patients was also addressed by Li et al., revealing that there is a gradient of T cell dysfunction and that the intensity of dysfunctional signature is related to tumor reactivity (Li et al., 2019).

In triple-negative breast cancer (TBNC) Karaayvaz et al., by scRNA seq analysis found in TME a specific epithelial cells subpopulation that could be associated with a good prognosis (Karaayvaz et al., 2018). TILs profile in TNBC was addressed by Savas et al., that found an improve patient survival is associated to a TME that is enriched in CD8 with feature of tissue-resident memory (Savas et al., 2018).

Several scRNA-seq analysis were done to determine tumor-infiltrating cells composition in lung cancer. In fact, thanks to scRNA-seq analysis was done the first-ever lung cancer TME cell atlas by the characterization of stromal cells derived from non-small cell lung cancer (NSCLC) patients (Lambrechts et al.,

2018). Studying tumor infiltrating T and myeloid cells profile in NSCLC were identify a subset of Tregs and tumor infiltrating myeloid cells defined by unique markers that correlate to poor prognosis (Guo et al., 2018; Zilionis et al., 2019). Moreover, investigation of cancer and TME cells of advanced NSCLC has allowed to identify specific cell population that are expressed based on tumor pathological types and degree of tumor heterogeneity (Wu et al., 2021).

An extensively work was done also to characterized TILs in colorectal cancer (CRC) by scRNA-seq and TCR analysis. Thanks to Zhang and its colleague in CRC patients with high microsatellite instability was found a specific T_H1-like-cells that maybe promote a better response to immunotherapy (L. Zhang et al., 2018). scRNA-seq study on myeloid composition of CRC patients and mouse models reveal the heterogeneity of these cells, their role in TME and also the translatability of therapies from mouse pre-clinical models to human cancer (L. Zhang et al., 2020). Furthermore, the characterization of single-cell of blood and gut cells in CRC patients conducted by Qi et al., permits to identify tumor specific innate lymphoid cells subsets that overexpress an anti-tumor biomarker SLAMF1 in patients with higher survival rates (Qi et al., 2021).

1.3.2 *Tumor infiltrating lymphocytes became dysfunctional during cancer progression*

In a disease context, naïve T cells are activated, differentiate into effector T cells and proliferate to better fight infection. After the resolution of inflammation most activated T cells die but a subset of them form the memory T cell pool to confer long-term protection against pathogens re-boot (Wherry & Kurachi, 2015). By contrast during chronic infections and cancers, T cells are continuously exposed to antigen and inflammatory signals. These events dysregulate the memory T cell differentiation program and T cells became “exhausted” (Wherry & Kurachi, 2015; Yi et al., 2010). If T cells remain endless exposed to an antigen for 2-4 weeks, in vivo, the state of exhaustion become established and does not revert with only the removal of the antigen (Brooks et al., 2006).

Tumor immunosurveillance theory arose already at the begging of 1900 by Paul Ehrlich and explained that cancer cells developed spontaneously in the organism and that immune system try to eliminate them to prevent the neoplastic formation (Center for History and New Media et al., 2015). Several studies, in subsequent years, demonstrated that the interplay between cancer cells and immune system is more complex and the immunoediting mechanism try to explain this relationship (Dupage et al., 2012).

Cancer shapes cancer development through three different stages: elimination, equilibrium and escape (Sim et al., 2019) (fig 7). During the first phase immune system is competent to destroy cancer cells producing an inflammatory environment. In the equilibrium phase tumor development is kept in check by adaptive immune system and there is a balance between tumor growth and regression. In the last phase cancer establishes itself and cancer cells are no longer recognized by immune system probably due to antigen loss, generating an immunosuppressive tumor environment (Escors, 2014).

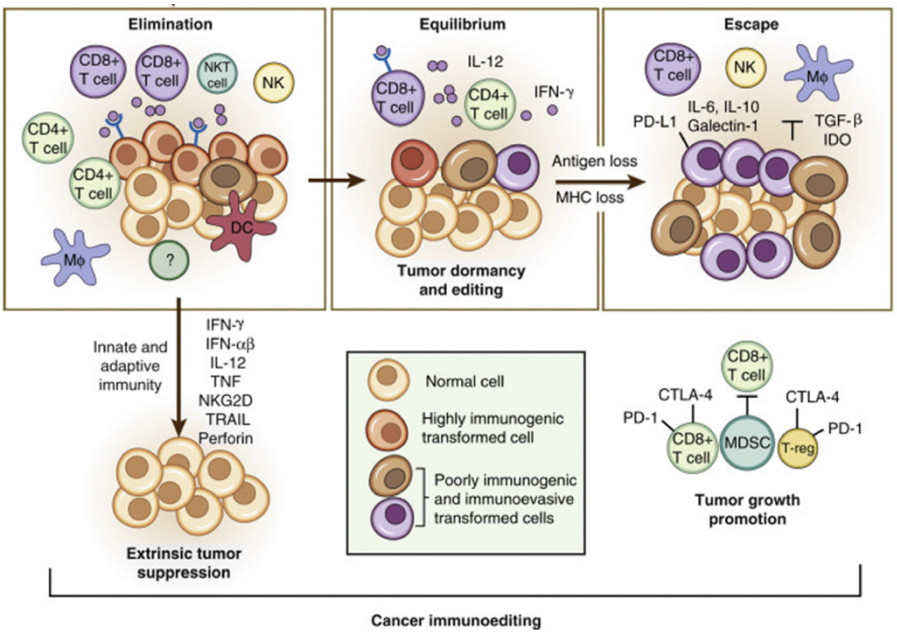


Figure 7. Cancer immunoediting mechanism. The interplay between Immune system and cancer cells build cancer progression through three phases: elimination, equilibrium and escape (Sim et al., 2019).

Cancer cells became able to escape to immune detection by the expression of immune checkpoint ligands (Escors, 2014). Immune checkpoint ligands activate inhibitory pathways to enhance immunosuppressive environment (Saleh et al., 2022). PD-1 ligand and CTLA4 are the most expressed among immune checkpoint molecules.

Exhaustion state is a specific lineage of differentiated T cells and is different, phenotypically and mechanistically, from other dysfunctional state of T cells as senescence and anergy (Wherry & Kurachi, 2015). Exhausted T cells are characterized by the progressive loss of effector functions and memory T cell properties, with a decrease of cytokines and chemokines secretion, usually IL-2 is the first cytokine that decrease, followed by TNF-α and IFN-γ; these cells have a decrease in their proliferative rate and the acquisition of multiple inhibitory receptors (Fig. 8). PD-1, T cell immunoglobulin and mucin domain 3 (TIM3), lymphocyte activate gene-3 (LAG-3) and cytotoxic T lymphocytes associated 4 (CTLA4) are the most frequent inhibitory receptors that define an exhausted status and provide T cell hyporesponsiveness to tumor-associated antigens (Sasada & Suekane, 2011; Wherry, 2011; Wherry & Kurachi, 2015).

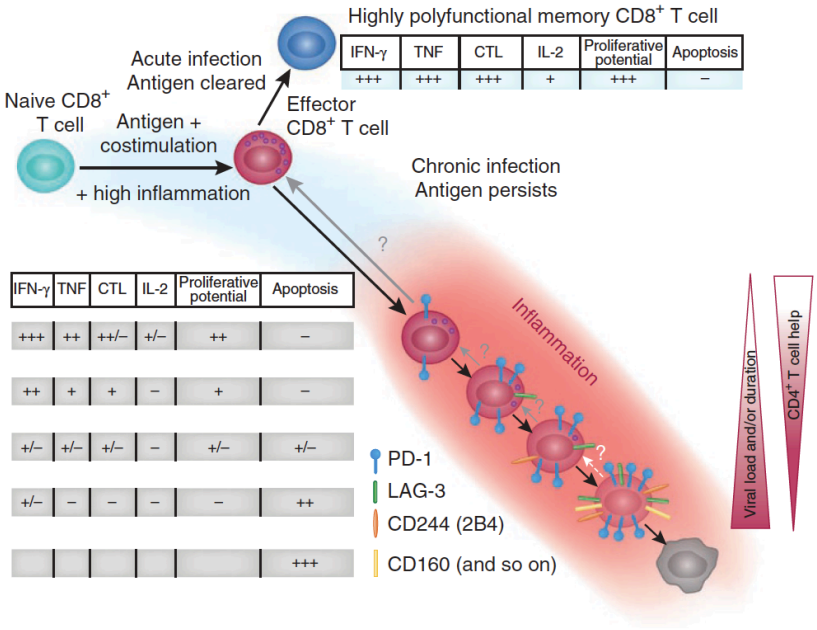


Figure 8. Development of T cell exhaustion during chronic infection. Naïve CD8⁺ T cells, after persistent antigen stimulation, acquired progressively an exhausted phenotype characterized by loss of effector functions, decrease of proliferative rate and increase of inhibitory receptors (Wherry, 2011).

Every inhibitory receptor blocks T cell activity with different mechanisms. CTLA4, prevents the second signal of activation competing with CD28 to bind B7, moreover induces trans-endocytosis of costimulatory ligands in order to restrict the possibility to T cell activation (Byun et al., 2017). PD-1 suppressive activity is not yet fully understood and there are five possible strategy proposed to explain how PD-1 suppress T cells functions, not mutually exclusive: downmodulation of TCR

signaling pathway (Sheppard et al., 2004); modulation of mTOR pathway (Staron et al., 2014); modulation of Ras pathway (Patsoukis et al., 2012); influencing T cell mobility (Zinselmeyer et al., 2013); inducing the expression of basic leucine zipper transcription factor that repress the expression of effector genes (Quigley et al., 2010).

Although PD-1 and CTLA4 are the most abundant and the most studied receptors, also TIM3 and LAG-3 are well established players in driving exhausted/dysfunctional phenotype. Once expressed, TIM3 binds its ligand, galectin-9, and inhibits T_H1 and T_H17 responses by hampering their expansion, inducing peripheral tolerance and promoting T_H1 cell death (Davoodzadeh Gholami et al., 2017; Sakuishi et al., 2010). LAG-3 instead, as CTLA4, binds CD3-TCR complex impeding co-receptor engagement, to negatively regulate proliferation, homeostasis and function of T cells (Workman et al., 2004). In human cancers it was found that TILs expressing more than one inhibitory receptor show a more dysfunctional effector functions in respect to TILs that express only PD-1 (Wherry & Kurachi, 2015).

Taking in advantage of this knowledge, in the past few years cancer immunotherapy has made a breakthrough in cancer treatment. Among cancer immunotherapies, Immune checkpoint therapy uses immune checkpoint inhibitors (ICIs) in order to reinvigorate immune response. Cancer immunotherapies foresee the usage of genetically engineered chimeric antigen receptor specific -T cells (CAR-T) to eliminate tumor cells by the

recognition of a cancer specific antigen (Escors, 2014; Saleh et al., 2022).

However, not all the patients positively respond to cancer immunotherapy; and the effectiveness of these treatments is limited to several intrinsic and extrinsic factors. The composition of TME, epigenetic and genetic alterations and tumor mutational load are some of the factors that compromise the treatment (Yu & Cui, 2018).

Despite all the evidences overmentioned, little is still known about the molecular mechanism that subtend T cell exhaustion and immune exclusion (Balkhi et al., 2018) and further studies are needed to understand how to build the best strategy to fight cancer progression.

One of the key aspects that would be useful to deepen is the understanding of the molecular mechanisms that influence at transcriptional level the immune response against tumor, in order to find novel therapeutic target to enhance TILs functions.

1.3.3 *Transposable elements RNAs are novel players in cancer immunity*

Immunotherapy against cancer progression try to restore tumor infiltrating lymphocytes functions to promote antitumor response but several studies demonstrated that TEs expression could stimulate immune response against cancer.

Tumorigenesis is associated to a deregulation of DNA methylation within the cells. Indeed it is known that a common feature among different tumors (CRC, lung, TNBC, ovarian, glioblastoma) is the hypermethylation within the promoter of tumor-suppressor genes, that cause their silencing (Klutstein et al., 2016). An alteration of DNA methylation is found also in repetitive DNA sequence; tumor cells through the modulation of lysine 9 and 27 on histone 3 methylation, are able to repress TEs transcription and block their immunostimulatory activity (Robbez-Masson et al., 2017). Tumor cells promoting DNA methyltransferase activity try to evade immune response, also by TEs inhibition.

Several works demonstrated that the use DNA methyltransferase inhibitors (DNMTis) could provide the expression of TEs promoting the induction of interferon-genes via dsRNA sensing pathways (Chiappinelli et al., 2015; Goel et al., 2017; Kong et al., 2019; Roulois et al., 2015; C. C. Smith et al., 2018). Roulois et al., and Chiappinelli et al., illustrated that in ovarian and CRC patients the use of DNMTi induce ERV RNAs expression that prompt the upregulation of interferon-responsive genes (Chiappinelli et al., 2015; Roulois et al., 2015). Moreover, Saito and its colleague found that the knockdown of DNMT1 suppress the proliferation of intestinal tumor organoids by anti-viral response via activation of dsRNA-containing ERVs (Saito et al., 2016).

Regulation of TEs are able to stimulate also adaptive immune response to fighting cancer progression. Smith et al., in 2016, proved that retroviral transcripts could produce neoantigens that activate adaptive immune response driving T and B cell functions (C. C. Smith et al., 2018). While Goel and its collegues , in 2017 demonstrated that increase of ERVs expression by inhibitors of cyclin-dependent kinase 4/6 not only provides more secretion of IFN type III but also suppresses $CD4^+$ Treg cells proliferation and promotes cytotoxic function of T cells enhancing tumor cells clearance (Goel et al., 2017).

All these studies underline the important role of TEs in regulate immune response also in disease context like cancer and outline TEs as novel possible molecules to be addressed in cancer immunotherapy.

1.4 Scope of the thesis

Tumor infiltrating lymphocytes (TILs) are the principal components of the tumor microenvironment and play a central role in antitumor immunity. During cancer immunoediting, TILs became dysfunctional, a phenotype associated to effector functions impairment, reduced cell growth and decreased killing capability. Indeed, cancer immunotherapies treatments try to revert TILs dysfunctional state in order to promote tumor clearance.

Immunotherapies nowadays represent the novel frontier in fighting cancer but still little is known regarding the epigenetic modulators responsible for TILs properties; the discovery of novel, possible targetable molecules and mechanisms could substantially improve knowledges regarding immunotherapies and patient responsiveness to them. For this reason, we are applying novel approaches and technologies in this filed, namely the investigation of transposable elements (TEs) functions as novel epigenetic players in TILs identity, plasticity and adaptability to the environmental cancer driven milieu.

TEs are interspersed repetitive DNA sequences that cover 40 - 45% of the human genome and growing evidence suggests that TEs exert a crucial function in epigenetic regulation both in cis and in trans, being a source of non-coding regulatory RNAs and participating to chromatin folding. Among TEs, we are interested in the possible epigenetic functions of LINE1 elements, that represents 18% of the human genome, largely accepted as novel key molecules involved in epigenetic regulation of cell identity. Until today the role and the dynamics of TEs-derived RNAs were investigated only in embryonic stem cells or during organism development, while there was no evidence of their possible functions in fully differentiated cells derived from adult tissues, as human T lymphocytes, that are plastic cells able to adapt and differentiate to diverse effector cells based on the cytokine milieu. We demonstrated that, among T cells subsets, there is a specific enrichment for LINE1 chromatin associated RNAs in naïve CD4+

T cells. Moreover, LINE1 RNAs show a peculiar and timely specific dynamic, being rapidly depleted from the nuclei after TCR activation. Notably, functional experiments suggested that these transcripts could regulate T cells effector functions.

Since these data, the aim of this thesis is to evaluate LINE1 involvement in the epigenetic regulation of cell identity and functions in TILs. We generated an in vitro model to study LINE1 dynamics in exhausted and dysfunctional T cells, moreover, we had the possibility to isolate ex vivo TILs from NSCL cancer, CRC and their normal counterpart derived from patients to perform functional experiments and assess LINE1 functions in a pathological context.

Finally, we aim to define LINE1 RNAs as novel TILs regulatory molecules to find novel targetable RNA molecules that could be used as adjuvants in therapy to reinforce patient's immune response against cancer.

1.5 Chapter 1 bibliography

Abbas, A. K., & Janeway, C. A. (2000). Immunology: Improving on nature in the twenty-first century. In *Cell*. https://doi.org/10.1016/S0092-8674(00)81689-X

Adoue, V., Binet, B., Malbec, A., Fourquet, J., Romagnoli, P., van Meerwijk, J. P. M., Amigorena, S., & Joffre, O. P. (2019). The Histone Methyltransferase SETDB1 Controls T Helper Cell Lineage Integrity by Repressing Endogenous Retroviruses. *Immunity*. https://doi.org/10.1016/j.immuni.2019.01.003

Anderson, N. M., & Simon, M. C. (2020). The tumor microenvironment. *Current Biology*. https://doi.org/10.1016/j.cub.2020.06.081

Attig, J., Agostini, F., Gooding, C., Chakrabarti, A. M., Singh, A., Haberman, N., Zagalak, J. A., Emmett, W., Smith, C. W. J., Luscombe, N. M., & Ule, J. (2018). Heteromeric RNP Assembly at LINEs Controls Lineage-Specific RNA Processing. *Cell*. https://doi.org/10.1016/j.cell.2018.07.001

Balkhi, M. Y., Wittmann, G., Xiong, F., & Junghans, R. P. (2018). YY1 Upregulates Checkpoint Receptors and Downregulates Type I Cytokines in Exhausted, Chronically Stimulated Human T Cells. *IScience*. https://doi.org/10.1016/j.isci.2018.03.009

Beck, S., Lee, B. K., Rhee, C., Song, J., Woo, A. J., & Kim, J. (2014). CpG island-mediated global gene regulatory modes in mouse embryonic stem cells. *Nature Communications*. https://doi.org/10.1038/ncomms6490

Belancio, V. P., Hedges, D. J., & Deininger, P. (2008). Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health. In *Genome Research*. https://doi.org/10.1101/gr.5558208

Bernstein, E., Duncan, E. M., Masui, O., Gil, J., Heard, E., & Allis, C. D. (2006). Mouse Polycomb Proteins Bind Differentially to Methylated Histone H3 and RNA and Are Enriched in Facultative Heterochromatin. *Molecular and Cellular Biology*. https://doi.org/10.1128/mcb.26.7.2560-2569.2006

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., Coussens, L. M., Gabrilovich, D. I., Ostrand-Rosenberg, S., Hedrick, C. C., Vonderheide, R. H., Pittet, M. J.,

Jain, R. K., Zou, W., Howcroft, T. K., Woodhouse, E. C., Weinberg, R. A., & Krummel, M. F. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nature Medicine*. https://doi.org/10.1038/s41591-018-0014-x

Bourc'his, D., & Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. In *Nature*. https://doi.org/10.1038/nature02886

Brisse, M., & Ly, H. (2019). Comparative structure and function analysis of the RIG-I-like receptors: RIG-I and MDA5. In *Frontiers in Immunology*. https://doi.org/10.3389/fimmu.2019.01586

Britten, R. J., & Davidson, E. H. (1971). Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *The Quarterly Review of Biology*. https://doi.org/10.1086/406830

Brooks, D. G., McGavern, D. B., & Oldstone, M. B. A. (2006). Reprogramming of antiviral T cells prevents inactivation and restores T cell activity during persistent viral infection. *Journal of Clinical Investigation*. https://doi.org/10.1172/JCI26856

Burgdorf, S., Kautz, A., Böhnert, V., Knolle, P. A., & Kurts, C. (2007). Distinct pathways of antigen uptake and intracellular routing in CD4 and CD8 T cell activation. *Science*. https://doi.org/10.1126/science.1137971

Byun, D. J., Wolchok, J. D., Rosenberg, L. M., & Girotra, M. (2017). Cancer immunotherapy-immune checkpoint blockade and associated endocrinopathies. In *Nature Reviews Endocrinology*. https://doi.org/10.1038/nrendo.2016.205

Caplan, L. R., Biller, J., Leary, M. C., Lo, E. H., Thomas, A. J., Yenari,

M., & Zhang, J. H. (2017). Primer on Cerebrovascular Diseases: Second Edition. In *Primer on Cerebrovascular Diseases: Second Edition*.

Casa, V., & Gabellini, D. (2012). A repetitive elements perspective in Polycomb epigenetics. *Frontiers in Genetics*. https://doi.org/10.3389/fgene.2012.00199

Center for History and New Media, The Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. R. M., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J. M., Miller, M. L., Reznik, E., Gauthier, N. P., Aksoy, B. A., Korkut, A., Gao, J., Ciriello, G., Schultz, N., … Lee, J. M. J.-S. J. (2015). Ueber den jetzigen Stand der Karzinomforschung. Vortrag gehalten vor den Studenten der Amsterdamer Universitaet, Vereinigung fuer wissenschaftliche Arbeit 1 June 1908. Printed in: P. Ehrlich. *Genome Biology*.

Chiappinelli, K. B., Strissel, P. L., Desrichard, A., Li, H., Henke, C., Akman, B., Hein, A., Rote, N. S., Cope, L. M., Snyder, A., Makarov, V., Buhu, S., Slamon, D. J., Wolchok, J. D., Pardoll, D. M., Beckmann, M. W., Zahnow, C. A., Mergoub, T., Chan, T. A., … Strick, R. (2015). Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell*. https://doi.org/10.1016/j.cell.2015.07.011

Chuong, E. B., Elde, N. C., & Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science*. https://doi.org/10.1126/science.aad5497

Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: From conflicts to benefits. In

Nature Reviews Genetics. https://doi.org/10.1038/nrg.2016.139

Clark, J., & Rager, J. E. (2020). Epigenetics: An overview of CpG methylation, chromatin remodeling, and regulatory/noncoding RNAs. In *Environmental Epigenetics in Toxicology and Public Health*. https://doi.org/10.1016/b978-0-12-819968-8.00001-9

Crespo, J., Sun, H., Welling, T. H., Tian, Z., & Zou, W. (2013). T cell anergy, exhaustion, senescence, and stemness in the tumor microenvironment. In *Current Opinion in Immunology*. https://doi.org/10.1016/j.coi.2012.12.003

Darby, M. M., & Sabunciyan, S. (2014). Repetitive elements and epigenetic marks in behavior and psychiatric disease. In *Advances in Genetics*. https://doi.org/10.1016/B978-0-12-800222-3.00009-7

Davoodzadeh Gholami, M., kardar, G. A., Saeedi, Y., Heydari, S., Garssen, J., & Falak, R. (2017). Exhaustion of T lymphocytes in the tumor microenvironment: Significance and effective mechanisms. In *Cellular Immunology*. https://doi.org/10.1016/j.cellimm.2017.10.002

de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A., & Pollock, D. D. (2011). Repetitive elements may comprise over Two-Thirds of the human genome. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1002384

Deininger, P. (2011). Alu elements: Know the SINEs. In *Genome Biology*. https://doi.org/10.1186/gb-2011-12-12-236

Deniz, Ö., Frost, J. M., & Branco, M. R. (2019). Regulation of transposable elements by DNA modifications. In *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-019-0106-6

Denli, A. M., Narvaiza, I., Kerman, B. E., Pena, M., Benner, C.,

Marchetto, M. C. N., Diedrich, J. K., Aslanian, A., Ma, J., Moresco, J. J., Moore, L., Hunter, T., Saghatelian, A., & Gage, F. H. (2015). Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell*. https://doi.org/10.1016/j.cell.2015.09.025

Dhuri, K., Bechtold, C., Quijano, E., Pham, H., Gupta, A., Vikram, A., & Bahal, R. (2020). Antisense oligonucleotides: An emerging area in drug discovery and development. In *Journal of Clinical Medicine*. https://doi.org/10.3390/JCM9062004

Doolittle, R. F., Feng, D. F., Johnson, M. S., & McClure, M. A. (1989). Origins and evolutionary relationships of retroviruses. In *The Quarterly review of biology*. https://doi.org/10.1086/416128

Duan, Q., Zhang, H., Zheng, J., & Zhang, L. (2020). Turning Cold into Hot: Firing up the Tumor Microenvironment. In *Trends in Cancer*. https://doi.org/10.1016/j.trecan.2020.02.022

Dupage, M., & Bluestone, J. A. (2016). Harnessing the plasticity of CD4+ T cells to treat immune-mediated disease. In *Nature Reviews Immunology*. https://doi.org/10.1038/nri.2015.18

Dupage, M., Mazumdar, C., Schmidt, L. M., Cheung, A. F., & Jacks, T. (2012). Expression of tumour-specific antigens underlies cancer immunoediting. *Nature*. https://doi.org/10.1038/nature10803

Elbarbary, R. A., Lucas, B. A., & Maquat, L. E. (2016). Retrotransposons as regulators of gene expression. In *Science*. https://doi.org/10.1126/science.aac7247

Escors, D. (2014). Tumour Immunogenicity, Antigen Presentation, and Immunological Barriers in Cancer Immunotherapy. *New Journal of Science*. https://doi.org/10.1155/2014/734515

Falzone, L., Salomone, S., & Libra, M. (2018). Evolution of cancer pharmacological treatments at the turn of the third millennium. In *Frontiers in Pharmacology*. https://doi.org/10.3389/fphar.2018.01300

Faulkner, G. J., Kimura, Y., Daub, C. O., Wani, S., Plessy, C., Irvine, K. M., Schroder, K., Cloonan, N., Steptoe, A. L., Lassmann, T., Waki, K., Hornig, N., Arakawa, T., Takahashi, H., Kawai, J., Forrest, A. R. R., Suzuki, H., Hayashizaki, Y., Hume, D. A., … Carninci, P. (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nature Genetics*. https://doi.org/10.1038/ng.368

Feng, Q., Moran, J. V., Kazazian, H. H., & Boeke, J. D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell*. https://doi.org/10.1016/S0092-8674(00)81997-2

Ferlay, J., Ervik, M., Lam, F., Colombet, M., Mery, L., Piñeros, M., Znaor, A., Soerjomataram, I., & Bray, F. (2020). *Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer*. Lyon, France: International Agency for Research on Cancer.

Ferrari, R., de Llobet Cucalon, L. I., Di Vona, C., Le Dilly, F., Vidal, E., Lioutas, A., Oliete, J. Q., Jochem, L., Cutts, E., Dieci, G., Vannini, A., Teichmann, M., de la Luna, S., & Beato, M. (2020). TFIIIC Binding to Alu Elements Controls Gene Expression via Chromatin Looping and Histone Acetylation. *Molecular Cell*. https://doi.org/10.1016/j.molcel.2019.10.020

Feschotte, Cédric. (2008). Transposable elements and the evolution of regulatory networks. In *Nature Reviews Genetics*.

https://doi.org/10.1038/nrg2337

Feschotte, Cleacuteldric, & Pritham, E. J. (2007). DNA transposons and the evolution of eukaryotic genomes. In *Annual Review of Genetics*. https://doi.org/10.1146/annurev.genet.40.110405.090448

Finn, O. J. (2017). Human tumor antigens yesterday, today, and tomorrow. *Cancer Immunology Research*. https://doi.org/10.1158/2326-6066.CIR-17-0112

Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C. A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A., Noro, Y., Wong, C. H., De Hoon, M., Andersson, R., Sandelin, A., Suzuki, H., Wei, C. L., Koseki, H., Hasegawa, Y., … Carninci, P. (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nature Genetics*. https://doi.org/10.1038/ng.2965

Gajewski, T. F., Schreiber, H., & Fu, Y. X. (2013). Innate and adaptive immune cells in the tumor microenvironment. In *Nature Immunology*. https://doi.org/10.1038/ni.2703

Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., Zinzindohoué, F., Bruneval, P., Cugnenc, P. H., Trajanoski, Z., Fridman, W. H., & Pagès, F. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*. https://doi.org/10.1126/science.1129139

Gao, R., Kim, C., Sei, E., Foukakis, T., Crosetto, N., Chan, L. K., Srinivasan, M., Zhang, H., Meric-Bernstam, F., & Navin, N. (2017). Nanogrid single-nucleus RNA sequencing reveals

phenotypic diversity in breast cancer. *Nature Communications*. https://doi.org/10.1038/s41467-017-00244-w

Gendrel, A. V., Lippman, Z., Yordan, C., Colot, V., & Martienssen, R. A. (2002). Dependence of heterochromatic histone H3 methylation patterns on the Arabidopsis gene DDM1. *Science*. https://doi.org/10.1126/science.1074950

Ghoneim, H. E., Fan, Y., Moustaki, A., Abdelsamed, H. A., Dash, P., Dogra, P., Carter, R., Awad, W., Neale, G., Thomas, P. G., & Youngblood, B. (2017). De Novo Epigenetic Programs Inhibit PD-1 Blockade-Mediated T Cell Rejuvenation. *Cell*. https://doi.org/10.1016/j.cell.2017.06.007

Goel, S., Decristo, M. J., Watt, A. C., Brinjones, H., Sceneay, J., Li, B. B., Khan, N., Ubellacker, J. M., Xie, S., Metzger-Filho, O., Hoog, J., Ellis, M. J., Ma, C. X., Ramm, S., Krop, I. E., Winer, E. P., Roberts, T. M., Kim, H. J., McAllister, S. S., & Zhao, J. J. (2017). CDK4/6 inhibition triggers anti-tumour immunity. *Nature*. https://doi.org/10.1038/nature23465

Guo, X., Zhang, Y., Zheng, L., Zheng, C., Song, J., Zhang, Q., Kang, B., Liu, Z., Jin, L., Xing, R., Gao, R., Zhang, L., Dong, M., Hu, X., Ren, X., Kirchhoff, D., Roider, H. G., Yan, T., & Zhang, Z. (2018). Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nature Medicine*. https://doi.org/10.1038/s41591-018-0045-3

Hall, L. L., Carone, D. M., Gomez, A. V., Kolpa, H. J., Byron, M., Mehta, N., Fackelmayer, F. O., & Lawrence, J. B. (2014). Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell*. https://doi.org/10.1016/j.cell.2014.01.042

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. In *Cell*. https://doi.org/10.1016/j.cell.2011.02.013

Hassan, G., & Seno, M. (2020). Blood and Cancer: Cancer Stem Cells as Origin of Hematopoietic Cells in Solid Tumor Microenvironments. In *Cells*. https://doi.org/10.3390/cells9051293

Hoelder, S., Clarke, P. A., & Workman, P. (2012). Discovery of small molecule cancer drugs: Successes, challenges and opportunities. In *Molecular Oncology*. https://doi.org/10.1016/j.molonc.2012.02.004

Hornung, V., Ablasser, A., Charrel-Dennis, M., Bauernfeind, F., Horvath, G., Caffrey, D. R., Latz, E., & Fitzgerald, K. A. (2009). AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature*. https://doi.org/10.1038/nature07725

Jachowicz, J. W., Bing, X., Pontabry, J., Bošković, A., Rando, O. J., & Torres-Padilla, M. E. (2017). LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nature Genetics*. https://doi.org/10.1038/ng.3945

Jacqueline, C., Bourfia, Y., Hbid, H., Sorci, G., Thomas, F., & Roche, B. (2016). Interactions between immune challenges and cancer cells proliferation: timing does matter! *Evolution, Medicine, and Public Health*. https://doi.org/10.1093/emph/eow025

Jansz, N., & Faulkner, G. J. (2021). Endogenous retroviruses in the origins and treatment of cancer. In *Genome Biology*. https://doi.org/10.1186/s13059-021-02357-4

Jerby-Arnon, L., Shah, P., Cuoco, M. S., Rodman, C., Su, M. J.,

Melms, J. C., Leeson, R., Kanodia, A., Mei, S., Lin, J. R., Wang, S., Rabasha, B., Liu, D., Zhang, G., Margolais, C., Ashenberg, O., Ott, P. A., Buchbinder, E. I., Haq, R., … Regev, A. (2018). A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. *Cell*. https://doi.org/10.1016/j.cell.2018.09.006

Jordan, I. K., Rogozin, I. B., Glazko, G. V., & Koonin, E. V. (2003). Origin of a substantial fraction of human regulatory sequences from transposable elements. In *Trends in Genetics*. https://doi.org/10.1016/S0168-9525(02)00006-9

Kambayashi, T., & Laufer, T. M. (2014). Atypical MHC class II-expressing antigen-presenting cells: Can anything replace a dendritic cell? In *Nature Reviews Immunology*. https://doi.org/10.1038/nri3754

Kapusta, A., Kronenberg, Z., Lynch, V. J., Zhuo, X., Ramsay, L. A., Bourque, G., Yandell, M., & Feschotte, C. (2013). Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. *PLoS Genetics*. https://doi.org/10.1371/journal.pgen.1003470

Karaayvaz, M., Cristea, S., Gillespie, S. M., Patel, A. P., Mylvaganam, R., Luo, C. C., Specht, M. C., Bernstein, B. E., Michor, F., & Ellisen, L. W. (2018). Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. *Nature Communications*. https://doi.org/10.1038/s41467-018-06052-0

Kassiotis, G., & Stoye, J. P. (2016). Immune responses to endogenous retroelements: Taking the bad with the good. In *Nature Reviews Immunology*. https://doi.org/10.1038/nri.2016.27

Kazazian, H. H. (2004). Mobile Elements: Drivers of Genome Evolution. In *Science*. https://doi.org/10.1126/science.1089670

Kazazian, H. H., & Moran, J. V. (1998). The impact of L1 retrotransposons on the human genome. In *Nature Genetics*. https://doi.org/10.1038/ng0598-19

Kelley, D., & Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biology*. https://doi.org/10.1186/gb-2012-13-11-r107

Klutstein, M., Nejman, D., Greenfield, R., & Cedar, H. (2016). DNA methylation in cancer and aging. In *Cancer Research*. https://doi.org/10.1158/0008-5472.CAN-15-3278

Kong, Y., Rose, C. M., Cass, A. A., Williams, A. G., Darwish, M., Lianoglou, S., Haverty, P. M., Tong, A. J., Blanchette, C., Albert, M. L., Mellman, I., Bourgon, R., Greally, J., Jhunjhunwala, S., & Chen-Harris, H. (2019). Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. *Nature Communications*. https://doi.org/10.1038/s41467-019-13035-2

Kubo, M. (2019). Interface between innate and adaptive immunity. *Japanese Journal of Allergology*. https://doi.org/10.15036/arerugi.68.132

Kumar, B. V., Connors, T. J., & Farber, D. L. (2018). Human T Cell Development, Localization, and Function throughout Life. In *Immunity*. https://doi.org/10.1016/j.immuni.2018.01.007

Lambrechts, D., Wauters, E., Boeckx, B., Aibar, S., Nittner, D., Burton, O., Bassez, A., Decaluwé, H., Pircher, A., Van den Eynde, K., Weynand, B., Verbeken, E., De Leyn, P., Liston, A., Vansteenkiste, J., Carmeliet, P., Aerts, S., & Thienpont, B.

(2018). Phenotype molding of stromal cells in the lung tumor microenvironment. *Nature Medicine*. https://doi.org/10.1038/s41591-018-0096-5

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., … Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature*. https://doi.org/10.1038/35057062

Li, H., van der Leun, A. M., Yofe, I., Lubling, Y., Gelbard-Solodkin, D., van Akkooi, A. C. J., van den Braber, M., Rozeman, E. A., Haanen, J. B. A. G., Blank, C. U., Horlings, H. M., David, E., Baran, Y., Bercovich, A., Lifshitz, A., Schumacher, T. N., Tanay, A., & Amit, I. (2019). Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. *Cell*. https://doi.org/10.1016/j.cell.2018.11.043

Liu, N., Lee, C. H., Swigut, T., Grow, E., Gu, B., Bassik, M. C., & Wysocka, J. (2018). Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature*. https://doi.org/10.1038/nature25179

Lu, J. Y., Chang, L., Li, T., Wang, T., Yin, Y., Zhan, G., Han, X., Zhang, K., Tao, Y., Percharde, M., Wang, L., Peng, Q., Yan, P., Zhang, H., Bi, X., Shao, W., Hong, Y., Wu, Z., Ma, R., … Shen, X. (2021). Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Research*. https://doi.org/10.1038/s41422-020-00466-6

Lu, J. Y., Shao, W., Chang, L., Yin, Y., Li, T., Zhang, H., Hong, Y., Percharde, M., Guo, L., Wu, Z., Liu, L., Liu, W., Yan, P.,

Ramalho-Santos, M., Sun, Y., & Shen, X. (2020). Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Reports*. https://doi.org/10.1016/j.celrep.2020.02.048

Luan, D. D., Korman, M. H., Jakubczak, J. L., & Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: A mechanism for non-LTR retrotransposition. *Cell*. https://doi.org/10.1016/0092-8674(93)90078-5

Macfarlan, T. S., Gifford, W. D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., Andrews, S. E., Franco, L., Rosenfeld, M. G., Ren, B., & Pfaff, S. L. (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes and Development*. https://doi.org/10.1101/gad.2008511

Mahnke, J., Schumacher, V., Ahrens, S., Käding, N., Feldhoff, L. M., Huber, M., Rupp, J., Raczkowski, F., & Mittrücker, H. W. (2016). Interferon Regulatory Factor 4 controls T H1 cell effector function and metabolism. *Scientific Reports*. https://doi.org/10.1038/srep35521

Maimela, N. R., Liu, S., & Zhang, Y. (2019). Fates of CD8+ T cells in Tumor Microenvironment. In *Computational and Structural Biotechnology Journal*. https://doi.org/10.1016/j.csbj.2018.11.004

Malik, H. S., Henikoff, S., & Eickbush, T. H. (2000). Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Research*. https://doi.org/10.1101/gr.145000

Man, K., Gabriel, S. S., Liao, Y., Gloury, R., Preston, S., Henstridge, D. C., Pellegrini, M., Zehn, D., Berberich-Siebelt, F., Febbraio,

M. A., Shi, W., & Kallies, A. (2017). Transcription Factor IRF4 Promotes CD8+ T Cell Exhaustion and Limits the Development of Memory-like T Cells during Chronic Infection. *Immunity*. https://doi.org/10.1016/j.immuni.2017.11.021

Marasca, F., Gasparotto, E., Polimeni, B., Vadalà, R., Ranzani, V., & Bodega, B. (2020a). The sophisticated transcriptional response governed by transposable elements in human health and disease. *International Journal of Molecular Sciences*, *21*(9). https://doi.org/10.3390/ijms21093201

Marasca, F., Gasparotto, E., Polimeni, B., Vadalà, R., Ranzani, V., & Bodega, B. (2020b). The sophisticated transcriptional response governed by transposable elements in human health and disease. In *International Journal of Molecular Sciences*. https://doi.org/10.3390/ijms21093201

Marian, A. J. (2014). Sequencing your genome: what does it mean? In *Methodist DeBakey cardiovascular journal*. https://doi.org/10.14797/mdcj-10-1-3

Martens, J. H. A., O'Sullivan, R. J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., & Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO Journal*. https://doi.org/10.1038/sj.emboj.7600545

McCLINTOCK, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.36.6.344

MCCLINTOCK, B. (1956). Intranuclear systems controlling gene action and mutation. *Brookhaven Symposia in Biology*.

Mirzaei, H. R. (2020). Adaptive Immunity. *Reference Module in*

Biomedical Sciences. https://doi.org/10.1016/B978-0-12-818731-9.00028-8

Molaro, A., & Malik, H. S. (2016). Hide and seek: How chromatin-based pathways silence retroelements in the mammalian germline. In *Current Opinion in Genetics and Development*. https://doi.org/10.1016/j.gde.2015.12.001

Monticelli, S. (2019). DNA (Hydroxy)Methylation in T Helper Lymphocytes. In *Trends in Biochemical Sciences*. https://doi.org/10.1016/j.tibs.2019.01.009

Morgan, H. D., Sutherland, H. G. E., Martin, D. I. K., & Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nature Genetics*. https://doi.org/10.1038/15490

Munoz-Lopez, M., & Garcia-Perez, J. (2010). DNA Transposons: Nature and Applications in Genomics. *Current Genomics*. https://doi.org/10.2174/138920210790886871

Natoli, G., & Ostuni, R. (2019). Adaptation and memory in immune responses. In *Nature Immunology*. https://doi.org/10.1038/s41590-019-0399-9

Ni, J. Z., Grate, L., Donohue, J. P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T. A., Blume, J. E., & Ares, M. (2007). Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes and Development*. https://doi.org/10.1101/gad.1525507

Padeken, J., Zeller, P., & Gasser, S. M. (2015). Repeat DNA in genome organization and stability. In *Current Opinion in Genetics and Development*. https://doi.org/10.1016/j.gde.2015.03.009

Pagès, F., Berger, A., Camus, M., Sanchez-Cabo, F., Costes, A., Molidor, R., Mlecnik, B., Kirilovsky, A., Nilsson, M., Damotte, D., Meatchi, T., Bruneval, P., Cugnenc, P.-H., Trajanoski, Z., Fridman, W.-H., & Galon, J. (2005). Effector Memory T Cells, Early Metastasis, and Survival in Colorectal Cancer. *New England Journal of Medicine*. https://doi.org/10.1056/nejmoa051424

Patsoukis, N., Brown, J., Petkova, V., Liu, F., Li, L., & Boussiotis, V. A. (2012). Selective effects of PD-1 on Akt and ras pathways regulate molecular components of the cell cycle and inhibit T cell proliferation. *Science Signaling*. https://doi.org/10.1126/scisignal.2002796

Payer, L. M., & Burns, K. H. (2019). Transposable elements in human genetic disease. In *Nature Reviews Genetics*. https://doi.org/10.1038/s41576-019-0165-8

Percharde, M., Lin, C. J., Yin, Y., Guan, J., Peixoto, G. A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., & Ramalho-Santos, M. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell*. https://doi.org/10.1016/j.cell.2018.05.043

Percharde, M., Sultana, T., & Ramalho-Santos, M. (2020). What Doesn't Kill You Makes You Stronger: Transposons as Dual Players in Chromatin Regulation and Genomic Variation. *BioEssays*. https://doi.org/10.1002/bies.201900232

Perepelitsa-Belancio, V., & Deininger, P. (2003). RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nature Genetics*. https://doi.org/10.1038/ng1269

Peschansky, V. J., & Wahlestedt, C. (2014). Non-coding RNAs as

direct and indirect modulators of epigenetic regulation. In *Epigenetics*. https://doi.org/10.4161/epi.27473

Plitas, G., & Rudensky, A. Y. (2020). Regulatory T Cells in Cancer. In *Annual Review of Cancer Biology*. https://doi.org/10.1146/annurev-cancerbio-030419-033428

Qi, J., Crinier, A., Escalière, B., Ye, Y., Wang, Z., Zhang, T., Batista, L., Liu, H., Hong, L., Wu, N., Zhang, M., Chen, L., Liu, Y., Shen, L., Narni-Mancinelli, E., Vivier, E., & Su, B. (2021). Single-cell transcriptomic landscape reveals tumor specific innate lymphoid cells associated with colorectal cancer progression. *Cell Reports Medicine*. https://doi.org/10.1016/j.xcrm.2021.100353

Quigley, M., Pereyra, F., Nilsson, B., Porichis, F., Fonseca, C., Eichbaum, Q., Julg, B., Jesneck, J. L., Brosnahan, K., Imam, S., Russell, K., Toth, I., Piechocka-Trocha, A., Dolfi, D., Angelosanto, J., Crawford, A., Shin, H., Kwon, D. S., Zupkosky, J., … Haining, W. N. (2010). Transcriptional analysis of HIV-specific CD8+ T cells shows that PD-1 inhibits T cell function by upregulating BATF. *Nature Medicine*. https://doi.org/10.1038/nm.2232

Robbez-Masson, L., Tie, C. H. C., & Rowe, H. M. (2017). Cancer cells, on your histone marks, get SET DB1, silence retrotransposons, and go! *Journal of Cell Biology*. https://doi.org/10.1083/jcb.201710068

Rodriguez-Terrones, D., Hartleben, G., Gaume, X., Eid, A., Guthmann, M., Iturbide, A., & Torres-Padilla, M. (2020). A distinct metabolic state arises during the emergence of 2-cell-like cells. *EMBO Reports*. https://doi.org/10.15252/embr.201948354

Roulois, D., Loo Yau, H., Singhania, R., Wang, Y., Danesh, A., Shen,

S. Y., Han, H., Liang, G., Jones, P. A., Pugh, T. J., O'Brien, C., & De Carvalho, D. D. (2015). DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell*. https://doi.org/10.1016/j.cell.2015.07.056

Roy-Engel, A. M., El-Sawy, M., Farooq, L., Odom, G. L., Perepelitsa-Belancio, V., Bruch, H., Oyeniran, O. O., & Deininger, P. L. (2005). Human retroelements may introduce intragenic polyadenylation signals. *Cytogenetic and Genome Research*. https://doi.org/10.1159/000084968

Saito, Y., Nakaoka, T., Sakai, K., Muramatsu, T., Toshimitsu, K., Kimura, M., Kanai, T., Sato, T., & Saito, H. (2016). Inhibition of DNA Methylation Suppresses Intestinal Tumor Organoids by Inducing an Anti-Viral Response. *Scientific Reports*. https://doi.org/10.1038/srep25311

Sakuishi, K., Apetoh, L., Sullivan, J. M., Blazar, B. R., Kuchroo, V. K., & Anderson, A. C. (2010). Targeting Tim-3 and PD-1 pathways to reverse T cell exhaustion and restore anti-tumor immunity. *Journal of Experimental Medicine*. https://doi.org/10.1084/jem.20100643

Saleh, R., Sasidharan Nair, V., Toor, S. M., & Elkord, E. (2022). Intrinsic and acquired cancer immunotherapy resistance. In *Cancer Immunology and Immunotherapy*. https://doi.org/10.1016/b978-0-12-823397-9.00014-4

Sasada, T., & Suekane, S. (2011). Variation of tumor-infiltrating lymphocytes in human cancers: Controversy on clinical significance. In *Immunotherapy*. https://doi.org/10.2217/imt.11.106

Savas, P., Virassamy, B., Ye, C., Salim, A., Mintoff, C. P., Caramia, F., Salgado, R., Byrne, D. J., Teo, Z. L., Dushyanthen, S., Byrne, A., Wein, L., Luen, S. J., Poliness, C., Nightingale, S. S., Skandarajah, A. S., Gyorki, D. E., Thornton, C. M., Beavis, P. A., … Loi, S. (2018). Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nature Medicine*. https://doi.org/10.1038/s41591-018-0078-7

Scacheri, C. A., & Scacheri, P. C. (2015). Mutations in the noncoding genome. In *Current Opinion in Pediatrics*. https://doi.org/10.1097/MOP.0000000000000283

Schmidt, D., Schwalie, P. C., Wilson, M. D., Ballester, B., Gonalves, Â., Kutter, C., Brown, G. D., Marshall, A., Flicek, P., & Odom, D. T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell*. https://doi.org/10.1016/j.cell.2011.11.058

Schreiber, R. D., Old, L. J., & Smyth, M. J. (2011). Cancer immunoediting: Integrating immunity's roles in cancer suppression and promotion. In *Science*. https://doi.org/10.1126/science.1203486

Schwarz, B. A., & Bhandoola, A. (2006). Trafficking from the bone marrow to the thymus: A prerequisite for thymopoiesis. In *Immunological Reviews*. https://doi.org/10.1111/j.0105-2896.2006.00350.x

Seeley, J. J., & Ghosh, S. (2017). Molecular mechanisms of innate memory and tolerance to LPS. *Journal of Leukocyte Biology*. https://doi.org/10.1189/jlb.3mr0316-118rr

Sheppard, K. A., Fitz, L. J., Lee, J. M., Benander, C., George, J. A.,

Wooters, J., Qiu, Y., Jussif, J. M., Carter, L. L., Wood, C. R., & Chaudhary, D. (2004). PD-1 inhibits T-cell receptor induced phosphorylation of the ZAP70/CD3ζ signalosome and downstream signaling to PKCθ. *FEBS Letters*. https://doi.org/10.1016/j.febslet.2004.07.083

Sim, F., Leidner, R., & Bell, R. B. (2019). Immunotherapy for Head and Neck Cancer. In *Oral and Maxillofacial Surgery Clinics of North America*. https://doi.org/10.1016/j.coms.2018.09.002

Slotkin, R. K., & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. In *Nature Reviews Genetics*. https://doi.org/10.1038/nrg2072

Smit, A. F. (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. In *Current Opinion in Genetics and Development*. https://doi.org/10.1016/S0959-437X(99)00031-3

Smith, C. C., Beckermann, K. E., Bortone, D. S., Cubas, A. A., Bixby, L. M., Lee, S. J., Panda, A., Ganesan, S., Bhanot, G., Wallen, E. M., Milowsky, M. I., Kim, W. Y., Rathmell, K., Swanstrom, R., Parker, J. S., Serody, J. S., Selitsky, S. R., & Vincent, B. G. (2018). Endogenous retroviral signatures predict immunotherapy response in clear cell renal cell carcinoma. *Journal of Clinical Investigation*. https://doi.org/10.1172/JCI121476

Smith, Z. D., Chan, M. M., Humm, K. C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., & Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. *Nature*. https://doi.org/10.1038/nature13581

Stark, J. M., Tibbitt, C. A., & Coquet, J. M. (2019). The metabolic requirements of Th2 cell differentiation. In *Frontiers in*

Immunology. https://doi.org/10.3389/fimmu.2019.02318

Staron, M. M., Gray, S. M., Marshall, H. D., Parish, I. A., Chen, J. H., Perry, C. J., Cui, G., Li, M. O., & Kaech, S. M. (2014). The Transcription Factor FoxO1 Sustains Expression of the Inhibitory Receptor PD-1 and Survival of Antiviral CD8+ T Cells during Chronic Infection. *Immunity*. https://doi.org/10.1016/j.immuni.2014.10.013

Starr, T. K., Jameson, S. C., & Hogquist, K. A. (2003). Positive and negative selection of T cells. In *Annual Review of Immunology*. https://doi.org/10.1146/annurev.immunol.21.120601.141107

Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P., Snyder, M. P., & Wang, T. (2014). Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Research*. https://doi.org/10.1101/gr.168872.113

Swain, S. L., McKinstry, K. K., & Strutt, T. M. (2012). Expanding roles for CD4 + T cells in immunity to viruses. In *Nature Reviews Immunology*. https://doi.org/10.1038/nri3152

Takaoka, A., Wang, Z., Choi, M. K., Yanai, H., Negishi, H., Ban, T., Lu, Y., Miyagishi, M., Kodama, T., Honda, K., Ohba, Y., & Taniguchi, T. (2007). DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature*. https://doi.org/10.1038/nature06013

Thornburg, B. G., Gotea, V., & Makałowski, W. (2006). Transposable elements as a significant source of transcription regulating signals. *Gene*. https://doi.org/10.1016/j.gene.2005.09.036

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G.,

Fallahi-Sichani, M., Dutton-Regester, K., Lin, J. R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G. K., … Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*. https://doi.org/10.1126/science.aad0501

Truffi, M., Sorrentino, L., & Corsi, F. (2020). Fibroblasts in the Tumor Microenvironment. In *Advances in Experimental Medicine and Biology*. https://doi.org/10.1007/978-3-030-37184-5_2

Tucker, J. M., & Glaunsinger, B. A. (2017). Host Noncoding Retrotransposons Induced by DNA Viruses: a SINE of Infection? *Journal of Virology*. https://doi.org/10.1128/jvi.00982-17

Unterholzner, L., Keating, S. E., Baran, M., Horan, K. A., Jensen, S. B., Sharma, S., Sirois, C. M., Jin, T., Latz, E., Xiao, T. S., Fitzgerald, K. A., Paludan, S. R., & Bowie, A. G. (2010). IFI16 is an innate immune sensor for intracellular DNA. *Nature Immunology*. https://doi.org/10.1038/ni.1932

Uzman, A., Lodish, H., Berk, A., Zipursky, L., & Baltimore, D. (2000). Molecular Cell Biology (4th edition) New York, NY, 2000, ISBN 0-7167-3136-3. *Biochemistry and Molecular Biology Education*. https://doi.org/10.1016/S1470-8175(01)00023-6

Vignali, D. A. A., Collison, L. W., & Workman, C. J. (2008). How regulatory T cells work. In *Nature Reviews Immunology*. https://doi.org/10.1038/nri2343

Viollet, S., Monot, C., & Cristofari, G. (2014). L1 retrotransposition: The snap-velcro model and its consequences. *Mobile Genetic Elements*. https://doi.org/10.4161/mge.28907

Vivier, E., & Malissen, B. (2005). Innate and adaptive immunity: Specificities and signaling hierarchies revisited. In *Nature*

Immunology. https://doi.org/10.1038/ni1153

Waddington, C. H. (2012). The epigenotype. 1942. *International Journal of Epidemiology*. https://doi.org/10.1093/ije/dyr184

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., Li, W., Zhou, Q., Aluru, N., Tang, F., He, C., Huang, X., & Liu, J. (2014). Programming and inheritance of parental DNA methylomes in mammals. *Cell*. https://doi.org/10.1016/j.cell.2014.04.017

Weaver, C. T., Harrington, L. E., Mangan, P. R., Gavrieli, M., & Murphy, K. M. (2006). Th17: An Effector CD4 T Cell Lineage with Regulatory T Cell Ties. In *Immunity*. https://doi.org/10.1016/j.immuni.2006.06.002

Wells, J. N., & Feschotte, C. (2020). A Field Guide to Eukaryotic Transposable Elements. In *Annual Review of Genetics*. https://doi.org/10.1146/annurev-genet-040620-022145

Wherry, E. J. (2011). T cell exhaustion. In *Nature Immunology*. https://doi.org/10.1038/ni.2035

Wherry, E. J., & Kurachi, M. (2015). Molecular and cellular insights into T cell exhaustion. In *Nature Reviews Immunology*. https://doi.org/10.1038/nri3862

Whitcomb, J. M., & Hughes, S. H. (1992). Retroviral reverse transcription and integration: Progress and problems. In *Annual Review of Cell Biology*. https://doi.org/10.1146/annurev.cb.08.110192.001423

Wilkinson, G. W., & Lowenstein, P. R. (1994). Introduction to gene transfer: viral vectors. *Gene Therapy*.

Wilson, C. B., Rowell, E., & Sekimata, M. (2009). Epigenetic control of T-helper-cell differentiation. In *Nature Reviews Immunology*.

https://doi.org/10.1038/nri2487

Wong, P., & Pamer, E. G. (2003). CD8 T cell responses to infectious pathogens. In *Annual Review of Immunology*. https://doi.org/10.1146/annurev.immunol.21.120601.141114

Workman, C. J., Cauley, L. S., Kim, I.-J., Blackman, M. A., Woodland, D. L., & Vignali, D. A. A. (2004). Lymphocyte Activation Gene-3 (CD223) Regulates the Size of the Expanding T Cell Population Following Antigen Activation In Vivo. *The Journal of Immunology*. https://doi.org/10.4049/jimmunol.172.9.5450

Wu, F., Fan, J., He, Y., Xiong, A., Yu, J., Li, Y., Zhang, Y., Zhao, W., Zhou, F., Li, W., Zhang, J., Zhang, X., Qiao, M., Gao, G., Chen, S., Chen, X., Li, X., Hou, L., Wu, C., … Zhou, C. (2021). Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nature Communications*. https://doi.org/10.1038/s41467-021-22801-0

Yang, F., & Wang, P. J. (2016). Multiple LINEs of retrotransposon silencing mechanisms in the mammalian germline. In *Seminars in Cell and Developmental Biology*. https://doi.org/10.1016/j.semcdb.2016.03.001

Yi, J. S., Cox, M. A., & Zajac, A. J. (2010). T-cell exhaustion: Characteristics, causes and conversion. In *Immunology*. https://doi.org/10.1111/j.1365-2567.2010.03255.x

Yu, Y., & Cui, J. (2018). Present and future of cancer immunotherapy: A tumor microenvironmental perspective. In *Oncology Letters*. https://doi.org/10.3892/ol.2018.9219

Yuyang Lu, J., Chang, L., Li, T., Wang, T., Yin, Y., Zhan, G., Zhang, K., Percharde, M., Wang, L., Peng, Q., Yan, P., Zhang, H., Han, X., Bi, X., Shao, W., Hong, Y., Wu, Z., Wang, P., Li, W., … Shen,

X. (2019). L1 and B1 repeats blueprint the spatial organization of chromatin. *BioRxiv*. https://doi.org/10.1101/802173

Zeng, M., Hu, Z., Shi, X., Li, X., Zhan, X., Li, X. D., Wang, J., Choi, J. H., Wang, K. W., Purrington, T., Tang, M., Fina, M., DeBerardinis, R. J., Moresco, E. M. Y., Pedersen, G., McInerney, G. M., Hedestam, G. B. K., Chen, Z. J., & Beutler, B. (2014). MAVS, cGAS, and endogenous retroviruses in T-independent B cell responses. *Science*. https://doi.org/10.1126/science.346.6216.1486

Zhang, L., Li, Z., Skrzypczynska, K. M., Fang, Q., Zhang, W., O'Brien, S. A., He, Y., Wang, L., Zhang, Q., Kim, A., Gao, R., Orf, J., Wang, T., Sawant, D., Kang, J., Bhatt, D., Lu, D., Li, C. M., Rapaport, A. S., … Yu, X. (2020). Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer. *Cell*. https://doi.org/10.1016/j.cell.2020.03.048

Zhang, L., Yu, X., Zheng, L., Zhang, Y., Li, Y., Fang, Q., Gao, R., Kang, B., Zhang, Q., Huang, J. Y., Konno, H., Guo, X., Ye, Y., Gao, S., Wang, S., Hu, X., Ren, X., Shen, Z., Ouyang, W., & Zhang, Z. (2018). Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature*. https://doi.org/10.1038/s41586-018-0694-x

Zhang, Y., Li, T., Preissl, S., Amaral, M. L., Grinstein, J. D., Farah, E. N., Destici, E., Qiu, Y., Hu, R., Lee, A. Y., Chee, S., Ma, K., Ye, Z., Zhu, Q., Huang, H., Fang, R., Yu, L., Izpisua Belmonte, J. C., Wu, J., … Ren, B. (2019). Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nature Genetics*. https://doi.org/10.1038/s41588-019-0479-7

Zhao, K., Du, J., Peng, Y., Li, P., Wang, S., Wang, Y., Hou, J., Kang, J., Zheng, W., Hua, S., & Yu, X. F. (2018). LINE1 contributes to autoimmunity through both RIG-I- and MDA5-mediated RNA sensing pathways. *Journal of Autoimmunity*. https://doi.org/10.1016/j.jaut.2018.02.007

Zheng, C., Zheng, L., Yoo, J. K., Guo, H., Zhang, Y., Guo, X., Kang, B., Hu, R., Huang, J. Y., Zhang, Q., Liu, Z., Dong, M., Hu, X., Ouyang, W., Peng, J., & Zhang, Z. (2017). Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell*. https://doi.org/10.1016/j.cell.2017.05.035

Zhu, J., & Paul, W. E. (2008). CD4 T cells: Fates, functions, and faults. *Blood*. https://doi.org/10.1182/blood-2008-05-078154

Zilionis, R., Engblom, C., Pfirschke, C., Savova, V., Zemmour, D., Saatcioglu, H. D., Krishnan, I., Maroni, G., Meyerovitz, C. V., Kerwin, C. M., Choi, S., Richards, W. G., De Rienzo, A., Tenen, D. G., Bueno, R., Levantini, E., Pittet, M. J., & Klein, A. M. (2019). Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. *Immunity*. https://doi.org/10.1016/j.immuni.2019.03.009

Zinselmeyer, B. H., Heydari, S., Sacristán, C., Nayak, D., Cammer, M., Herz, J., Cheng, X., Davis, S. J., Dustin, M. L., & McGavern, D. B. (2013). PD-1 promotes immune exhaustion by inducing antiviral T cell motility paralysis. *Journal of Experimental Medicine*. https://doi.org/10.1084/jem.20121416

Zitvogel, L., Tesniere, A., & Kroemer, G. (2006). Cancer despite immunosurveillance: Immunoselection and immunosubversion. In *Nature Reviews Immunology*. https://doi.org/10.1038/nri1936

Chapter 2. Thesis Results “LINE1 RNAs regulate Tumor infiltrating lymphocytes identity and plasticity”

2.1 Introduction

Innate and adaptive immune responses play a fundamental role in tumorigenesis; in the tumor microenvironment, the crosstalk between tumor infiltrating T lymphocytes (TILs) and tumor cells, results in either tumor elimination, equilibrium between immune response and residual tumor cell growth, or tumor escape from immune control (Quezada & Peggs, 2011). Tumor-dependent immunosuppressive mechanisms rely on the upregulation of modulatory molecules, called immune checkpoints, whose function is only partially characterized (Jamal-Hanjani et al., 2013). Nevertheless, these molecules (e.g. PD-1) are targeted by antibodies towards immune checkpoint inhibitors, that unleash the spontaneous anti-tumor immune responses in such a powerful way that it has created a paradigm shift in cancer therapy (Munn & Bronte, 2016). However, the fraction of patients that do not respond remains high, and the efforts in the field are mainly focused on searching specific inhibitors against novel surface markers (Guo et al., 2018; Zhang et al., 2018). Epigenetic mechanisms that govern the intratumoral dysfunctional state of TILs are still elusive, and nothing is reported on their targeting to reestablish TILs function. It is the basis of my PhD project to better comprehend molecular

mechanisms underlying at a transcriptional level the immune response against the tumor and to discover novel therapeutic targets for enhancing TILs function.

Despite human genome composition could estimate for up to 66-69% of repetitive DNA, the functions of this fraction are still largely ignored. Among the non-coding genome, Transposable Elements (TEs) (45% of the human genome) and particularly LINE1 (18% of the human genome) are boosting as novel key molecules involved in epigenetic regulation of cell identity (Chuong et al., 2017). Functionally, LINE1 can have several roles: they can provide novel regulatory regions, act as alternative promoters (Criscione et al., 2016; Denli et al., 2015), be binding sites for transcription factors (Bourque et al., 2008), and represent polyadenylation sites (Roy-Engel et al., 2005) or alternative splicing sites (Belancio et al., 2006). Mechanistically, chromatin LINE1 RNAs can regulate open chromatin accessibility during embryogenesis (Fadloun et al., 2013; Lavin et al., 2017), contributing to both maintenances of cell identity and 2-cell stage differentiation in complex with Nucleolin-Kap (Percharde et al., 2018). Although all these pieces of evidence underlined the importance of TEs, and in particular LINE1, in shape cellular transcriptional plasticity, their contribution to adult cell identity, plasticity and adaptation is unexplored. The study in which I was involved uncovers a novel epigenetic mechanism contributing to the enforcement of T-cell quiescence and suggests that LINE1 RNAs abundance is critical for T-cell

effector function in physiological and pathological contexts. In this broad study, my PhD project will be entirely focused in studying the LINE1 dynamic in shaping dysfunctional state of T cell in a pathological context like tumor. We demonstrated that dysfunctional T cells *in vitro* and TILs *ex vivo* reaccumulate LINE1 RNAs at chromatin and their downregulation reinforces T cell effector response. Therefore, the modulation of LINE1 RNAs has the potential of being an important mechanism in determining at transcriptional level TILs function in the tumor microenvironment.

2.2 Results

2.2.1 LINE1 RNAs are enriched in naïve CD4⁺ T cells and regulates their functions

At first, we investigated the expression and distribution of TEs in ex vivo sorted naïve and memory $CD4^+$ and $CD8^+$ T cell subsets derived from buffy coats of healthy donors. By quantitative PCR with reverse transcription (RT-qPCR) and RNA fluorescence in situ hybridization (RNA-FISH,) we assessed LINE1, HERV and Alu elements. Among all the three classes analyzed we found that LINE1 RNAs were enriched in the nuclei of naïve $CD4^+$ T cells in respect to all the other subsets (Fig. 1a,b,g); while Alu elements exhibit a broader and perinuclear distribution with higher expression in memory T cells subsets (Fig. 1c,d,h); lastly HERV elements were poorly expressed (Fig. 1 e,f,i). To

determine the LINE1 RNAs location within the nuclear compartment we performed a biochemical fractionation finding that LINE1 RNAs were entirely localized at the chromatin of naïve $CD4^+$ T cells (data not shown).

Since this peculiar enrichment of LINE1 RNAs in naïve $CD4^+$ T cells, we decided to test LINE1 dynamics upon naïve $CD4^+$ T cells in vitro activation and differentiation toward T_H1 phenotype. The activation was performed thanks to TCR stimulation using anti-CD3 and anti-CD28 beads while the differentiation was allowed thanks to a cytokines cocktail composed of IL-2, a pro-survival cytokine, IL-12 and anti-IFNγ to support T_H1 differentiation. We verify, by RT-qPCR, that LINE1 RNAs were downregulated just a few hours (8 hours) after TCR activation and their level remained low during differentiation, even after 7 days of culturing (Fig. 2a). We obtain the same results with the differentiation toward T_H2 and T_H17 phenotypes (data not shown). All these results pointed out a specific LINE1 RNAs pattern in naïve $CD4^+$ T cells. To better characterize which LINE1 elements are expressed in naïve $CD4^+$ T cells we performed an RNA sequencing of chromatin RNA of these cells. We found that among TEs classes, the chromatin of naïve $CD4^+$ T cells was enriched with LINE1M families. The majority of LINE1 elements were localized within the intron of protein-coding genes, promoting non-canonical transcripts variants, indeed they were spliced as new exons that contained LINE1 and an intronic region

(data not shown). We were able to validate the existence of these transcripts by RT-PCR in three different individuals (Fig. 2b).

To understand the role of LINE1-containing transcripts in naïve $CD4^+$ T cells we decided to downregulate LINE1 RNAs by antisense oligonucleotides (ASOs). These ASOs contain a chemical modification that permits their entry into the cell without the use of transfection agents and without inducing differentiation or toxic effect. These oligos can entry in the nucleus of the cells, bind their target RNAs inducing the cleavage by RNAse H activity. To perform control of these experiments we treated the cells with scr ASOs which is designed on non-human genomic region. Using specific LINE1 ASOs for 48 hours we were able to knock down LINE1 RNAs in naïve $CD4^+$ T cells as we observed by RT-qPCR and LINE1 RNA-FISH, observing a reduction of LINE1 RNAs expression at least 50% (Fig. 2 c-e). Moreover, using LINE1 ASOs in naïve $CD4^+$ T cells we detected a downregulation of LINE1-containing transcripts and an upregulation of canonical transcripts (data not shown). We activated and differentiated toward T_H1 phenotype naïve $CD4^+$ T cells treated with LINE1 ASOs or control scr ASOs. We verified that the knockdown of LINE1-containing transcripts increases cytokines production evaluating, by intracellular staining, the production of IFN-γ, that is the cytokines specifically expressed by T_H1 cells. Indeed, we detected around the double of IFN-γ production in LINE1 knockdown cells (Fig. 2 f-g), suggesting that

depletion of LINE1 RNAs can enhance the T cell effector response.

All these results suggest that the LINE1-containing transcripts expressed in the chromatin of naïve $CD4^+$ T cells can modulate the switch from quiescence to activation in these cells.

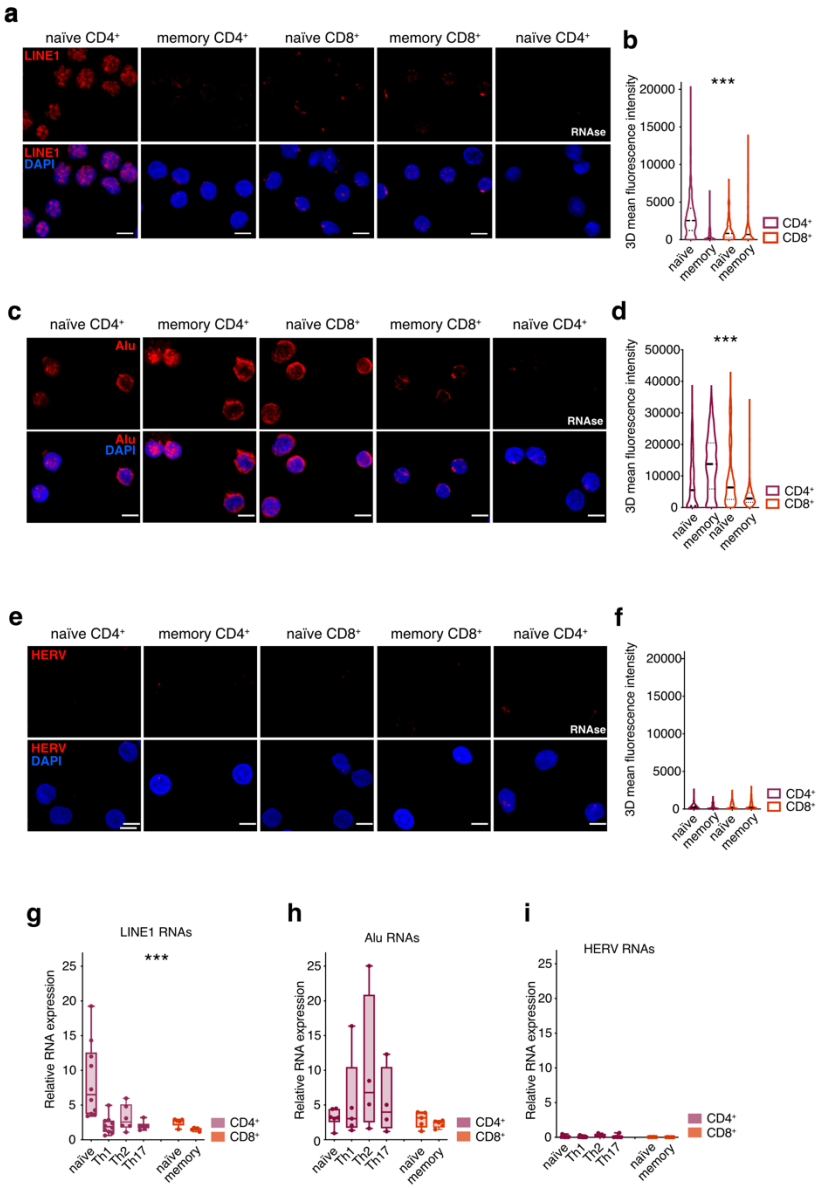


Figure 1. LINE1 RNAs are enriched in naïve CD4⁺ T cells

a, Confocal fluorescence microscopy images of LINE1 RNA-FISH (red) of naive and memory CD4⁺ and CD8⁺ T cells. LINE1 riboprobes were designed based on the LINE1 ORF2 region. Original magnification,

63×. Scale bar, 5 μm. **b**, Quantification of LINE1 RNA-FISH of at least 250 nuclei from four individuals (3D, three-dimensional). ***$P < 1 \times 10^{-15}$, one-way ANOVA. **c** Confocal fluorescence microscopy images of Alu RNA FISH (red) of naïve and memory $CD4^+$ and $CD8^+$ T cells. Original magnification 63×. Scale bar 5 μm. **d** Quantification of Alu RNA FISH, at least 220 nuclei, four individuals. *** $P < 1 \times 10^{-15}$, One-way ANOVA. **e** Confocal fluorescence microscopy images of HERV RNA FISH (red) performed on quiescent naïve and memory $CD4^+$ and $CD8^+$ T cells. Original magnification 63×. Scale bar 5 μm. **f** Quantification of HERV RNA FISH, at least 164 nuclei, three individuals. **g**, LINE1 expression measured by RT–qPCR in naive and memory T_H1, T_H2 and T_H17 $CD4^+$ T cells and in naive and memory $CD8^+$ T cells ($n = 10$ individuals for naive and T_H1 $CD4^+$ cells, $n = 6$ for T_H2 and T_H17 cells, $n = 5$ for $CD8^+$ subsets). ***$P = 0.0001$, one-way ANOVA. For box-and-whisker plots, the central line, the box and whiskers represent the median, the interquartile range (IQR) from first to third quartiles and 1.5 × IQR, respectively **h** Alu expression by RT-qPCR in quiescent naïve and memory T_H1, T_H2 and T_H17 $CD4^+$ T cells and in quiescent naïve and memory $CD8^+$ T cells, (n = 6 individuals for naïve $CD4^+$: n = 5 for T_H1 and $CD8^+$ subsets, n = 4 for T_H2 and T_H17). **i** HERV expression by RT-qPCR in quiescent naïve and memory T_H1, T_H2 and T_H17 $CD4^+$ T cells and in quiescent naïve and memory $CD8^+$ T cells, (n = 10 individuals for naïve and T_H1 $CD4^+$, n = 6 T_H17; n = 5 for T_H2 and $CD8^+$ subsets).

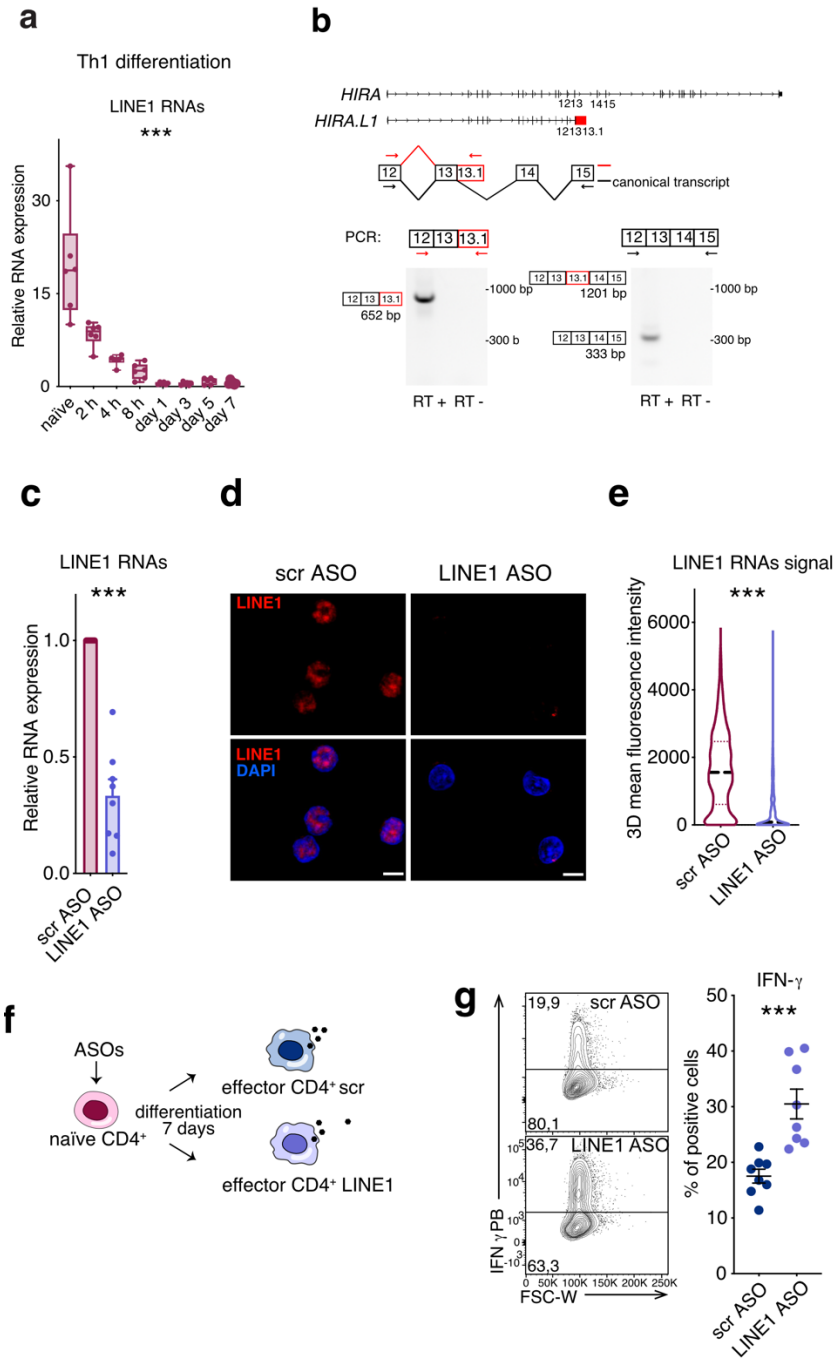


Figure 2. LINE1 RNAs regulate CD4⁺ T cells functions.

a LINE1 expression measured by RT–qPCR in quiescent and activated naive $CD4^+$ T cells ($n=6$ individuals). $^{***}P=1.16\times10^{-12}$, one-way ANOVA. **b** Scheme of canonical transcripts and LINE1-containing transcripts; the LINE1 exon is represented in red. In the middle, schemes of the PCR primers designed to verify the presence of the LINE1-containing transcripts and canonical transcripts are reported. Below, agarose gels representing RT-PCR results of LINE1-containing transcripts, HIRA. PCRs have been repeated on three individuals obtaining the same results. **c** LINE1 expression by RT-qPCR in quiescent naïve $CD4^+$ T cells treated with LINE1 or scr ASOs (n = 8 individuals). LINE1 ASOs were designed to target the ORF2 LINE1 region included in the LINE1-containing transcripts. Data are represented as mean ± s.e.m. Scr versus LINE1 ASO groups: *** $P = 3 \times 10^{-6}$, Two-tailed paired *t-test*. **d** Confocal fluorescence microscopy images of LINE1 RNA FISH (red) of quiescent naïve $CD4^+$ T cells treated with LINE1 or control (scr) ASOs. Original magnification 63×. Scale bar 5 μm. **e** Quantification of (**d**), at least 832 nuclei, two individuals. scr versus LINE1 ASO groups: *** $P < 1 \times 10^{-15}$, Two-tailed Mann Whitney test. **f** Scheme of LINE1-containing transcripts-knockdown (**g**) experiments in effector $CD4^+$ T cells. **g** IFN-γ-positive cells in effector $CD4^+$ T cells treated with LINE1 or scr ASOs ($n=8$ individuals). Data represent mean ± s.e.m. IFN-γ^+ cells in scr versus LINE1 ASO groups, $^{***}P=0.0002$, two-tailed paired *t*-test.

2.2.2 In vitro exhausted T cells show a re-expression of LINE1-containing transcripts.

Considering our results by which we demonstrated that LINE1-containing transcripts regulate T cell effector functions we decided to investigate LINE1 RNAs dynamic in T cells with dysfunctional effector properties. To this purpose, we set up an in vitro model in which we recapitulate features of the tumor microenvironment by the chronic stimulation of the TCR every two days of naïve $CD4^+$ T cells for 9 days (Fig. 3a) (Balkhi et al., 2018). Given the importance of $CD8^+$ T cells in the tumor microenvironment, we decided to test also the LINE1 RNAs dynamics in dysfunctional $CD8^+$ T cells (Fig. 4a). With these repetitive anti-CD3 stimulations we obtain $CD4^+$ and $CD8^+$ T cells that display an arrest of T cell growth as shown by cell growth curve, that we obtain with the cell count, and by cell trace that demonstrates a delay in proliferation (Fig. 3b and 4b). Indeed, cell trace, which enters by diffusion through the plasma membrane into the cells, permanently labels cells without affecting cells' morphology and permits to trace generations and divisions by their binding to amine groups of protein. We verified that exhausted $CD4^+$ T cells started to proliferate slowly after three TCR stimulation, from day 6, while exhausted $CD8^+$ T cells since proliferate less, started to decrease in growth after two

rounds of TCR stimulation, from day 4. We investigate also cell cycle progression but we didn't find any particular changes in cell cycle phases (Fig. 3e).

Then, we tested the presence of inhibitory receptors like PD-1, this receptor is a marker of exhausted phenotype and is the major target involved in immunotherapy. By receptor staining and RT-qPCR, we found the overexpression of PD-1 specifically in exhausted T cells both $CD4^+$ and $CD8^+$ T cells after three rounds of TCR stimulation (Fig. 3c and 4b). Moreover, we evaluated the effector properties by intracellular cytokines staining observing a decrease of IFN-γ production in exhausted $CD4^+$ T cells in respect to effector $CD4^+$ T cells, since we cultured the cells with specific T_H1 cytokines cocktails (Fig. 3d). Regarding the effector properties of effector $CD8^+$ T cells, we tested not only IFN-γ but also GrzB and PerfA production assessing the reduction of all these cytokines specifically in exhausted $CD8^+$ T cells (Fig. 4c). All these pieces of evidence demonstrated that, with the chronically stimulation of TCR, we were able to recapitulate a dysfunctional and exhausted phenotype of T cells as observed in the tumor microenvironment.

In these exhausted T cells, we investigated LINE1 RNAs dynamics by LINE1 RNA-FISH and RT-qPCR, and we assessed a statistically significant accumulation of LINE1 in the nuclei of exhausted T cells in respect to effector T cells, both for $CD4^+$ and $CD8^+$ T cells (Fig. 5a-b and Fig. 6a-b). Focusing on LINE1-containing transcripts we evaluated a consistent accumulation of

LINE1-containing transcripts accompanied by a reduction of canonical transcripts in exhausted $CD4^+$ and $CD8^+$ T cells (Fig. 5c and Fig. 6c).

These findings corroborate the role of LINE1-containing transcripts in regulating T cell identity and functions and demonstrate that the downregulation of LINE1 RNAs upon naïve $CD4^+$ T cells activation could be reverted.

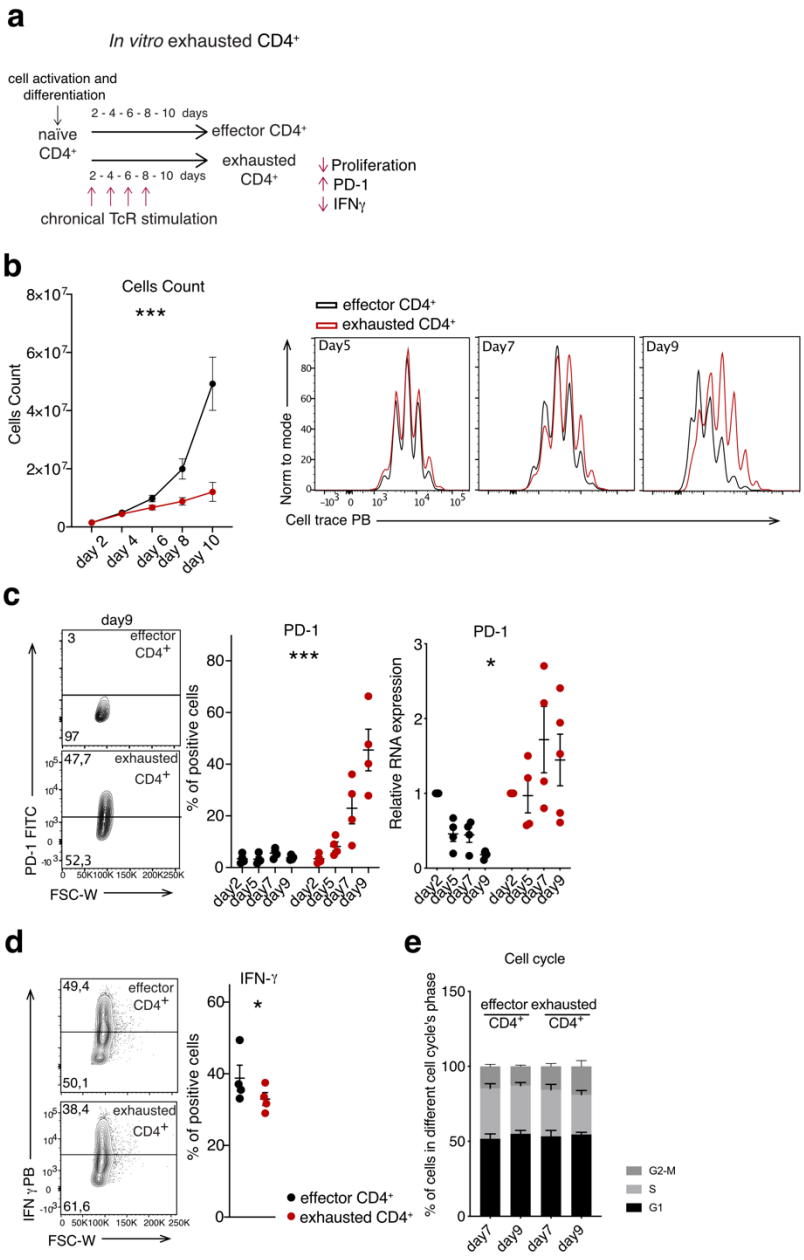


Figure 3. In vitro exhausted CD4+ T cells show a dysfunctional phenotype.

a Scheme of *in vitro* exhaustion of $CD4^+$ T cells. **b** On the left, cell count of effector (black) and exhausted (red) $CD4^+$ T cells (n = 5 individuals). Data are represented as mean ± s.e.m, *** $P < 2.7 \times 10^{-7}$, Two-way ANOVA. On the right histogram of cell trace staining on effector (black) and exhausted (red) $CD4^+$ T cells; from left to right histogram represent proliferation on day5, day7 and day9. **c** On the left, contour plot representation of PD-1 receptor staining; in the middle percentage of PD-1 positive cells (n = 4 individuals); on the right PD-1 expression by RT-qPCR. Data are represented as mean ± s.e.m, *** P = 0.00004, * P= 0.03 Two-way ANOVA. **d** IFN-γ positive cells (n = 4 individuals). Data are represented as mean ± s.e.m. IFN-γ: * P = 0.0319, One-tailed paired *t-test*. **e** Percentage of cells in different cell cycle phases at day 7 and day 9 (n=4 individuals) Data are represented as mean ± s.e.m

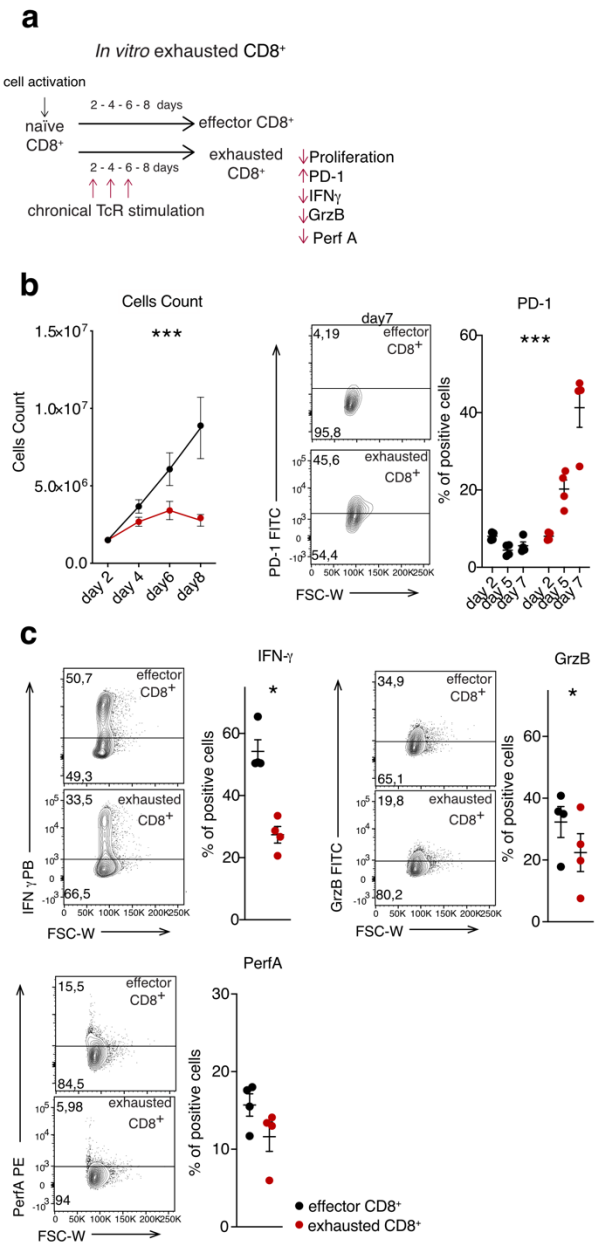


Figure 4. In vitro exhausted CD8⁺ T cells show a dysfunctional phenotype

a Scheme of *in vitro* exhaustion of CD8⁺ T cells. **b** On the left, cell count of effector and exhausted CD8⁺ T cells. Data are mean and ± s.e.m, N

= 4 individuals. *** P = 0.0003, Two-way ANOVA. In the middle, contour plot representation of PD-1 receptor staining, on the right percentage of PD-1 positive cells (n=4 individuals). Data are mean and ± s.e.m, *** P = 0.00003, Two-way ANOVA. **c** IFN-γ, GrzB and PerfA contour plot representation and percentage of IFN-γ, GrzB and PerfA positive cells (n=4 individuals). Data are mean and ± s.e.m. IFN-γ: * P = 0.0108, Two-tailed Paired *t-test*; GrzB * P = 0.0243, Two-tailed Paired *t-test*.

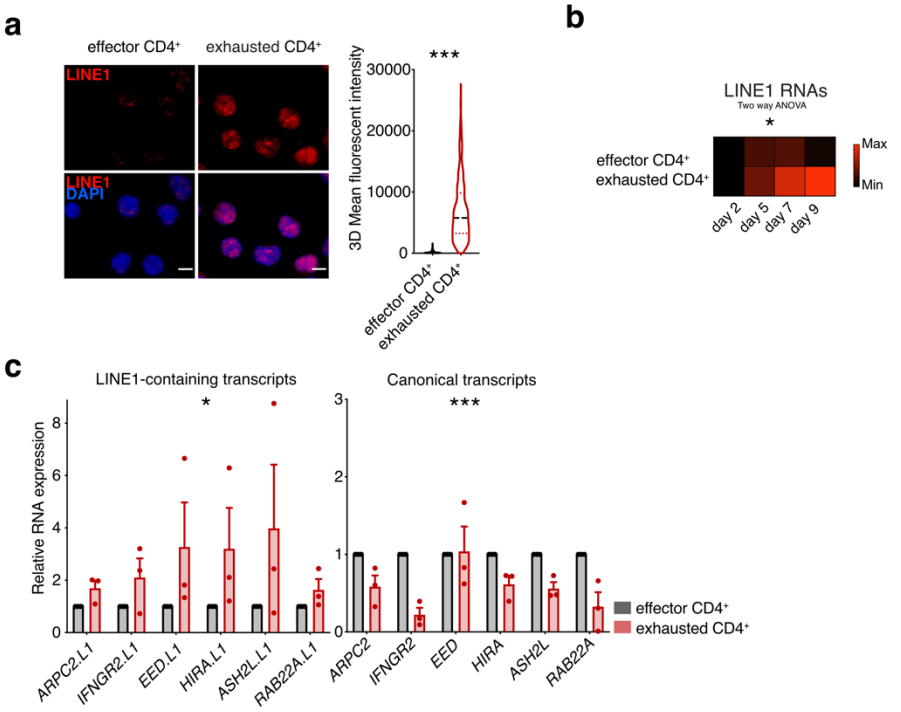


Figure 5. In vitro exhausted CD4+ T cells show a re-accumulation of LINE1-containing transcripts

a On the left confocal fluorescence microscopy images of LINE1 RNA-FISH (red) of effector and exhausted CD4+ T cells. Original magnification, 63×. Scale bar, 5 μm; on the right quantification of LINE1 RNA FISH was performed on at least 100 nuclei. Effector versus

exhausted CD4+ cells, ***$P < 1 \times 10^{-15}$, two-tailed Mann–Whitney test **b** Expression of LINE1 RNA by RT–qPCR in effector and exhausted CD4+ T cells ($n = 4$ individuals). Data represent mean as a heatmap. LINE1 RNA expression in effector versus exhausted CD4+ T cells, *$P = 0.0436$, two-way ANOVA. **c** Expression of LINE1-containing transcripts and canonical transcripts measured by RT–qPCR in effector and exhausted CD4+ T cells ($n = 3$ individuals). Data represent mean ± s.e.m. LINE1-containing transcript expression in effector versus exhausted CD4+ cells, *$P = 0.0148$, two-way ANOVA; canonical transcript expression in effector versus exhausted CD4+ cells, ***$P = 4.7 \times 10^{-4}$, two-way ANOVA

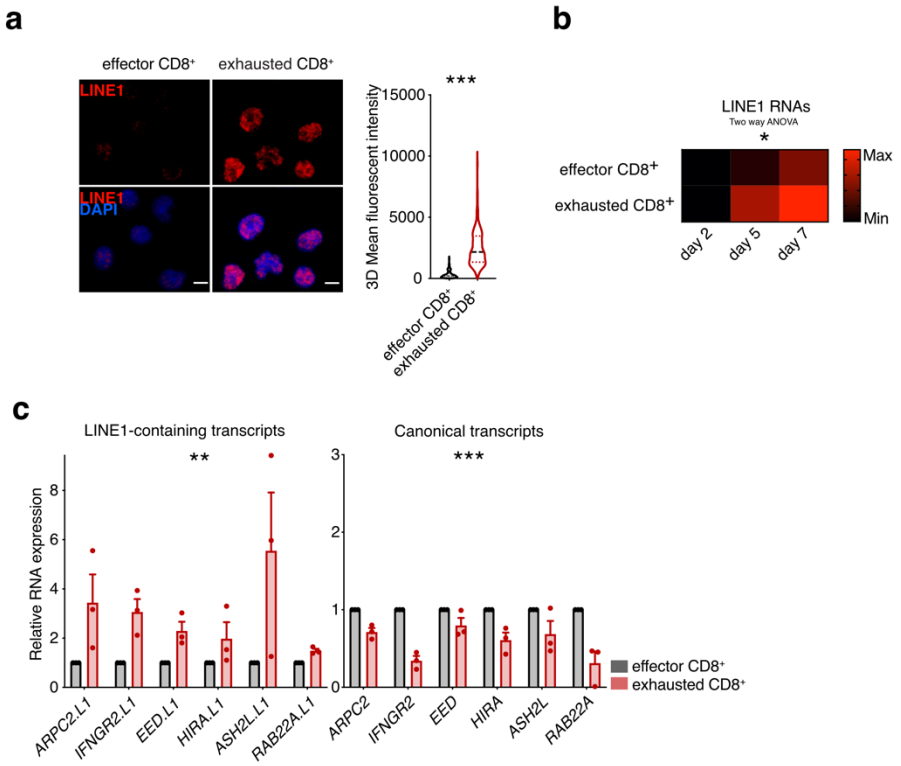


Figure 6. In vitro exhausted CD8⁺ T cells show a re-accumulation of LINE1-containing transcripts.

a Confocal fluorescence microscopy images of LINE1 RNA-FISH (red) of effector and exhausted CD8⁺ T cells. Original magnification, 63×. Scale bar, 5μm; quantification was performed on at least 100 nuclei. Effector versus exhausted CD8⁺ cells, ***$P<1\times10^{-15}$, two-tailed Mann–Whitney test. **b** Expression of LINE1 RNA by RT–qPCR in effector and exhausted CD8⁺ T cells ($n=4$ individuals). Data represent mean as a heatmap. LINE1 RNA expression in effector versus exhausted CD8⁺ T cells, *$P=0.0327$, two-way ANOVA. **c** Expression of LINE1-containing transcripts and canonical transcripts measured by RT–qPCR in effector and exhausted CD8⁺ T cells ($n=3$ individuals). Data represent mean ± s.e.m. LINE1-containing transcript expression

in effector versus exhausted CD8+ cells, $^{**}P = 0.0011$, two-way ANOVA; canonical transcript expression in effector versus exhausted CD8+ cells, $^{***}P = 9.3 \times 10^{-7}$, two-way ANOVA.

2.2.3 LINE1-containing transcripts downregulation increases T cell functions.

Considering the results obtained with the knockdown of LINE1 RNAs in activated naïve CD4+ T cells and the re-accumulation of LINE1-containing transcripts in exhausted CD4+ and CD8+ T cells we decided to deplete LINE1 RNAs in this in vitro exhausted T cells and to test immunological functions of these cells (Fig. 7a and Fig.8a). We treated T cells with ASOs concomitantly with every TCR stimulation starting from day 2.

At first, we verified, by RT-qPCR, the LINE1 ASOs treatment, as a control we used scr ASOs, finding the LINE1 depletion (around 50%) in exhausted CD4+ and CD8+ T cells treated with LINE1 ASOs (Fig. 7b and Fig.8b). Then, after 7 (for CD8) or 9 (for CD4) days from activation, we tested immunological functions by intracellular cytokines staining, killing ability and proliferation. Controlling cytokines secretion, we found that LINE1-containing transcripts depletion promotes an increase of IFN-γ and GrzB in exhausted CD4+ T cells and an increase of IFN-γ, GrzB and

PerfA in exhausted $CD8^+$ T cells (Fig. 7c and Fig.8c). Since we observed the increase of cytotoxic cytokines we decided to test the killing ability of these cells. We co-cultured exhausted $CD4^+$ and $CD8^+$ T cells treated with LINE1 and scr ASOs with heterologous monocytes derived from healthy donors for 12 hours and then we stained cells for live/dead detection to assess monocytes mortality. We decided to use monocytes since these cells express major histocompatibility complex of class I and to II (MHC class I and class II) and for this reason we could use these cells as killing targets for both $CD4^+$ and $CD8^+$ T cells that recognize MHC class II and I respectively. With these experiments, we verified that LINE1 depletion enhanced the killing ability of exhausted $CD4^+$ and $CD8^+$ T cells observing an increase in monocytes mortality (Fig. 7d and Fig. 8d). Lastly, we decided to follow the cell growth of exhausted $CD4^+$ and $CD8^+$ T cells treated with LINE1 and scr ASOs using cell trace staining but we didn't notice a restoration of proliferation capacity (Fig. 7 e and Fig.8e).

Our observations corroborate the anti-correlation between LINE1 RNAs accumulation in the nuclei and proper effector functions.

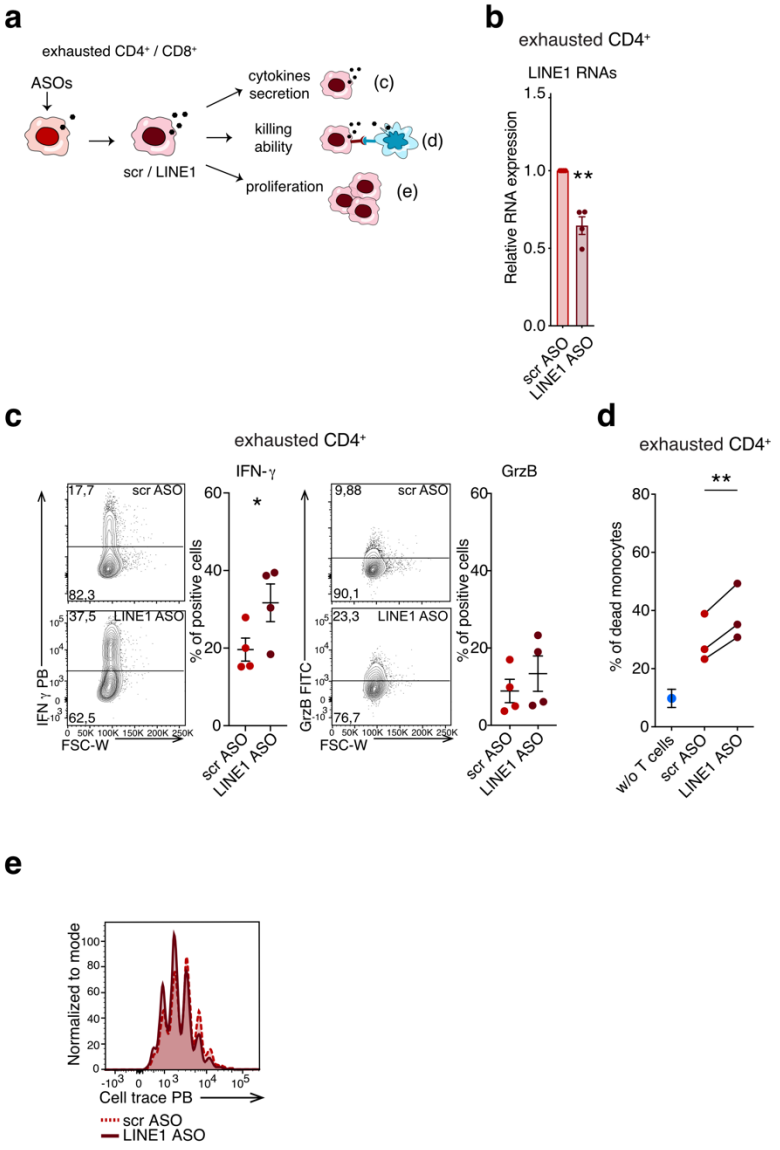


Figure 7. Downregulation of LINE1-containing transcripts in in vitro exhausted CD4+ T cells increase effector T cells functions.

a Scheme of the immunological assays performed on *in vitro* exhausted $CD4^+$ T cells treated with LINE1 or scr ASOs to evaluate their effector properties. **b** LINE1 expression by RT-qPCR in exhausted $CD4^+$ T cells (n = 4 individuals). Data are represented as mean ± s.e.m. scr versus LINE1 ASO exhausted $CD4^+$: ** P = 0.0083, Two-tailed paired *t-test* **c** IFN-γ^+ or $GrzB^+$ exhausted $CD4^+$ T cells (n = 4 individuals). Data are represented as mean ± s.e.m. IFN-γ in scr versus LINE1 ASO groups: * P = 0.0351 One-tailed paired *t-test*. **d** Percentage of dead heterologous monocytes co-cultured for 12 hours with exhausted $CD4^+$ T cells. Percentage of monocytes self-lysis is indicated (w/o T cells, in blue). n = 3 individuals for $CD4^+$, data for each individual are shown separately, monocytes self-lysis is represented as mean ± s.e.m. $CD4^+$ scr versus LINE1 ASO: ** P = 0.0092, Two-tailed paired *t-test*. **e** Proliferation assay with cell trace in exhausted $CD4^+$ T cells.

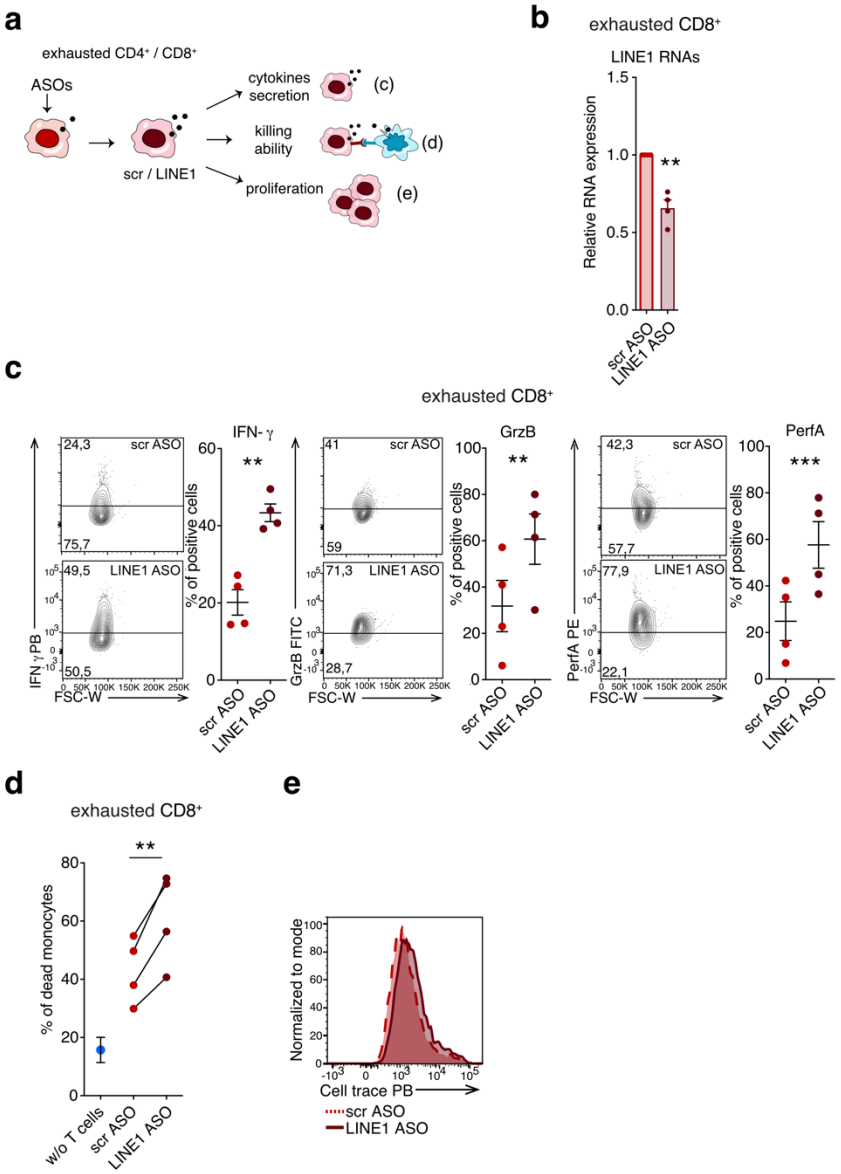


Figure 8. Downregulation of LINE1-containing transcripts in in vitro exhausted CD8⁺ T cells increase effector T cells functions.

a Scheme of the immunological assays performed on *in vitro* exhausted CD8⁺ T cells treated with LINE1 or scr ASOs to

evaluate their effector properties. **b** LINE1 expression by RT-qPCR in exhausted $CD8^+$ T cells (n = 4 individuals). Data are represented as mean ± s.e.m. scr versus LINE1 ASO exhausted $CD8^+$: ** P = 0.0073, Two-tailed paired *t-test*. **c** IFN-γ^+, $GrzB^+$ or $PerfA^+$ exhausted $CD8^+$ T cells (n = 4 individuals). Data are represented as mean ± s.e.m. IFN-γ in scr versus LINE1 ASO groups: ** P = 0.0016, Two-tailed paired *t-test*; GrzB in scr versus LINE1 ASO groups: ** P = 0.0039, PerfA in scr versus LINE1 ASO: *** P = 0.0002, Two-tailed paired *t-test*. **d** Percentage of dead heterologous monocytes co-cultured for 12 hours with exhausted $CD8^+$ T cells. Percentage of monocytes self-lysis is indicated (w/o T cells, in blue). n = 4 individuals for $CD8^+$, data for each individual are shown separately, monocytes self-lysis is represented as mean ± s.e.m. $CD8^+$ scr versus LINE1 ASO: ** P = 0.0084, Two-tailed paired *t-test*. **e** Proliferation assay with cell trace in exhausted $CD8^+$ T cells.

2.2.4 LINE-containing transcripts accumulate in tumor-infiltrating lymphocytes (TILs) and regulate TILs effector functions

Given the results obtained in exhausted T cells, we decided to observe LINE1 dynamic in ex vivo tumor infiltrating lymphocytes (TILs) to assess the importance of LINE1 elements in the establishment of the immunosuppressive tumor

microenvironment. Indeed, effector T cells infiltrating tumors became dysfunctional through poorly defined mechanisms (Franco et al., 2020; Jamal-Hanjani et al., 2013). For this purpose, we isolated memory TILs from normal adjacent and tumor tissues derived from lung and colorectal (CRC) cancer patients (Fig. 9a). By LINE1 RNA-FISH we found that memory $CD4^+$ and $CD8^+$ TILs derived from tumor tissues re-express LINE1 RNAs in respect to memory T cells derived from normal adjacent tissues (Fig. 9 b-c). We retrieve this peculiar result from all patients analyzed. Moreover, by PD-1 receptor staining, we verified that TILs derived from tumor tissues display a higher level of exhaustion surface marker PD-1 (Fig. 9 d-e). Considering these results and the ones obtained in exhausted T cells we decided to knock down LINE1 RNAs in memory TILs to control if we restore TIL effector functions using ASOs. Therefore, we knockdown LINE1-containing transcripts for 48 hours in memory $CD4^+$ and $CD8^+$ TILs, using scr ASOs as control, and we measure the expression of inhibitory receptors, the cytokines production and the killing capacity (Fig. 10a). Firstly, we controlled LINE1 reduction by RT-qPCR (Fig. 10b-c), then by receptor staining, we verified PD-1, TIM-3 and LAG-3 expressions, founding that all the three exhaustion surface markers were reduced upon 48 hours of silencing of LINE1-containing transcripts in memory $CD4^+$ and $CD8^+$ TILs (Fig. 10d and Fig. 11a). Analyzing the production of cytokines, we performed intracellular cytokine secretion of IFN-γ and GrzB in

memory $CD4^+$ TILs and of IFN-γ, GrzB and PerfA in memory $CD8^+$ TILs treated for 48 hours with LINE1 or scr ASOs and then activated for 48 hours. We observed an increase of effector cytokines production concomitantly with LINE1 depletion in memory $CD4^+$ and $CD8^+$ TILs (Fig. 10e and Fig. 11b). Lastly, after 48 hours of LINE1 or scr ASOs treatment, we co-cultured memory $CD4^+$ and $CD8^+$ TILs with heterologous monocytes derived from a healthy donor, assessing an increase of killing ability in TILs with LINE1 depletion (Fig. 10f and Fig.11c).

Our findings reveal that LINE1 RNAs can fine tune the effector response of TILs, and with LINE1 RNAs depletion we were able to restore dysfunctional TILs phenotype.

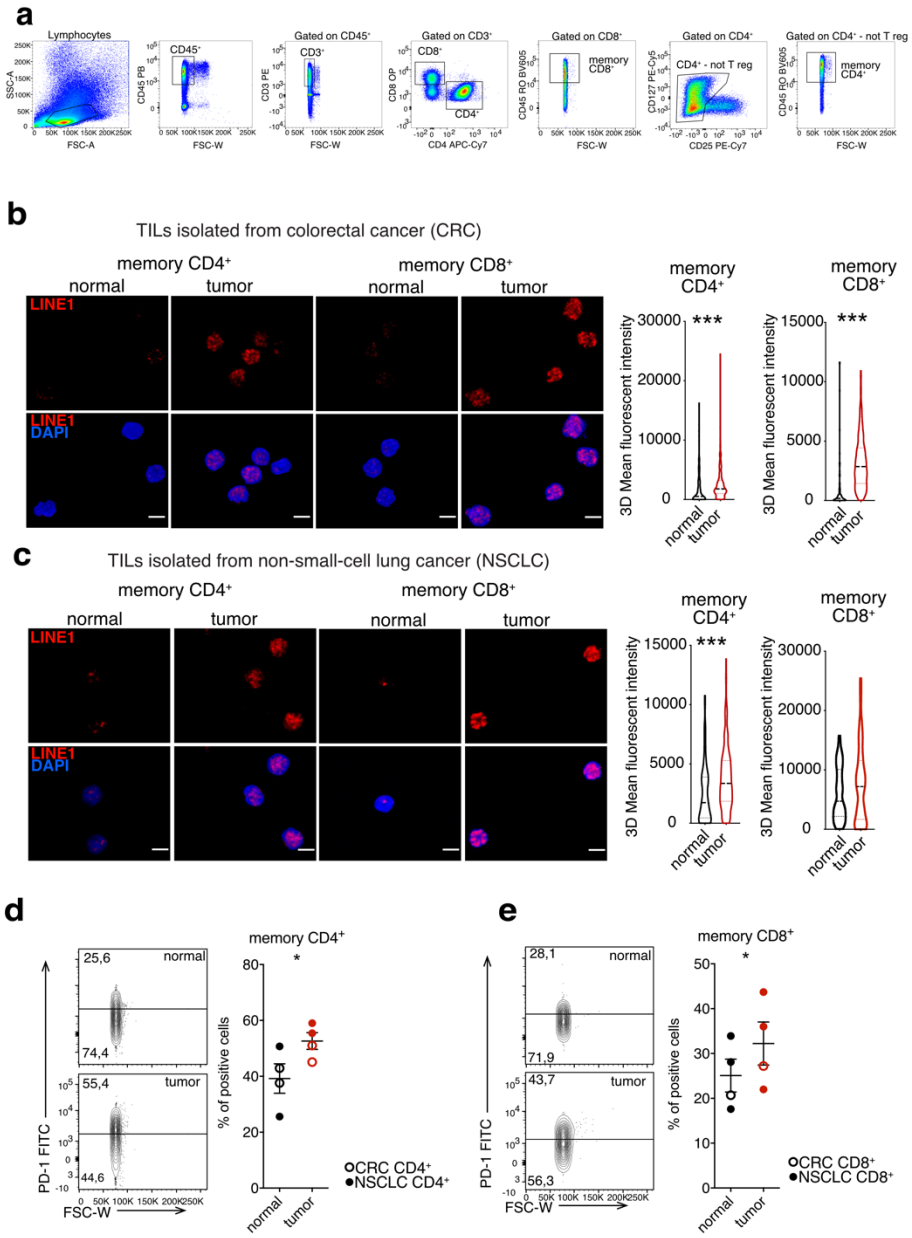


Figure 9. Memory CD4⁺ and CD8⁺ TILs re-accumulate LINE1-containing transcripts

a Sorting of tumor-infiltrating memory CD4⁺ and CD8⁺ T cells. **b** Top, confocal fluorescence microscopy images of LINE1 RNA-FISH (red) of

memory $CD4^+$ and $CD8^+$ T cells infiltrating normal adjacent tissue or CRC tumors. Original magnification, 63×. Scale bar, 5μm. Bottom, quantification of at least 100 nuclei from two patients. Normal versus tumor memory $CD4^+$ cells, ***$P = 7.3 \times 10^{-8}$, two-tailed Mann–Whitney test; normal versus tumor memory $CD8^+$ cells, ***exact $P < 1 \times 10^{-15}$, two-tailed Mann–Whitney test. **c** Top, confocal fluorescence microscopy images of LINE1 RNA-FISH (red) of memory $CD4^+$ and $CD8^+$ T cells infiltrating normal adjacent tissue or NSCLC tumors. Original magnification, 63×. Scale bar, 5μm. Bottom, quantification of at least 84 nuclei from at least two patients. Normal versus tumor memory $CD4^+$ cells, ***exact $P = 3.7 \times 10^{-4}$, two-tailed Mann–Whitney test. **d** On the left contour plot representation of PD-1 receptor staining, on the right percentage of PD-1 positive cells (n=4 individuals) of memory $CD4^+$ T cells infiltrating normal adjacent tissue or CRC and NSCL tumors. Data are mean and ± s.e.m, * P = 0.05, one-tailed t-test **e** On the left contour plot representation of PD-1 receptor staining, on the right percentage of PD-1 positive cells (n=4 individuals) of memory $CD8^+$ T cells infiltrating normal adjacent tissue or CRC and NSCL tumors. Data are mean and ± s.e.m, * P = 0.05, one-tailed t-test

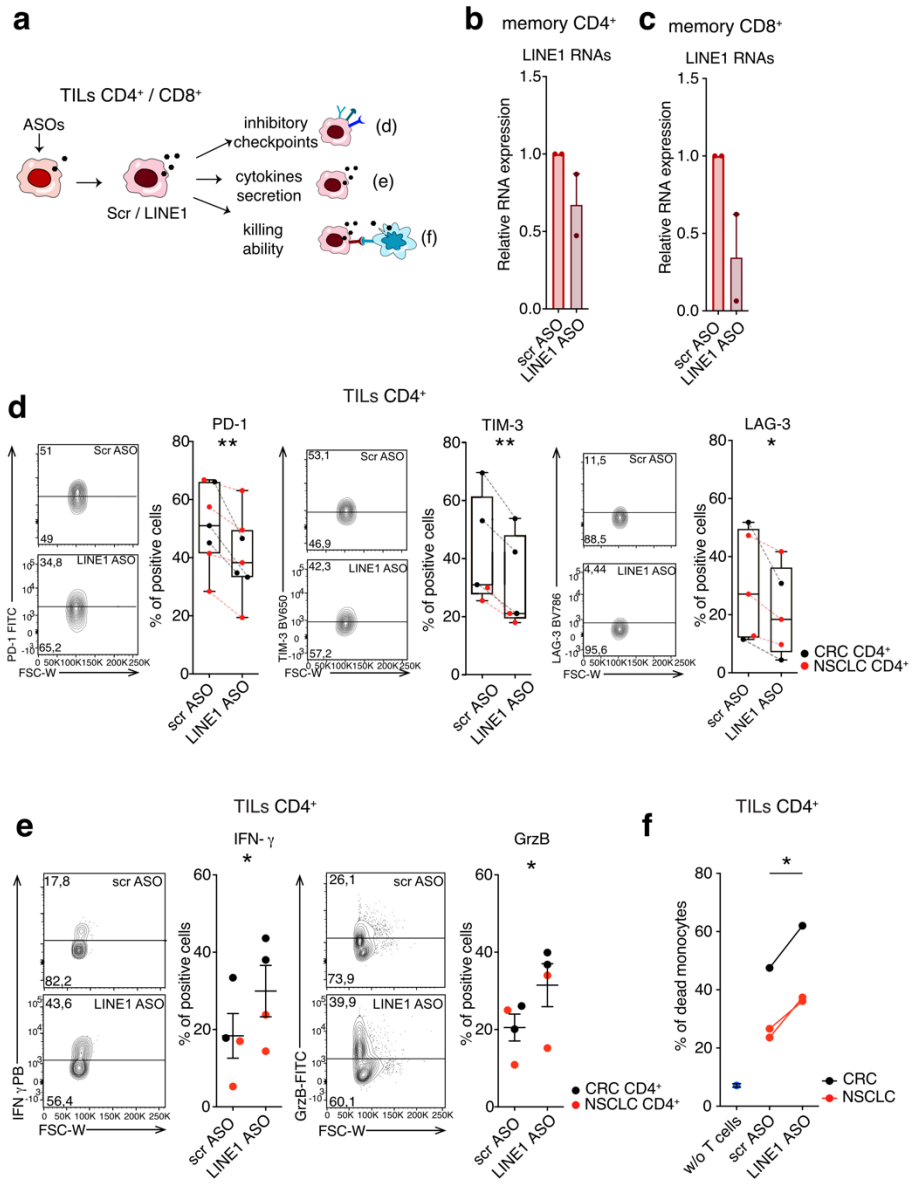


Figure 10. LINE1-containing transcripts modulate dysfunctional phenotype of CD4+ TILs

a Scheme of immunological assays performed on memory CD4+ and CD8+ TILs treated with LINE1 or control (scr) ASOs. **b-c** LINE1

expression by RT-qPCR in memory $CD4^+$ (**b**) and $CD8^+$ (**c**) TILs treated with LINE1 or control (scr) ASOs (n = 2 individuals). Data are represented as mean ± s.e.m. **d**, PD-1-, TIM3- or LAG-3-positive memory $CD4^+$ TILs isolated from CRC (black, $n=3$ individuals for PD-1 and TIM3; $n=2$ for LAG-3) or NSCLC (red, $n=4$ individuals for PD-1; $n=2$ for TIM3; $n=3$ for LAG-3) samples. PD-1^+ cells in scr versus LINE1 ASO groups, $^{**}P=0.0044$, two-tailed paired *t*-test; $TIM3^+$ cells in scr versus LINE1 ASO groups, $^{**}P=0.0016$, two-tailed paired *t*-test; LAG-3^+ cells in scr versus LINE1 ASO groups, $^{*}P=0.0442$, two-tailed paired *t*-test. For box-and-whisker plots, the central line, the box and whiskers represent the median, the IQR from first to third quartiles and 1.5 × IQR, respectively **e**, IFN-γ- or granzyme (Grz)B-positive memory $CD4^+$ TILs isolated from CRC (black, $n=2$ individuals) or NSCLC (red, $n=2$ individuals) samples. Data represent mean ± s.e.m. IFN-γ^+ cells in scr versus LINE1 ASO groups, $^{*}P=0.04795$, one-tailed paired *t*-test; $GrzB^+$ cells in scr versus LINE1 ASO groups, $^{*}P=0.0277$, two-tailed paired *t*-test. **f**, Percentage of dead heterologous monocytes co-cultured for 12 h with memory $CD4^+$ TILs from CRC (black, $n=1$ individual) or NSCLC (red, $n=2$ individuals) samples. The percentage of monocyte self-lysis is indicated (without T cells, in blue). $CD4^+$ cells, scr versus LINE1 ASO groups, $^{*}P=0.0150$, two-tailed paired *t*-test.

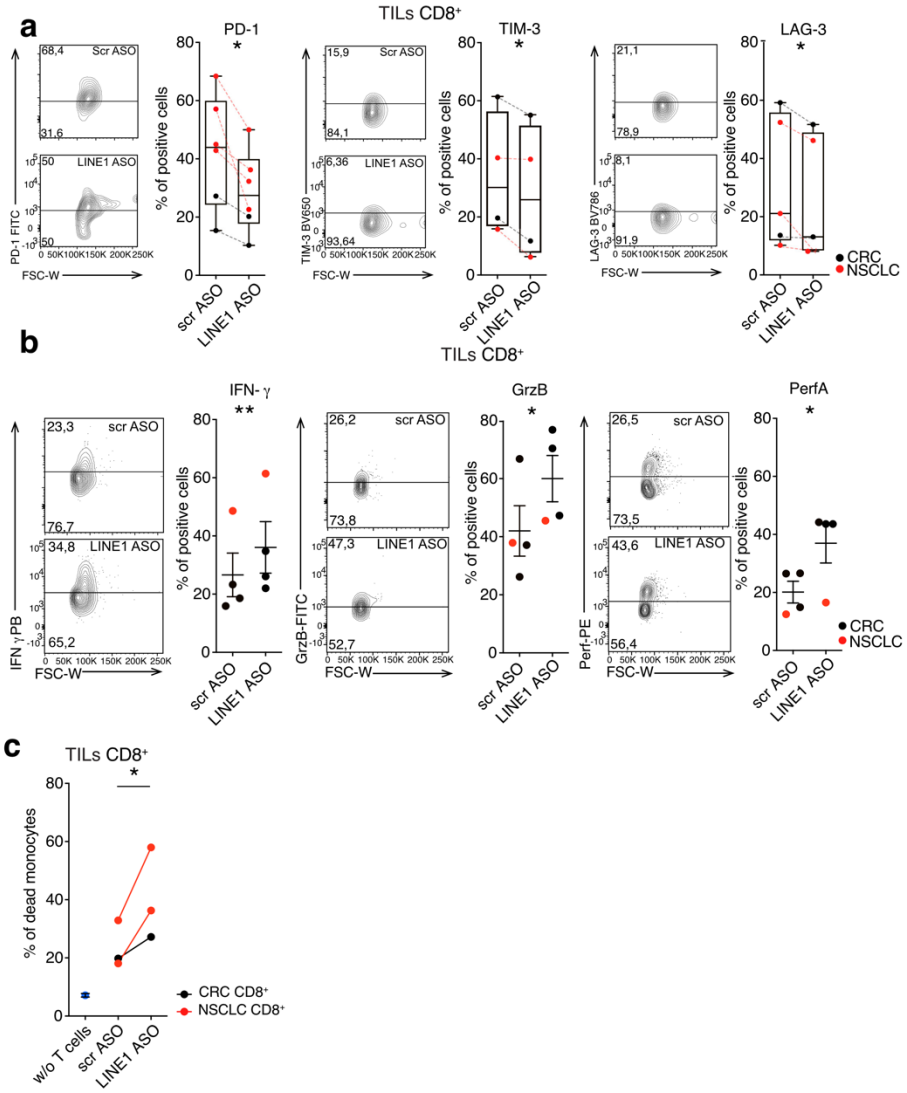


Figure 11. LINE1-containing transcripts modulate dysfunctional phenotype of CD8+ TILs

a PD-1-, TIM3- or LAG-3-positive memory CD8+ TILs isolated from CRC (black, $n=2$ individuals) or NSCLC (red, $n=4$ individuals for PD-1; $n=2$ for TIM3 and LAG-3) samples. PD-1+ cells in scr versus LINE1 ASO groups, $^{*}P=0.02676$, two-tailed paired *t*-test; TIM3+ cells in scr

versus LINE1 ASO groups, *P=0.02707, one-tailed paired *t*-test; LAG-3$^+$ cells in scr versus LINE1 ASO groups, *P=0.03166, one-tailed paired *t*-test. **b**, IFN-γ-, GrzB- or perforin (Perf)A-positive memory CD8$^+$ TILs isolated from CRC (black, n=3 individuals) or NSCLC (red, n=1 individual) samples. Data represent mean±s.e.m. IFN-γ$^+$ cells in scr versus LINE1 ASO groups, **P=0.0095, two-tailed paired *t*-test; GrzB$^+$ cells in scr versus LINE1 ASO groups, *P=0.0275, one-tailed paired *t*-test; PerfA$^+$ cells in scr versus LINE1 ASO groups, *P=0.02225, one-tailed paired *t*-test. **c** Percentage of dead heterologous monocytes co-cultured for 12h with memory CD8$^+$ TILs from CRC (black, n=1 individual) or NSCLC (red, n=2 individuals) samples. The percentage of monocyte self-lysis is indicated (without T cells, in blue). CD8$^+$ cells, scr versus LINE1 ASO groups, *P=0.0408, two-tailed paired *t*-test.

2.2.5 scRNA seq and PrimeFlow RNA Assay try to depict TEs' signature in TILs.

All these data demonstrated that LINE1 transcripts could represent a novel hallmark to define the phenotype of exhausted T cells. To better define how LINE1 transcripts are distributed among dysfunctional T cells we want to perform a single-cell RNA seq with a full-length technology (Gao et al., 2017). This specific approach could allow us to capture all TEs RNA isoform that we would otherwise lose using 5' or 3' scRNA-seq approach. By this technology, we aim to combine the presence or absence

of LINE1 RNAs with the dysfunctional state of TILs trying to evaluate if LINE1 transcripts are present also in the absence of inhibitory receptors to define T cell dysfunction.

Moreover, we are applying a scRNA-seq approach to determine how TEs are distributed among TILs subsets to possibly identify a TEs' transcriptional signature that could discriminate specific TILs subsets. Indeed, we are performing scRNA-seq on $CD3^+$ TILs derived from normal adjacent and tumor tissues of lung and CRC cancer patients (Fig. 12a). With this approach, we managed to discriminate TILs subset not only by gene-expression but also by TEs expression (Fig. 12b). We are planning to further investigate these preliminary data and to validate with PrimeFlow RNA Assay specific TEs expression in specific TILs subsets (Fig. 12a). Indeed, PrimeFlow RNA Assay is a novel technology that allows, by flow cytometry, to evaluate receptor markers and RNA transcripts. This technique involves the staining with receptor markers to distinguish all the different T cell subset and the hybridization with a target probe, that identify the RNA of interest, and several passages to amplify the target probe signal (Fig. 12a and Fig. 14c).

We have already set up this technology analyzing LINE1 RNA content in PBMCs derived from healthy donors, in exhausted T cells and TILs (Fig. 13-14). Firstly, we analyzed PBMCs derived from healthy donors assessing that with receptor staining we discriminated all T cell subsets and with LINE1 RNA staining we evaluated LINE1 RNA content among these cells (Fig. 13 a,b).

Indeed, we found that naïve CD4+ T cells are enriched of LINE1 RNA in respect to memory CD4+ and naïve and memory CD8+ T cells (Fig. 13a,b), as we already demonstrated by other techniques like RNA FISH and RT-qPCR (Fig. 1a,b,g).

Moreover, by PrimeFlow RNA Assay we validate the re-accumulation of LINE1 transcripts in exhausted CD4+ T cells in respect to effector CD4+ T cells (Fig. 13c). With this protocol, we also could appreciate the broadness of LINE1 transcripts in exhausted T cells (Fig. 13d), since we found that exhausted CD4+ T cells show LINE1 RNA positive cells that are not positive for exhausted markers like PD1 and TIM3.

Lastly, we analyzed with PrimeFlow RNA Assay also TILs isolated from normal adjacent and tumor tissue derived from NSCL cancer patient. To perform this experiment, we used T cells isolated from peripheral blood of healthy donors as positive control of the technique. We verified that memory CD4+ and CD8+ T cells derived from tumor tissue have a higher expression of LINE1 RNAs in respect to T cells isolated from normal adjacent tissue (Fig. 14a), as we already demonstrated by LINE1 RNA-FISH (Fig. 9 b,c). Moreover, we tested also memory CD4+ and CD8+ T cells isolated from peripheral blood, normal adjacent and tumor tissue of CRC treated patients. Surprisingly we detected a lower expression of LINE1 RNA in TILs in respect to memory T cells derived from normal adjacent tissue of the same patient (Fig. 14b). This result was encouraging but, at the moment, we have the chance to test only one treated patient but we hopefully

want to corroborate this data by analyzing other treated and responder patients.

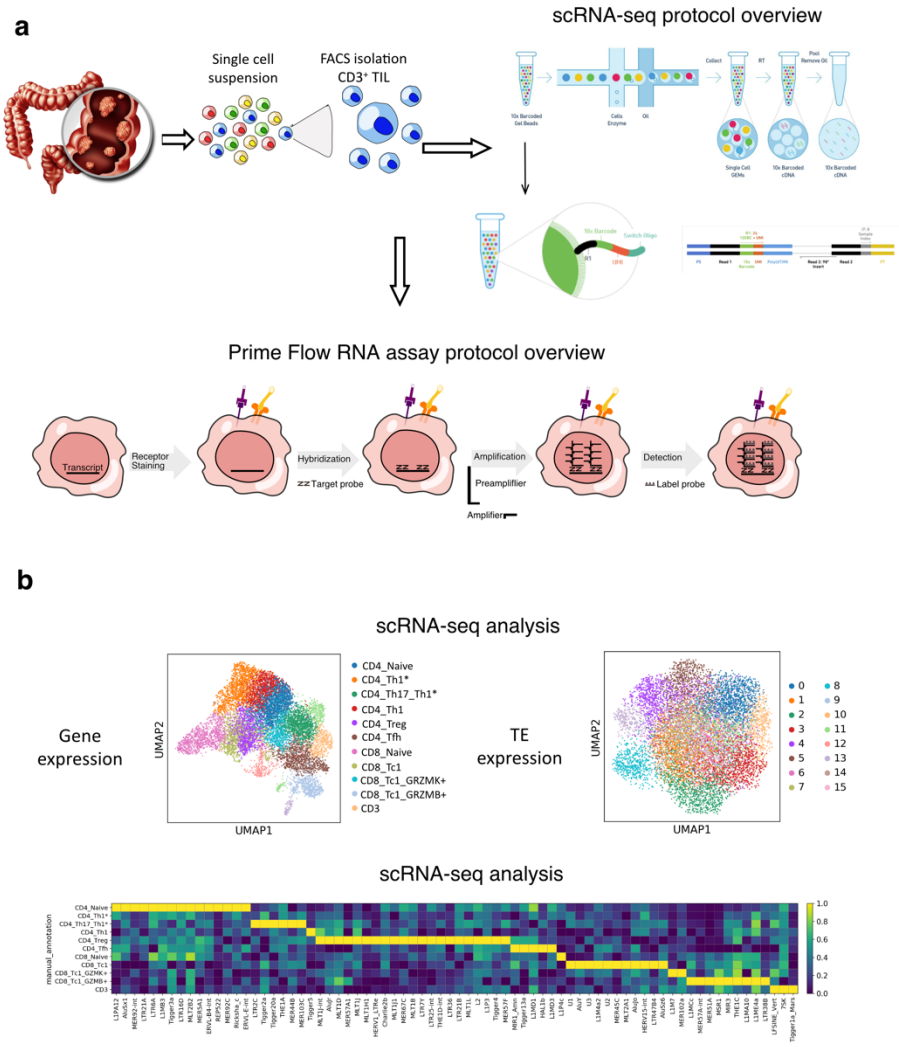


Figure 12. scRNA-seq and Prime Flow RNA assay try to depict signature of TEs transcripts in TILs.

a Schematic representation of scRNA-seq and PrimeFlow RNA assay in TILs. **b** On the left (top) UMAP representation of TILs subsets discriminated by gene expression and (bottom) UMAP representation

of TILs cluster based on TEs expression; on the right heatmap of TEs expression in TILs subset

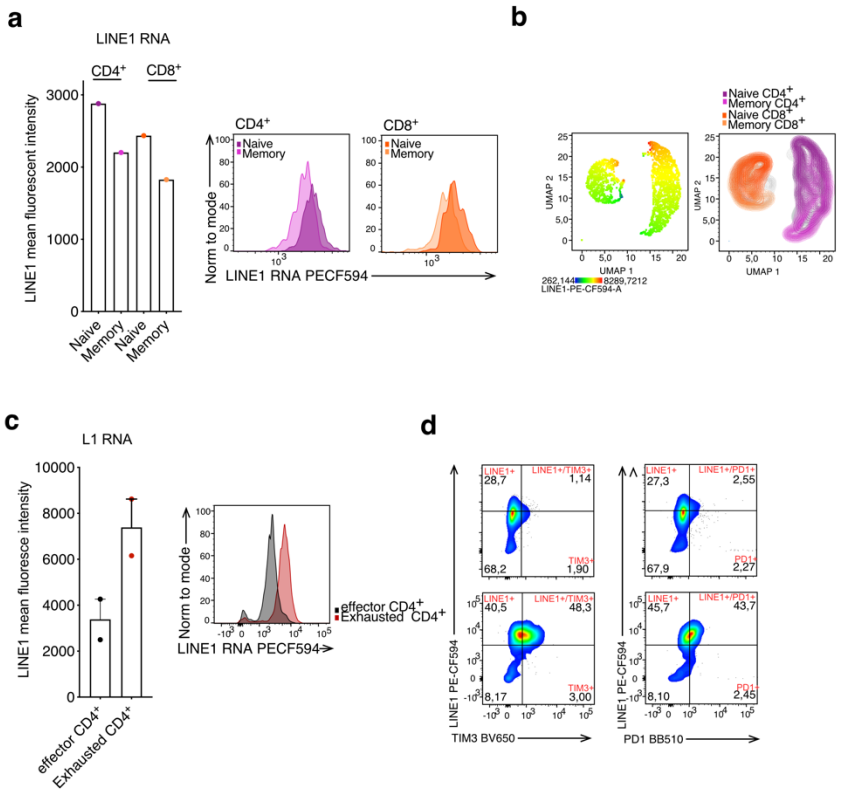
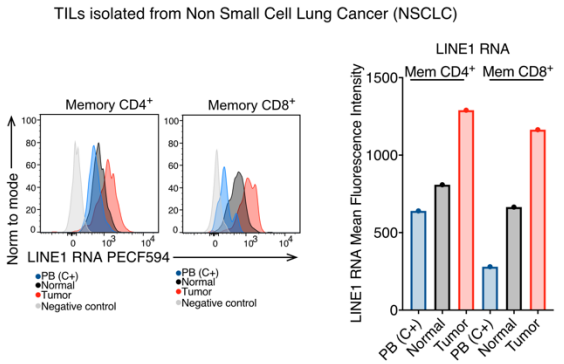


Figure 13. Prime Flow RNA assay depicts LINE1 RNAs content in PBMCs and in vitro exhausted T cells.

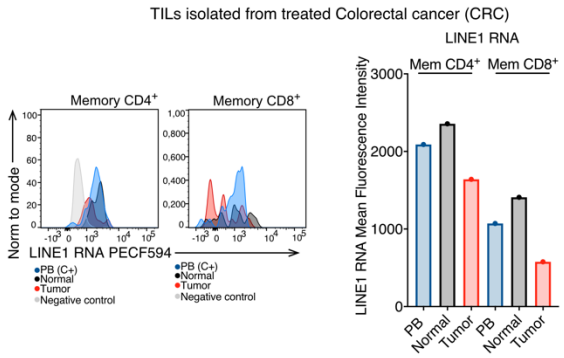
a On the left, bar plot of LINE1 RNA mean fluorescent intensity in naïve and memory CD4+ and CD8+ T cells derived from healthy donor; on the right graphic histogram of LINE1 RNA mean fluorescent intensity in naïve and memory CD4+ and CD8+ T cells. **b** UMAP represents T cells subset distribution based on CD4, CD8 and LINE1 RNA marker, on the left LINE1 RNA heat map representation, on the right identification of different T cell subsets. **c** On the left, bar plot of LINE1 RNA mean

fluorescent intensity in effector (black) and exhausted (red) $CD4^+$ T cells; on the right graphic histogram of LINE1 RNA mean fluorescent intensity in effector (black) and exhausted (red) $CD4^+$ T cells (n=1). **d** Contour plot representation of LINE1 RNA and TIM3 (left) or PD1 (right) positive cells in effector and exhausted $CD4^+$ T cells (n=2).

a



b



c

Table of TILs subsets' antibody

Receptor	Fluorophore
CD45	BUV805-BD
CD3	BUV395-BD
CD4	BUV496-BD
CD8	BUV563-BD
CD25	APC-R700-BD
CD127	PE-Cy5-BD
CD45RO	BV695-BD
CCR7	BV711-BD
CCR6	APC-BD
Foxp3	PE-BD
IL-10	BB700-BD
IL-23R	BV650-BD
CTLA-4	BV421-BD
CX3CR1	BV750-BD
LAYN	FITC-Homemade
PD-1	PE-Cy7-BD
TIM-3	BV480-BD
LAG-3	BV786-BD
CD69	BUV737-BD
LINE1	PE-CF594-Invitrogen

Table of selection strategy of TILs subsets

Subset	Selection strategy
CD4 Treg cluster	CD45+, CD3+, CD4+,CD25+,CD127low, Foxp3+,IL-10+,CTLA4+
CD4 Th17	CD45+, CD3+, CD4+,CD25+/-,CD127+/- (not CD4+Treg), CD45RO+, CCR6+, IL-23R+
CD4 Recently activate effector memory (Temra)	CD45+, CD3+, CD4+,CD25+/-,CD127+/- (not CD4+Treg), CD45RO-, CCR7-, CX3CR1+
CD8 Temra	CD45+, CD3+, CD8+,CD45RO-, CCR7-, CX3CR1+
CD4 Dysfunctional	CD45+, CD3+, CD4+,CD25+/-,CD127+/- (not CD4+Treg), CD45RO+, LAYN+, PD1+, TIM-3+,LAG-3+, CTLA4+
CD8 Dysfunctional	CD45+, CD3+, CD8+, CD45RO+, LAYN+, PD1+, TIM-3+,LAG-3+, CTLA4+
CD4 Central memory	CD45+, CD3+, CD4+,CD25+/-,CD127+/- (not CD4+Treg), CD45RO+, CCR7+
CD8 Tissue resident memory	CD45+, CD3+, CD8+, CD45RO-, CCR7+,CD69+,CD103+

Figure 14. Prime Flow RNA assay depicts LINE1 RNAs content ex vivo TILs cells.

a On the left graphic histogram of LINE1 RNA mean fluorescent intensity, on the left bar plot of LINE1 RNA mean fluorescent intensity

in memory $CD4^+$ and $CD8^+$ T cells derived from peripheral blood of healthy donors and normal adjacent and tumor tissue of NSCL cancer patient (n=1). **b** On the left graphic histogram of LINE1 RNA mean fluorescent intensity, on the left bar plot of LINE1 RNA mean fluorescent intensity in memory $CD4^+$ and $CD8^+$ T cells derived from peripheral blood, normal adjacent and tumor tissue of treated CRC patient (n=1). **c** Table of strategy to select by receptor marker TILs by PromeFlow RNA Assay

2.3 Methods

Human blood and tissue samples

Blood from anonymous healthy donors was provided by the Fondazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Cà Granda Ospedale Maggiore Policlinico in Milan. Age and the sex of healthy donors were unknown (privacy reasons). CRC and NSCLC samples were provided by the European Institute of Oncology (IEO), non-tumoral samples were obtained from normal adjacent tissue at least 10 cm distal from the lesion; no patients received palliative surgery or neo-adjuvant chemotherapy and/or radiotherapy. Ethics committees of the hospitals approved the use of human samples for research purposes, and informed consent was obtained from all subjects (authorization nos. R807/18 IEO 849 and 708_2020).

T cell purification and sorting and monocyte purification

Human peripheral blood mononuclear cells (PBMCs) were purified from blood samples by density gradient centrifugation with Ficoll-Paque PLUS. T cells were negatively selected from PBMCs with a magnetic separator (autoMACS Pro Separator, Miltenyi Biotec) using the Pan T Cell Isolation kit (Miltenyi Biotec) or the $CD4^+$ T Cell Isolation kit (Miltenyi Biotec). T cells were stained with antibodies specific for surface markers, and T cell subsets were sorted by flow cytometry as follows: naive $CD4^+$ cells as $CD4^+CD25^-CD127^{-/hi}CD45RO^-$ cells, $CD4^+$ T_H1 cells as $CD4^+CD25^-CD127^{-/hi}CD45RO^+CXCR3^+CCR6^-$ cells, $CD4^+$ T_H2 cells as $CD4^+CD25^-CD127^{-/hi}CD45RO^+CRTH2^+$ cells and $CD4^+$ T_H17 cells as $CD4^+CD25^-CD127^{-/hi}CD45RO^+CCR6^+CXCR3^-$ cells; naive $CD8^+$ cells as $CD4^-CD8^+CD45RO^-$ cells and memory $CD8^+$ cells as $CD4^-CD8^+CD45RO^+$ cells. Lymphocytes infiltrating normal and tumor tissues were obtained using a procedure indicated in the Supplementary Information. Viability of T cells was first assessed with the LIVE/DEAD Fixable Green Dead Cell Stain kit (Invitrogen by Life Technologies, L34969); live T cell subsets were then sorted by flow cytometry as follows: memory $CD4^+$ cells as $CD45^+CD3^+CD4^+CD25^-CD127^{-/hi}CD45RO^+$ cells and memory $CD8^+$ cells as $CD45^+CD3^+CD8^+CD45RO^+$ cells. The following antibodies were used for flow cytometry-based sorting: anti-CD4–APC-Cy7 (BD Biosciences, clone RPA-T4) or anti-CD4–VioGreen (Miltenyi Biotec, clone VIT4); anti-CD8–

VioGreen (Miltenyi Biotec, clone REA734) or anti-CD8–VioBlue (Miltenyi Biotec, clone REA734); anti-CD25–PECy7 (Invitrogen by Life Technologies, clone BC96); anti-CD127–PE-Cy5 (BioLegend, clone A019D5) or anti-CD127–PE (Miltenyi Biotec, clone MB15-18C9); anti-CD45RO–BV605 (BioLegend, clone UCHL1) or anti-CD45RO–APC (Miltenyi Biotec, clone UCHL1); anti-CD3–PE (BD Biosciences, clone UCHT1); anti-CD45–Pacific Blue (BioLegend, clone 2D1); anti-CD183–PE-Cy5 (BD Biosciences, clone 1C6/CXCR3); anti-CD294(CRTH2)–APC-Vio770 (Miltenyi Biotec, clone REA598); anti-CCR6–FITC (BioLegend, clone G034E3). Staining were performed using 1 μl of each antibody for every 1×10^6 cells in 10 μl PBS for 30 min at 37 °C. Cell sorting was performed using the FACSAria III (BD Biosciences) using BD FACSDiva Software version 8.0.3. The purity of sorted cells was >97.5%. Monocytes were isolated from PBMCs by positive selection with a magnetic separator (autoMACS Pro Separator, Miltenyi Biotec) using CD14-specific microbeads (Miltenyi Biotec).

Knockdown experiments

Knockdown experiments were performed using FANA (2′-deoxy-2′-fluoro-β-D-arabinonucleic
acid, https://www.aumbiotech.com)–ASOs. For LINE1 RNA species, five ASOs were designed based on the ORF2 region of the LINE1 consensus sequence. An unrelated scr ASO was used as a control. ASOs were mixed in equimolar proportions and

administered without any transfection reagent (by gymnosis) following the manufacturer's instruction at a final concentration of 10 μM.

Quiescent naive CD4$^+$ T cells isolated from healthy donors were cultured for 48 h in complete medium supplemented with 200 IU ml^{-1} recombinant interleukin (IL)-2 and 10 μM ASOs to knock down LINE1. For LINE1-knockdown experiments, naive CD4$^+$ cells treated with ASOs were then activated with anti-CD3–anti-CD28 beads in T_H1 medium, kept in culture in the presence of 10 μM ASOs and collected after 7 d (effector CD4$^+$ cells). Exhausted CD4$^+$ and CD8$^+$ T cells were treated starting from day 2 with 10 μM ASOs for 5–7 d to knock down LINE1. Memory CD4$^+$ and CD8$^+$ TILs isolated from tumor samples were cultured for 48 h in complete medium supplemented with 200 IU ml^{-1} recombinant IL-2 and 10 μM ASOs to knock down LINE1; after 48 h of ASO treatment, TILs were subjected to surface marker staining and T cell killing or activated for 48 h and subjected to intracellular cytokine staining. Knockdown efficiency was controlled by RT–qPCR and/or RNA-FISH and by flow cytometry analysis (described below).

CD4$^+$ T-cells in vitro differentiation

Quiescent naïve CD4$^+$ T-cells have been plated at 1,5 x 10^6/mL, stimulated with Dynabeads Human T-activator anti-CD3/anti-CD28 beads (Gibco; cat. num. 1131D) and cultured for hours or

days with the appropriate T helper medium of differentiation. T helper medium of differentiation consists in complete medium composed by RPMI 1640 with GlutaMAX-I (Gibco) supplemented with 10% (v/v) Fetal Bovine Serum (FBS) (Gibco), 1% (v/v) non-essential amino acids, 1mM sodium pyruvate, 50 U/mL penicillin, 50 μg/mL streptomycin, plus T helper specific cytokines. Th1 cytokines: 20 IU/mL recombinant IL-2 (cat. num. 130-097-744), 10 ng/mL recombinant IL-12 (cat. num. 130-0976-704), 2 μg/mL neutralizing anti-IL-4 (cat. num. 130-095-753). Naïve $CD8^+$ T cells have been plated at 1,5 x 10^6/mL, stimulated with Dynabeads Human T-activator anti-CD3/anti-CD28 beads and were cultured in T helper medium plus 20 IU/mL recombinant IL-2 (cat. num. 130-097-744). Cells were maintained at 37°C in a 5% CO2 humidified incubator, were counted and split every 2-3 days.

In vitro exhausted $CD4^+$ and $CD8^+$ T-cells

Naïve $CD4^+$ T-cells were activated and differentiate to Th1 phenotype, every 2 days T-cells were counted and exposed to stimulatory anti-CD3/anti-CD28 beads. Naïve $CD8^+$ T-cells were activated and differentiate in T helper medium plus IL-2, every 2 days T-cells were counted and exposed to stimulatory anti-CD3/anti-CD28 beads. Exhausted T-cells were tested for proliferation reduction, for PD-1 marker increase and for T-cell effector properties assessed with intracellular staining for lineage specific cytokines and killing ability.

T cell surface and intracellular staining, proliferation assay and cell cycle analysis.

Surface marker staining was performed by incubating 1 μl antibody for every 5×10^4 cells in PBS at 37 °C for 30 min. T cells were washed with PBS and then analyzed. The following antibodies were used: Alexa Fluor 488 anti-CD279 (PD-1) (BioLegend, clone EH12.2H7), anti-CD366 (TIM3-1)–BV650 (BioLegend, clone F38-2E2), anti-CD223 (LAG-3)–BV785 (BioLegend, clone 11C3C65). For intracellular cytokine staining, 5×10^4 T cells were stimulated with 50 ng ml^{-1} phorbol 12-myristate 13-acetate and with 0.5 μg ml^{-1} ionomycin for 2 h at 37 °C; subsequently, 100 μg ml^{-1} brefeldin A (Merck) was added for an additional 2 h at 37 °C. Cells were washed, fixed and permeabilized for 30 min at 4 °C with the Foxp3 Transcription Factor Fixation/Permeabilization kit (Invitrogen by Life Technologies) according to the manufacturer's instructions. Cytokines were stained by incubating 1 μl antibody for every 5×10^4 cells diluted in Permeabilization Buffer (Invitrogen by Life Technologies) for 20 min at room temperature. T cells were washed with PBS and then analyzed. For cytokine staining, the following antibodies were used: anti-IFN-γ–V450 (clone B27), anti-GrzB–FITC (clone GB11), anti-PerfA–APC (clone deltaG9), anti-PerfA–PE (clone deltaG9). Proliferation assays of chronically stimulated cells were performed using cell trace

(C34557); naive $CD4^+$ and $CD8^+$ T cells were incubated with 1 μl cell trace for every 1×10^6 cells in PBS at 37 °C for 20 min. Cells were then washed with 10% FBS for 5 min at 37 °C and activated as reported previously; proliferation was assessed 5–7 d after activation. For cell cyce analysis T cells were fixed for 10 minutes at room temperature in cold ethanol 70%, then fixed cells were washed and stained in PBS with 0,1% of NP-40, RNAse A (100 μg/mL), and propidium iodide (50 μg/mL). For all above-mentioned analyses, an average of 10^4 cells were acquired with the FACSCanto I (BD Biosciences) using BD FACSDiva Software version 8.0.3, and data were analyzed using FlowJo version 10.6.1 software.

Killing assay

In vitro exhausted effector $CD4^+$ and $CD8^+$ T cells or TILs were treated with ASOs and co-cultured for 12 h with heterologous monocytes at a 1:1 ratio. After co-culturing, cells were stained with the LIVE/DEAD Fixable Green Dead Cell Stain kit (Invitrogen by Life Technologies, L34969) for 20 min at room temperature, washed with PBS and stained with anti-CD14–APC antibody (clone M5E2) to recognize monocytes. Monocytes were identified as CD14-positive cells, and their viability was assessed as the percentage of dead monocytes. To assess the spontaneous lysis of monocytes, we cultured monocytes without T cells as a control. An average of 10^4 cells were acquired with the FACSCanto I (BD Biosciences) using BD FACSDiva

Software version 8.0.3, and data were analyzed using FlowJo version 10.6.1 software.

Prime-Flow RNA Assay

Prime-Flow RNA assay was performed according to the manufacturer's instructions with minor adaptation. T cells were stained for surface markers for 30 minutes at 37°C. In particular, $CD3^+$ T cells derived from healthy donors and TILs derived from CRC and lung cancer patients, were stained with anti- CD45, CD3, CD4, CD8, CD45RO and CD45RA to discriminate $CD4^+$ and $CD8^+$ naïve and memory subset; while exhausted $CD4^+$ T cells were stained with anti- PD1 and TIM3 to appreciate the exhausted status of the cells. After surface marker staining T cells were washed with PBS1X by centrifugation at 500 rpm for 5 minutes at 4°C. The cells were fixed with fixation solution I for 30 minutes in agitation at 4°C, were washed and then permeabilized with Permeabilization buffer 10X diluted in ddH_2O for 30 minutes in agitation at 4°C, according to protocol. The cells were then re-fixed with fixation solution II for 1 hour at room temperature in agitation. After these two rounds of fixing, the cells were washed with Prime-Flow RNA wash buffer and ibridize for 2 hours at 40°C in agitation with LINE1 probe diluted 1:20 in target probe diluent. The cells were washed and maintained overnight at 4°C. The day after the probe's signal were amplified incubating T cells with Pre-Amplification and Amplification

solution for 1.30 hours each at 40°C in agitation. Then the cells were labelled for 1 hours at 40°C in agitation, with Prime-Flow RNA label probe (100X) in PrimeFlow RNA Label Probe diluent. An average of 10^5 cells were acquired with the FACSSymphony (BD Biosciences) using BD FACSDiva Software version 8.0.3, and data were analyzed using FlowJo version 10.6.1 software.

RNA isolation and RT–qPCR

Total RNA was isolated using the RNeasy Mini kit (Qiagen) and the QIAshredder system (Qiagen) according to the manufacturer's instructions. During the extraction, DNase treatment with the RNase-free DNase Set (QIAGEN) was performed. Total RNA was reverse transcribed using the SuperScript III First-Strand Synthesis SuperMix kit (Invitrogen by Life Technologies) following the manufacturer's instructions. Real-time quantitative PCR was performed on the StepOnePlus Real-Time PCR System (Applied Biosystems by Life Technologies) using Power SYBR Green PCR Master Mix (Applied Biosystems by Life Technologies). All gene expression data were normalized to the expression of two independent housekeeping genes (*RNA18SN*1 and *GAPDH*). Normalized Ct values were calculated as $2^{-\Delta Ct}$ or $2^{-\Delta\Delta Ct}$.

RNA-FISH

Briefly, antisense biotinylated riboprobes for LINE1, AluY and HERVK were transcribed in vitro using the MAXIscript T7

transcription kit (Invitrogen) and Biotin RNA Labeling Mix (Roche). In total, 50–100 ng antisense biotinylated riboprobes per experiment were used. T cells fixed with 3% paraformaldehyde were washed with 0.05% Triton X-100 in PBS, permeabilized with 0.5% Triton X-100 in PBS and maintained in 20% glycerol–PBS. Cells were frozen and thawed with dry ice and deproteinized with 0.1 N HCl. T cells were hybridized with riboprobes at 52.5 °C for 3.5 min and incubated overnight at 37 °C in a water bath. Glasses were washed with 50% formamide–2× SSC (saline-sodium citrate), 2× SSC, 1× SSC and 4× SSC–0.2% Tween-20. T cells were blocked with BSA and then incubated with streptavidin HRP (1:1,000, PerkinElmer by Akoya Biosciences) diluted in TNT–BSA (0.1 M Tris-HCl, pH 8, 0.150 M NaCl, 0.1% NP-40 and 4% BSA in DEPC). T cells were washed four times with TNT, and the signal was amplified by incubating TSA working solution (1:150) in 1× amplification buffer for 3 min (TSA Plus Fluorescent kit Cy3.5, PerkinElmer). T cells were washed four times with TNT; nuclei were counterstained with 1 μg ml^{-1} DAPI. Images of RNA-FISH were acquired with a Leica TCS SP5 Confocal microscope (Leica Microsystems) using an HCX PL APO ×63, 1.40-NA oil-immersion objective (Leica Microsystems) with additional 2× zoom, using 0.29-μm *Z* stacks at randomly chosen fields. To obtain images comparable between different T cells isolated from the same individuals and/or within the same experiment, we performed the RNA-FISH protocol at the same time for all conditions and images were

acquired with equivalent parameters and processed similarly; at least ~100 nuclei per condition were acquired.

Quantification of 3D mean fluorescence intensity and colocalization.

To quantify mean fluorescence intensity of RNA (AluYa5, HERV-K and LINE1) signals in 3D reconstructed nuclei, raw images were analyzed with a proper pipeline constructed with NIS-Elements analysis AR imaging software (Nikon). The pipeline generates at first a 2D mask to identify the ROI of the nuclei through DAPI's signals. To clean DAPI signals the pipeline foreseen binary processes: filter on dimension to avoid doublets and fill holes to close the holes of chromatin. The background signal was subtracted to better recognize the nuclei. Then, another mask was generated to identify the RNA (AluYa5, HERV-K and LINE1) thanks to Cy3.5 signals, maintaining an open threshold to keep the largest amount of signals. The two masks were combined and the software integrate all the confocal z stack and generate a 3D image. On this 3D projection, mean fluorescent intensity of RNA signals was calculated in the nuclei previously identified

Indeed, the colocalization of RNA and histone mark signals was measured by ImageJ Software. At first, a ROI of the nuclei was designed through DAPI's signals, then using "Colocalization

Threshold" the software calculates the Pearson Correlation of RNA and histone marks signals for every nucleus.

Chromatin and nucleoplasm RNA extraction

Briefly 5-10 x106 of quiescent naive CD4+ T-cells were resuspended in 60 μL of Buffer A (10 mM HEPES pH 7.5, 10 mM KCl, 10% (v/v) glycerol, 340 mM sucrose, 4 mM MgCl2, 1 mM DTT, 1X Protease Inhibitor Cocktail (PIC)), an equal volume of Buffer A 0.2% (v/v) Triton X-100 was added and T-cells were lysed for 12 min on ice. T-cells were centrifugated at 1200 g for 5 min at 4°C, the supernatant was collected representing cytosolic RNA fraction. The nuclear pellet was washed in 120 μL of NRB Buffer (20 mM HEPES pH 7.5, 50 % (v/v) glycerol, 75 mM NaCl, 1 mM DTT, 1X PIC) at 900g for 5 min at 4°C and resuspended in 60 μL of NRB Buffer, an equal volume of NUN Buffer was added (20 mM HEPES pH 7.5, 300 mM NaCl, 1 M Urea, 1 % (v/v) NP-40, 1 mM MgCl2, 1 mM DTT) and T-cells were lysed for 5 min on ice. The lysate was centrifugated at 1200 g, 5 min at 4°C, the supernatant was collected representing nucleoplasmatic RNA fraction. The chromatin pellet was washed in 500 μL of Buffer A at 1200g for 5 min at 4°C and then the pellet was resuspended in 50 μL of Buffer A representing the chromatin RNA fraction. Total, nucleoplasm and chromatin associated RNA was extracted using Maxwell RSC miRNA Tissue kit (Promega,

cat. num. AS1460) following manufacturer's instructions with minor adaptation.

RNA library preparation and sequencing

Further details on chromatin and nucleoplasm RNA extraction are provided in the Supplementary Methods. RNA integrity was checked with the TapeStation system (High Sensitivity RNA ScreenTape assay), and 15–75 ng total RNA was used to prepare libraries. RNA was ribodepleted with the RiboGone-Mammalian kit (Takara, 634846) following the manufacturer's instruction, and libraries were prepared with the SMARTer Stranded RNA-Seq kit (Takara, 634836) according to the manufacturer's instructions. Libraries were sequenced as paired 100-bp or 150-bp reads on the Illumina NextSeq 500. RNA-seq libraries were prepared for chromatin and nucleoplasm RNA from quiescent naive $CD4^+$ T cells (four individuals).

LINE1 transcripts validation by RT-PCR

LINE1 transcripts were validated by RT-PCR with GoTaq G2 Flexi DNA polymerase (Promega, cat. num. M7806). PCR reactions were performed on naïve CD4+ T-cells cDNA (RT minus was used to verify the splicing of the novel transcriptional variants). Primers were designed on HIRA.L1 transcripts and on the corresponding canonical mRNAs. PCR amplicons were

controlled by electrophoresis on 1.6% agarose gel. All the transcripts have been validated in at least 3 different individuals.

Statistical analysis

Statistical details including the number of individuals and nuclei acquired in imaging experiments, statistical methods used, *F* values and exact *P* values are reported in each figure legend, in the text. Results presented in this article were obtained from a minimum of three individuals; exceptions are indicated in figure legends (experiments performed on exhausted T cells, TILs or patients with transplantation).

2.4 Discussion

During my PhD project, I have the possibility to contribute to a huge study in which we described a novel function for chromatin LINE1 RNAs in human T lymphocytes. Indeed, we found that LINE1 RNAs are enriched in naïve $CD4^+$ T cells and these elements are spliced as non-canonical splicing variants of genes involved in T cell activation. We discovered that these non-canonical LINE1-containing transcripts are required to maintain

naïve $CD4^+$ T cell quiescence, in fact, their downregulation promotes and increases $CD4^+$ T cells effector functions.

Specifically, I was involved in discovering a possibly LINE1 RNA role in dysfunctional T cells that showed an impaired effector function. Unexpectedly we found that in vitro exhausted T cells with decreased effector properties exhibit upregulation of LINE1-containing transcripts and that the modulation of these RNAs can restore a proper T cell functionality. We described this mechanism also in ex vivo TILs isolated from NSLC and CRC patients in which we observed a re-accumulation of LINE1 RNAs and with the downregulation of these elements we revert the exhausted phenotype. All these pieces of evidence show that LINE1 RNAs could be a potential target for cancer therapy. For this purpose, we are planning to better characterize LINE1 RNAs distribution among dysfunctional T cells with scRNA-seq approach in order to corroborate their potential role in affecting the transcriptional plasticity of dysfunctional TILs subsets.

2.5 Chapter 2 bibliography

Balkhi, M. Y., Wittmann, G., Xiong, F., & Junghans, R. P. (2018). YY1 Upregulates Checkpoint Receptors and Downregulates Type I Cytokines in Exhausted, Chronically Stimulated Human T Cells. *IScience*. https://doi.org/10.1016/j.isci.2018.03.009

Belancio, V. P., Hedges, D. J., & Deininger, P. (2006). LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Research*. https://doi.org/10.1093/nar/gkl027

Bourque, G., Leong, B., Vega, V. B., Chen, X., Yen, L. L., Srinivasan, K. G., Chew, J. L., Ruan, Y., Wei, C. L., Huck, H. N., & Liu, E. T. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Research*. https://doi.org/10.1101/gr.080663.108

Chuong, E. B., Elde, N. C., & Feschotte, C. (2017). Regulatory activities of transposable elements: From conflicts to benefits. In *Nature Reviews Genetics*. https://doi.org/10.1038/nrg.2016.139

Criscione, S. W., Theodosakis, N., Micevic, G., Cornish, T. C., Burns, K. H., Neretti, N., & Rodić, N. (2016). Genome-wide characterization of human L1 antisense promoter-driven transcripts. *BMC Genomics*. https://doi.org/10.1186/s12864-016-2800-5

Denli, A. M., Narvaiza, I., Kerman, B. E., Pena, M., Benner, C., Marchetto, M. C. N., Diedrich, J. K., Aslanian, A., Ma, J., Moresco, J. J., Moore, L., Hunter, T., Saghatelian, A., & Gage, F. H. (2015). Primate-Specific ORF0 Contributes to Retrotransposon-Mediated Diversity. *Cell*. https://doi.org/10.1016/j.cell.2015.09.025

Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E., Carninci, P., & Torres-Padilla, M. E. (2013). Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nature Structural and Molecular Biology*. https://doi.org/10.1038/nsmb.2495

Franco, F., Jaccard, A., Romero, P., Yu, Y. R., & Ho, P. C. (2020). Metabolic and epigenetic regulation of T-cell exhaustion. In *Nature Metabolism*. https://doi.org/10.1038/s42255-020-00280-9

Guo, X., Zhang, Y., Zheng, L., Zheng, C., Song, J., Zhang, Q., Kang,

B., Liu, Z., Jin, L., Xing, R., Gao, R., Zhang, L., Dong, M., Hu, X., Ren, X., Kirchhoff, D., Roider, H. G., Yan, T., & Zhang, Z. (2018). Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nature Medicine*. https://doi.org/10.1038/s41591-018-0045-3

Jamal-Hanjani, M., Thanopoulou, E., Peggs, K. S., Quezada, S. A., & Swanton, C. (2013). Tumour heterogeneity and immune-modulation. In *Current Opinion in Pharmacology*. https://doi.org/10.1016/j.coph.2013.04.006

Lavin, Y., Kobayashi, S., Leader, A., Amir, E. ad D., Elefant, N., Bigenwald, C., Remark, R., Sweeney, R., Becker, C. D., Levine, J. H., Meinhof, K., Chow, A., Kim-Shulze, S., Wolf, A., Medaglia, C., Li, H., Rytlewski, J. A., Emerson, R. O., Solovyov, A., … Merad, M. (2017). Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. *Cell*. https://doi.org/10.1016/j.cell.2017.04.014

Munn, D. H., & Bronte, V. (2016). Immune suppressive mechanisms in the tumor microenvironment. In *Current Opinion in Immunology*. https://doi.org/10.1016/j.coi.2015.10.009

Percharde, M., Lin, C. J., Yin, Y., Guan, J., Peixoto, G. A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., & Ramalho-Santos, M. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell*. https://doi.org/10.1016/j.cell.2018.05.043

Quezada, S. A., & Peggs, K. S. (2011). Tumor-reactive CD4 + T cells: Plasticity beyond helper and regulatory activities. In *Immunotherapy*. https://doi.org/10.2217/imt.11.83

Roy-Engel, A. M., El-Sawy, M., Farooq, L., Odom, G. L., Perepelitsa-

Belancio, V., Bruch, H., Oyeniran, O. O., & Deininger, P. L. (2005). Human retroelements may introduce intragenic polyadenylation signals. *Cytogenetic and Genome Research*. https://doi.org/10.1159/000084968

Zhang, L., Yu, X., Zheng, L., Zhang, Y., Li, Y., Fang, Q., Gao, R., Kang, B., Zhang, Q., Huang, J. Y., Konno, H., Guo, X., Ye, Y., Gao, S., Wang, S., Hu, X., Ren, X., Shen, Z., Ouyang, W., & Zhang, Z. (2018). Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature*. https://doi.org/10.1038/s41586-018-0694-x

Chapter 3.

Published Paper “LINE1 are spliced in non-canonical transcript variants to regulate T cell quiescence and exhaustion”

LINE1 are spliced in non-canonical transcript variants to regulate T cell quiescence and exhaustion

Federica Marasca[1,2,14], Shruti Sinha[1,14], Rebecca Vadalà[1,3], Benedetto Polimeni[1,3], Valeria Ranzani[1], Elvezia Maria Paraboschi[4,5], Filippo Vittorio Burattin[1], Marco Ghilotti[1], Mariacristina Crosti[1], Maria Luce Negri[1], Susanna Campagnoli[6], Samuele Notarbartolo[1], Andrea Sartore-Bianchi[7,8], Salvatore Siena[7,8], Daniele Prati[9], Giovanni Montini[2,10], Giuseppe Viale11, Olga Torre[12], Sergio Harari[2,12], Renata Grifantini[1,6], Giulia Soldà[4,5], Stefano Biffo[1,13], Sergio Abrignani[1,2] and Beatrice Bodega[1,13]

[1]INGM, Istituto Nazionale di Genetica Molecolare ‘Romeo ed Enrica Invernizzi’, Milan, Italy.
[2]Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy.
[3]Ph.D. Program in Translational and Molecular Medicine, DIMET, University of Milan-Bicocca, Monza, Italy.
[4]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy.

[5]Humanitas Clinical and Research Center, IRCCS, Rozzano, Milan, Italy.
[6]CheckmAb Srl, Milan, Italy.
[7]Niguarda Cancer Center, Grande Ospedale Metropolitano Niguarda, Milan, Italy.
[8]Department of Oncology and Hematology–Oncology, University of Milan, Milan, Italy.
[9]Department of Transfusion Medicine and Hematology, Fondazione IRCCS Cà Granda, Ospedale
Maggiore Policlinico, Milan, Italy.
[10]Pediatric Nephrology and Dialysis Unit, Fondazione IRCCS Cà Granda, Ospedale Maggiore Policlinico, Milan, Italy.
[11]University of Milan, European Institute of Oncology IRCCS, Milan, Italy.
[12]Department of Medical Sciences, San Giuseppe Hospital MultiMedica
IRCCS, Milan, Italy.
[13]Department of Biosciences, University of Milan, Milan, Italy.
[14]These authors contributed equally: Federica Marasca, Shruti Sinha.
e-mail: abrignani@ingm.org; bodega@ingm.org

Nature Genetics https://doi.org/10.1038/s41588-021-00989-7

nature genetics

ARTICLES

https://doi.org/10.1038/s41588-021-00989-7

Check for updates

LINE1 are spliced in non-canonical transcript variants to regulate T cell quiescence and exhaustion

Federica Marasca[1,2,14], Shruti Sinha[1,14], Rebecca Vadalà[1,3], Benedetto Polimeni[1,3], Valeria Ranzani[1], Elvezia Maria Paraboschi[4,5], Filippo Vittorio Burattin[1], Marco Ghilotti[1], Mariacristina Crosti[1], Maria Luce Negri[1], Susanna Campagnoli[6], Samuele Notarbartolo[1], Andrea Sartore-Bianchi[7,8], Salvatore Siena[7,8], Daniele Prati[9], Giovanni Montini[2,10], Giuseppe Viale[11], Olga Torre[12], Sergio Harari[2,12], Renata Grifantini[1,6], Giulia Soldà[4,5], Stefano Biffo[1,13], Sergio Abrignani[1,2] ✉ and Beatrice Bodega[1,13] ✉

How gene expression is controlled to preserve human T cell quiescence is poorly understood. Here we show that non-canonical splicing variants containing long interspersed nuclear element 1 (LINE1) enforce naive CD4$^+$ T cell quiescence. LINE1-containing transcripts are derived from CD4$^+$ T cell-specific genes upregulated during T cell activation. In naive CD4$^+$ T cells, LINE1-containing transcripts are regulated by the transcription factor IRF4 and kept at chromatin by nucleolin; these transcripts act in *cis*, hampering levels of histone 3 (H3) lysine 36 trimethyl (H3K36me3) and stalling gene expression. T cell activation induces LINE1-containing transcript downregulation by the splicing suppressor PTBP1 and promotes expression of the corresponding protein-coding genes by the elongating factor GTF2F1 through mTORC1. Dysfunctional T cells, exhausted in vitro or tumor-infiltrating lymphocytes (TILs), accumulate LINE1-containing transcripts at chromatin. Remarkably, depletion of LINE1-containing transcripts restores TIL effector function. Our study identifies a role for LINE1 elements in maintaining T cell quiescence and suggests that an abundance of LINE1-containing transcripts is critical for T cell effector function and exhaustion.

Transposable elements (TEs) are mobile repetitive elements that account for approximately half of the human genome. The majority of TEs are retrotransposons, which have acquired the ability to move throughout the genome via a 'copy-and-paste' mechanism, using their RNA as an intermediate. Retrotransposons are divided into three main classes, one made by human endogenous retroviruses (HERVs) and long terminal repeats and the other two classes made by non-long terminal repeat retrotransposons such as long and short interspersed nuclear elements, and they are organized in superfamilies (for example, LINE1, ERVL and Alu) and families (for example, L1Hs, L1M, HERVK, AluY)[1,2]. Most TEs are present in the genome as retrotransposition-inactive fossils, resulting from mobilizations that occurred millions of years ago and were remodeled through evolution[3]; they are transcribed in a tissue-specific manner[4]. Among TEs, LINE elements comprise the largest class, and LINE1 elements alone cover almost 18% (~500,000 copies) of the human genome. Functionally, LINE1 elements can have several roles uncoupled from retrotransposition[5]: they can provide new regulatory regions, act as alternative promoters[6,7], serve as binding sites for transcription factors[8] and represent polyadenylation[9] or alternative splicing sites[10]. Mechanistically, chromatin-bound LINE1 RNA species can regulate chromatin accessibility during embryogenesis[11,12]. In mouse embryonic stem cells (mESCs), they contribute to epigenetic maintenance of both cell identity and two-cell-stage differentiation[13,14]. However, the role of RNA species transcribed from LINE1 elements in cells derived from adult tissues is still unexplored; therefore, we sought to investigate their expression and function in human T lymphocytes.

T lymphocytes are a major component of adaptive immune responses that provide protection against pathogens and transformed cells while maintaining self-tolerance. In particular, naive CD4$^+$ T cells, when activated by the recognition of cognate antigens, exit from quiescence and initiate activation and differentiation programs[15], enabling the acquisition of specialized effector functions that can be associated with distinct helper T cell subsets (for example, T_H1, T_H2, T_H17, T_{FH})[16,17]. Although T cell quiescence is actively enforced at transcriptional, translational and metabolic levels[18–21], there is little information regarding its epigenetic maintenance.

[1]INGM, Istituto Nazionale di Genetica Molecolare 'Romeo ed Enrica Invernizzi', Milan, Italy. [2]Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy. [3]Ph.D. Program in Translational and Molecular Medicine, DIMET, University of Milan-Bicocca, Monza, Italy. [4]Department of Biomedical Sciences, Humanitas University, Pieve Emanuele, Milan, Italy. [5]Humanitas Clinical and Research Center, IRCCS, Rozzano, Milan, Italy. [6]CheckmAb Srl, Milan, Italy. [7]Niguarda Cancer Center, Grande Ospedale Metropolitano Niguarda, Milan, Italy. [8]Department of Oncology and Hematology–Oncology, University of Milan, Milan, Italy. [9]Department of Transfusion Medicine and Hematology, Fondazione IRCCS Cà Granda, Ospedale Maggiore Policlinico, Milan, Italy. [10]Pediatric Nephrology and Dialysis Unit, Fondazione IRCCS Cà Granda, Ospedale Maggiore Policlinico, Milan, Italy. [11]University of Milan, European Institute of Oncology IRCCS, Milan, Italy. [12]Department of Medical Sciences, San Giuseppe Hospital MultiMedica IRCCS, Milan, Italy. [13]Department of Biosciences, University of Milan, Milan, Italy. [14]These authors contributed equally: Federica Marasca, Shruti Sinha. ✉e-mail: abrignani@ingm.org; bodega@ingm.org

In this study, we report a mechanism of gene expression control in T cells that involves splicing of distinct intronic LINE1 elements belonging to L1M families as exons of new transcript variants (hereafter called LINE1-containing transcripts). We show that LINE1-containing transcripts are specifically generated in naive CD4+ T cells under the control of the transcription factor IRF4. They are retained at chromatin by nucleolin, where they hinder H3K36me3 deposition in *cis* and keep gene expression paused. T cell activation induces LINE1-containing transcript downregulation through splicing suppression via PTBP1 and elongation of canonical transcripts by GTF2F1. We demonstrate that, in tumor-infiltrating dysfunctional T cells, LINE1-containing transcripts aberrantly accumulate at chromatin due to the impairment of this mechanism. Strikingly, TIL effector function could be restored by targeting LINE1-containing transcripts with antisense oligonucleotides (ASOs). We suggest that LINE1-containing transcripts are fundamental for maintaining naive T cell quiescence with a mechanism that is re-established in dysfunctional T cells, in which it dampens their effector function.

Results

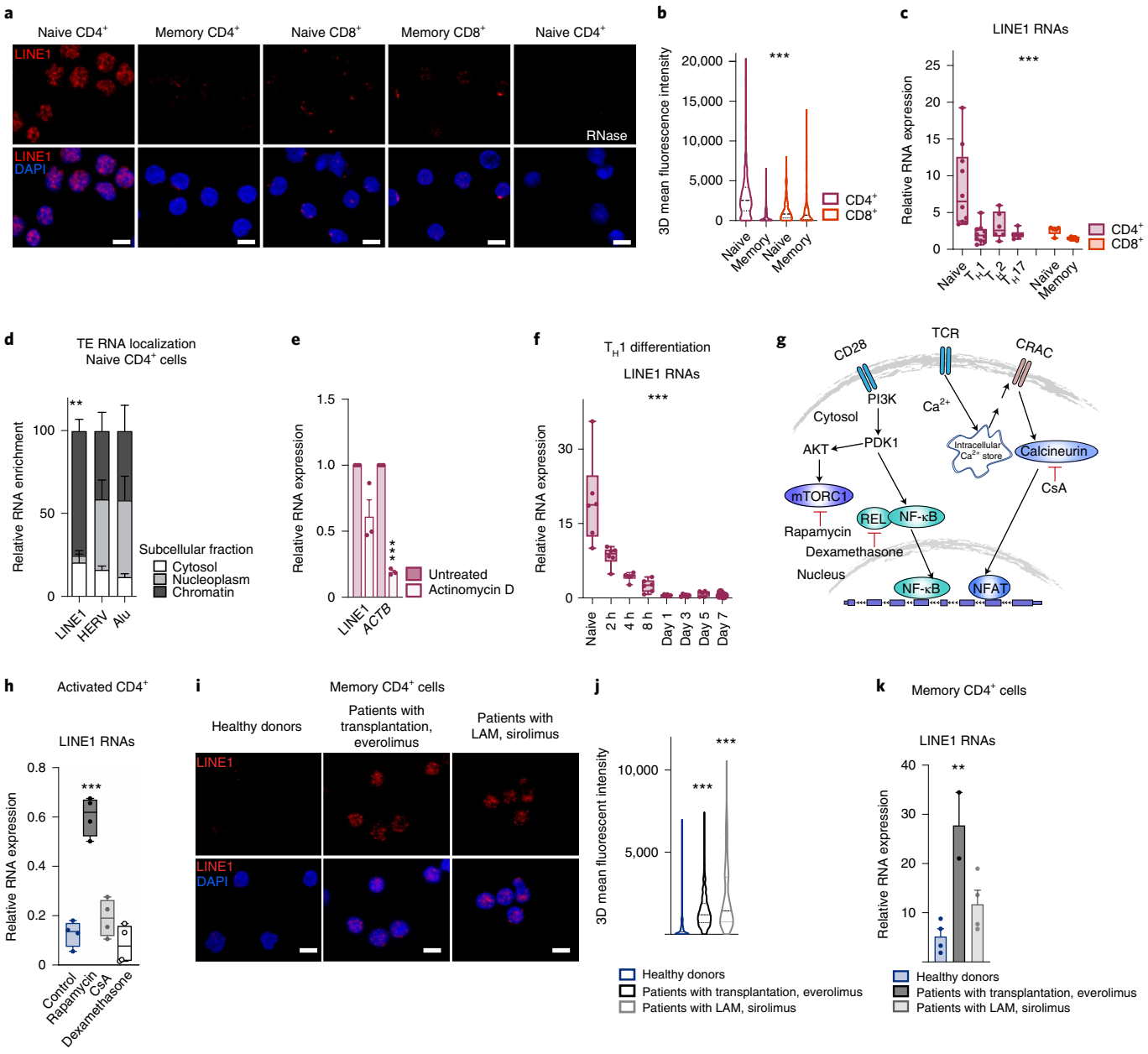
LINE1 elements are expressed in naive CD4+ T cells and are associated with chromatin. To investigate the expression of TEs in human T lymphocytes, we analyzed LINE1, Alu and HERV elements by RNA fluorescence in situ hybridization (RNA-FISH) and quantitative PCR with reverse transcription (RT–qPCR) in quiescent naive and memory CD4+ and CD8+ T cells isolated from healthy individuals (Extended Data Fig. 1a). We observed that LINE1 elements were specifically expressed in the nuclei of naive CD4+ T cells (Fig. 1a–c). By contrast, Alu RNA species showed a broad perinuclear distribution in all T cell subsets (Extended Data Fig. 1b–d), while HERV elements were poorly expressed (Extended Data Fig. 1e–g). LINE1 RNA species were highly enriched in the chromatin fraction of naive CD4+ T cells (Fig. 1d and Extended Data Fig. 1h) and were associated with active chromatin regions as determined by colocalization with H3K4me3 (Extended Data Fig. 1i,j). Prolonged actinomycin D treatment moderately affected LINE1 expression levels, indicating that these RNA species were not transcribed at a high rate in naive CD4+ T cells (Fig. 1e). We then analyzed LINE1 RNA dynamics in naive CD4+ T cell activation and differentiation and found that their expression was rapidly downregulated upon activation and remained at low levels during differentiation under polarizing conditions (that is, under T_H1, T_H2 and T_H17 conditions; Fig. 1f and Extended Data Fig. 1k,l).

We hypothesized that LINE1 expression is regulated by T cell-specific signaling pathways. Thus, we treated activated or differentiated naive CD4+ T cells with different immunosuppressive drugs that target mTORC1, calcineurin or nuclear factor (NF)-κB pathways (Fig. 1g), finding that the mTORC1 inhibitor rapamycin restored LINE1 RNA levels in activated and differentiated T cells (Fig. 1h and Extended Data Fig. 1m–o). To strengthen these in vitro data, we investigated whether LINE1 expression was affected by mTORC1 inhibition in vivo. Thus, we examined LINE1 expression in memory CD4+ T cells isolated from the blood of patients with kidney transplantation treated with the mTORC1 inhibitor everolimus and that of patients with lymphangioleiomyomatosis (LAM) (MIM, 606690)[22] treated for life with the rapamycin analog sirolimus[23] (Supplementary Table 1). Consistent with the inhibitory role of mTORC1 on LINE1 expression in vitro, we found that, at variance with healthy individuals, memory CD4+ T cells of these patients re-expressed LINE1 (Fig. 1i–k). Therefore, we demonstrate that LINE1 RNA species are localized at chromatin in CD4+ T cells, where their expression is rapidly downregulated following T cell activation in a mTORC1-dependent manner.

LINE1 elements are spliced in T cell-activation gene isoforms. To determine which LINE1 elements are expressed and how LINE1-containing transcripts are constituted, we sequenced chromatin-bound and nucleoplasm RNA from quiescent naive CD4+ T cells (Supplementary Table 2). First, we obtained read counts of TE classes, superfamilies and families in chromatin and nucleoplasmic fractions. We found that, among TE classes, LINE and, in particular, L1M families (evolutionarily old, present in primates and widely in other mammals) were the most expressed and chromatin enriched, whereas the L1P and L1H families were expressed at low levels and enriched in the nucleoplasm (Extended Data Fig. 2a,b). As nuclear LINE1 RNA was already described to regulate early development and cell identity in mESCs[13], we asked whether the same LINE1 families were expressed in mESCs and human T cells. We found that, in mESCs, the evolutionarily young,

Fig. 1 | LINE1 elements are expressed and bound to chromatin in quiescent naive CD4+ T cells and downregulated by mTORC1 upon T cell activation. **a**, Confocal fluorescence microscopy images of LINE1 RNA-FISH (red) of naive and memory CD4+ and CD8+ T cells. LINE1 riboprobes were designed based on the LINE1 ORF2 region (Extended Data Fig. 4b and Supplementary Table 11). Original magnification, 63×. Scale bar, 5 µm. **b**, Quantification of LINE1 RNA-FISH of at least 250 nuclei from four individuals (3D, three dimensional). $^{***}P<1\times10^{-15}$, one-way ANOVA, $F=149.7$. **c**, LINE1 expression measured by RT–qPCR in naive and memory T_H1, T_H2 and T_H17 CD4+ T cells and in naive and memory CD8+ T cells ($n=10$ individuals for naive and T_H1 CD4+ cells, $n=6$ for T_H2 and T_H17 cells, $n=5$ for CD8+ subsets). LINE1 primers were designed based on the LINE1 ORF2 region (Extended Data Fig. 4b and Supplementary Table 11). $^{***}P=0.0001$, one-way ANOVA, $F=7.1$. For box-and-whisker plots, the central line, the box and whiskers represent the median, the interquartile range (IQR) from first to third quartiles and 1.5 × IQR, respectively. **d**, Expression of LINE1, HERV and Alu in the cytosol, nucleoplasm and chromatin of naive CD4+ T cells ($n=4$ individuals for LINE1, $n=3$ for HERV and Alu) (primers are described in Supplementary Table 11). Data represent mean ± s.e.m. $^{**}P=0.0062$, one-way ANOVA, $F=32.58$. **e**, LINE1 and *ACTB* expression measured by RT–qPCR in naive CD4+ T cells with or without actinomycin D ($n=3$ individuals). Data represent mean ± s.e.m. *ACTB*, untreated versus actinomycin D, treated groups, $^{***}P=0.0003$, two-tailed paired *t*-test. **f**, LINE1 expression measured by RT–qPCR in quiescent and activated naive CD4+ T cells ($n=6$ individuals). $^{***}P=1.16\times10^{-12}$, one-way ANOVA, $F=24.77$. For box-and-whisker plots, the central line, the box and whiskers represent the median, the IQR from first to third quartiles and 1.5 × IQR, respectively. **g**, Scheme of pathways and drugs used for their inhibition. CRAC, calcium release-activated channels; CsA, cyclosporin A; NFAT, nuclear factor of activated T cells; PDK1, phosphoinositide-dependent kinase 1; PI3K, phosphoinositide 3-kinase; TCR, T cell receptor. **h**, LINE1 expression measured by RT–qPCR in activated naive CD4+ T cells treated with the inhibitors indicated in **g** ($n=4$ individuals). Control versus rapamycin, $^{***}P=0.0009$, two-tailed paired *t*-test. For box-and-whisker plots, the central line, the box and whiskers represent the median, the IQR from first to third quartiles and 1.5 × IQR, respectively. **i**, Confocal fluorescence microscopy images of LINE1 RNA-FISH (red) of memory T_H1 CD4+ T cells isolated from healthy individuals, patients with kidney transplantation treated with everolimus or patients with LAM treated with sirolimus. Original magnification, 63×. Scale bar, 5 µm. **j**, Quantification of LINE1 RNA-FISH of at least 138 nuclei from five healthy donors, two patients with kidney transplantation and two patients with LAM. Healthy donors versus patients with transplantation, $^{***}P<1\times10^{-15}$, two-tailed Mann–Whitney test; healthy donors versus patients with LAM, $^{***}P<1\times10^{-15}$, two-tailed Mann–Whitney test. **k**, LINE1 expression measured by RT–qPCR in memory T_H1 CD4+ T cells isolated from four healthy individuals, two patients with kidney transplantation treated with everolimus and four patients with LAM treated with sirolimus. Data represent mean ± s.e.m. $^{**}P=0.0070$, ordinary one-way ANOVA, $F=10.94$.

retrotransposition-competent, L1Md_T and L1Md_A subfamilies were more expressed with respect to the evolutionarily old L1_Mus1 and L1_Mus3 subfamilies (ref. [13] and Extended Data Fig. 2c). In T cells, almost half of LINE1 reads were chimeric (that is, mapping to both LINE1 and a nonrepetitive region), and the majority of reads derived from LINE1 localized to protein-coding genes (Extended Data Fig. 2d,e), thus most likely included in non-canonical transcript variants. In mESCs, the majority of LINE1 reads were entirely derived from LINE1 elements with a broader genomic distribution (Extended Data Fig. 2f,g), indicating that different LINE1 elements are expressed in mouse pluripotent cells than those in human T cells. To identify transcript variants containing LINE1, we used a de novo stranded genome-guided transcriptome assembly with two algorithms, Trinity[24] and StringTie[25] (Methods). We identified 3,072 multi-exonic transcripts containing at least one exon with LINE1. To obtain a catalog of LINE1-containing transcripts reconstructed with high confidence, we filtered them based on their consistent expression in different individuals and evidence of LINE1 exon transcription at the chromatin level (H3K36me3/H3 lysine 9 trimethyl (H3K9me3) ratio[26,27]). We retrieved 461 LINE1-containing transcripts that were non-canonical spicing variants originating from 407 protein-coding genes (Supplementary Table 3). We validated and accurately reconstructed 88% of these transcripts with long-read sequencing performed on the chromatin fraction of naive CD4+ T cells (Extended Data Fig. 3a and Supplementary Table 3), and several were confirmed by RT–PCR in naive CD4+ T cells isolated from different individuals (Extended Data Fig. 3b). The spliced LINE1 elements were short in length (371 bp on average), consisting mainly of open reading frame (ORF)2-truncated elements and enriched in distinct L1M subfamilies (that is, L1ME4a, L1MC4 and L1ME4b) (Extended Data Fig. 4a–c). In particular, 80% of these LINE1 elements were located within an intron (Extended Data Fig. 4d) and spliced as new exons that contained LINE1 and an intronic fragment (Extended Data Fig. 4e). Fig. 2a,b and Extended Data Fig. 4f show how representative LINE1-containing transcripts (for example, *HIRA.L1*, *RAB22A.L1* and *ARPC2.L1*) were reconstructed



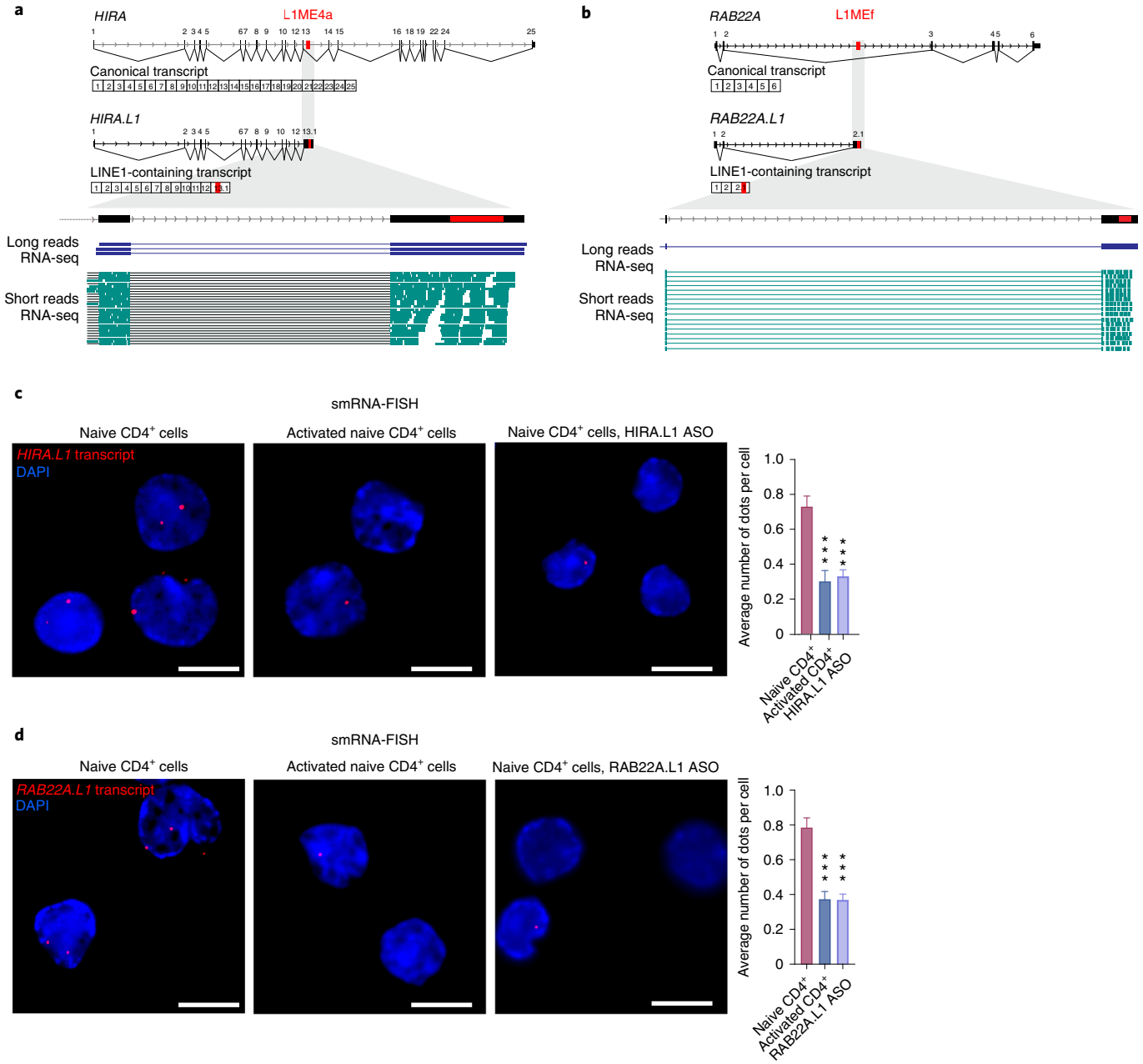


Fig. 2 | Intronic LINE1 elements are spliced in non-canonical transcript variants of cell-activation genes in naive CD4⁺ T cells. a,**b**, Scheme of *HIRA.L1* and *RAB22A.L1* transcripts. A magnified view of the exon containing a LINE1 element is shown. Long reads (blue) and short reads (green) of chromatin RNA-seq that support the exon reconstruction are shown. **c**,**d**, Wide-field fluorescence microscopy images of smRNA-FISH for *HIRA.L1* and *RAB22A.L1* (red) performed on quiescent and activated naive CD4⁺ T cells and on naive CD4⁺ T cells treated with HIRA.L1 (**c**) or RAB22A.L1 (**d**) ASOs. Images were processed with Richardson–Lucy two-dimensional deconvolution using NIS-Elements analysis AR imaging software. Nuclei were counterstained with 4,6-diamidino-2-phenylindole (DAPI). Original magnification, 100×. Scale bar, 5 µm. Bar plot represents the average smRNA-FISH dot number per nucleus. Data are shown as mean ± s.e.m. **c**, A total of 215 nuclei from naive cells ($n = 2$ individuals), 96 nuclei from activated cells ($n = 1$ individual) and 293 nuclei from naive cells depleted of *HIRA.L1* ($n = 2$ individuals) were used. **d**, A total of 223 nuclei from naive cells, 182 nuclei from activated cells and 300 nuclei from naive cells depleted of *RAB22A.L1* ($n = 2$ individuals for each condition) were used. *HIRA.L1* in naive versus activated CD4⁺ cells, ***exact $P = 0.000020$, two-tailed Mann–Whitney test; *HIRA.L1* in naive versus naive, HIRA.L1 ASO groups, $^{***}P = 1.8 \times 10^{-8}$, two-tailed Mann–Whitney test; *RAB22A.L1* in naive versus activated CD4⁺ cells, $^{***}P = 1.82 \times 10^{-7}$, two-tailed Mann–Whitney test; *RAB22A.L1* in naive versus naive, RAB22A.L1 ASO groups, $^{***}P = 2.52 \times 10^{-9}$, two-tailed Mann–Whitney test.

with short and long reads, split on the spliced LINE1 exon. The presence of *HIRA.L1* and *RAB22A.L1* in naive CD4⁺ T cells and their downregulation upon activation was further confirmed with single-molecule RNA-FISH (smRNA-FISH), detecting the unique portion of the LINE1 exon (Fig. 2c,d).

To understand the functional relevance of the 407 protein-coding genes from which LINE1-containing transcripts are derived, we used Ingenuity Pathway Analysis network analysis and found a statistically significant enrichment of genes associated with cell activation (for example, gene expression, cell signaling and cell-to-cell

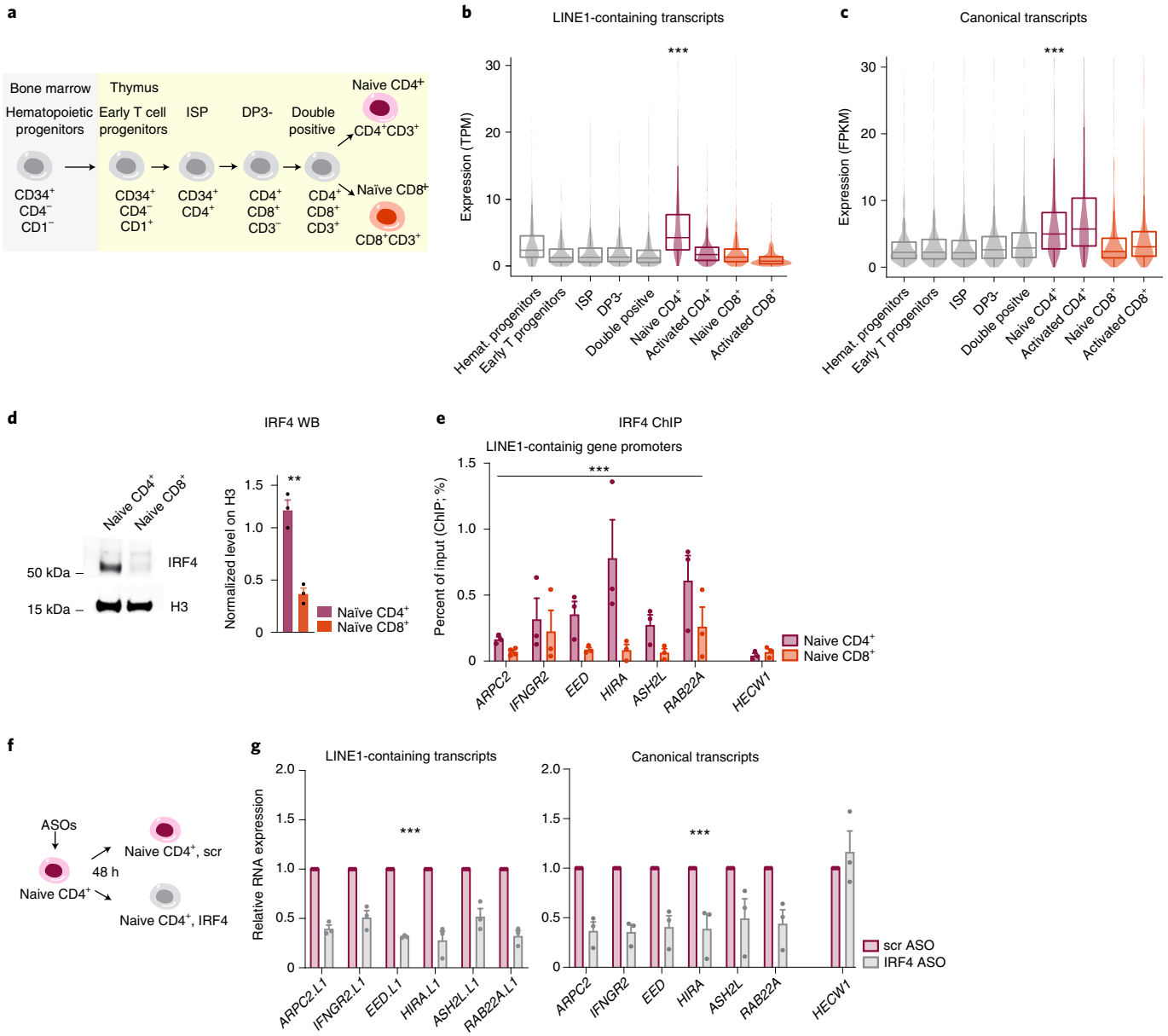


Fig. 3 | IRF4 regulates expression of LINE1-containing transcripts and canonical transcripts in CD4⁺ T cells. a, Scheme of CD4⁺ and CD8⁺ T cell development. **b**,**c**, Expression levels of LINE1-containing transcripts and canonical transcripts in RNA-seq datasets of T cells represented in **a**. Expression levels of 461 LINE1-containing transcripts presented in **b** were obtained using Salmon (Methods), ($n=1$ individual for hematopoietic progenitors, early T cell progenitors, immature single positive (ISP) cells, double positive CD3- (DP3-) cells and double-positive cells; $n=3$ individuals for naive CD4⁺ and activated CD4⁺ cells; $n=2$ individuals for naive CD8⁺ and activated CD8⁺ cells). $^{***}P<0.001$, two-tailed Wilcoxon rank-sum test of naive CD4⁺ T cells versus each cell type and $^{***}P<0.001$, two-tailed Wilcoxon signed-rank test with Bonferroni correction, naive versus activated CD4⁺ T cells (*P* values in **b** from left to right, 9.65×10^{-27}, 3.57×10^{-86}, 6.19×10^{-81}, 4.25×10^{-77}, 4.00×10^{-90}, 1.54×10^{-182}, 1.43×10^{-126}, 1.16×10^{-222}; in **c** from left to right, 8.32×10^{-52}, 5.25×10^{-48}, 1.60×10^{-50}, 8.73×10^{-34}, 2.61×10^{-24}, 6.51×10^{-15}, 2.33×10^{-62}, 7.59×10^{-29}). For box-and-whisker plots, the central line, the box and whiskers represent the median, the IQR from first to third quartiles and 1.5×IQR, respectively. FPKM, fragments per kb per million; hemat., hematopoietic; TPM, transcripts per million. **d**, IRF4 protein levels by western blot (WB) in naive CD4⁺ and naive CD8⁺ T cells and quantification ($n=3$ individuals). Data represent mean ± s.e.m. $^{**}P=0.0085$, two-tailed paired *t*-test. **e**, ChIP assays for IRF4 at promoters of LINE1-containing genes and the *HECW1* promoter (negative control) in naive CD4⁺ and CD8⁺ T cells ($n=3$ individuals). Primers were designed to amplify IRF4-binding sites (Supplementary Table 11). Data represent the mean percent of input ± s.e.m. Promoters of LINE1-containing genes in naive CD4⁺ versus CD8⁺ cells, $^{***}P=0.00054$, $F=21.8$, two-way ANOVA. **f**, Scheme of *IRF4* knockdown in quiescent naive CD4⁺ T cells. **g**, Expression of LINE1-containing transcripts and canonical transcripts measured by RT–qPCR in quiescent naive CD4⁺ T cells treated with IRF4 or control (scramble (scr)) ASOs ($n=3$ individuals). Primers are specific for the nonrepetitive region of LINE1 exons (Supplementary Table 11). Data represent mean ± s.e.m. LINE1-containing transcripts in scr versus IRF4 ASO groups, $^{***}P=2.11\times10^{-11}$, $F=552$, two-way ANOVA; canonical transcripts in scr versus IRF4 ASO groups, $^{***}P=1.56\times10^{-7}$, $F=116$, two-way ANOVA.

interactions and cell cycle) (Supplementary Table 4). Together, the above experiments identify a large set of non-canonical transcript variants of T cell-activation genes.

LINE1-containing transcripts are regulated by IRF4. As LINE1-containing transcripts are derived from genes involved in T cell activation, we investigated why naive CD8⁺ T cells, unlike the

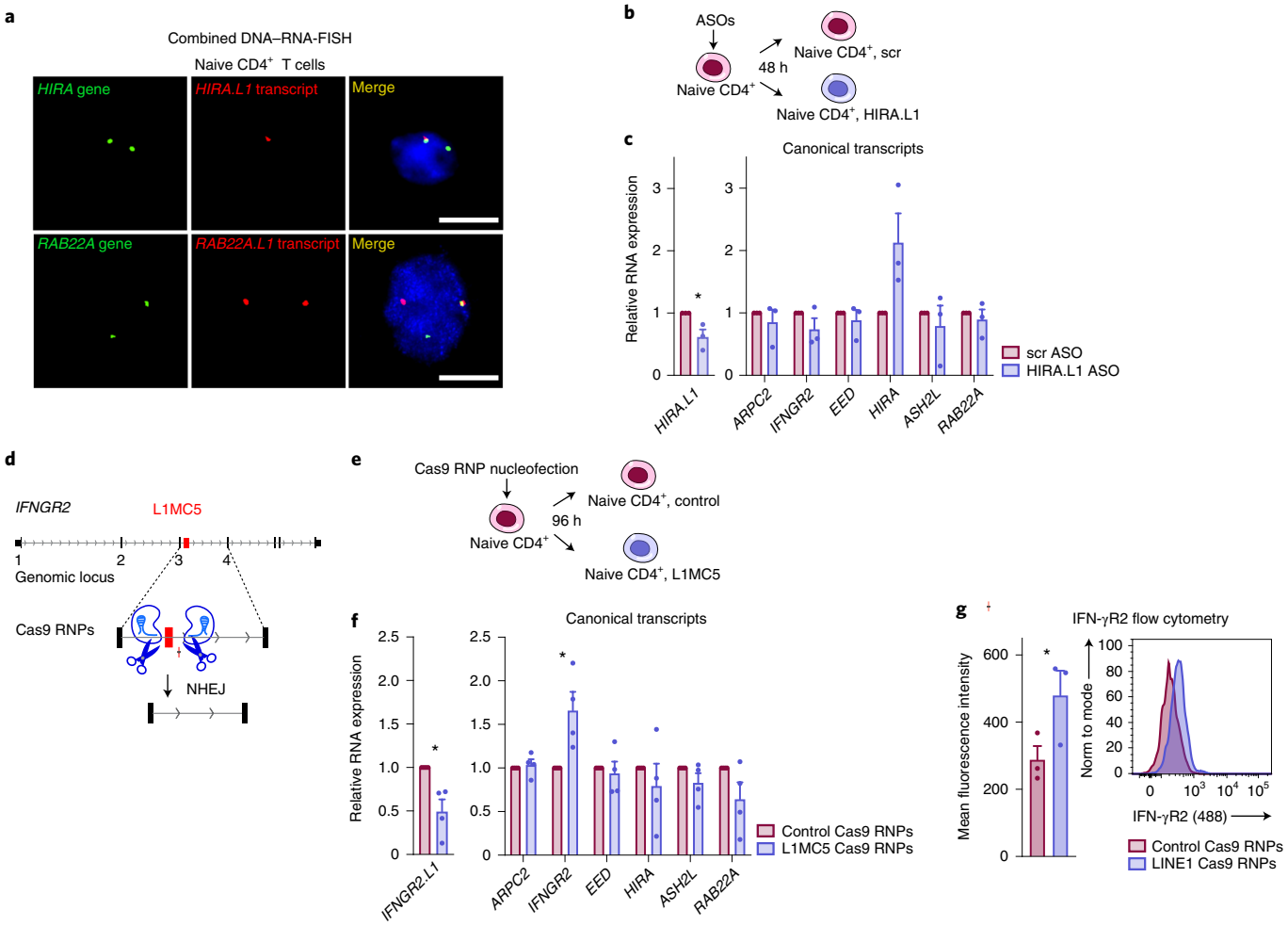


Fig. 4 | LINE1-containing transcripts are localized in *cis* and halt the expression of corresponding genes. a, Wide-field fluorescence microscopy images of combination DNA–RNA-FISH of quiescent naive CD4+ T cells, probing the *HIRA* genomic locus (top, green) and the *HIRA.L1* transcript (top, red) or the *RAB22A* genomic locus (bottom, green) and the *RAB22A.L1* transcript (bottom, red). Images were processed with Richardson–Lucy two-dimensional deconvolution using NIS-Elements analysis AR imaging software. Nuclei were counterstained with DAPI. Original magnification, 100×. Scale bar, 5 µm. **b**, Scheme of *HIRA.L1*-knockdown experiments in naive CD4+ T cells. **c**, Expression of *HIRA.L1* and of *ARPC2*, *IFNGR2*, *EED*, *HIRA*, *ASH2L* and *RAB22A* canonical transcripts measured by RT-qPCR in naive CD4+ T cells treated with HIRA.L1 or control (scr) ASOs ($n = 3$ individuals). Data represent mean ± s.e.m. *HIRA.L1* expression in scr versus HIRA.L1 ASO groups, $^*P = 0.0424$, one-tailed paired *t*-test. **d**, Design of the L1MC5 element deletion in the *IFNGR2* gene with CRISPR–Cas9 RNP complexes: the intronic L1MC5 element was deleted using sgRNA species complementary to nonrepetitive regions flanking the repeat. NHEJ, nonhomologous end joining. **e**, Scheme of the Cas9 RNP experiment in quiescent naive CD4+ T cells. **f**, Expression of *IFNGR2.L1* and of *ARPC2*, *IFNGR2*, *EED*, *HIRA*, *ASH2L* and *RAB22A* canonical transcripts measured by RT-qPCR in quiescent naive CD4+ T cells nucleofected with control Cas9 RNPs or Cas9 RNPs with sgRNA species targeting the flanking regions of the L1MC5 element in the intron of the *IFNGR2* gene (Supplementary Table 11) ($n = 4$ individuals). Data represent mean ± s.e.m. *IFNGR2.L1* expression in control versus L1MC5 Cas9 RNPs groups, $^*P = 0.0363$, two-tailed paired *t*-test; *IFNGR2* expression in control versus L1MC5 Cas9 RNPs, $^*P = 0.0275$, one-tailed paired *t*-test. **g**, IFN-γ receptor 2 (IFN-γR2) protein levels determined by flow cytometry analysis in quiescent naive CD4+ T cells nucleofected with control Cas9 RNPs or Cas9 RNPs with sgRNA species targeting flanking regions of the L1MC5 element in the intron of the *IFNGR2* gene and activated ($n = 3$ individuals). Norm, normalized. Data represent mean ± s.e.m. IFN-γR2 expression in control versus L1MC5 Cas9 RNPs, $^*P = 0.0397$, one-tailed paired *t*-test.

developmentally close CD4+ T cells, do not express them. Therefore, we profiled the expression of 461 LINE1-containing transcripts and of the corresponding protein-coding genes (hereafter named canonical transcripts) using RNA-seq data from T cell progenitors and from mature naive and activated CD4+ and CD8+ T cells (Fig. 3a and Methods). Analysis showed that LINE1-containing transcripts were selectively expressed by naive CD4+ T cells, supporting previous results (Figs. 1a–c and 3b and Supplementary Table 5). Interestingly, canonical transcripts were also preferentially expressed in CD4+ T cells; in particular, in activated CD4+ T cells, expression of LINE1-containing transcripts was downregulated, whereas canonical transcript expression was upregulated compared to that in quiescent cells (Fig. 3b,c, Extended Data Fig. 5a,b and Supplementary Table 6).

To identify transcriptional regulators of LINE1-containing transcripts, we searched in RNA-seq (Methods) and proteomic datasets[28] for transcription factors expressed more highly in CD4+ T cells than in CD8+ T cells and for which the DNA-binding motifs are enriched at the promoters of LINE1-containing genes. IRF4, a key factor for CD4+ T cell activation[29,30], ranked first in this analysis (Extended Data Fig. 5c and Supplementary Table 7). We verified that IRF4 expression was low in CD8+ T cells (Fig. 3d) and upregulated in CD4+ T cells upon activation (Extended Data Fig. 5d). Chromatin immunoprecipitation (ChIP) analysis demonstrated

that IRF4 binding to promoters of LINE1-containing genes was higher in naive CD4+ T cells than in CD8+ T cells (Fig. 3e). Finally, *IRF4* depletion in naive CD4+ T cells by specific ASOs (Fig. 3f and Extended Data Fig. 5e,f) resulted in the downregulation of both LINE1-containing and canonical transcripts (Fig. 3g), indicating that IRF4 controls their expression. Overall, these data suggest that IRF4 directly regulates the expression of LINE1-containing transcripts and the corresponding canonical transcripts that are selectively expressed in CD4+ T cells.

LINE1-containing transcripts act with nucleolin, hampering expression of the corresponding genes. As LINE1-containing transcripts derive from alternative splicing and are localized at chromatin, we asked whether they could regulate expression of the corresponding canonical transcripts. First, we observed that LINE1-containing transcripts were localized in *cis* in roughly 50% of nuclei analyzed, as exemplified for *HIRA.L1* and *RAB22A.L1* by combined DNA–RNA-FISH experiments (Fig. 4a and Supplementary Table 8).

Next, we depleted naive CD4+ T cells of *HIRA.L1* or *RAB22A.L1* transcripts with ASOs designed on the nonrepetitive region of the LINE1 exon (Fig. 2c,d) and found selective upregulation of the corresponding canonical transcript (Fig. 4b,c and Extended Data Fig. 6a,b). Finally, we deleted the LINE1 element from the *IFNGR2* or *ARPC2* gene (L1MC5 or L1MC5a, respectively) using Cas9 ribonucleoprotein (RNP) complexes[31] with single-guide (sg)RNA species complementary to the intronic region flanking the repeat and showed that, in naive CD4+ T cells, (1) LINE1 elements are necessary to generate LINE1-containing transcripts, (2) the regulatory role of LINE1-containing transcripts occurs in *cis*, and (3) upregulation of canonical transcripts observed in quiescent T cells results in increased protein levels in activated T cells when translational machinery is switched on[19] (Fig. 4d–g and Extended Data Fig. 6c–g). Our results suggest that LINE1-containing transcripts keep the expression of corresponding canonical transcripts paused.

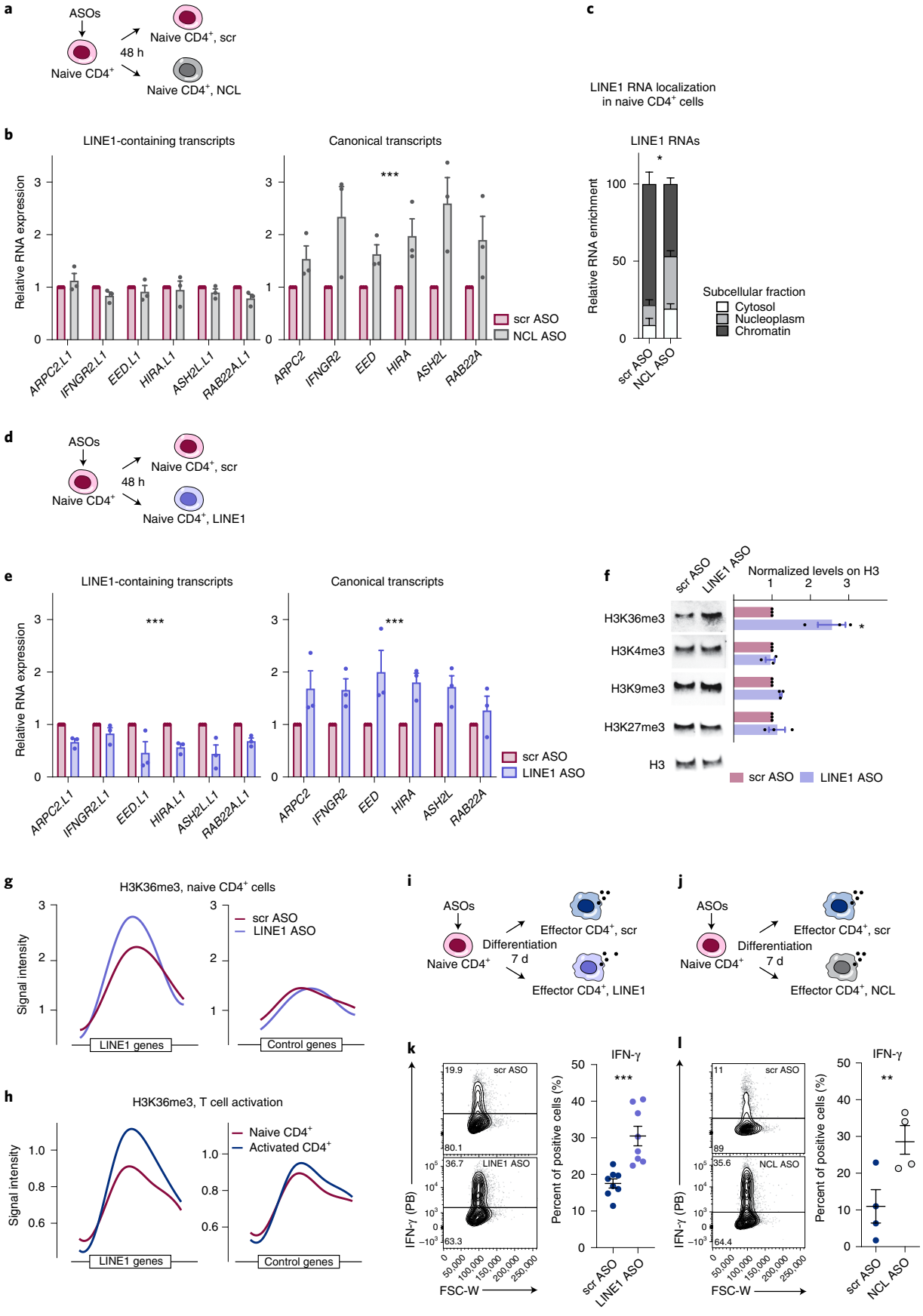
We speculated that this mechanism might occur at the chromatin level and investigated the role of the nucleolin–LINE1-containing transcript partnership in T cells. Indeed, nucleolin is a LINE1 RNA-binding protein[32] reported to regulate cell identity and two-cell-stage differentiation genes in complex with the regulator KAP1 in mESCs[13]. First, we verified by RNA-immunoprecipitation (RIP) experiments that LINE1-containing transcripts were bound by nucleolin (Extended Data Fig. 7a,b) and by co-immunoprecipitation experiments that nucleolin was also in complex with KAP1 in CD4+ T cells (Extended Data Fig. 7c). Next, we depleted nucleolin in naive CD4+ T cells with specific ASOs, observing that, although LINE1-containing transcript levels remained unchanged, LINE1 transcript were significantly dissociated from the chromatin fraction, and this was sufficient to increase expression of the corresponding canonical transcripts in the absence of cell activation (Fig. 5a–c and Extended Data Fig. 7d–f). The same result was obtained by depleting all LINE1-containing transcripts in naive CD4+ T cells by means of ASOs designed to target a common LINE1 region (Fig. 5d,e and Extended Data Figs. 4b and 7g–i). As LINE1 RNA species can regulate chromatin condensation and gene silencing[12–14], we asked whether knocking down all LINE1-containing transcripts affected chromatin organization in quiescent T cells. Thus, we assessed the level of several histone marks (that is, H3K36me3, H3 lysine 4 trimethyl (H3K4me3), H3K9me3 and H3 lysine 27 trimethyl (H3K27me3)) by quantitative western blot of histone extracts[33] and immunostaining naive T cells treated with LINE1 or nucleolin ASOs and observed a marked increase in H3K36me3 levels (Fig. 5f and Extended Data Fig. 7j–m), indicating chromatin remodeling toward active transcription. Interestingly, we performed H3K36me3 ChIP sequencing (ChIP–seq) of naive T cells depleted of LINE1-containing transcripts and found that the increase in H3K36me3 levels was specific for LINE1-containing genes (Fig. 5g and Supplementary Table 9). As these genes are involved in cell activation, we analyzed their H3K36me3 levels by ChIP–seq in quiescent and activated naive CD4+ T cells and observed a similar trend (Fig. 5h and Supplementary Table 9). On the contrary, H3K36me3 levels on randomly selected genes did not increase either upon LINE1-containing transcript depletion nor in response to cell activation (Fig. 5g,h, Extended Data Fig. 8a–d and Supplementary Table 9). Finally, as the regulation of cell-activation genes is instrumental for T cell effector function, we measured interferon (IFN)-γ levels in naive CD4+ T cells depleted of LINE1-containing transcripts or *NCL* and differentiated for 7 d under T_H1-polarizing conditions and observed that these cells increased IFN-γ production (Fig. 5i–l), suggesting that downregulation of LINE1-containing transcript expression can enhance the T cell effector response.

Altogether, these results indicate that LINE1-containing transcripts modulate the switch from quiescence to activation in naive CD4+ T cells, controlling expression of corresponding canonical transcripts in complex with nucleolin via H3K36me3 chromatin remodeling.

PTBP1 downregulates LINE1-containing transcripts, and GTF2F1 upregulates canonical transcripts. How is LINE1-containing transcript expression downregulated in activated CD4+ T cells? Several heteromeric RNA-binding proteins, such as PTBP1 and MATR3, have been reported to bind intronic LINE1 elements, influencing their lineage-specific splicing[34]. We speculated that a similar mechanism could be responsible for downregulation of LINE1-containing transcript expression in activated CD4+ T cells. Consistently, we found that the PTBP1

Fig. 5 | LINE1-containing transcripts in complex with nucleolin control gene expression, hampering H3K36me3 deposition in quiescent CD4+ T cells. **a**, Scheme of *NCL*-knockdown experiments in naive CD4+ T cells. **b**, Expression of LINE1-containing transcripts and canonical transcripts measured by RT–qPCR in quiescent naive CD4+ T cells treated with NCL or control (scr) ASOs ($n=3$ individuals). Data represent mean ± s.e.m. Canonical transcripts in scr versus NCL ASO groups, $^{***}P=0.00006$, $F=36.2$, two-way ANOVA. **c**, Expression of LINE1 in the cytosol, nucleoplasm and chromatin of naive CD4+ T cells treated with NCL or scr ASOs ($n=3$ individuals). Data represent mean ± s.e.m. LINE1 RNA species in scr versus NCL ASO groups, $^{*}P=0.0345$, $F=8.772$, two-way ANOVA. **d**, Scheme of LINE1 transcript knockdown in naive CD4+ T cells. **e**, Expression of LINE1-containing transcripts and canonical transcripts measured by RT–qPCR in quiescent naive CD4+ T cells treated with LINE1 or scr ASOs ($n=3$ individuals). LINE1 ASOs target the ORF2 LINE1 region of LINE1-containing transcripts (Extended Data Fig. 4b). Data represent mean ± s.e.m. LINE1-containing transcripts in scr versus LINE1 ASO groups, $^{***}P=7\times10^{-6}$, $F=56.4$, two-way ANOVA; canonical transcripts in scr versus LINE1 ASO groups, $^{***}P=0.00006$, $F=35.3$, two-way ANOVA. **f**, Quantitative western blot of H3K36me3, H3K4me3, H3K9me3 and H3K27me3 levels in quiescent naive CD4+ T cells treated with LINE1 or scr ASOs ($n=3$ individuals). Data represent mean ± s.e.m. H3K36me3 levels in scr versus LINE1 ASO groups, $^{*}P=0.0495$, two-tailed paired *t*-test. **g**,**h**, Positional distribution of the median of H3K36me3 ChIP–seq signals (ChIP/input fold enrichment) on gene bodies of LINE1-containing genes or control genes in naive CD4+ T cells treated with LINE1 (**g**) or scr ASOs and in quiescent or activated naive CD4+ T cells (**h**). **i**,**j**, Scheme of LINE1-containing transcripts-knockdown (**i**) or *NCL*-knockdown experiments (**j**) in effector CD4+ T cells. **k**,**l**, IFN-γ-positive cells in effector CD4+ T cells treated with LINE1 or scr ASOs ($n=8$ individuals) or with NCL or scr ASOs ($n=4$ individuals). Data represent mean ± s.e.m. IFN-γ+ cells in scr versus LINE1 ASO groups, $^{***}P=0.0002$, two-tailed paired *t*-test; IFN-γ+ cells in scr versus NCL ASO groups, $^{**}P=0.0041$, two-tailed paired *t*-test.

binding motif was enriched in LINE1 exons (Supplementary Table 10). Moreover, as we reported that LINE1 downregulation is under the control of mTORC1 (Fig. 1), we compared the Attig et al.[34] dataset with that of Hsu et al.[35]; the latter thoroughly describes the proteins regulated by mTORC1. We identified only one protein, GTF2F1, a transcription elongation factor that binds



intronic LINE1 and is also regulated by mTORC1, to be phosphorylated upon T cell activation[36]. Thus, we investigated the role of PTBP1 and GTF2F1 in the regulation of LINE1-containing transcripts in activated CD4+ T cells. We performed RIP experiments with PTBP1 and GTF2F1, demonstrating that these two factors bind LINE1 exons specifically in activated CD4+ T cells (Fig. 6a–d). In detail, as exemplified for *RAB22A*, PTBP1 binds only pre-mRNA, in line with its splicing-suppressive role, whereas GTF2F1 binds both pre-mRNA and the spliced canonical transcript, as expected for a transcriptional elongating factor (Fig. 6e–g). When we depleted PTBP1 or GTF2F1 with ASOs in naive CD4+ T cells and then activated them, we found that LINE1-containing transcript expression was specifically upregulated in the absence of PTBP1, while canonical transcripts were less transcribed, knocking down the two factors (Fig. 6h,i and Extended Data Fig. 8e–h). These results demonstrate that LINE1-containing transcript expression is downregulated in T cell activation through a splicing-suppression mechanism that occurs to promote expression of canonical transcripts in activated T cells.

LINE1-containing transcripts are re-expressed in exhausted T cells. Given that LINE1-containing transcripts regulate T cell effector function (Fig. 5i–l), we explored their dynamics and role in dysfunctional CD4+ and CD8+ T cells, exhausted by repetitive anti-CD3 stimulation[37]. As expected, exhausted T cells displayed growth arrest, programmed cell death protein (PD)-1 upregulation and reduced effector cytokine secretion (Extended Data Fig. 9a,d). We observed a consistent accumulation of LINE1-containing transcripts in the nuclei of exhausted CD4+ T cells but notably also in those of CD8+ T cells, and the corresponding canonical transcripts were downregulated (Fig. 7a–d).

These results suggest that the mechanism leading to LINE1-containing transcript downregulation upon CD4+ T cell activation could be reverted in exhausted T cells. Thus, we analyzed protein levels of IRF4, nucleolin, GTF2F1 and PTBP1 by western blotting, finding that both CD4+ and CD8+ exhausted T cells expressed high levels of IRF4 and nucleolin and downregulated GTF2F1 expression (Extended Data Fig. 9e). Moreover, we found by RIP assays that, in exhausted T cells, LINE1-containing transcripts were bound by nucleolin, while GTF2F1 and PTBP1 binding was reduced compared to that in functional effector cells (Fig. 7e and Extended Data Fig. 9f). Finally, we depleted exhausted T cells of *IRF4* or *NCL* and retrieved results similar to what was observed in naive CD4+ T cells: (1) expression of LINE1-containing transcripts and canonical transcripts was downregulated by *IRF4* depletion (Fig. 7f,g and Extended Data Fig. 9g); and (2) canonical transcript expression was upregulated, and LINE1-containing transcript levels remained unchanged upon *NCL* knockdown (Fig. 7h, i and Extended Data Fig. 9h). Therefore, we concluded that exhausted T cells re-express LINE1-containing transcripts via IRF4, accumulate these transcripts at chromatin through nucleolin and lack the PTBP1–GTF2F1 splicing-suppression machinery that is necessary for their downregulation.

LINE1-containing transcripts suppress the effector function of TILs. Considering the results obtained in exhausted T cells, we asked whether our findings could be relevant in the establishment of the immunosuppressive tumor microenvironment, in which effector T cells are often rendered dysfunctional through poorly defined mechanisms[38,39]. Indeed, new immunotherapy approaches aim to revert the dysfunctional state of TILs to promote tumor clearance[40]. We evaluated LINE1 expression in memory CD4+ and CD8+ TILs isolated from colorectal cancer (CRC) and non-small cell lung cancer (NSCLC) samples and from the corresponding colon and lung non-tumoral adjacent tissues. We detected reproducible and uniform LINE1 re-expression in CD4+ and CD8+ TILs from all patients analyzed (Fig. 8a,b and Extended Data Fig. 10a) and hypothesized that LINE1-containing transcripts targeting by ASOs could improve TIL effector function.

Thus, we isolated intratumoral CD4+ and CD8+ TILs, knocked down LINE1-containing transcripts (Extended Data Fig. 10b,c) and measured the expression of co-inhibitory receptors, the production of effector cytokines and cytotoxic capacity (Fig. 8c). Of note, we found that the frequency of PD-1-, LAG-3- and TIM3-positive cells was reduced (Fig. 8d,e) and the production of effector cytokines increased both in CD4+ and CD8+ TILs depleted of LINE1-containing transcripts (Fig. 8f,g). Concordantly, LINE1-containing transcript depletion significantly boosted the ability of CD4+ and CD8+ T cells to kill heterologous monocytes (Fig. 8h,i). Similar results were also observed in exhausted CD4+ and CD8+ T cells depleted of LINE1-containing transcripts, as revealed by increased effector cytokine production and enhanced killing ability (Extended Data Fig. 10d–j). However, the proliferation capacity of exhausted T cells depleted of LINE1-containing transcripts was not restored (Extended Data Fig. 10k,l). Therefore, we have characterized a mechanism of gene expression control impaired in dysfunctional T cells (exhausted cells or TILs) that could be efficiently targeted to restore the effector response.

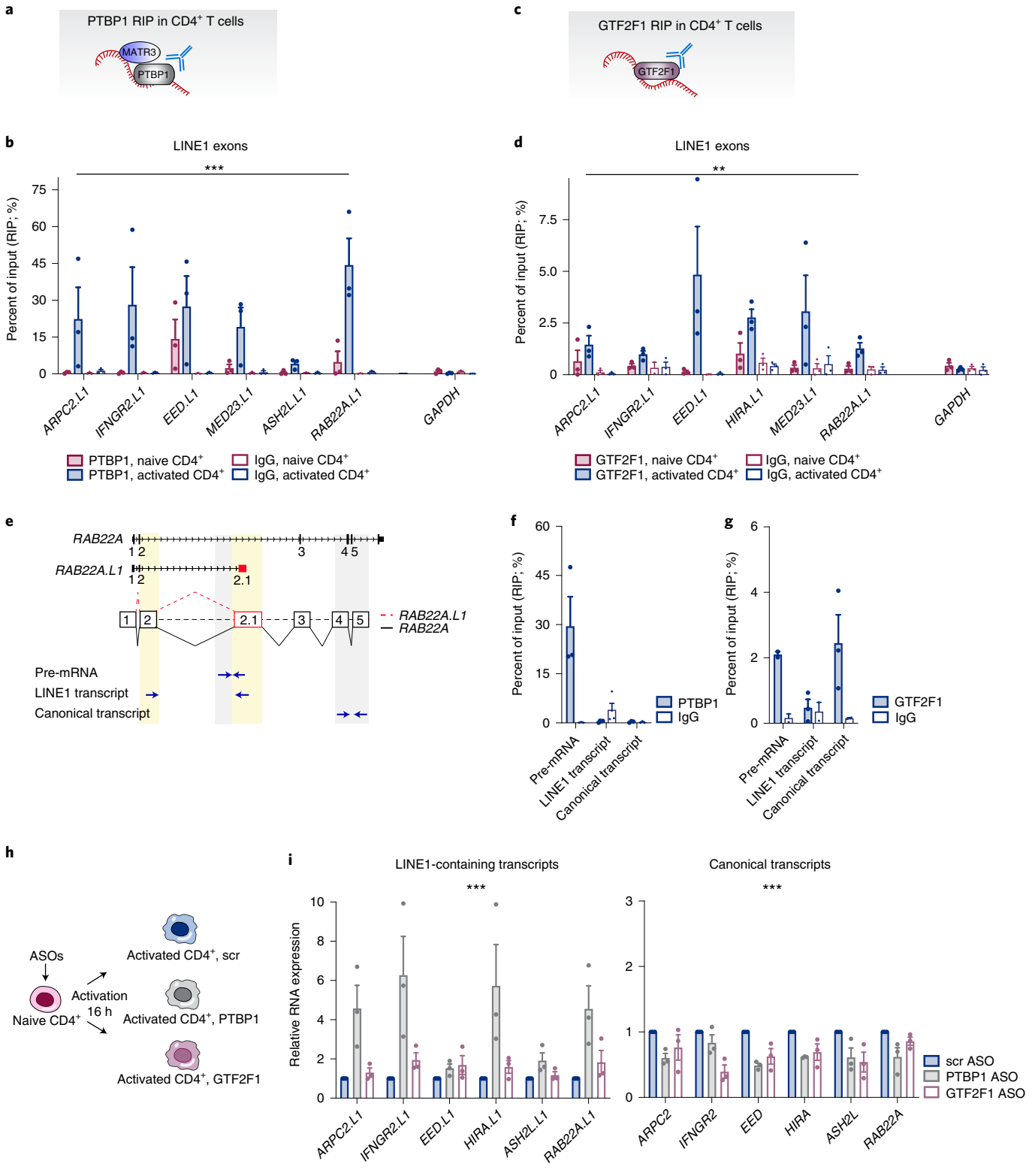
Discussion

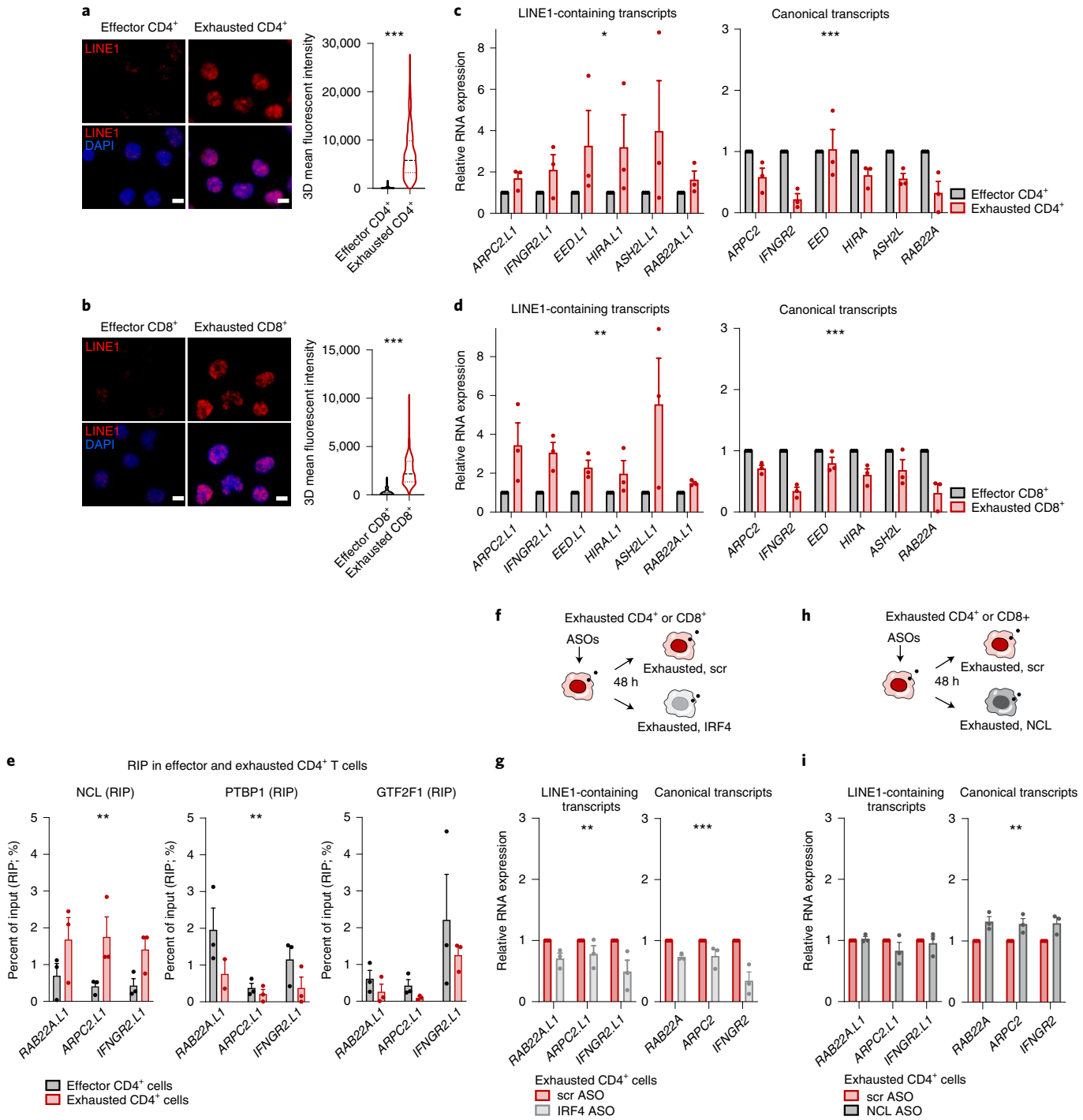
In this work, we report that RNA transcribed from LINE1 elements is involved in maintenance of the quiescent state of naive CD4+ T cells. Although LINE1 elements are the most abundant class of TEs, LINE1 elements expressed in T cells represent only about 0.1–1% of total LINE1 genomic insertions. These elements belong to L1M families, are ORF2-truncated regions (retrotransposition incompetent) and are spliced as exons of non-canonical transcript variants of genes involved in T cell activation. This suggests a diversification of molecular functions acquired by specific

Fig. 6 | Upon T cell activation, LINE1-containing transcripts are suppressed by PTBP1, while GTF2F1 favors expression of canonical transcripts. **a**–**d**, RIP assays in quiescent and activated naive CD4+ T cells with anti-PTBP1 (**a**,**b**) and anti-GTF2F1 (**c**,**d**) antibodies ($n=3$ individuals). Data represent mean percent of input ± s.e.m. PTBP1 RIP in naive versus activated CD4+ T cells, $^{***}P=0.0002$, $F=27.57$, two-way ANOVA. GTF2F1 RIP in naive versus activated CD4+ T cells, $^{**}P=0.00168$, $F=16.2$, two-way ANOVA. IgG, immunoglobulin G. **e**, Scheme of RT–qPCR assays of *RAB22A* transcripts to detect PTBP1 or GTF2F1 binding with pre-mRNA, LINE1-containing transcripts or canonical transcripts. **f**,**g**, *RAB22A* transcripts were amplified by RT–qPCR in PTBP1 (**f**) or GTF2F1 (**g**) RIP assays performed on activated CD4+ T cells ($n=3$ individuals). Data represent mean percent of input ± s.e.m. **h**, Scheme of *PTBP1*- or *GTF2F1*-knockdown experiments in activated CD4+ T cells. **i**, Expression of LINE1-containing transcripts and canonical transcripts measured by RT–qPCR in naive CD4+ T cells treated with PTBP1 or GTF2F1 or control (scr) ASOs and then activated for 16 h ($n=3$ individuals). Data represent mean ± s.e.m. LINE1-containing transcripts in scr versus PTBP1 versus GTF2F1 ASO groups, $^{***}P=0.0003$, $F=22.31$, two-way ANOVA; canonical transcripts in scr versus PTBP1 versus GTF2F1 groups, $^{***}P=2.5\times10^{-7}$, two-way ANOVA, $F=35.44$; LINE1-containing transcripts in scr versus PTBP1 groups, $^{***}P=0.00013$, $F=30.38$, two-way ANOVA; canonical transcripts in scr versus PTBP1 groups, $^{***}P=1.04\times10^{-6}$, $F=81.85$, two-way ANOVA; LINE1-containing transcripts in scr versus GTF2F1 ASO groups, $^{**}P=0.00328$, $F=13.37$, two-way ANOVA; canonical transcripts in scr versus GTF2F1 ASO groups, $^{***}P=0.000023$, $F=44.18$, two-way ANOVA.

TEs during evolution and represents an additional layer of transcriptional complexity in human cells.

We described in detail the biogenesis and function of LINE1-containing transcripts. In quiescent naive T cells, these transcripts are expressed under the control of the IRF4 transcription factor. They act at the chromatin level to pause expression of genes from which they originate, hampering the deposition of H3K36me3. In activated cells, LINE1-containing transcript expression is downregulated by PTBP1–GTF2F1, thus involving chromatin remodeling and splicing suppression, acting in *cis* to promote expression of canonical transcripts (Extended Data Fig. 10m). Hence, in activated cells, despite higher IRF4 levels, the mTORC1 pathway, which is turned on, alters the splicing pattern of LINE1-containing genes and removes the fraction of transcripts containing LINE1 exons. These





dynamics promote expression of canonical transcripts. We have demonstrated that LINE1-containing transcripts are maintained at chromatin in complex with nucleolin. The LINE1 RNA–nucleolin partnership has been already reported to regulate cell identity in mESCs, suggesting that this mechanism is partially evolutionarily conserved, although it involves different LINE1 families (evolutionarily young in mice and old in humans), in line with TE specialization throughout evolution. LINE1-containing transcripts directly control expression of their corresponding genes, although we have noticed in combined DNA–RNA-FISH experiments that some LINE1-containing transcript dots do not colocalize in *cis*, suggesting that they might similarly regulate *trans* gene expression in naive T cells, a feature that has been already demonstrated for long non-coding RNA species[41]. The nucleolin–KAP1 complex is reported to regulate deposition of H3K9me2 and H3K9me3 repressive marks in mESCs[13], while we observed that LINE1-containing transcript knockdown mainly results in increased H3K36me3 levels. Thus, further studies will be required to investigate in-depth chromatin dynamics and regulation by LINE1-containing transcripts in quiescent naive T cells.

LINE1-containing transcripts are fundamental to control cell-activation genes in naive T cells, avoiding inappropriate T cell activation and modulating T cell effector function. On the other hand, we observed that the same mechanism is active in

Fig. 7 | IRF4 in cooperation with nucleolin cause re-accumulation of LINE1-containing transcripts in exhausted T cells. a,**b**, Confocal fluorescence microscopy images of LINE1 RNA-FISH (red) of effector and exhausted CD4$^+$ (**a**) or CD8$^+$ (**b**) T cells. Original magnification, 63×. Scale bar, 5 µm; quantification was performed on at least 100 nuclei. Effector versus exhausted CD4$^+$ cells, $^{***}P<1\times10^{-15}$, two-tailed Mann–Whitney test; effector versus exhausted CD8$^+$ cells, $^{***}P<1\times10^{-15}$, two-tailed Mann–Whitney test. **c**,**d**, Expression of LINE1-containing transcripts and canonical transcripts measured by RT–qPCR in effector and exhausted CD4$^+$ T cells (**c**) or CD8$^+$ T cells (**d**) ($n=3$ individuals). Data represent mean ± s.e.m. LINE1-containing transcript expression in effector versus exhausted CD4$^+$ cells, $^{*}P=0.0148$, $F=8.09$, two-way ANOVA; canonical transcript expression in effector versus exhausted CD4$^+$ cells, $^{***}P=4.7\times10^{-4}$, $F=38.1$, two-way ANOVA; LINE1-containing transcript expression in effector versus exhausted CD8$^+$ cells, $^{**}P=0.0011$, $F=17.9$, two-way ANOVA; canonical transcript expression in effector versus exhausted CD8$^+$ cells, $^{***}P=9.3\times10^{-7}$, $F=83.7$, two-way ANOVA. **e**, RIP assays for nucleolin (NCL), PTBP1 and GTF2F1 in effector and exhausted CD4$^+$ T cells ($n=3$ individuals). Data represent mean percent of input ± s.e.m. Nucleolin RIP in effector versus exhausted CD4$^+$ cells, $^{**}P=0.0052$, $F=18.31$, two-way ANOVA; PTBP1 RIP in effector versus exhausted CD4$^+$ cells, $^{**}P=0.0076$, $F=18.6$, two-way ANOVA (see also Extended Data Fig. 9f). **f**, Scheme of *IRF4*-knockdown experiments in exhausted CD4$^+$ and CD8$^+$ T cells. **g**, Expression of LINE1-containing transcripts and canonical transcripts measured by RT–qPCR in exhausted CD4$^+$ T cells treated with IRF4 or control (scr) ASOs ($n=3$ individuals). Data represent mean ± s.e.m. LINE1-containing transcript expression in scr versus IRF4 ASO groups, $^{**}P=0.0059$, $F=17.34$, two-way ANOVA; canonical transcript expression in scr versus IRF4 ASO groups, $^{***}P=0.0006$, $F=42.74$, two-way ANOVA (see also Extended Data Fig. 9g). **h**, Scheme of *NCL* knockdown in exhausted CD4$^+$ and CD8$^+$ T cells. **i**, Expression levels of LINE1-containing transcripts and canonical transcripts measured by RT–qPCR in exhausted CD4$^+$ T cells treated with NCL or scr ASOs ($n=3$ individuals). Data represent mean ± s.e.m. Canonical transcript expression in scr versus NCL ASO groups, $^{**}P=0.0011$, $F=34.11$, two-way ANOVA (see also Extended Data Fig. 9h).

dysfunctional T cells, in which the accumulation of LINE1-containing transcripts led to impairment of their effector function (Extended Data Fig. 10m). The regulatory mechanisms governing T cell exhaustion are still poorly understood; here we identify a transcriptional regulatory mechanism driven by LINE1-containing transcripts to control an exhaustion program in T cells. IRF4 has been described to retain a dual function, contributing either to the effector or exhaustion phenotype. This dualism depends on the microenvironment and on association with other transcription factors[42,43]. Further investigation will be needed to identify additional players that, in cooperation with IRF4, might promote LINE1-containing transcript re-expression in exhaustion conditions.

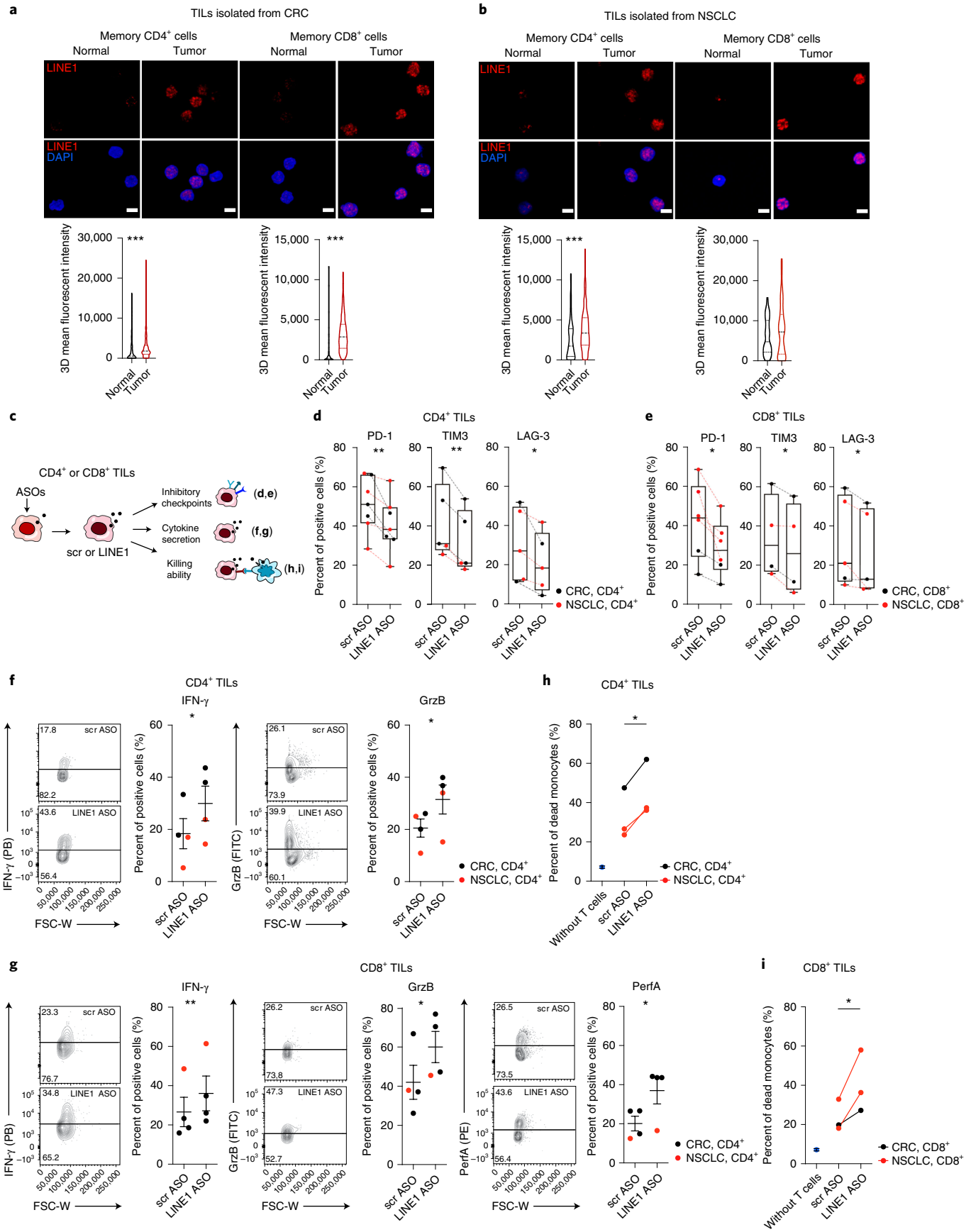
Notably, we show that the TIL effector response can be restored by depletion of LINE1-containing transcripts, indicating that their accumulation is a reversible mechanism, and that LINE1-driven dysfunction of effector T cells may be a potential target for therapeutic intervention. Therefore, we suggest that characterization of LINE1 dynamics in TILs, also at the single-cell level, will provide more insights into their potential role in shaping the transcriptional plasticity of specific TIL subsets.

In conclusion, this work establishes that regulatory evolution of abundant, repetitive LINE1 elements has been engaged in human T lymphocytes to allow a quick and effective switch from quiescence to activation and to regulate T effector function in physiological and pathological contexts. We surmise that the mechanism that we propose may be broadly applicable to other differentiated cells and tissues, serving as a general system to ensure molecular innovation and transcriptional plasticity.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00989-7.

Fig. 8 | LINE1-containing transcripts modulate the dysfunctional phenotype of CD4$^+$ and CD8$^+$ TILs. a, Top, confocal fluorescence microscopy images of LINE1 RNA-FISH (red) of memory CD4$^+$ and CD8$^+$ T cells infiltrating normal adjacent tissue or CRC tumors. Original magnification, 63×. Scale bar, 5 µm. Bottom, quantification of at least 100 nuclei from two patients. Normal versus tumor memory CD4$^+$ cells, $^{***}P=7.3\times10^{-8}$, two-tailed Mann–Whitney test; normal versus tumor memory CD8$^+$ cells, ***exact $P<1\times10^{-15}$, two-tailed Mann–Whitney test. **b**, Top, confocal fluorescence microscopy images of LINE1 RNA-FISH (red) of memory CD4$^+$ and CD8$^+$ T cells infiltrating normal adjacent tissue or NSCLC tumors. Original magnification, 63×. Scale bar, 5 µm. Bottom, quantification of at least 84 nuclei from at least two patients. Normal versus tumor memory CD4$^+$ cells, ***exact $P=3.7\times10^{-4}$, two-tailed Mann–Whitney test. **c**, Scheme of immunological assays performed on memory CD4$^+$ and CD8$^+$ TILs treated with LINE1 or control (scr) ASOs. **d**, PD-1-, TIM3- or LAG-3-positive memory CD4$^+$ TILs isolated from CRC (black, $n=3$ individuals for PD-1 and TIM3; $n=2$ for LAG-3) or NSCLC (red, $n=4$ individuals for PD-1; $n=2$ for TIM3; $n=3$ for LAG-3) samples. PD-1$^+$ cells in scr versus LINE1 ASO groups, $^{**}P=0.0044$, two-tailed paired *t*-test; TIM3$^+$ cells in scr versus LINE1 ASO groups, $^{**}P=0.0016$, two-tailed paired *t*-test; LAG-3$^+$ cells in scr versus LINE1 ASO groups, $^{*}P=0.0442$, two-tailed paired *t*-test. For box-and-whisker plots, the central line, the box and whiskers represent the median, the IQR from first to third quartiles and 1.5 × IQR, respectively. **e**, PD-1-, TIM3- or LAG-3-positive memory CD8$^+$ TILs isolated from CRC (black, $n=2$ individuals) or NSCLC (red, $n=4$ individuals for PD-1; $n=2$ for TIM3 and LAG-3) samples. PD-1$^+$ cells in scr versus LINE1 ASO groups, $^{*}P=0.02676$, two-tailed paired *t*-test; TIM3$^+$ cells in scr versus LINE1 ASO groups, $^{*}P=0.02707$, one-tailed paired *t*-test; LAG-3$^+$ cells in scr versus LINE1 ASO groups, $^{*}P=0.03166$, one-tailed paired *t*-test. For box-and-whisker plots, the central line, the box and whiskers represent the median, the IQR from first to third quartiles and 1.5 × IQR, respectively. **f**, IFN-γ- or granzyme (Grz)B-positive memory CD4$^+$ TILs isolated from CRC (black, $n=2$ individuals) or NSCLC (red, $n=2$ individuals) samples. Data represent mean ± s.e.m. IFN-γ$^+$ cells in scr versus LINE1 ASO groups, $^{*}P=0.04795$, one-tailed paired *t*-test; GrzB$^+$ cells in scr versus LINE1 ASO groups, $^{*}P=0.0277$, two-tailed paired *t*-test. **g**, IFN-γ-, GrzB- or perforin (Perf)A-positive memory CD8$^+$ TILs isolated from CRC (black, $n=3$ individuals) or NSCLC (red, $n=1$ individual) samples. Data represent mean ± s.e.m. IFN-γ$^+$ cells in scr versus LINE1 ASO groups, $^{**}P=0.0095$, two-tailed paired *t*-test; GrzB$^+$ cells in scr versus LINE1 ASO groups, $^{*}P=0.0275$, one-tailed paired *t*-test; PerfA$^+$ cells in scr versus LINE1 ASO groups, $^{*}P=0.02225$, one-tailed paired *t*-test. **h**,**i**, Percentage of dead heterologous monocytes co-cultured for 12 h with memory CD4$^+$ (**h**) or CD8$^+$ (**i**) TILs from CRC (black, $n=1$ individual) or NSCLC (red, $n=2$ individuals) samples. Percentage of monocyte self-lysis is indicated (without T cells, in blue). CD4$^+$ cells, scr versus LINE1 ASO groups, $^{*}P=0.0150$, two-tailed paired *t*-test; CD8$^+$ cells, scr versus LINE1 ASO groups, $^{*}P=0.0408$, two-tailed paired *t*-test.


a
TILs isolated from CRC
Memory CD4+ cells
Memory CD8+ cells
Normal
Tumor
LINE1
LINE1 DAPI
3D mean fluorescent intensity
b
TILs isolated from NSCLC
c
CD4+ or CD8+ TILs
ASOs
scr or LINE1
Inhibitory checkpoints
Cytokine secretion
Killing ability
(d,e)
(f,g)
(h,i)
d
CD4+ TILs
PD-1
TIM3
LAG-3
Percent of positive cells (%)
scr ASO
LINE1 ASO
CRC, CD4+
NSCLC, CD4+
e
CD8+ TILs
CRC, CD8+
NSCLC, CD8+
f
IFN-γ
GrzB
IFN-γ (PB)
GrzB (FITC)
FSC-W
g
PerfA
PerfA (PE)
h
Percent of dead monocytes (%)
Without T cells
i

Received: 9 December 2020; Accepted: 18 November 2021;
Published online: 17 January 2022

References

1. de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A. & Pollock, D. D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **7**, e1002384 (2011).
2. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
3. Lanciano, S. & Cristofari, G. Measuring and interpreting transposable element expression. *Nat. Rev. Genet.* **21**, 721–736 (2020).
4. Faulkner, G. J. et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* **41**, 563–571 (2009).
5. Marasca, F. et al. The sophisticated transcriptional response governed by transposable elements in human health and disease. *Int. J. Mol. Sci.* **21**, 32012 (2020).
6. Criscione, S. W. et al. Genome-wide characterization of human L1 antisense promoter-driven transcripts. *BMC Genomics* **17**, 463 (2016).
7. Denli, A. M. et al. Primate-specific ORF0 contributes to retrotransposon-mediated diversity. *Cell* **163**, 583–593 (2015).
8. Bourque, G. et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* **18**, 1752–1762 (2008).
9. Roy-Engel, A. M. et al. Human retroelements may introduce intragenic polyadenylation signals. *Cytogenet. Genome Res.* **110**, 365–371 (2005).
10. Belancio, V. P., Hedges, D. J. & Deininger, P. LINE-1 RNA splicing and influences on mammalian gene expression. *Nucleic Acids Res.* **34**, 1512–1521 (2006).
11. Fadloun, A. et al. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* **20**, 332–338 (2013).
12. Jachowicz, J. W. et al. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* **49**, 1502–1510 (2017).
13. Percharde, M. et al. A LINE1–nucleolin partnership regulates early development and ESC identity. *Cell* **174**, 391–405 (2018).
14. Lu, J. Y. et al. Genomic repeats categorize genes with distinct functions for orchestrated regulation. *Cell Rep.* **30**, 3296–3311 (2020).
15. Zhu, J., Yamane, H. & Paul, W. E. Differentiation of effector CD4 T cell populations*. *Annu. Rev. Immunol.* **28**, 445–489 (2010).
16. Luckheeram, R. V., Zhou, R., Verma, A. D. & Xia, B. CD4$^+$T cells: differentiation and functions. *Clin. Dev. Immunol.* **2012**, 925135 (2012).
17. DuPage, M. & Bluestone, J. A. Harnessing the plasticity of CD4$^+$ T cells to treat immune-mediated disease. *Nat. Rev. Immunol.* **16**, 149–163 (2016).
18. Hwang, S. S. et al. mRNA destabilization by BTG1 and BTG2 maintains T cell quiescence. *Science* **367**, 1255–1260 (2020).
19. Ricciardi, S. et al. The translational machinery of human CD4$^+$ T cells is poised for activation and controls the switch from quiescence to metabolic remodeling. *Cell Metab.* **28**, 895–906 (2018).
20. Feng, X. et al. Transcription factor Foxp1 exerts essential cell-intrinsic regulation of the quiescence of naive T cells. *Nat. Immunol.* **12**, 544–550 (2011).
21. Ouyang, W., Beckett, O., Flavell, R. A. & Li, M. O. An essential role of the Forkhead-box transcription factor Foxo1 in control of T cell homeostasis and tolerance. *Immunity* **30**, 358–371 (2009).
22. Moir, L. M. Lymphangioleiomyomatosis: current understanding and potential treatments. *Pharmacol. Ther.* **158**, 114–124 (2016).
23. Sehgal, S. N. Sirolimus: its discovery, biological properties, and mechanism of action. *Transplant. Proc.* **35**, 7S–14S (2003).
24. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
25. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
26. Kolasinska-Zwierz, P. et al. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* **41**, 376–381 (2009).
27. Spies, N., Nielsen, C. B., Padgett, R. A. & Burge, C. B. Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* **36**, 245–254 (2009).
28. Howden, A. J. M. et al. Quantitative analysis of T cell proteomes and environmental sensors during T cell differentiation. *Nat. Immunol.* **20**, 1542–1554 (2019).
29. Henriksson, J. et al. Genome-wide CRISPR screens in T helper cells reveal pervasive crosstalk between activation and differentiation. *Cell* **176**, 882–896 (2019).
30. Huber, M. & Lohoff, M. IRF4 at the crossroads of effector T-cell fate decision. *Eur. J. Immunol.* **44**, 1886–1895 (2014).
31. Hultquist, J. F. et al. CRISPR–Cas9 genome engineering of primary CD4$^+$ T cells for the interrogation of HIV–host factor interactions. *Nat. Protoc.* **14**, 1–27 (2019).
32. Peddigari, S., Li, P. W., Rabe, J. L. & Martin, S. L. hnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition. *Nucleic Acids Res.* **41**, 575–585 (2013).
33. Bodega, B. et al. A cytosolic Ezh1 isoform modulates a PRC2–Ezh1 epigenetic adaptive response in postmitotic cells. *Nat. Struct. Mol. Biol.* **24**, 444–452 (2017).
34. Attig, J. et al. Heteromeric RNP assembly at LINEs controls lineage-specific RNA processing. *Cell* **174**, 1067–1081 (2018).
35. Hsu, P. P. et al. The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. *Science* **332**, 1317–1322 (2011).
36. Tan, H. et al. Integrative proteomics and phosphoproteomics profiling reveals dynamic signaling networks and bioenergetics pathways underlying T cell activation. *Immunity* **46**, 488–503 (2017).
37. Dunsford, L. S., Thoirs, R. H., Rathbone, E. & Patakas, A. A human in vitro T cell exhaustion model for assessing immuno-oncology therapies. In *Immuno-Oncology* https://doi.org/10.1007/978-1-0716-0171-6_6 (Springer, 2020).
38. Jamal-Hanjani, M., Thanopoulou, E., Peggs, K. S., Quezada, S. A. & Swanton, C. Tumour heterogeneity and immune-modulation. *Curr. Opin. Pharmcol.* **13**, 497–503 (2013).
39. Franco, F., Jaccard, A., Romero, P., Yu, Y. R. & Ho, P. C. Metabolic and epigenetic regulation of T-cell exhaustion. *Nat. Metab.* **2**, 1001–1012 (2020).
40. Galon, J. & Bruni, D. Approaches to treat immune hot, altered and cold tumours with combination immunotherapies. *Nat. Rev. Drug Discov.* **18**, 197–218 (2019).
41. Carmona, S., Lin, B., Chou, T., Arroyo, K. & Sun, S. LncRNA *Jpx* induces *Xist* expression in mice using both *trans* and *cis* mechanisms. *PLoS Genet.* **14**, e1007378 (2018).
42. Man, K. et al. Transcription factor IRF4 promotes CD8$^+$ T cell exhaustion and limits the development of memory-like T cells during chronic infection. *Immunity* **47**, 1129–1141 (2017).
43. Seo, H. et al. BATF and IRF4 cooperate to counter exhaustion in tumor-infiltrating CAR T cells. *Nat. Immunol.* **22**, 983–995 (2021).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

Methods

Human blood and tissue samples. Blood from anonymous healthy donors was provided by the Fondazione Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS) Cà Granda Ospedale Maggiore Policlinico in Milan. Age and the sex of healthy donors were unknown (privacy reasons). Peripheral blood from patients with LAM was obtained from the Ospedale San Giuseppe-MultiMedica IRCCS in Milan. Peripheral blood from patients with kidney transplantation was obtained from the Fondazione IRCCS Cà Granda, Ospedale Maggiore Policlinico, Milan. CRC and NSCLC samples were provided by the European Institute of Oncology (IEO), non-tumoral samples were obtained from normal adjacent tissue at least 10 cm distal from the lesion; no patients received palliative surgery or neo-adjuvant chemotherapy and/or radiotherapy. Ethics committees of the hospitals approved the use of human samples for research purposes, and informed consent was obtained from all subjects (authorization nos. R807/18 IEO 849 and 708_2020). Available characteristics of patients are described in Supplementary Table 1.

T cell purification and sorting and monocyte purification. Human peripheral blood mononuclear cells (PBMCs) were purified from blood samples by density gradient centrifugation with Ficoll-Paque PLUS. T cells were negatively selected from PBMCs with a magnetic separator (autoMACS Pro Separator, Miltenyi Biotec) using the Pan T Cell Isolation kit (Miltenyi Biotec) or the CD4+ T Cell Isolation kit (Miltenyi Biotec). T cells were stained with antibodies specific for surface markers, and T cell subsets were sorted by flow cytometry as follows: naive CD4+ cells as $CD4^+CD25^-CD127^{-/hi}CD45RO^-$ cells, CD4+ T_H1 cells as $CD4^+CD25^-CD127^{-/hi}CD45RO^+CXCR3^+CCR6^-$ cells, CD4+ T_H2 cells as $CD4^+CD25^-CD127^{-/hi}CD45RO^+CRTH2^+$ cells and CD4+ T_H17 cells as $CD4^+CD25^-CD127^{-/hi}CD45RO^+CCR6^+CXCR3^-$ cells; naive CD8+ cells as $CD4^-CD8^+CD45RO^-$ cells and memory CD8+ cells as $CD4^-CD8^+CD45RO^+$ cells. Lymphocytes infiltrating normal and tumor tissues were obtained using a procedure indicated in the Supplementary Information. Viability of T cells was first assessed with the LIVE/DEAD Fixable Green Dead Cell Stain kit (Invitrogen by Life Technologies, L34969); live T cell subsets were then sorted by flow cytometry as follows: memory CD4+ cells as $CD45^+CD3^+CD4^+CD25^-CD127^{-/hi}CD45RO^+$ cells and memory CD8+ cells as $CD45^+CD3^+CD8^+CD45RO^+$ cells. The following antibodies were used for flow cytometry-based sorting: anti-CD4–APC-Cy7 (BD Biosciences, clone RPA-T4) or anti-CD4–VioGreen (Miltenyi Biotec, clone VIT4); anti-CD8–VioGreen (Miltenyi Biotec, clone REA734) or anti-CD8–VioBlue (Miltenyi Biotec, clone REA734); anti-CD25–PECy7 (Invitrogen by Life Technologies, clone BC96); anti-CD127–PE-Cy5 (BioLegend, clone A019D5) or anti-CD127–PE (Miltenyi Biotec, clone MB15-18C9); anti-CD45RO–BV605 (BioLegend, clone UCHL1) or anti-CD45RO–APC (Miltenyi Biotec, clone UCHL1); anti-CD3–PE (BD Biosciences, clone UCHT1); anti-CD45–Pacific Blue (BioLegend, clone 2D1); anti-CD183–PE-Cy5 (BD Biosciences, clone 1C6/CXCR3); anti-CD294(CRTH2)–APC-Vio770 (Miltenyi Biotec, clone REA598); anti-CCR6–FITC (BioLegend, clone G034E3). Staining were performed using 1 µl of each antibody for every 1×10^6 cells in 10 µl PBS for 30 min at 37 °C. Cell sorting was performed using the FACSAria III (BD Biosciences) using BD FACSDiva Software version 8.0.3. The purity of sorted cells was >97.5%. Monocytes were isolated from PBMCs by positive selection with a magnetic separator (autoMACS Pro Separator, Miltenyi Biotec) using CD14-specific microbeads (Miltenyi Biotec). Further details on CD4+ and CD8+ T cell differentiation in vitro, exhausted CD4+ and CD8+ T cells in vitro and T cell treatments are reported in the Supplementary Information.

Knockdown experiments. Knockdown experiments were performed using FANA (2′-deoxy-2′-fluoro-β-D-arabinonucleic acid, https://www.aumbiotech.com)–ASOs. For LINE1 RNA species, five ASOs were designed based on the ORF2 region of the LINE1 consensus sequence, while, for *HIRA.L1* or *RAB22A.L1*, three ASOs were designed based on a unique and specific portion of the sequence of LINE1-containing transcripts. For *GTF2F1*, *NCL* and *PTBP1* mRNA, four ASOs were used; and, for *IRF4* mRNA, two ASOs were used. An unrelated scr ASO was used as a control. ASOs were mixed in equimolar proportions and administered without any transfection reagent (by gymnosis) following the manufacturer's instruction at a final concentration of 10 µM.

Quiescent naive CD4+ T cells isolated from healthy donors were cultured for 48 h in complete medium supplemented with 200 IU ml^{-1} recombinant interleukin (IL)-2 and 10 µM ASOs to knock down LINE1, *IRF4*, *HIRA.L1*, *RAB22A.L1* or *NCL*. For LINE1- and *NCL*-knockdown experiments, naive CD4+ cells treated with ASOs were then activated with anti-CD3–anti-CD28 beads in T_H1 medium, kept in culture in the presence of 10 µM ASOs and collected after 7 d (effector CD4+ cells). For *GTF2F1*- and *PTBP1*-knockdown experiments, naive CD4+ cells treated with ASOs were then activated with anti-CD3–anti-CD28 beads in T_H1 medium, kept in culture in the presence of 10 µM ASOs and collected after 16 h (activated CD4+ cells). In vitro exhausted CD4+ and CD8+ T cells were treated starting from day 6 and collected after 48 h to knock down *IRF4* and *NCL*. Exhausted CD4+ and CD8+ T cells were treated starting from day 2 with 10 µM ASOs for 5–7 d to knock down LINE1. Memory CD4+ and CD8+ TILs isolated from tumor samples were cultured for 48 h in complete medium supplemented with 200 IU ml^{-1} recombinant IL-2 and 10 µM ASOs to knock down LINE1; after 48 h of ASO treatment, TILs were subjected to surface marker staining and T cell killing or activated for 48 h and subjected to intracellular cytokine staining. Knockdown efficiency was controlled by RT–qPCR and/or RNA-FISH and by western blot or flow cytometry analysis (described below).

T cell surface and intracellular staining and proliferation assay. Surface marker staining was performed by incubating 1 µl antibody for every 5×10^4 cells in PBS at 37 °C for 30 min. T cells were washed with PBS and then analyzed. The following antibodies were used: Alexa Fluor 488 anti-CD279 (PD-1) (BioLegend, clone EH12.2H7), anti-CD366 (TIM3-1)–BV650 (BioLegend, clone F38-2E2), anti-CD223 (LAG-3)–BV785 (BioLegend, clone 11C3C65) and anti-IFNGR2 (Ab224197, Abcam; this antibody was visualized with a conjugated secondary antibody, see below). For intracellular cytokine staining, 5×10^4 T cells were stimulated with 50 ng ml^{-1} phorbol 12-myristate 13-acetate and with 0.5 µg ml^{-1} ionomycin for 2 h at 37 °C; subsequently, 100 µg ml^{-1} brefeldin A (Merck) was added for an additional 2 h at 37 °C. Cells were washed, fixed and permeabilized for 30 min at 4 °C with the Foxp3 Transcription Factor Fixation/Permeabilization kit (Invitrogen by Life Technologies) according to the manufacturer's instructions. Cytokines were stained by incubating 1 µl antibody for every 5×10^4 cells diluted in Permeabilization Buffer (Invitrogen by Life Technologies) for 20 min at room temperature. T cells were washed with PBS and then analyzed. For cytokine staining, the following antibodies were used: anti-IFN-γ–V450 (clone B27), anti-GrzB–FITC (clone GB11), anti-PerfA–APC (clone deltaG9), anti-PerfA–PE (clone deltaG9). For PTBP1, IRF4 and ARPC2 staining by flow cytometry, T cells were fixed and permeabilized as described above for 30 min at 4 °C. Next, cells were incubated with 1 µl primary antibody for every 5×10^4 cells diluted in Permeabilization Buffer (Invitrogen by Life Technologies) for 1 h at room temperature. T cells were washed with PBS and stained with secondary antibody for 30 min at room temperature. T cells were washed with PBS and then analyzed. The following primary antibodies were used: anti-PTBP1 (Abcam, Ab133734), anti-IRF4 (BioLegend, 646412) and anti-ARCP2 (Ab133315, Abcam). Alexa Fluor 488 goat anti-rabbit (Invitrogen Life Technologies) and Alexa Fluor 647 goat anti-rat (Invitrogen Life Technologies) were used as secondary antibodies. Proliferation assays of chronically stimulated cells were performed using cell trace (C34557); naive CD4+ and CD8+ T cells were incubated with 1 µl cell trace for every 1×10^6 cells in PBS at 37 °C for 20 min. Cells were then washed with 10% FBS for 5 min at 37 °C and activated as reported previously; proliferation was assessed 5–7 d after activation. For all above-mentioned analyses, an average of 10^4 cells were acquired with the FACSCanto I (BD Biosciences) using BD FACSDiva Software version 8.0.3, and data were analyzed using FlowJo version 10.6.1 software.

Killing assay. In vitro exhausted effector CD4+ and CD8+ T cells or TILs were treated with ASOs and co-cultured for 12 h with heterologous monocytes at a 1:1 ratio. After co-culturing, cells were stained with the LIVE/DEAD Fixable Green Dead Cell Stain kit (Invitrogen by Life Technologies, L34969) for 20 min at room temperature, washed with PBS and stained with anti-CD14–APC antibody (clone M5E2) to recognize monocytes. Monocytes were identified as CD14-positive cells, and their viability was assessed as the percentage of dead monocytes. To assess the spontaneous lysis of monocytes, we cultured monocytes without T cells as a control. An average of 10^4 cells were acquired with the FACSCanto I (BD Biosciences) using BD FACSDiva Software version 8.0.3, and data were analyzed using FlowJo version 10.6.1 software.

RNA isolation and RT–qPCR. Total RNA was isolated using the RNeasy Mini kit (Qiagen) and the QIAshredder system (Qiagen) according to the manufacturer's instructions. During the extraction, DNase treatment with the RNase-free DNase Set (QIAGEN) was performed. Total RNA was reverse transcribed using the SuperScript III First-Strand Synthesis SuperMix kit (Invitrogen by Life Technologies) following the manufacturer's instructions. Real-time quantitative PCR was performed on the StepOnePlus Real-Time PCR System (Applied Biosystems by Life Technologies) using Power SYBR Green PCR Master Mix (Applied Biosystems by Life Technologies). All gene expression data were normalized to the expression of two independent housekeeping genes (*RNA18SN1* and *GAPDH*). Normalized Ct values were calculated as $2^{-\Delta Ct}$ or $2^{-\Delta\Delta Ct}$. For actinomycin D treatment, spike-in *Drosophila melanogaster* RNA was used for normalization. See Supplementary Table 11 for primer sequences.

RNA-FISH and RNA-FISH–immunofluorescence. RNA-FISH and RNA-FISH–immunofluorescence were performed as described in ref. [44]. Briefly, antisense biotinylated riboprobes for LINE1, AluY and HERVK were transcribed in vitro using the MAXIscript T7 transcription kit (Invitrogen) and Biotin RNA Labeling Mix (Roche) (see Supplementary Table 11 for primer sequences). In total, 50–100 ng antisense biotinylated riboprobes per experiment were used. T cells fixed with 3% paraformaldehyde were washed with 0.05% Triton X-100 in PBS, permeabilized with 0.5% Triton X-100 in PBS and maintained in 20% glycerol–PBS. Cells were frozen and thawed with dry ice and deproteinized with 0.1 N HCl. T cells were hybridized with riboprobes at 52.5 °C for 3.5 min and incubated overnight at 37 °C in a water bath. Glasses were washed with 50%

formamide–2× SSC (saline-sodium citrate), 2× SSC, 1× SSC and 4× SSC–0.2% Tween-20. T cells were blocked with BSA and then incubated with streptavidin HRP (1:1,000, PerkinElmer by Akoya Biosciences) diluted in TNT–BSA (0.1 M Tris-HCl, pH 8, 0.150 M NaCl, 0.1% NP-40 and 4% BSA in DEPC). T cells were washed four times with TNT, and the signal was amplified by incubating TSA working solution (1:150) in 1× amplification buffer for 3 min (TSA Plus Fluorescent kit Cy3.5, PerkinElmer). T cells were washed four times with TNT; nuclei were counterstained with 1 μg ml^{-1} DAPI. When RNA-FISH was coupled with immunofluorescence, T cells were incubated with primary antibodies specific for H3K4me3 (1:250, Millipore, 07-473), H3K36me3 (1:250, Abcam, 9050) or H3K9me3 (1:500, Abcam, Ab8898) in a solution of 2% BSA, 10% goat serum, 0.1% Tween and PBS overnight at 4 °C. A secondary antibody conjugated with Alexa Fluor 647 was used (1:1,000). Glasses were mounted in ProLong Diamond antifade mounting medium (Thermo Fisher). Images of RNA-FISH and RNA-FISH–immunofluorescence were acquired with a Leica TCS SP5 Confocal microscope (Leica Microsystems) using an HCX PL APO ×63, 1.40-NA oil-immersion objective (Leica Microsystems) with additional 2× zoom, using 0.29-μm *Z* stacks at randomly chosen fields. Supplementary Table 8 reports for each experiment the number of individuals and the number of nuclei acquired and analyzed and parameters of acquisition (laser power, scan speed, gain and offset). To obtain images comparable between different T cells isolated from the same individuals and/or within the same experiment, we performed the RNA-FISH protocol at the same time for all conditions and images were acquired with equivalent parameters and processed similarly; at least ~100 nuclei per condition were acquired. Further details on analysis of RNA-FISH and RNA-FISH–immunofluorescence are provided in the Supplementary Methods.

Single-molecule RNA-FISH protocol and analysis. smRNA-FISH was performed using HuluFISH technology. Antisense riboprobes were designed by Pixelbio based on specific and unique regions of *HIRA.L1* or *RAB22A.L1* LINE1-containing transcripts (Supplementary Table 11); anti-*RAB22A.L1* probes were directly labeled with ATTO-568, while anti-*HIRA.L1* probes were directly labeled with ATTO-647. Quiescent naive, 8-h activated CD4$^+$ T cells or naive CD4$^+$ T cells knocked down for *HIRA.L1* or *RAB22A.L1* transcripts were fixed in 4% paraformaldehyde, washed with 135 mM glycine and kept in 70% ethanol overnight. T cells were rinsed with 20% glycerol for 1 h and then treated with 0.025% pepsin in 0.01 N HCl for 3.5 min. *RAB22A.L1* probes were diluted 1:40 in a solution of 20% formamide, 2× SSC and 10% dextran sulfate, and *HIRA.L1* probes were diluted 1:40 in a solution of 10% formamide, 2× SSC and 10% dextran sulfate; probes were hybridized with T cells overnight at 37 °C in a water bath. Glasses were washed three times for 5 min in 10% formamide–2× SSC for *HIRA.L1* probes or three times for 5 min in 20% formamide–2× SSC and for 5 min with 2× SSC for *RAB22A.L1* probes; nuclei were counterstained with 1 μg ml^{-1} DAPI. Glasses were mounted in ProLong Glass antifade mounting medium (Thermo Fisher). Images of smRNA-FISH were acquired with the Eclipse Ti-E microscope (Nikon Instruments) and the Plan Apo λ objective microscope (×100 oil, Nikon Instruments) using 0.3-μm *Z* stacks at randomly chosen fields. Supplementary Table 8 reports for each experiment the number of individuals and the number of nuclei acquired and analyzed and parameters of acquisition (LED power and camera exposure time). To obtain comparable images between different T cells within the same individual and/or experiment, we performed the smRNA-FISH protocol at the same time for all conditions and images were acquired with equivalent parameters and processed similarly. At least ~100 nuclei were analyzed, and the number of dots per cell were counted. Imaging analysis was performed with Fiji 2.1.0 or 1.53C or ImageJ 1.50i and NIS-Elements AR Analysis 5.11.01 (Nikon).

DNA–RNA-FISH. The DNA–RNA-FISH protocol was adapted from ref. [45] and from refs. [44,46]. A detailed protocol is provided in the Supplementary Information.

CRISPR–Cas9 deletion of LINE1 elements in quiescent naive CD4$^+$ T cells. LINE1 element deletions were performed taking advantage of Cas9 ribonuclear complex (RNP) nucleofection with a protocol adapted from refs. [31,47]. Briefly, two sgRNA species were designed to target nonrepetitive flanking sites of the LINE1 element contained in *ARPC2.L1* or *IFNGR2* (Supplementary Table 11). Quiescent naive CD4$^+$ T cells were nucleofected with Cas9–sgRNA RNP complexes; for each sgRNA, a Cas9–sgRNA complex was prepared at a ratio of 1:3 by gently mixing 40 μM Alt-R S.p. HiFi Cas9 Nuclease V3 (IDT, 1081061) with 120 μM sgRNA (Merck); the two complexes were allowed to form separately for 15 min at 37 °C. Both Cas9–sgRNA complexes were then mixed and added to 1×10^6 naive CD4$^+$ T cells that were resuspended in 20 μl primary cell nucleofection solution (P3 Primary Cells 4D-Nucleofector X kit S, Lonza); quiescent naive CD4$^+$ T cells were previously sorted and maintained in culture for 24 h in complete medium supplemented with 200 IU ml^{-1} recombinant IL-2. Naive CD4$^+$ T cells were electroporated using a 4D nucleofector, the P3 Primary Cells 4D-Nucleofector X kit S (Lonza, LOV4XP3032), with the EH115 pulse program. After nucleofection, cells were resuspended in complete medium supplemented with 200 IU ml^{-1} recombinant IL-2 and kept in culture for 4 d at 37 °C in a humidified incubator with 5% CO_2. Deletion was assessed by PCR with GoTaq G2 Flexi DNA Polymerase of genomic DNA purified from nucleofected naive CD4$^+$ T cells. PCR products were subjected to TA cloning and Sanger sequencing similarly to the method in ref. [48]. Primer sequences and sgRNA species used are listed in Supplementary Table 11.

Western blotting. The histone-extraction protocol and subsequent western blot analysis were performed as described in ref. [33]. Nuclear protein extraction was performed as described in ref. [33]. Total protein extraction was performed as described in ref. [19]. Further details on protein extraction and western blotting analysis and co-immunoprecipitation are provided in the Supplementary Methods.

RNA library preparation and sequencing. Further details on chromatin and nucleoplasm RNA extraction are provided in the Supplementary Methods. RNA integrity was checked with the TapeStation system (High Sensitivity RNA ScreenTape assay), and 15–75 ng total RNA was used to prepare libraries. RNA was ribodepleted with the RiboGone-Mammalian kit (Takara, 634846) following the manufacturer's instruction, and libraries were prepared with the SMARTer Stranded RNA-Seq kit (Takara, 634836) according to the manufacturer's instructions. Libraries were sequenced as paired 100-bp or 150-bp reads on the Illumina NextSeq 500. RNA-seq libraries were prepared for (1) chromatin and nucleoplasm RNA from quiescent naive CD4$^+$ T cells (four individuals) and (2) total RNA from quiescent naive CD4$^+$ T cells and naive CD4$^+$ T cells activated with anti-CD3–anti-CD28 beads in T_H1 medium for 16 h (three individuals). Further details on processing and alignment of RNA-seq datasets, principal-component analysis, TE subfamily expression quantification in RNA-seq datasets and LINE1 read characterization are provided in the Supplementary Information.

De novo reconstruction of LINE1-containing transcripts. A comprehensive catalog of LINE1-containing transcripts in the quiescent naive CD4$^+$ T cell chromatin compartment was generated by combining two different approaches for de novo transcript assembly. Briefly, chromatin RNA-seq reads mapped in proper pairs (SAM flags 99, 147, 83 and 163) from four biological replicates were pooled together, amounting to a total of 113 million proper read pairs. To reconstruct transcripts containing TEs with greater confidence, two independent algorithms were used: Trinity 2.8.4 (ref. [24]) in genome-guided mode ('--SS_lib_type FR --genome_guided_bam --genome_guided_max_intron 10000 --genome_guided_min_reads_per_partition 3') in tandem with PASA 2.3.3 (ref. [49]) ('-C -R --ALT_SPLICE --ALIGNERS blat,gmap --CPU 1 --transcribed_is_aligned_orient') and StringTie 2.0 (ref. [25]) ('--rf -a 3'). Mono-exonic transcripts were removed from further analysis as already performed in refs. [50,51] to filter out possible artifactual transcripts due to transcriptional noise or low polymerase fidelity; furthermore, these transcripts are difficult to assess bioinformatically and need extensive manual curation. Multi-exonic transcripts intersecting with TEs (UCSC RepeatMasker) were selected. To obtain a new and consistent catalog of nonredundant transcripts, only those transcripts sharing the TE-containing exon (BEDtools 2.29.2, 'intersectBed -f 0.8 -r -s') identified by both assemblers were selected. A unified set of TE transcripts was obtained by merging the selected transcripts using StringTie 2.0 ('merge -i -f 0'). TE transcripts were annotated using GffCompare 0.11.2 against transcripts from the GENCODE version 25 GTF file. Finally, de novo reconstructed TE transcripts having at least 20 bp of overlap between the TE-containing exon and a LINE1 locus were annotated as LINE1-containing transcripts, retrieving 3,072 transcripts. Genes containing LINE1-containing transcripts within their genomic position are referred to as 'LINE1-containing genes'. Further details on expression quantification of LINE-containing transcripts, filtering of LINE1-containing transcripts and LINE1-containing transcript characterization are provided in the Supplementary Methods.

Gene expression quantification and controls in RNA-seq datasets. Aligned reads were used to generate read counts per gene using HTSeq 0.12.4 ('htseq-count -s yes --nonunique all') against GENCODE version 25 and normalized to FPKM values using the total number of reads mapping within the coordinates of gene models as the library size. Expression values of canonical transcripts were selected from all quantified genes. To demonstrate that canonical transcripts are particularly preferentially regulated in naive CD4$^+$ T cells, we evaluated distribution quantiles of the fold change of the naive versus activated condition ($\log_2$(FC of mean FPKM of activated cells/mean FPKM of naive T cells)) of canonical transcripts, considering very detailed probabilities (0, 0.02, 0.05, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 0.98, 1), and we applied these ranges to all genes expressed in T cells (12,675 genes). This approach allows selection of genes upregulated during TCR-driven activation similar to those of canonical transcripts. We then selected genes that did not include LINE1 insertions (3,250 genes) or that contained LINE1 insertions that differed from those that make LINE1-containing transcripts (7,642 genes), and we randomly selected three non-overlapping sets of 407 genes respectively according to their range of fold change probability distribution. For Ingenuity Pathway Analysis, refer to the Supplementary Information.

Nanopore sequencing data analysis. For nanopore cDNA library preparation and sequencing, refer to the Supplementary Methods. Nanopore cDNA signals were

processed into demultiplexed reads using Guppy basecalling software version 5.0.7 with parameters 'guppy_basecaller --flowcell FLO-MIN106 --kit SQK-PCB109 --barcode_kits SQK-PCB109 –trim_barcodes'. Reads from three biological replicates were aligned to the reference transcriptome containing GENECODE 25 data and reconstructed TE-containing transcripts (De novo reconstruction of LINE1-containing transcripts) using minimap2 version 2.17-r941 with the parameter '-ax map-ont'. The presence of LINE1-containing transcripts in the Nanopore data was tested by selecting transcripts that were uniquely aligned. Further details on LINE1-containing transcript validation by PCR are provided in the Supplementary Methods.

Chromatin immunoprecipitation. The ChIP assay was performed as described in ref. [52] with modifications. Further details are provided in the Supplementary Information.

ChIP–seq data analysis. Furthermore, H3K36me3 ChIP–seq data were generated to inspect the chromatin of LINE1-containing genes in T cell activation and in naive $CD4^+$ cells depleted of LINE1-containing transcripts. Reads from technical replicates were pooled together, and read quality before and after trimming was assessed using FastQC 0.11.9. Reads were trimmed for low-quality base calls using Trimmomatic 0.39 in paired-end mode with parameters 'ILLUMINACLIP:TruSeq3-PE.fa:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:50′ or in single-end mode with adaptor reference file 'TruSeq3-SE.fa' and the same parameters as above. Trimmed reads were aligned to the human genome assembly hg38 using Bowtie 1.2.3 with parameters '-m 1 --best --strata -v 3′ and '-X 2000 –fr' for paired-end reads only. After alignment, paired-end reads not mapped in a proper pair as well as duplicated reads were removed using SAMtools 1.9 (ref. [53]). ChIP peaks were called using the MACS 2.2.6 callpeak module, giving as input the alignment file of the ChIP target and its relative control input with parameters '--keep-dup all -g 3049315783 -B -p 0.01—broad', paired-end read-specific parameters '-f BAMPE' and single-end read-specific parameters '-f BAM --nomodel --extsize 200'. Coverage tracks were calculated by subtracting the background signal from the fragment pileup using the MACS2 bdgcmp module with parameters '-m FE'. As a control for LINE1-containing genes, a set of the same number of genes was randomly sampled from a pool of protein-coding genes marked by at least one H3K36me3 peak using the GNU coreutils built-in 'shuf' command. The positional distribution of H3K36me3 on LINE1-containing genes and control genes was obtained by dividing gene models into 40 bins, while −1.5 kb and +3 kb flanking regions were smoothed with 150-bp long bins using deepTools 3.4.1 computeMatrix with parameters '-m 6000 -b 3000 -a 3000 -bs 150', the average between replicates and the median between genes was calculated, and a cubic smoothing spline was fitted into the data using the R 3.6.2 built-in 'smooth.spline' function. Further details on ChIP–seq data analysis are provided in the Supplementary Information.

Motif-enrichment analysis. Transcription factor-binding motifs provided by the HOCOMOCO Human (version 11, CORE) database were searched for putative promoter sequences of LINE1-containing genes (for further details, please refer to the Supplementary Methods). RNA-binding protein motifs provided by Ray et al.[54] were searched for in LINE1 element sequences contained in LINE1-containing transcripts (Supplementary Table 10).

Statistical analysis. Statistical details including the number of individuals and nuclei acquired in imaging experiments, statistical methods used, *F* values and exact *P* values are reported in each figure legend, in the text or in the Supplementary Tables. Results presented in this article were obtained from a minimum of three individuals; exceptions are indicated in figure legends (experiments performed on exhausted T cells, TILs or patients with transplantation). Further details on statistical analysis are provided in the Supplementary Information.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data are available in the main text or in the Supplementary Information. Data that support the findings of this study have been deposited and are publicly available at ArrayExpress: E-MTAB-9572 and E-MTAB-10798 for ChIP–seq data, E-MTAB-9574 for short-read RNA-seq data and E-MTAB-10797 for long-read RNA-seq data. RNA-seq data from Bediaga et al.[55] can be download from https://doi.org/10.1038/s41598-020-80165-9. RNA-seq data from Buratin et al.[56] can be downloaded from https://doi.org/10.1182/bloodadvances.2020002337. RNA-seq data from mESCs from the ENCODE Project Consortium can be downloaded from https://doi.org/10.1038/nature11247. ChIP–seq data from the ENCODE Project Consortium can be downloaded from https://doi.org/10.1038/s41467-020-14743-w. All sequencing datasets are listed in Supplementary Table 12. Source data are provided with this paper.

References

44. Marasca, F., Cortesi, A. & Bodega, B. 3D COMBO chrRNA–DNA–ImmunoFISH. *Methods Mol. Biol.* **2157**, 281–297 (2021).
45. Shibayama, Y., Fanucchi, S. & Mhlanga, M. M. Visualization of enhancer-derived noncoding RNA. *Methods Mol. Biol.* **1468**, 19–32 (2017).
46. Cortesi, A. et al. 4q-D4Z4 chromatin architecture regulates the transcription of muscle atrophic genes in facioscapulohumeral muscular dystrophy. *Genome Res.* **29**, 883–895 (2019).
47. Seki, A. & Rutz, S. Optimized RNP transfection for highly efficient CRISPR/Cas9-mediated gene knockout in primary T cells. *J. Exp. Med.* **215**, 985–997 (2018).
48. Fontana, C. et al. Early maternal care restores LINE-1 methylation and enhances neurodevelopment in preterm infants. *BMC Med.* **19**, 42 (2021).
49. Haas, B. J. et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
50. Ramsay, L. et al. Conserved expression of transposon-derived non-coding transcripts in primate stem cells. *BMC Genomics* **18**, 214 (2017).
51. Ranzani, V. et al. The long intergenic noncoding RNA landscape of human lymphocytes highlights the regulation of T cell differentiation by linc-MAF-4. *Nat. Immunol.* **16**, 318–325 (2015).
52. Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP–seq for histones and transcription factors. *Nat. Methods* **12**, 963–965 (2015).
53. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
54. Ray, D. et al. A compendium of RNA-binding motifs for decoding gene regulation. *Nature* **499**, 172–177 (2013).
55. Bediaga, N. G. et al. Multi-level remodelling of chromatin underlying activation of human T cells. *Sci. Rep.* **11**, 528 (2021).
56. Buratin, A. et al. Large-scale circular RNA deregulation in T-ALL: unlocking unique ectopic expression of molecular subtypes. *Blood Adv.* **4**, 5902–5914 (2020).

Acknowledgements

We are grateful to A. Lanzavecchia, V. Costanzo, P. Della Bona and C. Lanzuolo for discussions and critical revision of the manuscript. We acknowledge the scientific and technical assistance of the INGM Imaging Facility, in particular, C. Cordiglieri and A. Fasciani (Istituto Nazionale di Genetica Molecolare 'Romeo ed Enrica Invernizzi' (INGM), Milan, Italy). This work has been supported by the following grants: the Fondazione Cariplo (Bando Giovani, grant no. 2018-0321 to F.M.), the Fondazione AIRC under 5 per Mille 2018 (ID, AIRC 5x1000 programs: project 21147 'ISM' and ID 21091 to S.A.), the Fondazione Regionale per la Ricerca Biomedica (FRRB CP2_12/2018) and the Fondazione Cariplo (grant no. 2019-3416 to B.B.).

Author contributions

F.M. designed and performed experiments, analyzed and interpreted data and wrote and finalized the manuscript. S. Sinha designed and performed bioinformatic experiments, analyzed data and revised the manuscript. R.V. set up, performed and analyzed experiments. B.P. set up the strategy for LINE1-containing transcript selection, performed RNA-seq, ChIP–seq and Nanopore long-read sequencing analyses. V.R. supervised bioinformatic analyses and revised the manuscript. E.M.P. performed NGS sequencing. F.V.B. and M.G. performed experimental validations. M.C. supervised flow cytometry sorting and immunological assays. M.L.N. performed ChIP experiments. S.C. provided technical support for TIL isolation from tissues. S.N. supervised immunological assays of TILs. A.S.-B. and S. Siena provided tumor samples. D.P. provided buffy coats. G.M. provided blood samples from patients with kidney transplantation. G.V. provided tumor samples. O.T. and S.H. provided blood samples from patients with LAM. R.G. provided technical support for TIL isolation from tissues. G.S. conceived the protocol for RNA-seq library preparation. S.B. conceived the study of mTORC1 signaling, analyzed experiments and revised the manuscript. B.B. and S.A. conceived and conceptualized the study, designed experiments, analyzed and interpret data and wrote and finalized the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

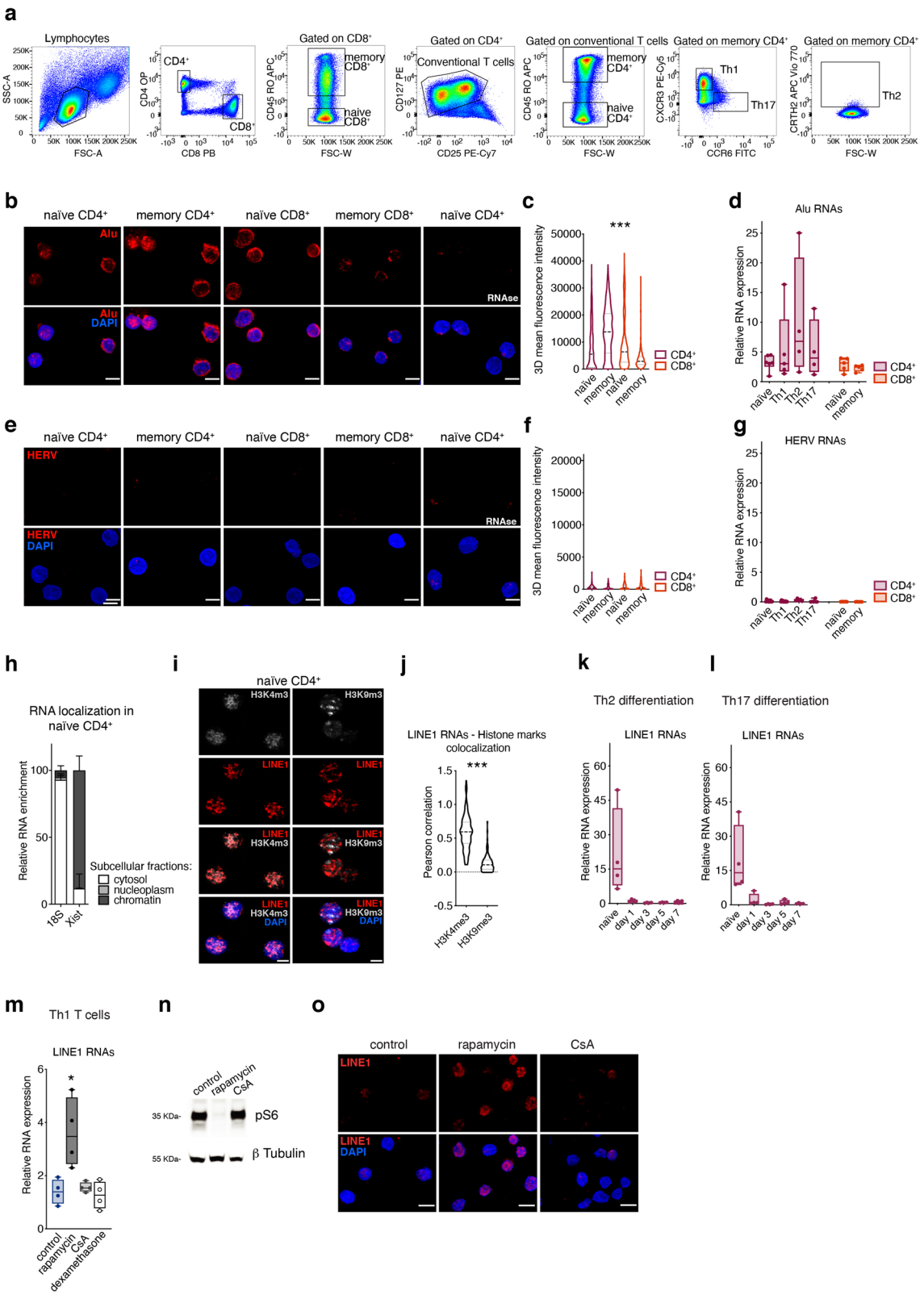
Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00989-7.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00989-7.

Correspondence and requests for materials should be addressed to Sergio Abrignani or Beatrice Bodega.

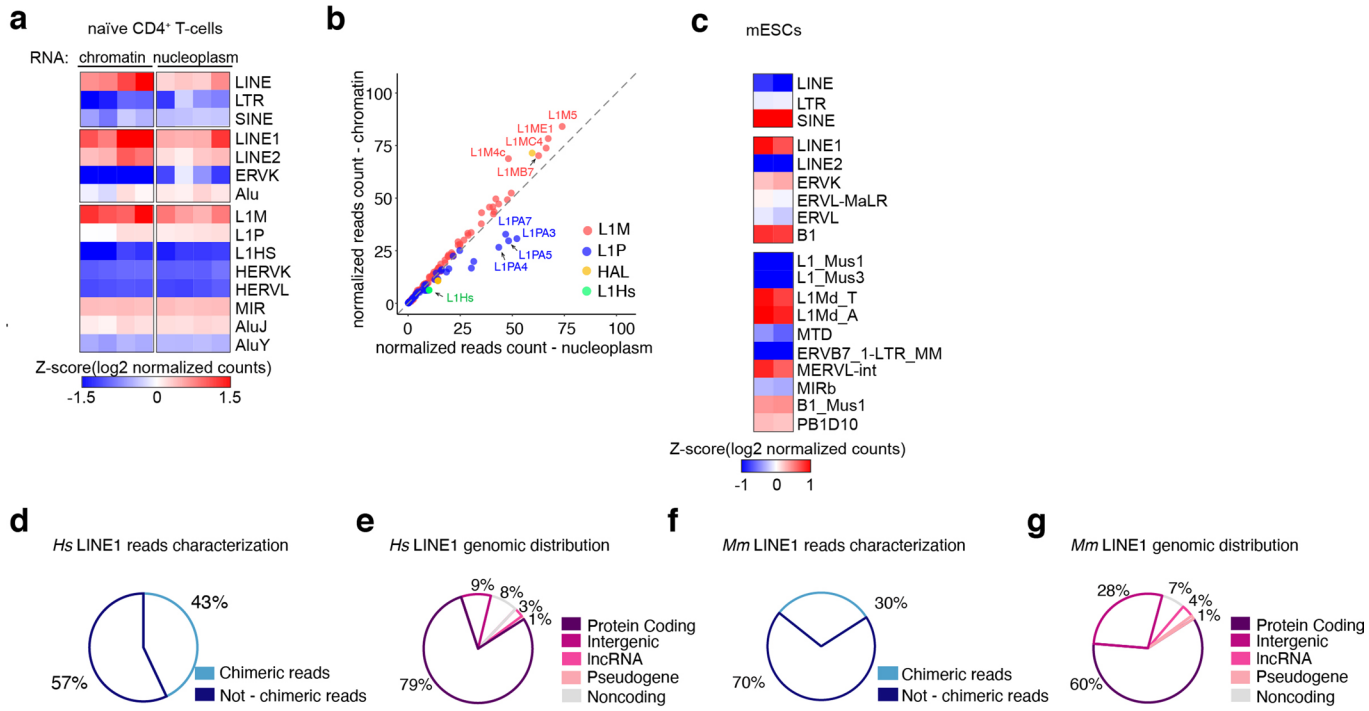
Peer review information *Nature Genetics* thanks Musa Mhlanga and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

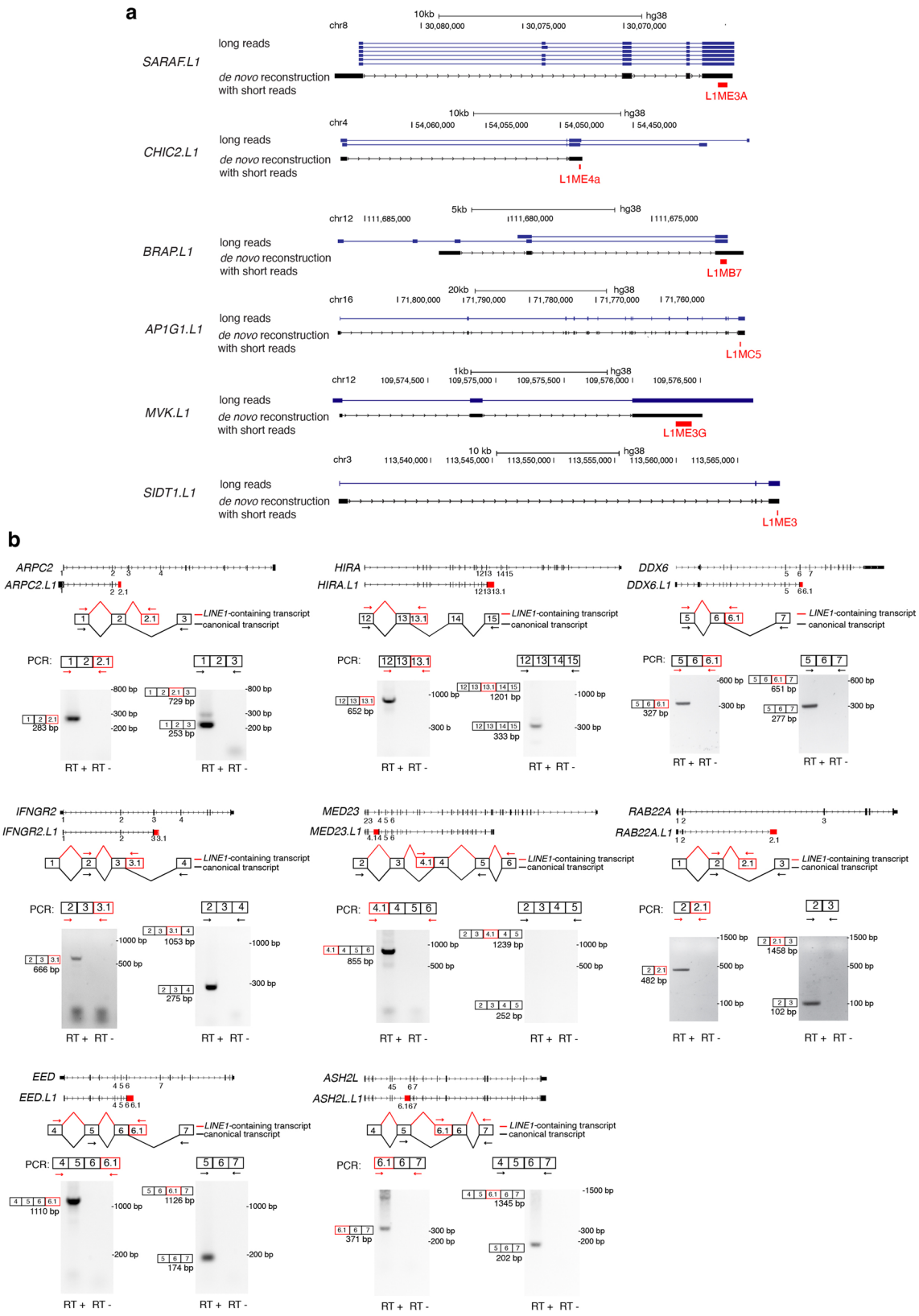


Extended Data Fig. 1 | See next page for caption.

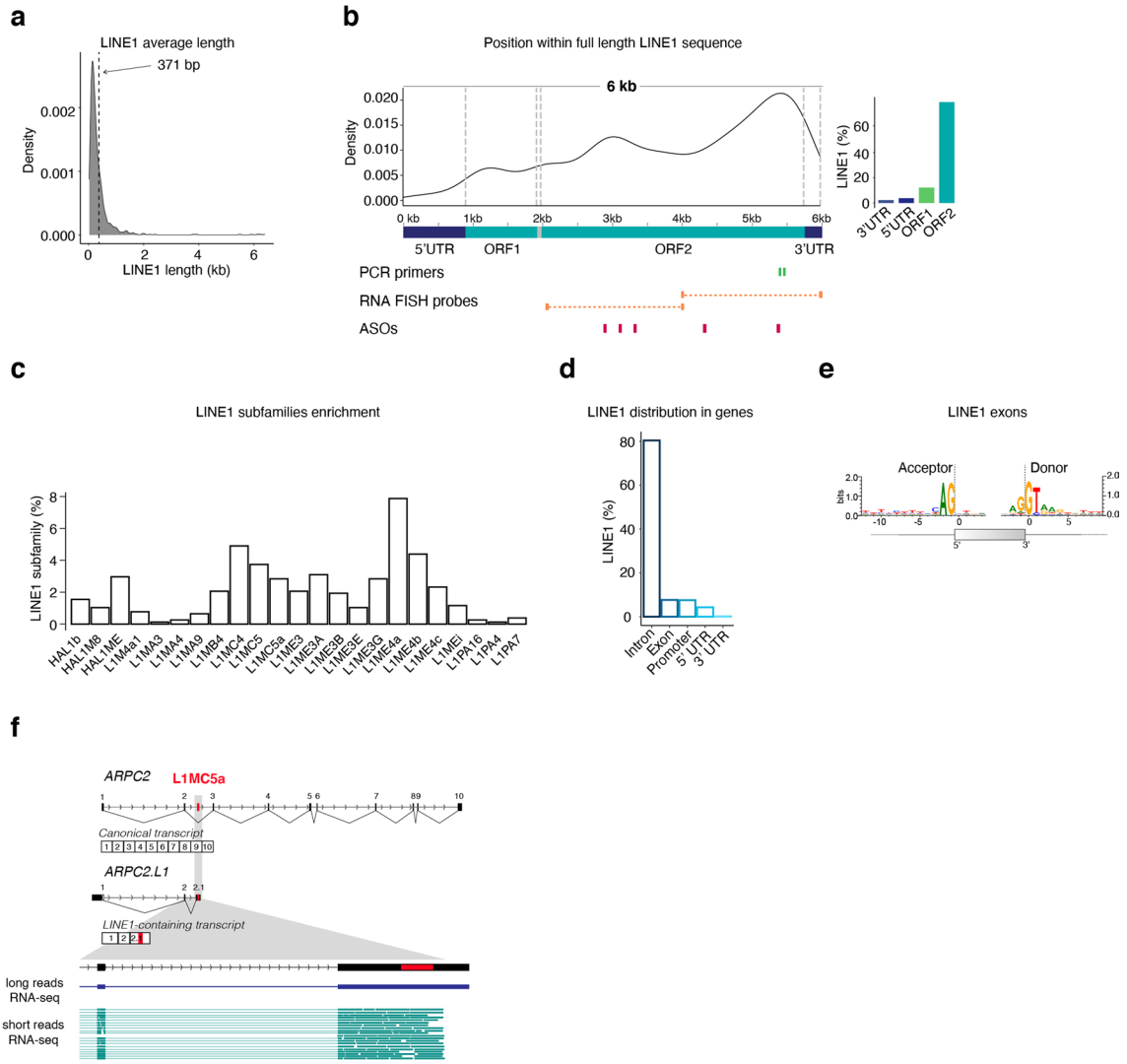
Extended Data Fig. 1 | LINE1 RNAs are enriched in open chromatin regions in naïve CD4⁺ T cells. (**a**) Sorting of naïve and memory CD4⁺ and CD8⁺ T cells, T_H1, T_H2 and T_H17 CD4⁺ T helper subsets from blood samples. (**b**) Confocal fluorescence microscopy images of Alu RNA FISH (red) of naïve and memory CD4⁺ and CD8⁺ T cells. Original magnification 63×. Scale bar 5 μm. (**c**) Quantification of Alu RNA FISH, at least 220 nuclei, four individuals. *** $P < 1 \times 10^{-15}$, One-way ANOVA, F = 47.5. (**d**) Alu expression by RT-qPCR in quiescent naïve and memory T_H1, T_H2 and T_H17 CD4⁺ T cells and in quiescent naïve and memory CD8⁺ T cells, (n = 6 individuals for naïve CD4⁺: n = 5 for T_H1 and CD8⁺ subsets, n = 4 for T_H2 and T_H17). For box-and-whisker plots, the central line, box and whiskers represent the median, interquartile range (IQR) from first to third quartiles, and 1.5 × IQR, respectively. (**e**) Confocal fluorescence microscopy images of HERV RNA FISH (red) performed on quiescent naïve and memory CD4⁺ and CD8⁺ T cells. Original magnification 63×. Scale bar 5 μm. (**f**) Quantification of HERV RNA FISH, at least 164 nuclei, three individuals. (**g**) HERV expression by RT-qPCR in quiescent naïve and memory T_H1, T_H2 and T_H17 CD4⁺ T cells and in quiescent naïve and memory CD8⁺ T cells, (n = 10 individuals for naïve and T_H1 CD4⁺, n = 6 T_H17; n = 5 for T_H2 and CD8⁺ subsets). (**h**) Expression of *18S* and *Xist*, in cytosol, nucleoplasm and chromatin of naïve CD4⁺ T cells (n = 3 individuals). Data are represented as mean ± s.e.m. (**i**) Confocal fluorescence microscopy images of LINE1 RNA FISH (red) and immunofluorescent staining (gray) for H3K4me3 and H3K9me3 of naïve CD4⁺ T cells. Original magnification 63×. Scale bar 5 μm. (**j**) Pearson correlation of colocalization relative to (**i**), at least 103 nuclei, three individuals. H3K9me3 versus H3K4me3 colocalization: *** $P < 1 \times 10^{-15}$, Two-tailed Mann Whitney test. (**k-l**) Naïve CD4⁺ T cells were activated with TCR engagement and differentiated to (**k**) T_H2 or (**l**) T_H17. LINE1 expression was analyzed by RT-qPCR, (n = 4 individuals). For box-and-whisker plots, the central line, box and whiskers represent the median, interquartile range (IQR) from first to third quartiles, and 1.5 × IQR, respectively. (**m**) LINE1 expression by RT-qPCR in 72 hours activated naive CD4⁺ T treated with signaling pathway inhibitors (n = 4 individuals). Control versus rapamycin: * P = 0.02678 Two-tailed paired *t* test. For box-and-whisker plots, the central line, box and whiskers represent the median, interquartile range (IQR) from first to third quartiles, and 1.5 × IQR, respectively. (**n**) Phosphorylated S6 protein (pS6) levels by western blot in 72 hours activated naïve CD4⁺ T treated with rapamycin or CsA. (**o**) Confocal fluorescence microscopy images of LINE1 RNA FISH (red) of naive CD4⁺ T cells that were activated for 72 hours and then treated with rapamycin or CsA. Original magnification 63×. Scale bar 10 μm.



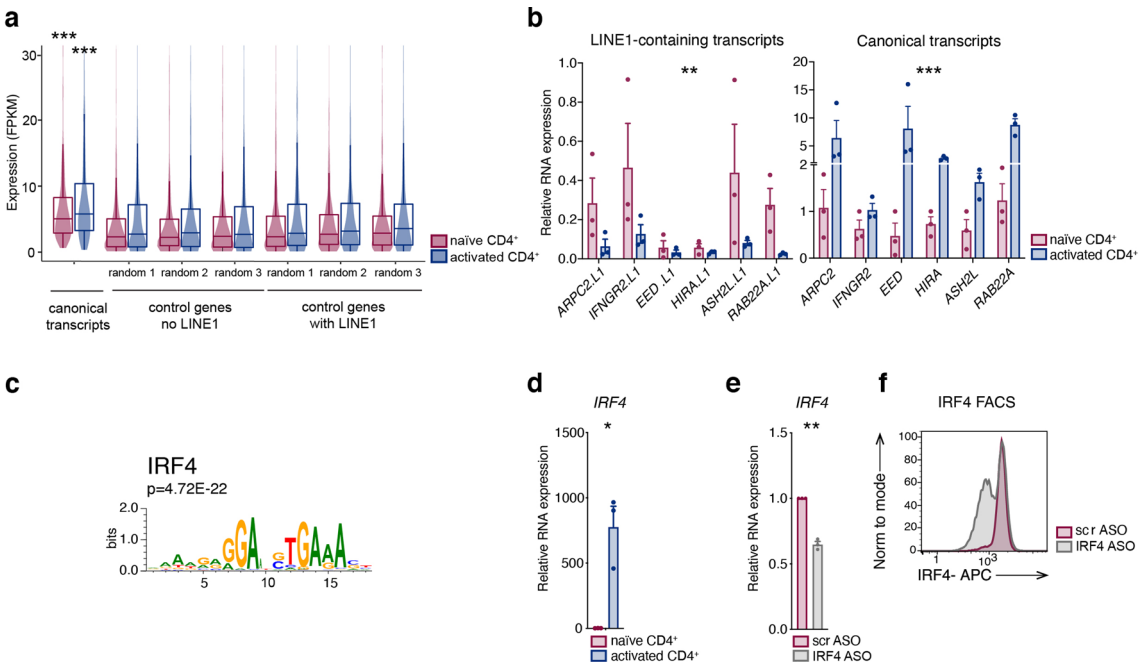
Extended Data Fig. 2 | Human CD4⁺ T cells express evolutionarily old LINE1 elements, whereas mESCs express evolutionarily young LINE1. (**a**) Heatmap of TEs expression at class, family and subfamily level in chromatin and nucleoplasm RNA-seq of naïve CD4⁺ T cells, four individuals are plotted. Z-score was computed on the log2 transformed normalized read count using DESeq2. (**b**) Scatter plot of LINE1 subfamilies expressions in nucleoplasm (x-axis) and chromatin (y-axis) RNA-seq of naïve CD4⁺ T cells. Subfamilies are color coded based on evolutive origin: mammalian-specific (L1M, orange), primate-specific (L1P, blue), human-specific (L1Hs, green), HAL (yellow). (**c**) Heatmap of TEs expression at class, superfamily and subfamily level in mESCs RNA-seq, two biological replicates are plotted. Z-score was computed on the log2 transformed normalized read count using DESeq2 (**d**) Pie-chart representing distribution of LINE1 chimeric and not chimeric reads in human CD4⁺ T cells. (**e**) Pie chart representing genomic distribution of LINE1 reads among intergenic regions, protein coding, lncRNAs, pseudogenes and noncoding RNAs transcriptional units in human CD4⁺ T cells. (**f**) Pie-chart representing distribution of LINE1 chimeric and not chimeric reads in mESCs. (**g**) Pie chart representing genomic distribution of LINE1 reads among intergenic regions, protein coding, lncRNAs, pseudogenes and noncoding RNAs transcriptional units in mESCs.



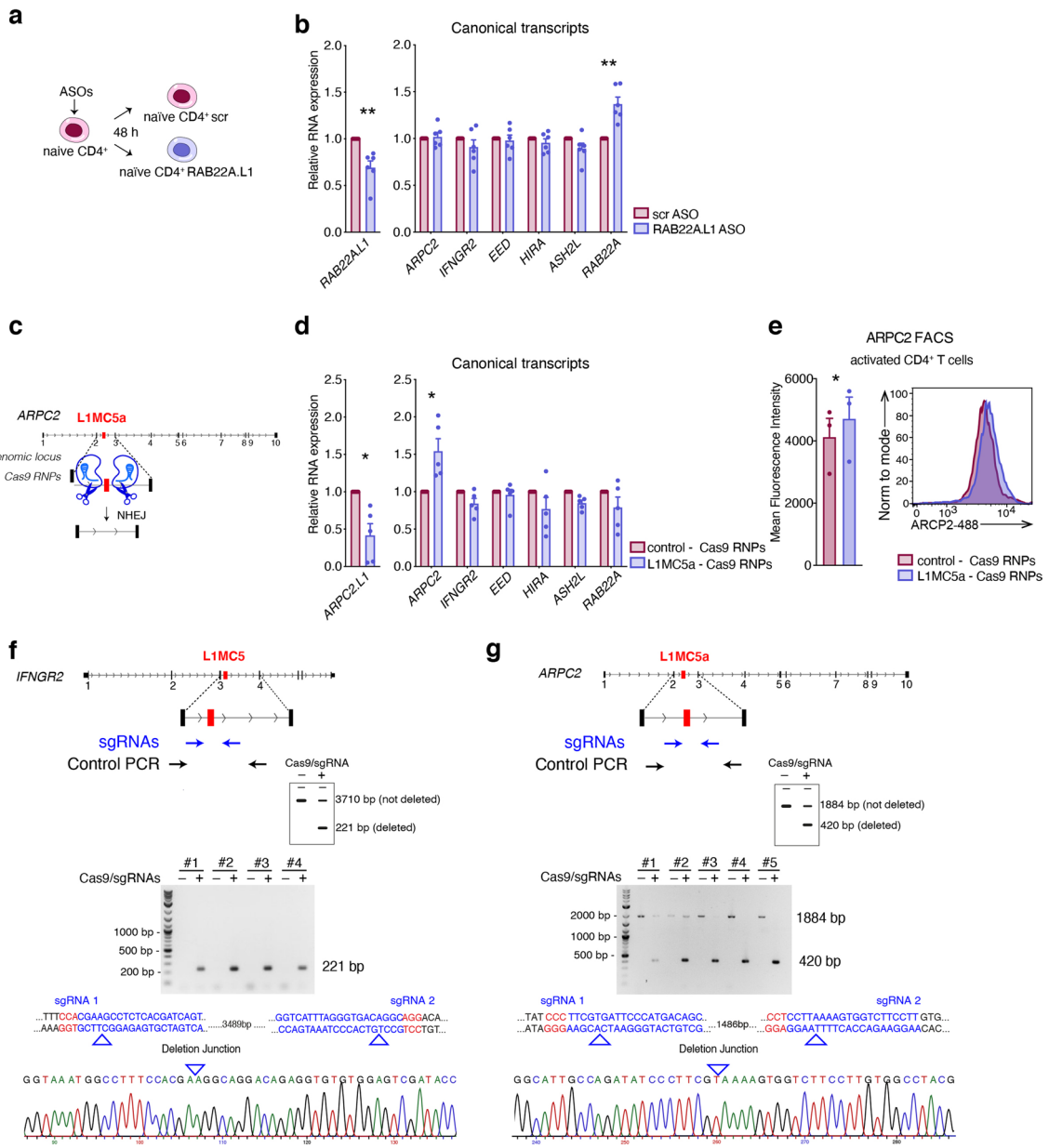
Extended Data Fig. 3 | Validation of LINE1-containing transcripts by long reads sequencing and by RT-PCR. (**a**) Examples of long reads covering the LINE1-containing transcripts reconstructed by *de novo* transcriptome assembly. (**b**) Scheme of canonical transcripts and LINE1-containing transcripts; the LINE1 exon is represented in red. In the middle, schemes of the PCR primers designed to verify the presence of the LINE1-containing transcripts and canonical transcripts are reported. Below, agarose gels representing RT-PCR results of LINE1-containing transcripts. PCRs have been repeated on three individuals obtaining same results.



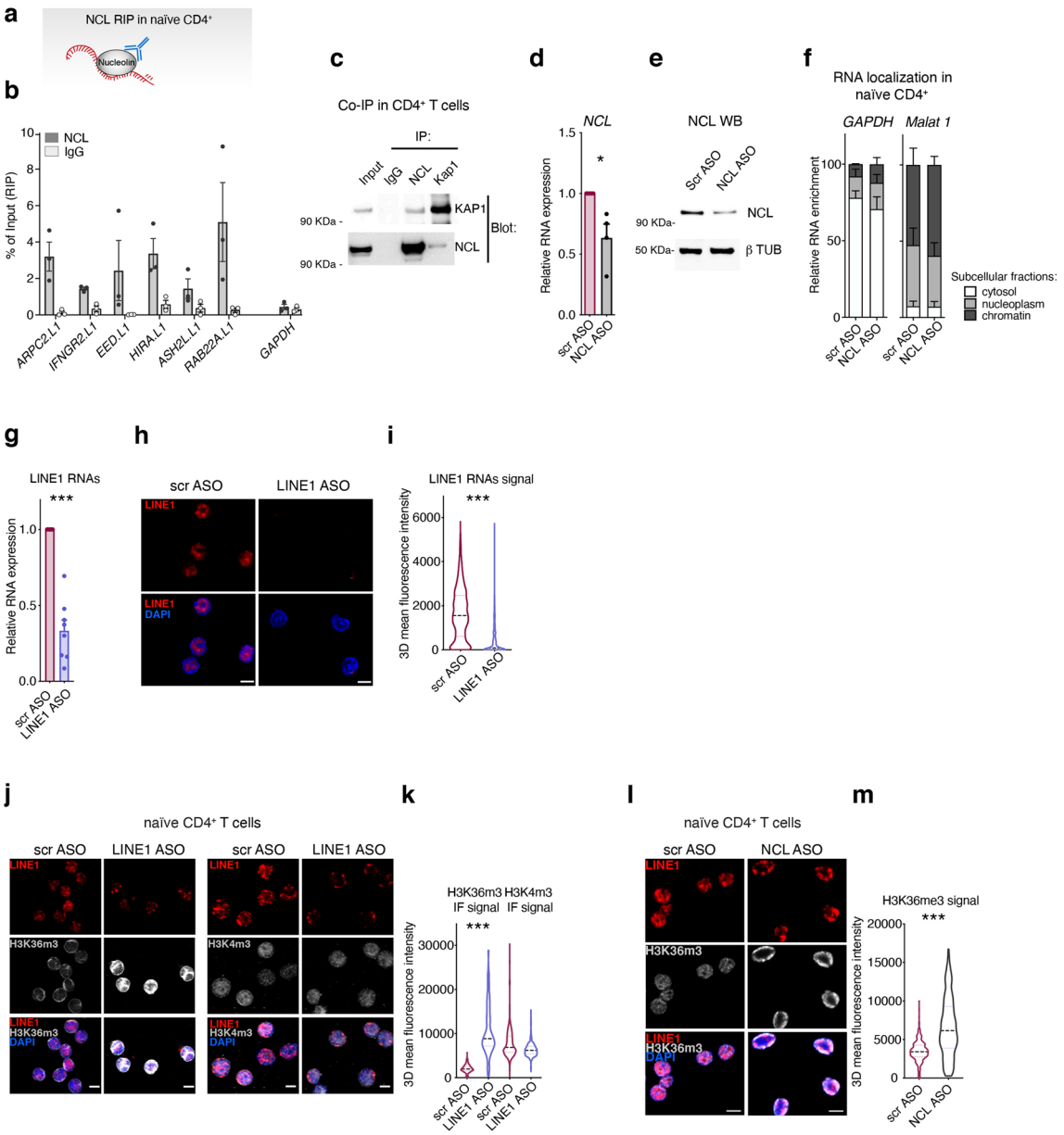
Extended Data Fig. 4 | Characterization of LINE1 elements spliced in novel exons of non-canonical splicing variants. (**a**) Length distribution of the LINE1 elements spliced in the LINE1-containing transcripts. (**b**) Distribution and enrichment of the expressed LINE1 regions in respect to a full length LINE1 sequence (6kb). Primers used for LINE1 RT-qPCR, probes for LINE1 RNA FISH and antisense oligonucleotide (ASOs) for LINE1 knock down experiments are shown. Right, percentage of the expressed LINE1 elements complementary to ORF1, ORF2, 5′UTR and 3′UTR regions of the full length LINE1. (**c**) Bar plot showing the percentage of the most enriched LINE1 subfamilies in the LINE1-containing transcripts. (**d**) LINE1 distribution among introns, exons, promoters, 5′UTR and 3′UTR of the LINE1 containing protein coding genes. (**e**) Consensus motifs of the donor and acceptor splicing sites of the LINE1 exons. (**f**) Scheme of *ARPC2.L1* transcript. The novel exon containing LINE1 element is zoomed. Long reads (blue) and short reads (green) of chromatin RNA-seq that support the novel exon reconstruction are shown.



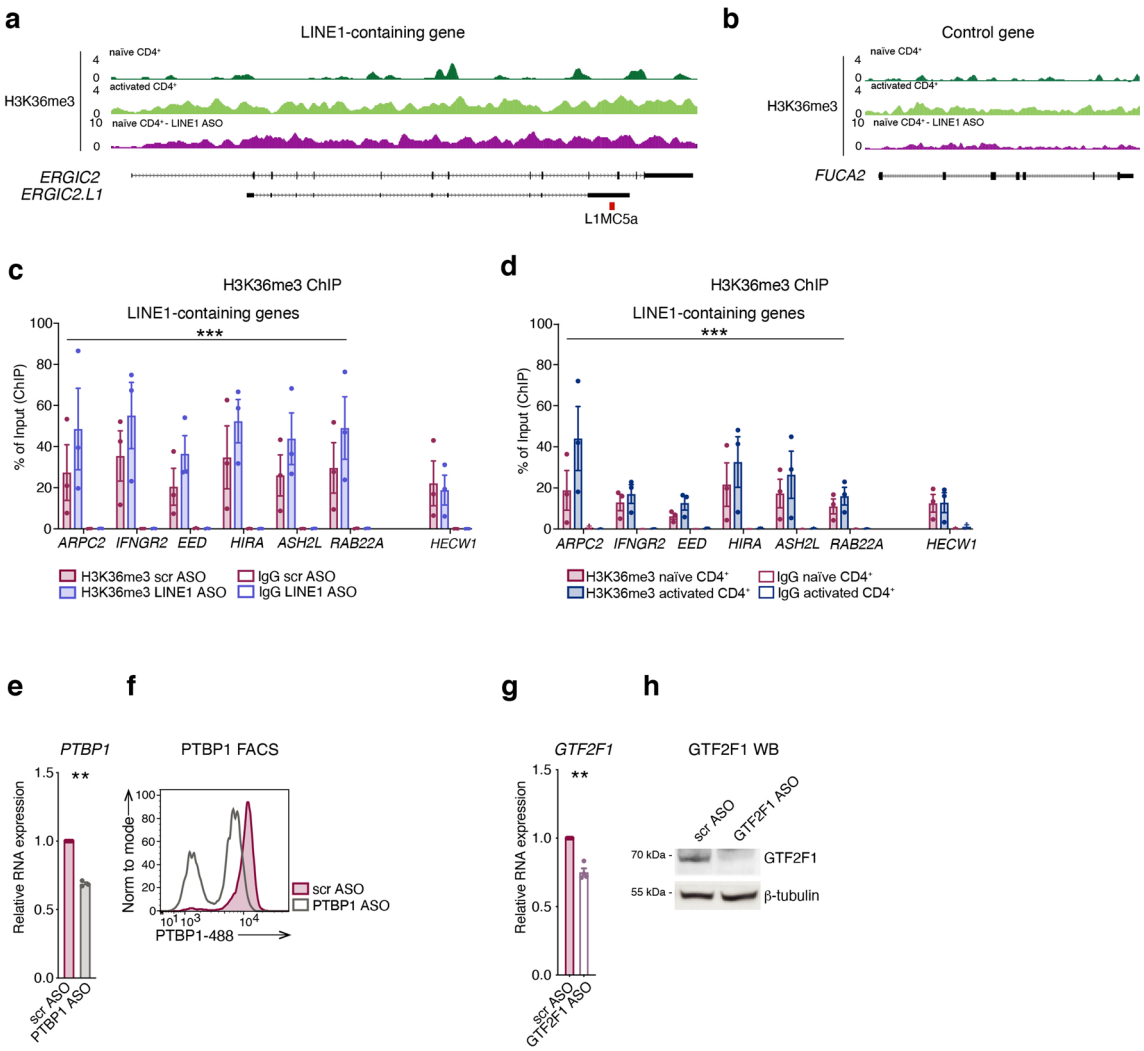
Extended Data Fig. 5 | Upon T cell activation, LINE1-containing transcripts are downregulated while canonical transcripts are upregulated. (**a**) Expression of the canonical transcripts and of three random set of control genes that i) do not retain genomic LINE1 elements (control genes no LINE1) or ii) retain LINE1 elements but do not generate LINE1-containing transcripts (control genes with LINE1) in RNA-seq datasets of quiescent and activated naïve CD4+ T cells (n = 3 individuals). *** P < 0.001, Two-tailed Wilcoxon signed rank test for paired samples, corrected accordingly to Bonferroni for multiple testing, was performed to compare canonical transcripts with control genes in naïve (P values from left to right: 1.89 x 10^{-12}, 6.42 x 10^{-13}, 1.91 x 10^{-12}, 1.27 x 10^{-13}, 1.04 x 10^{-12}, 3.13 x 10^{-10}) or activated T cells (P values from left to right: 3.06 x 10^{-07}, 3.20 x 10^{-08}, 1.15 x 10^{-06}, 6.41 x 10^{-11}, 1.55 x 10^{-07}, 4.98 x 10^{-05}). For box-and-whisker plots, the central line, box and whiskers represent the median, interquartile range (IQR) from first to third quartiles, and 1.5×IQR, respectively. (**b**) Expression of LINE1-containing transcripts and canonical transcripts by RT-qPCR in quiescent and 16 hours activated naïve CD4+ T cells (n = 3 individuals). Data are represented as mean ± s.e.m. LINE1-containing transcripts in naïve versus activated CD4+ T cells: ** P = 0.0096, F = 9.647, Two-way ANOVA; canonical transcripts in naïve versus activated CD4+ T cells: *** P = 0.00045, F = 22.85, Two-way ANOVA. (**c**) IRF4 Transcription Factors (TFs) binding motif research was performed on promoter regions of the LINE1-containing genes, selecting TFs upregulated in CD4+ naïve T cells. (**d**) *IRF4* expression by RT-qPCR in quiescent and 16 hours activated CD4+ T cells (n = 3 individuals). Data are represented as mean ± s.e.m. naïve versus activated CD4+ T cells: * P = 0.03975, Two-tailed paired *t* test. (**e**) *IRF4* expression by RT-qPCR and (**f**) IRF4 protein by FACS analysis in quiescent naïve CD4+ T cells treated with IRF4 or control (scr) ASOs. Data are represented as mean ± s.e.m (n = 3 individuals). scr versus IRF4 ASO groups: ** P = 0.0042, Two-tailed paired *t* test.



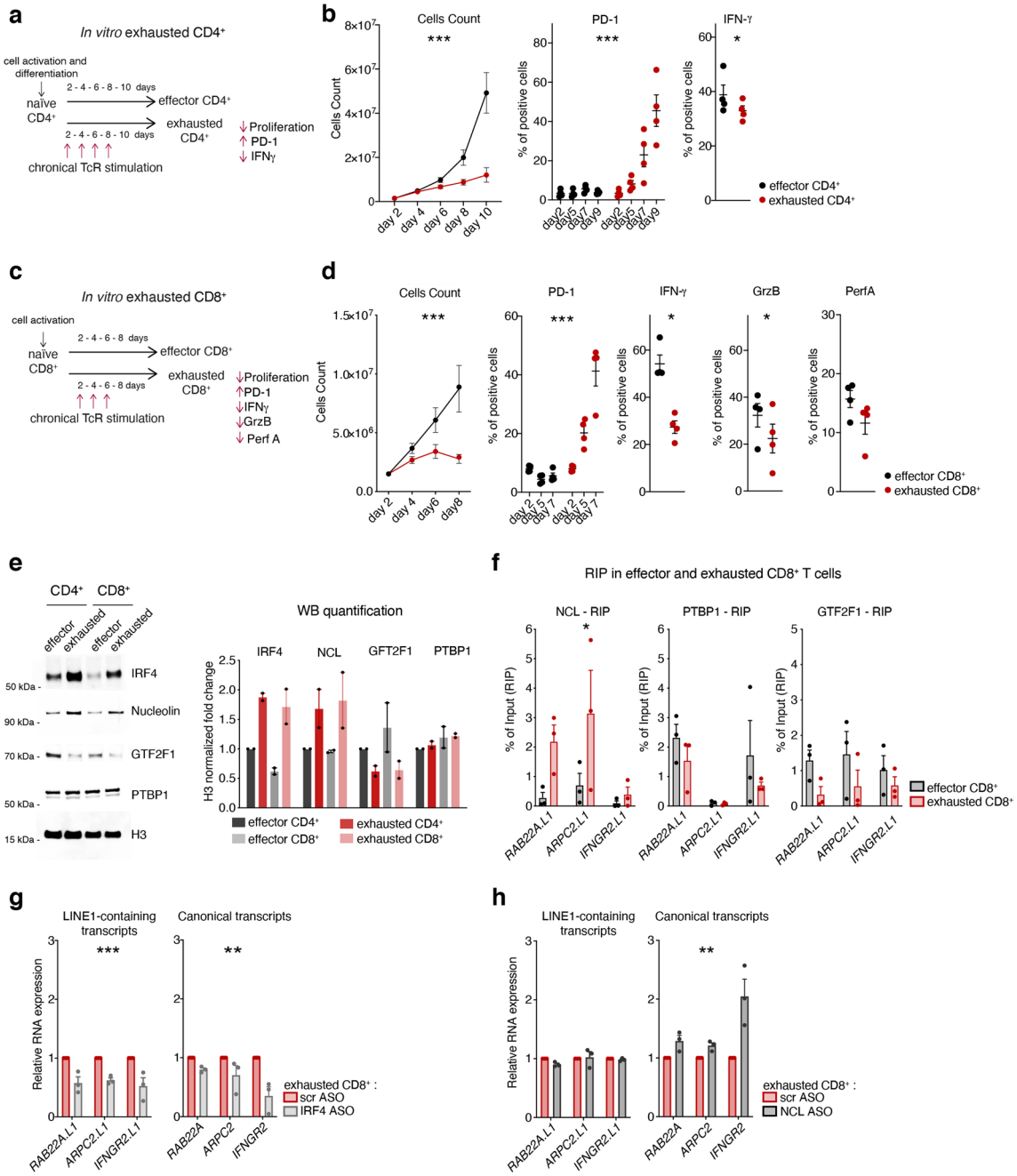
Extended Data Fig. 6 | LINE1-containing transcripts keep the expression of the corresponding canonical transcripts paused. (**a**) Scheme of *RAB22A.L1* knock down experiments in quiescent naïve CD4+ T cells. (**b**) Expression of *RAB22A.L1* and of *ARPC2*, *IFNGR2*, *EED*, *HIRA*, *ASH2L*, *RAB22A* canonical transcripts in quiescent naïve CD4+ T cells treated with RAB22A.L1 or control (scr) ASOs (n = 6 individuals). *RAB22A.L1* in scr versus RAB22A.L1 ASO groups: ** P = 0.00798, Two-tailed paired *t* test; *RAB22A* in scr versus RAB22A.L1 ASO groups: ** P = 0.0041, Two-tailed paired *t* test. (**c**) Design of the L1MC5a element deletion in *ARPC2* gene with CRISPR-Cas9 RNP complexes: the intronic L1MC5a element was deleted using sgRNAs complementary to the non-repetitive regions flanking the repeat. (**d**) Expression of *ARPC2.L1* and of *ARPC2*, *IFNGR2*, *EED*, *HIRA*, *ASH2L*, *RAB22A* canonical transcripts in quiescent naïve CD4+ T cells nucleofected with control Cas9 RNPs – or Cas9-RNPs with sgRNAs targeting L1MC5a in *ARPC2* (see Supplementary Table 6) (n = 5 individuals). Data are represented as mean ± s.e.m. *ARPC2.L1* in control versus L1MC5a Cas9 RNPs groups * P = 0.0201, Two-tailed paired *t* test; *ARPC2* in control versus L1MC5a Cas9 RNPs groups * P = 0.0310, Two-tailed paired *t* test. (**e**) ARPC2 protein by FACS analysis in quiescent naïve CD4+ T cells nucleofected with control Cas9 RNPs – or Cas9-RNPs with sgRNAs targeting L1MC5a in *ARPC2* gene and then activated (n = 3 individuals). Data are represented as mean ± s.e.m. ARPC2 in control versus groups L1MC5a Cas9 RNPs * P = 0.0273, Two-tailed paired *t* test. (**f-g**) Sanger sequencing of the PCR amplification products performed to assess L1MC5 and L1MC5a deletion from *IFNGR2* and *ARPC2*. PCRs have been performed on four individuals in (**f**) and on five individuals in (**g**).



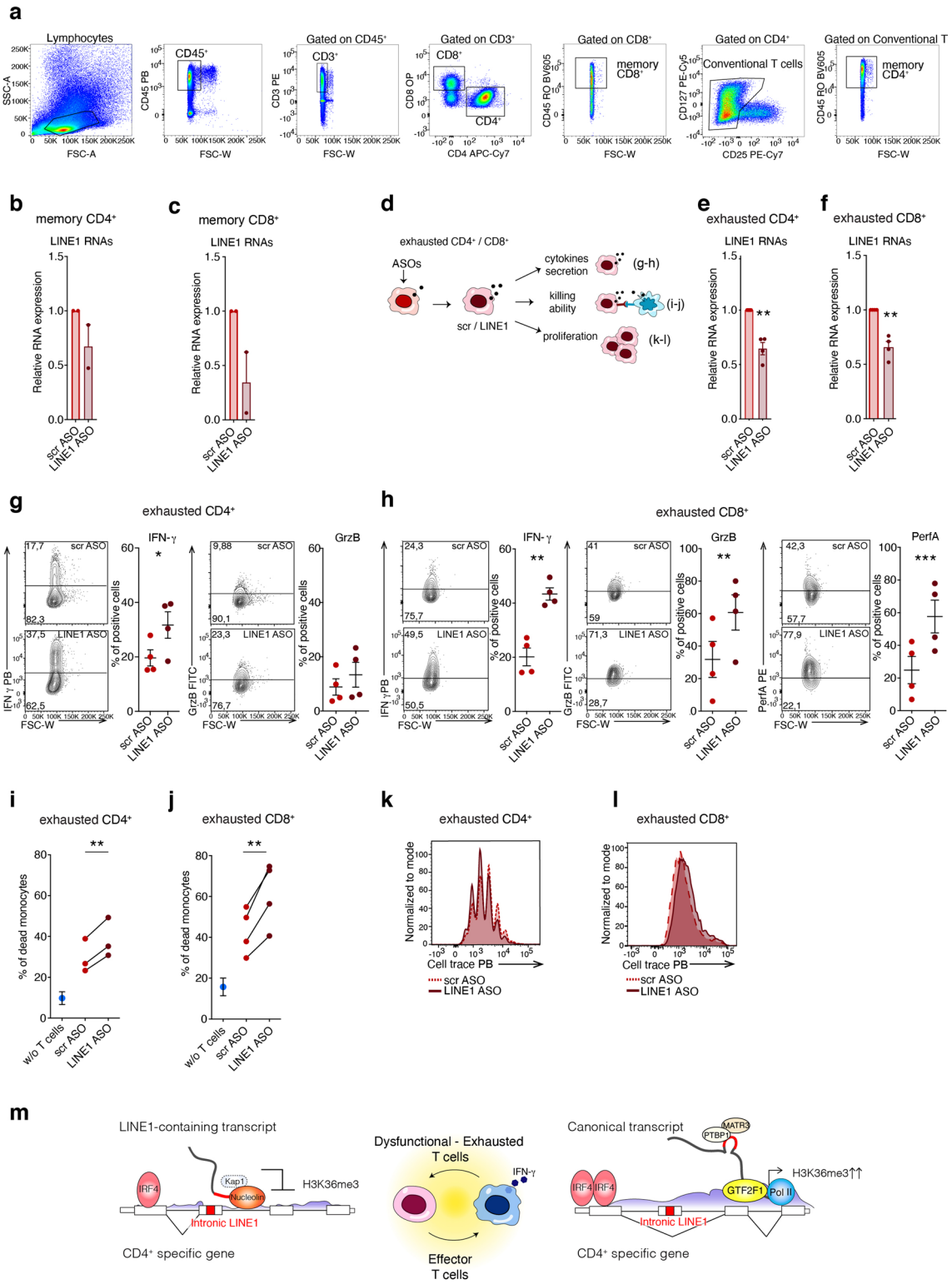
Extended Data Fig. 7 | LINE1-containing transcripts are bound by Nucleolin and regulate H3K36me3 levels in naïve CD4⁺ T cells. (**a-b**) RIP assays in naïve CD4⁺ T cells with anti NCL antibody (n = 3 individuals). Data are represented as mean % of input ± s.e.m. (**c**) Co-IP and western blots of NCL and KAP1 in CD4⁺ T cells. Co-IP has been repeated on two individuals obtaining similar results. (**d**) *NCL* expression by RT-qPCR and (**e**) NCL western blot in quiescent naïve CD4⁺ T cells treated with NCL or control (scr) ASOs. Data are represented as mean ± s.e.m (n = 4 individuals). *NCL* expression in scr versus NCL ASO groups: * P = 0.0482 Two-tailed paired *t* test. (**f**) Expression of *GAPDH* and *MALAT1* in cytoplasm, nucleoplasm and chromatin of naïve CD4⁺ T cells treated with NCL or scr ASOs (n = 3 individuals). Data are represented as mean ± s.e.m. (**g**) LINE1 expression by RT-qPCR in quiescent naïve CD4⁺ T cells treated with LINE1 or scr ASOs (n = 8 individuals). LINE1 ASOs were designed to target the ORF2 LINE1 region included in the LINE1-containing transcripts (see Extended Data Fig. 4b). Data are represented as mean ± s.e.m. Scr versus LINE1 ASO groups: *** P = 3 x 10^{-6}, Two-tailed paired *t* test. (**h**) Confocal fluorescence microscopy images of LINE1 RNA FISH (red) of quiescent naïve CD4⁺ T cells treated with LINE1 or control (scr) ASOs. LINE1 ASOs target the ORF2 LINE1 region of LINE1-containing transcripts. Original magnification 63x. Scale bar 5 µm. (**i**) Quantification of (**h**), at least 832 nuclei, two individuals. scr versus LINE1 ASO groups: *** P < 1 x 10^{-15}, Two-tailed Mann Whitney test. (**j**) Confocal fluorescence microscopy images of LINE1 RNA FISH (red) and immunofluorescent staining (gray) for H3K36me3 or H3K4me3 of naïve CD4⁺ T cells treated with LINE1 or control (scr) ASOs. Original magnification 63x. Scale bar 5 µm. (**k**) Quantification of (**j**), at least 267 nuclei, two individuals. H3K36me3 in scr versus LINE1 ASO groups: *** P < 1 x 10^{-15}, Two-tailed Mann-Whitney test. (**l**) Confocal fluorescence microscopy images of LINE1 RNA FISH (red) and immunofluorescent staining (gray) for H3K36me3 of quiescent naïve CD4⁺ T cells treated with NCL ASOs or control (scr) ASOs. Original magnification 63x. Scale bar 5 µm. (**m**) Quantification of (**l**), at least 259 nuclei, three individuals. H3K36me3 in scr versus LINE1 ASO groups: *** P < 1 x 10^{-15}, Two-tailed Mann Whitney test.



Extended Data Fig. 8 | Controls for ChIP-seq and for PTBP1 and GTF2F1 knock down. (**a-b**) Representative H3K36me3 ChIP-seq tracks for (**a**) *ERGIC2* LINE1 containing gene and (**b**) *FUCA2* control gene in quiescent and 16 hours activated naïve CD4$^+$ T cells and naïve CD4$^+$ T cells treated with LINE1 ASOs. (**c-d**) ChIP assays for H3K36me3 at LINE1 containing genes and HECW1 (control) in (**c**) quiescent naïve CD4$^+$ T cells treated with LINE1 or control (scr) ASOs and in (**d**) quiescent and 16 hours activated naïve CD4$^+$ T cells (n = 3 individuals). Data are represented as mean % of input ± s.e.m. H3K36me3 ChIP in scr versus LINE1 ASO groups: *** P = 6.2 x 10^{-7}, F = 90.24, Two-way ANOVA. H3K36me3 ChIP in naïve CD4$^+$ versus activated CD4$^+$: *** P = 3.6 x 10^{-4}, F = 40.45 Two-way ANOVA. (**e**) *PTBP1* expression by RT-qPCR and (**f**) PTBP1 protein by FACS analysis in quiescent naïve CD4$^+$ T cells treated with PTBP1 or scr ASOs and then activated for 16 hours. Data are represented as mean ± s.e.m (n = 3 individuals). *PTBP1* expression in scr versus PTBP1 ASO groups: ** P = 0.0014 Two-tailed paired *t* test. (**g**) *GTF2F1* expression by RT-qPCR and (**h**) GTF2F1 protein by western blot in quiescent naïve CD4$^+$ T cells treated with GTF2F1 or scr ASOs and then activated for 16 hours. Data are represented as mean ± s.e.m (n = 3 individuals). *GTF2F1* expression in scr versus GTF2F1 ASO groups: ** P = 0.0031, Two-tailed paired *t* test.



Extended Data Fig. 9 | IRF4 and Nucleolin are overexpressed in *in vitro* exhausted CD4⁺ and CD8⁺ T cells while GTF2F1 is downregulated. (**a**) Scheme of *in vitro* exhaustion of CD4⁺ T cells. (**b**) On the left, cell count of effector and exhausted CD4⁺ T cells (n = 5 individuals). Data are represented as mean ± s.e.m, *** $P < 2.7 \times 10^{-7}$, F = 57.22 Two-way ANOVA. In the middle, PD-1 positive cells (n = 4 individuals). Data are represented as mean ± s.e.m, *** P = 0.00004, F = 40.3 Two-way ANOVA. On the right, IFN-γ positive cells (n = 4 individuals). Data are represented as mean ± s.e.m. IFN-γ: * P = 0.0319, One-tailed paired *t* test. (**c**) Scheme of *in vitro* exhaustion of CD8⁺ T cells. (**d**) On the left, cell count of effector and exhausted CD8⁺ T cells. Data are mean and ± s.e.m, N = 4 individuals. *** P = 0.0003, F = 26.05 Two-way ANOVA. In the middle, PD-1 positive cells. Data are mean and ± s.e.m, N = 4 individuals. *** P = 0.00003, F = 58.1, Two-way ANOVA. On the right, IFN-γ, GrzB and PerfA positive cells. Data are mean and ± s.e.m, N = 4 individuals. IFN-γ: * P = 0.0108, Two-tailed Paired *t* test; GrzB * P = 0.0243, Two-tailed Paired *t* test. (**e**) IRF4, NCL, GTF2F1 and PTBP1 by western blot performed in nuclear extracts of effector and exhausted CD4⁺ and CD8⁺ T cells. Data are represented as mean ± s.e.m (n = 2 individuals). (**f**) RIP assays for NCL, PTBP1 and GTF2F1 in effector and exhausted CD8⁺ T cells (n = 3 individuals). Data are represented as mean % of input ± s.e.m. NCL RIP in effector versus exhausted: * P = 0.0488, F = 5.54, Two-way ANOVA. (**g-h**) LINE1-containing transcripts and canonical transcripts expression by RT-qPCR in exhausted CD8⁺ T cells treated with (**g**) IRF4 or control (scr) ASOs or with (**h**) NCL or scr ASOs (n = 3 individuals). Data are represented as mean ± s.e.m. LINE1-containing transcripts in scr versus IRF4 ASO groups: *** P = 0.0004, F = 50.6. Two-way ANOVA; canonical transcripts in scr versus IRF4 ASO groups: ** P = 0.0024, F = 25.3. Two-way ANOVA. Data are represented as mean ± s.e.m. Canonical transcripts in scr versus NCL ASO groups: ** P = 0.0026, F = 24.5. Two-way ANOVA.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | LINE1-containing transcripts regulate effector functions of exhausted CD4$^+$ and CD8$^+$ T cells. (**a**) Sorting of tumor-infiltrating memory CD4$^+$ and CD8$^+$ T cells. (**b-c**) LINE1 expression by RT-qPCR in memory CD4$^+$ (a) and CD8$^+$ (b) TILs treated with LINE1 or control (scr) ASOs (n = 2 individuals). Data are represented as mean ± s.e.m. (**d**) Scheme of the immunological assays performed on *in vitro* exhausted CD4$^+$ and CD8$^+$ T cells treated with LINE1 or scr ASOs to evaluate their effector properties (**g-l**). (**e-f**) LINE1 expression by RT-qPCR in exhausted (**e**) CD4$^+$ and (**f**) CD8$^+$ T cells (n = 4 individuals). Data are represented as mean ± s.e.m. scr versus LINE1 ASO exhausted CD4$^+$: ** P = 0.0083, Two-tailed paired *t* test; scr versus LINE1 ASO exhausted CD8$^+$: ** P = 0.0073, Two-tailed paired *t* test. (**g**) IFN-γ$^+$ or GrzB$^+$ exhausted CD4$^+$ T cells (n = 4 individuals). Data are represented as mean ± s.e.m. IFN-γ in scr versus LINE1 ASO groups: * P = 0.0351 One-tailed paired *t* test. (**h**) IFN-γ$^+$, GrzB$^+$ or PerfA$^+$ exhausted CD8$^+$ T cells (n = 4 individuals). Data are represented as mean ± s.e.m. IFN-γ in scr versus LINE1 ASO groups: ** P = 0.0016, Two-tailed paired *t* test; GrzB in scr versus LINE1 ASO groups: ** P = 0.0039, PerfA in scr versus LINE1 ASO: *** P = 0.0002, Two-tailed paired *t* test. (**i-j**) Percentage of dead heterologous monocytes co-cultured for 12 hours with exhausted (**i**) CD4$^+$ or (**j**) CD8$^+$ T cells. Percentage of monocytes self-lysis is indicated (w/o T cells, in blue). n = 3 individuals for CD4$^+$ and n = 4 individuals for CD8$^+$, data for each individual are shown separately, monocytes self-lysis is represented as mean ± s.e.m. CD4$^+$ scr versus LINE1 ASO: ** P = 0.0092, Two-tailed paired *t* test; CD8$^+$ scr versus LINE1 ASO: ** P = 0.0084, Two-tailed paired *t* test. (**k-l**) Proliferation assay with cell trace in exhausted (**j**) CD4$^+$ or (**k**) CD8$^+$ T cells. (**m**) Scheme of how LINE1-containing transcripts control gene expression in T cell quiescence and exhaustion.

Chapter 4

Review "The sophisticated transcriptional response governed by transposable elements in human health and disease"

The Sophisticated Transcriptional Response Governed by Transposable Elements in Human Health and Disease

Federica Marasca[1], Erica Gasparotto[1], Benedetto Polimeni[1], Rebecca Vadalà[1,2], Valeria Ranzani[1] and Beatrice Bodega[1]*

[1]Fondazione INGM, Istituto Nazionale di Genetica Molecolare "Enrica e Romeo Invernizzi", 20122 Milan, Italy; marasca@ingm.org(F.M.); gasparotto@ingm.org(E.G.); polimeni@ingm.org(B.P.); vadala@ingm.org(R.V.); ranzani@ingm.org(V.R.)

[2] Translational and Molecular Medicine, DIMET, University of Milan-Bicocca, 20900 Monza, Italy

* Correspondence: bodega@ingm.org

Received: 7 April 2020; Accepted: 29 April 2020; Published: 30 April 2020

Int. J. Mol. Sci. 2020, 21, 3201; doi:10.3390/ijms21093201

International Journal of
Molecular Sciences

MDPI

Review

The Sophisticated Transcriptional Response Governed by Transposable Elements in Human Health and Disease

Federica Marasca [1], Erica Gasparotto [1], Benedetto Polimeni [1], Rebecca Vadalà [1,2], Valeria Ranzani [1] and Beatrice Bodega [1,*]

[1] Fondazione INGM, Istituto Nazionale di Genetica Molecolare "Enrica e Romeo Invernizzi", 20122 Milan, Italy; marasca@ingm.org (F.M.); gasparotto@ingm.org (E.G.); polimeni@ingm.org (B.P.); vadala@ingm.org (R.V.); ranzani@ingm.org (V.R.)
[2] Translational and Molecular Medicine, DIMET, University of Milan-Bicocca, 20900 Monza, Italy
* Correspondence: bodega@ingm.org

Received: 7 April 2020; Accepted: 29 April 2020; Published: 30 April 2020

Abstract: Transposable elements (TEs), which cover ~45% of the human genome, although firstly considered as "selfish" DNA, are nowadays recognized as driving forces in eukaryotic genome evolution. This capability resides in generating a plethora of sophisticated RNA regulatory networks that influence the cell type specific transcriptome in health and disease. Indeed, TEs are transcribed and their RNAs mediate multi-layered transcriptional regulatory functions in cellular identity establishment, but also in the regulation of cellular plasticity and adaptability to environmental cues, as occurs in the immune response. Moreover, TEs transcriptional deregulation also evolved to promote pathogenesis, as in autoimmune and inflammatory diseases and cancers. Importantly, many of these findings have been achieved through the employment of Next Generation Sequencing (NGS) technologies and bioinformatic tools that are in continuous improvement to overcome the limitations of analyzing TEs sequences. However, they are highly homologous, and their annotation is still ambiguous. Here, we will review some of the most recent findings, questions and improvements to study at high resolution this intriguing portion of the human genome in health and diseases, opening the scenario to novel therapeutic opportunities.

Keywords: transposable elements; co-option; genome plasticity; immune system response; cancer progression; next generation sequencing approaches

1. Introduction

1.1. Transposable Elements (TEs) Account for Genome Evolution and Inter-Individual Genetic Variability

Two thirds of the human genome are composed of repetitive elements (66%), among which transposable elements (TEs) alone account for the 40–45% of human genome composition [1,2]. One fascinating question for genome biologists is to untangle the functions of this "dark side" of the genome, that still represents "alive matter" which evolution can influence to generate novel functions. It is clear nowadays that TEs capability of regulating the genome resides mainly in generating a sophisticated plethora of RNA regulatory networks, which in turn influence the transcriptional output of the cell [3–5]. TEs are organized into four different classes and, with the exception of DNA transposons, are mainly retrotransposons, which have acquired the ability by using RNA as intermediate to move via a 'copy and paste' mechanism. Retrotransposons include long interspersed elements (LINEs), short interspersed elements (SINEs), and long terminal repeat (LTR) retrotransposons.

They are further classified as autonomous or non-autonomous depending on whether they have open reading frames (ORFs) that encode for the machinery required for the retrotransposition [6].

LINE is a class of transposon that is very ancient and evolutionary successful. Three LINE superfamilies are found in the human genome, namely LINE1, LINE2 and LINE3, of which only LINE1 is still active. Full-length LINE1 (L1) elements are approximately 6 kb long and constitute an autonomous component of the genome. A LINE1 element has an internal polymerase II promoter and encodes for two open reading frames, ORF1 and ORF2 (Figure 1) [7]. Once the L1 RNA is transcribed, it is exported to the cytoplasm for translation, and subsequently assembled with the chaperone RNA- binding proteins ORF1 and the endonuclease and reverse transcriptase ORF2. These ribonucleoparticles are then reimported into the nucleus, where ORF2 makes a single-stranded nick and primes reverse transcription from the 3' end of the L1 RNA. Reverse transcription frequently results in many truncated, nonfunctional insertions, and for this reason, most of the LINE-derived repeats are short, with an average size around 900–1000 bp. The L1s are estimated to be present in more than 500,000 copies in the human genome [7].

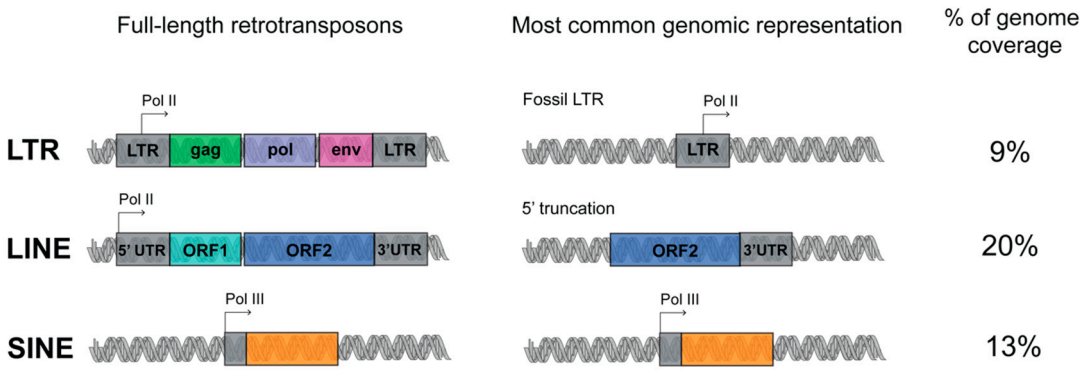


Figure 1. Schematic representation of retrotransposons classes organization. Retrotransposons are divided in three major classes: long interspersed elements (LINE), short interspersed elements (SINE) and long terminal repeat (LTR). Left, full length retrotransposons: the regulatory sequences are represented in grey; RNA Pol II and Pol III promoters are indicated with arrows; the protein coding sequences are indicated with colors. Middle, most common transposable elements (TEs) in the human genome. Right, retrotransposon coverage of the human genome (see the main text for details).

The L1 machinery is also responsible for the retrotransposition of the SINEs (which can be classified into three superfamilies: Alu, MIR, MIR3), non-autonomous retroelements without any coding potential, short in length (around 300 bp) and transcribed from polymerase III promoter (Figure 1). The most represented human specific SINE superfamily, the Alu, is represented in 1,090,000 copies in the human genome [8].

The LTR retrotransposons are initiated and terminated by long terminal direct repeats embedded by transcriptional regulatory elements. The autonomous LTR retrotransposons contain *gag* and *pol* genes, which encode a reverse transcriptase, integrase, protease and RNAse H (Figure 1). Four superfamilies of LTR exist: ERV- class I, ERV(K) class II, ERV(L) class III, and MalR. MalR is the most represented superfamily of LTR, present in 240,000 copies [9].

Evolutionary biologists hypothesize that self-replicating RNA genomes were the basis of early life on earth, and that the advent of reverse transcription had a pivotal function in the evolution of the first DNA genomes, the more stable deoxyribose-based polymers [6,10]. From this perspective, multiple rounds of reverse transcription could have helped to expand both the size and complexity of the human genome. It is particularly evident in both mammals and plants that retrotransposons have massively accumulated, driving genome evolution. It is reported that L1 and Alu represent the most prominent catalysts of the human genome evolution [11] and that homologous recombination between TEs could have driven/drives mutations, chromosome rearrangement, deletions, inversions and translocations [12]. TEs are a major source of somatic genomic diversity and interindividual variability [13] and TE insertions

have been documented as physiological occurrences [14–16]. In particular L1 retrotransposition has been extensively described as taking place in neurons, from fly to man [17–19], a mechanism that is fine-tuned and epigenetically regulated in neural progenitor development and differentiation, contributing to the somatic diversification of neurons in the brain [13,20]. The deregulation of TEs activity is nowadays emerging as an important contributor to many different diseases, as it occurs in neurological and inflammatory diseases and cancers [21–23].

The hosts have developed many systems to control TEs expression and expansion [24] (thus, epigenetic modification and noncoding RNAs (ncRNA) such Piwi interacting-RNAs) to contain the possible detrimental effects of their retrotransposition. This expansion has achieved a balance between detrimental and beneficial effects, possibly becoming a novel regulatory mechanism to promote genomic functions acquired through evolution [3]. It is nowadays accepted, both in mouse and in human, that TEs have been co-opted into multiple regulatory functions for the accommodation of the host genomes metabolisms and transcription, mediated both by their DNA elements and by their transcribed RNAs counterparts.

1.2. Not Just Transposition: TEs RNAs Are a Prolific Source for Novel Regulatory Functions

TEs were first discovered in maize by Barbara McClintock almost 80 years ago. She suggested these elements as "controlling elements" able to regulate the genes activity [25,26]. Her theories, even if dismissed for a long time, were pioneering and with the advent of next generation sequencing (NGS) technologies have been thoroughly revised. Currently emerging is the concept that TEs interact with the transcriptional regulatory functions of the hosts genomes [3,4,27,28].

Although a massive portion of the literature has been centered on the study of the retrotransposition and the effects of the de novo insertions, it is worth noting that TEs can have RNA regulatory functions decoupled from their retrotransposition.

International decade-long projects such as ENCODE (Encyclopedia of DNA Elements) and FANTOM (Functional Annotation of the Mammalian Genome) have produced and bioinformatically analyzed a vast number of datasets opening the way for studying TEs. These results revealed that TEs have precise functions in establishing and influencing the cell type specific transcriptional programs, creating regulatory networks that are fostered both by their genomic elements and the derived transcripts [3,28], revealing that the RNAs transcribed from this elements could have a myriad of functions, definitely changing the way in which many genomic concepts were written in textbooks [29].

These studies clarified that TEs can create novel or alternative promoters [30], promote the assembly of transcription factors [31] and epigenetic modifiers and favor their spreading [32] and the regulation of gene expression. Further, TEs in particular SINEs and HERVs, have been demonstrated to have functions in 3D genome folding, as the binding sites for chromatin organizers [33–35].

In the 2009 Faulkner et al. [36], demonstrated for the first time that TEs are widely expressed in human and mouse cell types with tissue-specific patterns of expression, suggesting a specific spatiotemporal activation of retrotransposons. Faulkner et al. further demonstrated that up to the 30% of the transcripts initiate within repetitive elements [36]. It is interesting to notice that tissues of embryonic origin contain the highest proportion of transposable element-derived sequences in their transcriptomes, with specific expression of LTR in placenta and oocytes [37]. In accordance, it was recently found that different classes of repeats are specifically enriched in genes with a definite spatiotemporal expression, further dictating their timing and magnitude of expression in development [38].

Within this scenario, TEs magnify the transcriptome complexity in different ways: generating antisense transcripts, usually in proximity to gene promoters [36], acting on the maturation of mRNAs via nursing alternative splicing sites for tissue specific exonization [39,40], and providing alternative polyadenylation signals [41,42] and sites for the RNA-mediated decoy [43]. Furthermore, TEs contribute to RNA regulatory sequences within introns and untranslated regions (UTRs) [36]. It is important to notice that TEs are major contributors to long noncoding RNAs (lncRNAs) [44,45]. In this scenario,

an enhancer RNAs function was proposed for LTR derived transcripts, as required for pluripotency maintenance in mouse and human embryonic stem (ES) cells [46,47]. Further, it has been demonstrated that LINEs and SINEs are expressed as RNAs tightly associated to the chromatin compartment, where they localized at euchromatin, suggesting a possible function of these RNAs in 3D genome folding [48]. L1s have been described also as chromatin associated RNAs both in embryogenesis, regulating open chromatin accessibility [49,50], and in mouse ES cells, where they are involved in the regulation of genes required for cell identity maintenance and two-cell stage differentiation [51].

Although these seminal papers have increased the awareness and knowledge of the functions of TEs, highlighting important epigenetic roles for transposons in embryogenesis and development, the contribution of TEs to adult cell plasticity and diseases occurrence and progression is still poorly investigated. This is a result of the intrinsic difficulties in studying TEs, which due to their repetitive nature, high degree of homology, sequence divergence, and degeneration, render almost unfeasible the application of the technologies established for biallelic genes, in particular in bioinformatic.

Here, we will revise the TEs mediated multi-faced functions in promoting the establishment of a sophisticated plethora of RNA regulatory networks, which in turn influence the transcriptional plasticity of the cells. We will show how TEs transcriptional deregulation in pathological context is instead instrumental to fuel diseases. In particular we will review how TEs RNA can become a key player in the regulation of the immune response, using cell intrinsic specific pathways to directly control the regulation of interferon production and the activation of the immune cells; the alteration of these phenomena occurs in autoimmune and inflammatory diseases. Similarly, transcriptional deregulation of TEs represents a hallmark of cells that have lost identity, such as in cancer cells, where TEs onco-co-optation represents an important way to evolve cancer specific functions to promote tumor fitness and survival. Many of these findings have been achieved through the employment of NGS technologies with the application of bioinformatic pipelines that are in continuous evolution. Within this frame, an unambiguous TE identification and expression quantification of TEs at the genomic instance level would allow the precise and systematic definition of their contribution to RNA regulatory networks. We will review advances in the field and the challenges that should be addressed in this direction.

2. The Transcriptional Role of TEs in Shaping the Innate and Adaptive Immune Response

2.1. TEs RNAs Boost Innate and Adaptive Immune Response

The immune system is able to protect our organism against pathogens and foreign substances thanks to an innate and adaptive immune response [52]. During evolution, TEs have established transcriptional networks acting as regulatory DNA elements and also as signaling molecules for the immune system. The RNAs transcribed from TEs and/or the corresponding reverse transcribed cDNA are used by the host sensing pathways to promote expression of interferons that further solicits the immune response [53–55] (Figure 2A).

Alu can act as regulatory elements for *IFN-β* genes providing the binding site for the key transcription factor NF-κB [56]. Similarly, Thomson et al., [57] in the 2009 discovered that *IFN-γ* locus is highly enriched for TEs, where a cluster of ERVL-Alu Sx-ERVL is required for the NF-κB dependent activation of *IFN-γ* expression in response to LPS [57]. Recently, Choung et al. [58] demonstrated that elements originating from LTR work as IFN-inducible enhancers, a function conserved among different mammalian species, suggesting an evolutive co-option of TEs elements to regulate the expression of genes related to immunity [58].

A viral infection can promote the immune response through different mechanism, that include not only the recognition of the viral capsid protein or surface glycoprotein of the viruses, but also involve the sensing pathways for cytosolic foreign RNA and DNA, both single- and double-stranded (i.e., ssRNA, dsRNA, ssDNA, dsDNA) [59]. These molecules activate the immune response binding the host pattern recognition receptors (PRRs), that are transmembrane receptors as Toll-like receptors

(TLRs) or the cytosolic receptors as RIG-I and MDA5. ERVs can promote the antiviral immune response through the encoded env protein, the transcribed RNA or reverse transcribed cDNA that are recognized by the host PRRs receptors [60]. In 2006, Rolland et al. [61] found that the envelope of HERV-W, a member of specific superfamily of ERV elements, binds the TLR4 located on the cellular membrane, inducing the secretion of proinflammatory cytokines and the stimulation of monocytes and dendritic cells that in turns promote the $CD4^+$ T effector cells response [61] (Figure 2A). In 2004 Heil et al. [62] demonstrated that other isoforms of TLR, as TLR7 and TLR8, localized in the endosomal membranes, are able to bind cytosolic HERV ssRNA. Cytosolic PRRs are very sensitive in detecting cytosolic RNA and Chiappinelli et al. [63] in 2015 demonstrated that HERV dsRNA binds MDA5, promoting IFN-β production (Figure 2A). However, in analogy to ERVs, other TEs can also regulate the activation of the immune response, stimulating the same pathways. L1 RNA, due to its AU-rich sequence, can be-recognized by RIG-I and MDA5 [64]. Cytosolic L1 RNA is recognized by RIG-I through its 5′ terminal triphosphate form (a feature common to TEs) and L1 dsRNA binds MDA5. These interactions promote *IFN-β* expression [65]. Similarly, also the cytosolic dsRNA derived from Alu could induce the transcription of inflammatory genes and inhibit viral protein synthesis [66,67].

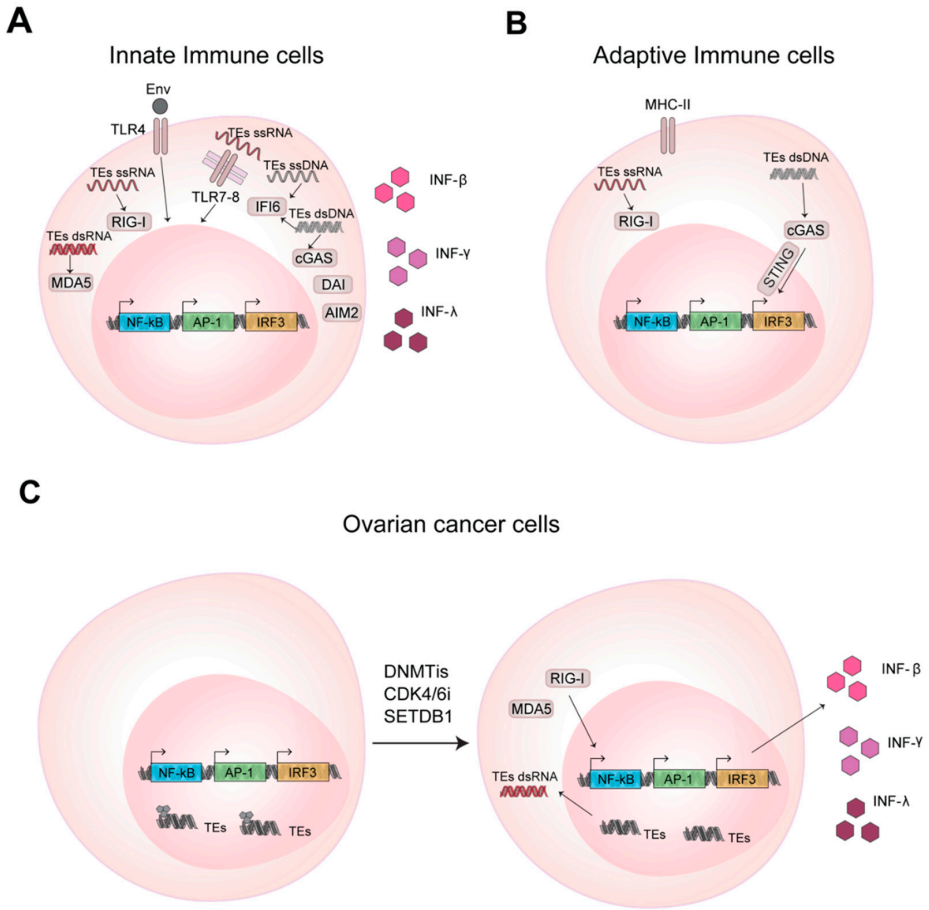


Figure 2. TEs promote innate and adaptive immune response activation in health and disease through RNA and DNA sensing pathways. (**A**) Nucleic acids of TEs bind and activate the transmembrane Toll-like receptors (TLRs) and cytosolic pattern recognition receptors (PRRs) activating transcription factors that promotes *INF* gene transcription and IFNs production. (**B**) TEs in T and B lymphocytes activate adaptive immune response through RNA and DNA sensing pathways, as mentioned in (**A**). (**C**) In cancer cells the inhibition of DNA methylation, promotes TEs expression and enhances cytokines production.

Besides the RNA-sensing pathways, TEs are able to elicit the immune response also through DNA-sensing pathways, often stimulated by their cytosolic, reverse transcribed ssDNA or cDNA [60,68–70] (Figure 2A). In B cells, it has been demonstrated that ERVs could promote cell activation and the production of antigen-specific antibodies through both their cytosolic RNAs and cDNAs, the former activating RIG-I-MAVS and the latter cGAS-STING (GMP-AMP synthase (cGAS), adaptor stimulator of IFN genes (STING)) pathways [71] (Figure 2B). In the same context, L1 retrotransposition is instead inhibited by the cytidine deaminase AID (activation-induced cytidine deaminase) that, by reducing ORF1 protein level, promotes a strict surveillance of retrotransposon accumulation in the cytoplasm. Importantly, mutation in AID promotes the increase of cytosolic L1 RNA and cDNA, contributing to the autoimmune phenotype typical of diseases that show defects in the *AID* gene as hyper-IgM syndrome [72].

Overall, we provide evidence that TEs, taking advantage of the machinery used by the viral genomes, are intriguing novel players able to fine tune and regulate the immune response, shaping TEs as possible novel targets for immunological approaches.

2.2. Deregulation of the Expression Levels of TEs is Implicated in Autoimmunity and Inflammation

Several studies have demonstrated the function of TEs in innate and adaptive immune response via IFN regulation with multiple mechanisms. This concept envisages the deregulation of TEs as a possible key component in the development of inflammation and autoimmune diseases.

In 2016, Manghera et al. [73] demonstrated that ERVK overexpression in amyotrophic lateral sclerosis (ALS) could represent a connection between the neuronal damage in ALS and the impaired signaling by proinflammatory cytokines signaling. In ALS, it has been demonstrated that ERVK reactivation occurs in the neurons of the motor cortex in ALS. ERVK promoter retains two conserved interferon-stimulated response elements (ISREs) that are activated in the motor neuron by the proinflammatory cytokines signaling typical of ASL. ERVK expression in turn contributes to a neurodegenerative phenotype nursing the inflammatory response using the above-mentioned sensing pathways, identifying ERVK as novel players of the pathology [73,74]. Similarly, another study reported that ERVK expression or the env protein translation cause retraction and beading of neurites in human neurons. A mice transgenic model expressing the ERVK env indeed developed motor dysfunction and a loss of volume in the motor cortex, impaired synaptic activity in pyramidal neurons, defects in the dendritic spine and DNA damage increase [75]. These studies define that ERVK reactivation can contribute to degeneration of motor neuron, possibly via different mechanisms, identifying ERVK as novel biomarker of the ASL pathology.

This mechanism can also be extended to other inflammatory disease and the connection of TEs upregulation with inflammatory signaling could be targeted for specific therapies. Systemic lupus erythematosus (SLE) and Sjögren's syndrome are autoimmune diseases in which the overexpression of Alu elements enhances the inflammatory autoimmune response. It has been described that, in these syndromes, autoantibodies are produced to target the RNA binding protein Ro60. Ro60 is able to bind an RNA motif derived from endogenous Alu retroelements, and its disfunction results in enhanced expression of Alu RNAs that then promotes *IFN*- type I regulated genes upregulation, feeding the inflammatory phenotype of the diseases [76]. Alu expression was discovered as directly proportional with interferon signature metric (ISM) level of SLE patients. This study attributes the pathogenicity of anti-Ro60 autoantibodies and type I interferon in SLE and Sjögren's syndrome to Alu retroelements. However, further studies are required to evaluate the potential of Alu and other Ro60-associated RNAs to activate the IFN response in health and disease [77].

Aicardi-Goutières syndrome (AGS) is an inflammatory disorder, most typically affecting brain, characterized by the dysregulation of type I IFN levels due to mutation in several factors like TREX1 (DNA sensing pathways 3'repair exonuclease) that plays a critical role in restricting the amount of endogenous DNA in the cytosol, and ADAR1 (adenosine deaminase), acting on RNA sensing pathway [78]. Thomas et al. demonstrated in AGS that *TREX1* mutation permits extranuclear

accumulation of L1 reverse transcribed cDNA (ssDNA), that triggers inflammation by IFN type I secretion, identifying these molecules as source of neuroinflammation [79]. L1 accumulation has been shown to induce neurotoxicity in neurons and astrocytes; the use of reverse transcriptase inhibitors in AGS neurons and organoids model rescued this phenotype, suggesting the potential use of these inhibitors in treating AGS and related disorders. Similarly, the mutation in *ADAR1* improves cytosolic L1 ssRNA levels increasing the IFN production [65]. Moreover, a recent study performed in a cellular model of senescence and inflammation proposed that L1 become transcriptionally de-repressed in late senescence, the RNA is retrotranscribed and the derived cDNA (ssDNA) activates IFN-I response; this work revealed a contribution of transposable elements to the senescence-associated inflammatory secretory phenotype, suggesting that L1 reverse transcriptase inhibition could be a therapeutic target for the age-associated disorders [80].

Overall, this evidence suggests that TEs are fine regulators of the immune response and that, in particular, their deregulation is associated with pro-inflammatory and autoimmune phenotypes, suggesting that TEs, being able to impose an aberrant activation of the immune response, could represent important yet under-investigated players in complex and multifactorial inflammatory diseases.

2.3. TEs RNAs are Novel Players in Cancer Immunity

The tumor microenvironment is represented by malignant, stromal and immune cells, the latter actively involved in tumor fight. Indeed, at first, the immune system is able to destroy and kill cancer cells, but with tumor progression, cancer develops a specific transcriptional program that escapes or attenuates the immune surveillance, rendering the immune cells dysfunctional [81–85]. The immune system regulation acquired a great relevance in cancer therapy and the immunotherapy based on the principle that the immune cells infiltrating the tumor can be reactivated in order to unleash antitumor response [86].

However, it is worth to notice that the tumor, by the above-mentioned mechanisms, can in principle stimulate the immune response against cancer by promoting TEs expression. TEs are under control of DNA methylation and methylation of the lysine 9 or 27 of histone H3 that repress their transcription and block the TEs mediated immunostimulatory activity [87] (Figure 2C). It has been demonstrated that DNA methyltransferases activity, by inhibiting the expression of TEs, permits the evasion of immune surveillance in cancer [63,88,89]. Conversely, the inhibition of DNA methylation reactivates TEs and promotes the innate and adaptive immune response against cancer cells [63,88–90] (Figure 2C). In 2015, Chiappinelli et al. demonstrated that, in ovarian cancer, the use of DNA methyltransferase inhibitors (DNMTis) promotes ERVs expression, whose cytosolic dsRNA is recognized by RIG-I and MDA5, further inducing IFN-β production and triggering the immune response [63] (Figure 2C). Similarly, it has been proven that the colorectal cancer initiating cells (CICs, promoting tumor relapse and affecting patient survival) treated with low doses of DNMTis experimentally accumulate cytosolic ERVs dsRNA. ERVs dsRNA are recognized by MDA5 receptor, support the downstream activation of IRFs and the upregulation of interferon-responsive genes. This reduces the proliferation of CICs displaying anti-cancer effects [89] (Figure 2C). In agreement, Goel and colleagues found that inhibitors of cyclin-dependent kinase 4/6 (CDK4/6) increase ERVs expression, to promote the cytoplasmic accretion of ERV dsRNA and increase IFN type III secretion [88] (Figure 2C). Further, these inhibitors suppress CD4+ T regulatory cells proliferation, increase tumor immunogenicity and promote the cytotoxic response by T cells enhancing tumor cells clearance [88]. In acute myeloid leukemia (AML), Cuellar and colleagues [90] demonstrated that the silencing of H3K9 methyltransferase SETDB1 leads to the overexpression of different TEs, promoting IFN antiviral response through dsRNA-sensing pathway [90] (Figure 2C). Other epigenetic histone marks as the repressive H3K27me3 are involved similarly in the regulation of cancer—promoted immune response. A novel subclass of ERVs, the SPARCS, are silenced by EZH2. The inhibition of EZH2, from one side promotes SPARCS expression that in turn activates the dsRNA-sensing pathway and the IFN response, and also on the others induces old MHC class I upregulation and neoantigens presentation, with an overall stimulation of the tumor

T cells infiltration and immune activation, suggesting a possible role for TEs in cancer therapy such as adjuvant molecules used in combination with immunotherapies [91].

Very recently, Smith et al. [92] proposed that, besides the ERVs mediated dsRNA promotion of the innate immune response involving the RIG-I pathways, novel retroviral epitopes expressed by the tumor cells drive T and B cell responses, promoting the adaptive immune activation. Importantly, Smith et al. demonstrated that ERVs expression could be used as a prognostic biomarker for outcome of immunotherapies [92].

These studies corroborate the idea that in TEs could reside novel, regulatory molecules able to modulate tumor immunogenicity and anti-tumor immune response, contouring TEs as novel molecules to be investigated in cancer immunotherapy.

3. TEs Transcriptional Landscape in Cancer Tissue

3.1. The Expression of TEs is Widely Dysregulated in Cancer Tissue

Genetic alterations are recognized as major causes of neoplasia, being able to further promote transcriptional alteration in cancer [93]. In such complexity, TEs are widely dysregulated during cancer development [94] and in many different tumor types [95]. The expression of TEs can promote retrotransposition and the human genome has evolved mechanisms, at both the transcriptional and post-transcriptional level, to avoid detrimental effects on the host genome [24]. While physiologically DNA repetitive sequences and transposons are highly methylated and repressed [96], human cancers can display hypomethylation specifically in definite genomic regions, as in certain classes of TEs promoters, increasing their expression and retrotransposition [95,97]. Other epigenetic modifications can be altered in cancer promoting TEs expression: glioma patients with pervasive H3K27 acetylation display ERV overexpression [98], and as mentioned above, mutation in the epigenetic modifiers for H3K9me3 and H3K27me3 could induce the expression of different classes of TEs [90,91].

Interestingly, the majority of the actively expressed TEs in cancer are those evolutionary youngest such as piggyBac, L1HS, HERVK, and HERVH [95]. Being often full-length elements, these retain intact regulatory regions and behave as active binding sites for transcription factors, possibly further contributing to the transcriptional alteration of cancer tissues [99]. Importantly, it is emerging that TEs expression and dynamics are cancer-tissue specific, similar to what it has been demonstrated for normal tissues [95,100]. Kong et al. [95] have analyzed 7345 TCGA RNA-seq deriving from 25 cancer types in comparison with their normal adjacent tissue counterpart, retrieving an increased TEs expression in certain tumors (stomach, bladder, liver, and head and neck) and a reduced one in others (thyroid, breast, kidney chromophobe, and lung adenocarcinoma). Further, TE subfamilies show a specific expression across all the tumor types. Out of the 19,057 TE subfamilies, 587 display a different expression in at least one TCGA cancer, 80% of which were overexpressed and belonged to LTR and LINE classes. The overexpression of LTRs, particularly HERV, has been found in different epithelial tumor types, including colorectal cancer (CRC) [101], melanoma [102,103], renal cell carcinoma [104], pancreatic adenocarcinoma [105], glioma [98], as well as breast [106] and ovarian cancer [107]. L1 has been reported as aberrantly expressed in many cancers (breast, head and neck, lung [108]), and CRC [109]. Interestingly, in CRC patients, L1 expression depends on DNA damage repair ability, where Microsatellite Instable (MSI, mutated in the DNA repair machinery) cancers show a lower expression of L1 in respect to Microsatellite Stable (MSS, not mutated), reflecting different DNA methylation levels [109].

TEs deregulation in cancer can contribute to TEs use as promoters [110]. The promoter activation can further lead to the expression of the genes surrounding TEs, which may contribute with TEs to tumorigenesis in a synergistic or cooperative manner [110] (Figure 3A). The specific mechanism by which TEs are able to promote the tumorigenesis and tumor progression remain mostly unknown [111]. Within this frame, a possibly critical function for L1 expression has been suggested in early phases of cancer formation, setting up the gene expression profiles favorable to tumor development [112].

Indeed, in CRC a correlation has been described between disease stage and L1 hypomethylation and expression [113]. This finding has also been confirmed for ERV in the endometrial carcinoma growth [114], suggesting a correlation between TEs and the establishment of a cancer specific program.

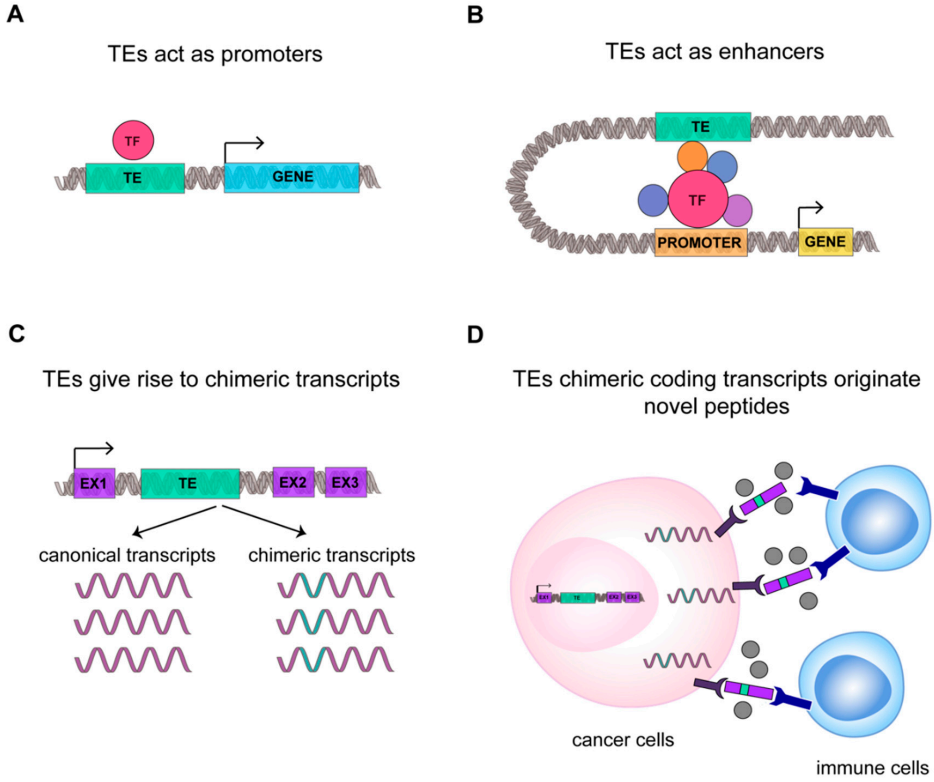


Figure 3. TEs transcriptome contributes to cancer transcriptional fingerprint. A schematic representation of new function mediated by TEs in cancer: (**A**) TE (in green) can act as promoter sequence or (**B**) enhancer sequence. Transcription Factor and cofactors (TF) are highlighted in red and violet. (**C**) TEs can generate new chimeric transcripts, (**D**) giving origin to new oncogene transcripts and peptides that can be recognized by immune system as not-self, improving cancer immunogenicity.

Collectively, these findings suggest that TEs deregulation can be specifically involved in the establishment of cancer-specific transcriptional programs, suggesting that transposons can be co-opted for cancer fitness and survival, and that these elements could be used in defining novel molecular classifications.

3.2. TEs RNAs Improve Cancer Specific Transcriptional Complexity and Plasticity

TEs can contribute to cancer specific functions acting at different level of transcriptional regulation. It has been reported that TEs, dispersed across the human genome, represent a huge reservoir of gene regulatory modules, both promoters and enhancers [110,115,116] and that transposons can mediate the genesis of new transcripts [116–118], possibly contributing to the translation of new cancer-specific peptides [95].

Cancer-specific co-option of transposable elements takes the name of onco-exaptation, a term coined by Babaian et al., to describe the mechanism by which epigenetically repressed TEs have been harnessed to promote human oncogenesis (Figure 3) [115]. Babaian et al., analyzed RNA-seq datasets from nine Hodgkin lymphomas (HL), finding that proinflammatory transcription factor IRF5 was upregulated in HL-derived cell lines due to the transcriptional activation of the retroviral LOR1a LTR as regulatory enhancer (Figure 3B) [115]. Similarly, in 2010, Wolff et al. found in bladder cancer that the demethylation of a specific L1 promoter induces activation of an alternative transcript of the *MET* gene, that codifies for a permanently active MET protein, a tyrosine kinase receptor that

promotes tumor growth (Figure 3A) [110]. For ALK-negative anaplastic large cell lymphoma it was demonstrated that the reactivation of LTRs causes the expression of a novel isoforms of the oncogene *ERBB4*, a type I receptor tyrosine kinase [119]. In agreement, Jang et al. characterized the activation of TE-derived cryptic promoter in 7769 tumors and 625 normal tissues, found TE-derived oncogene transcripts in 15 cancer types, and proposed that this mechanism contributes to oncogene activation in about half of all malignant diseases [116]. This result suggests that transposons can contribute to the genesis of new chimeric transcripts (Figure 3C). Likewise, the activation of L1 promoters, due to the loss of DNA methylation, can promote the transcription of nearby regions, generating cancer-specific L1 chimeric transcripts [117]. The L1-chimeric transcript *LCT13*, transcribed from the L1 antisense promoter, behaves as ncRNA silencing the tumor suppressor TFPI-2 and promoting cancer progression [118]. In diffuse large B-cell lymphomas 98 TE-gene chimeric transcripts have been found and the expression of *LTR2-FABP7* chimeric transcript was suggested to code for a novel protein able to positively influence diffuse large B-cell lymphoma cell proliferation [120]. Notably, in different cancers, 83 unique peptides derived from TEs chimeric transcripts have been identified, among which 39 were common to different tumor tissues [95]. We can hypothesize that these new peptides originating from TEs chimeric transcripts could increase the number of cancer-associated neoantigens, possibly rendering tumor more susceptible to immunotherapy (Figure 3D) [95].

Collectively, these data suggest that TEs can be deeply involved in orchestrating cancer type–specific regulatory networks, increasing cancer transcriptional complexity and plasticity, and further promoting tumor adaptability and fitness.

3.3. TEs Regulate Cancer Tumorigenicity and Progression

It is accepted that TE mediated retrotransposition can act at the genomic level, promoting genome instability and cancer progression [16,121,122]. However, very little is known regarding the functional correlation between TEs expression and cancer establishment and progression.

In pancreatic cancer cell lines, it has been demonstrated that L1 and HERVK silencing reduce the tumorigenicity of the cells inoculated in nude mice [105]. Aschacher et al. [123] show evidence that downregulation of L1 in different cancer cell lines induces telomere shortening and consequently slower spheroid cancer cell development, promoting a G2 phase cell cycle arrest and suggesting L1 involvement in cancer cell proliferation. In hepatocellular carcinoma it has been proposed that L1 RNAs can have a function in cellular transformation, through a splicing-mediated regulation of the protooncogene G antigen 6 (GAGE6). This suggests that endogenous L1 RNA may display regulatory functions in the process of tumorigenesis [124]. Finally, functional analyses in leukemic stem cells revealed a specific contribution of TEs classes to inflammation, the expression of SINE and LTR positively correlating with inflammation, and L1 anticorrelating [125]. The authors add clinical values to this discovery, hypothesizing that TEs can be used and targeted to modulate the immune response with tumor microenvironment.

Other possible mechanisms by which TEs can promote cancer progression are described for HERVs elements, that could initiate the transcription of ncRNA and lncRNAs with oncogenic properties [126]. The 5′ end of LTR7 element induces the expression of the pro-oncogenic lncRNA ROR [45] in different cancer types [127] while a HERVK 11 ncRNA binds to polypyrimidine tract-binding protein-associated splicing factor, inhibiting the repression of proto-oncogene transcription and consequently leading to cell transformation and tumorigenesis [128,129]. However, HERVs RNAs can display also tumor suppressive function, as for the antisense transcript of ERV-9 LTR that in normal cells physically binds transcription factors involved in cell proliferation, and that is downregulated in malignant cells sustaining uncontrolled cancer growth [130,131].

An additional evolutionary way by which TEs RNAs drive tumor progression are the microvesicles, small lipid bilayer extracellular vesicles released from cells that could contain RNAs, DNA fragments, peptides, and lipids [132]. Retroviral-like microvesicles have been found in the plasma of cancer patients [133–135]. In particular, in vitro studies showed that tumor-derived microvesicles are enriched

in HERV, L1, and Alu DNA and RNAs, that could be transmitted to other cancer and normal cells thanks to macrovesicles fusion with the cellular membranes [136]. Microvesicles are recently shown to drive cancer growth and proliferation and to regulate near or distant healthy cells within tumor microenvironment [126].

Combined, these data strongly suggest that TEs can be specifically involved in promoting tumorigenesis and cancer progression in a wide set of cancer types acting via different mechanisms, that, beside the more obvious genomic effects of the retrotransposition, surprisingly also involve the RNA counterpart of these elements.

4. Next-Generation Sequencing (NGS) Approaches for the Analysis of TEs

4.1. Dealing with Ambiguity in RNA-Seq Reads Alignment: A Challenge to Resolve TEs Expression Quantification

TEs have been co-opted in different biological scenarios representing novel molecules able to regulate the tissue specific transcriptional networks that establish in physiological and pathological context. The advent of evolving NGS technologies, the formation of international consortia that produced a multitude of datasets and developed bioinformatic tools have been indispensable for realizing how broad is TEs involvement in mammalian biology, and depicting precise function for certain classes, superfamilies and subfamilies of TEs in a given spatiotemporal frame. However, in order to precisely define the contribution of a given TEs locus to the regulatory networks of specific genes, it is important to identify and characterize TEs at the genomic instance resolution. A systematic and unambiguous analysis of TEs (that are repeated in several highly homologous interspersed genomic loci) at the genomic instance level or within genes containing TEs using RNA-seq is a non-trivial task (Figure 4), due to the limitations of mapping algorithms, which do not allow the assignment of multi-mapping reads to a precise genomic locus [137].

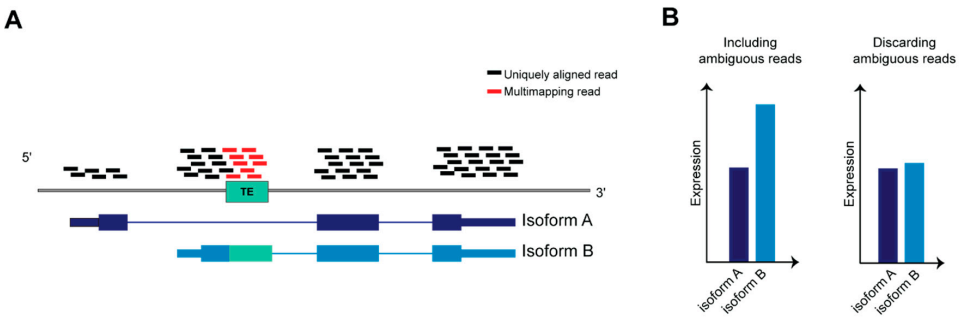


Figure 4. Ambiguous reads in transcript quantification. (**A**) Schematic representation of RNA-seq reads aligned on a gene on the reference genome, the gene is transcribed in two transcript isoforms, A and B. (**B**) Isoform B is twice more abundant than A; however, if ambiguous reads are discarded from reads count, the difference between A and B will be negligible after normalizing read counts against transcript length.

Here, we provide a comprehensive overview of the technological progresses in NGS technologies and computational methods, from the sequencing design (e.g., read length and pairing) to the development of specific tools for the downstream analysis of TEs annotation and expression. Also provided is an outline on the contribution of the knowledge that we have acquired and previously summarized on TE functions in genome biology.

Some precautions in the library preparation can help mitigating the amount of multi-mapping TE-derived reads, such as using a paired-end layout and a longer read length to make more likely that the read will contain a unique genomic sequence that can be mapped. However, long repeat instances, such as LTR and LINE retrotransposons, can span from hundreds to thousands of nucleotides, challenging an unambiguous identification via the current, state-of-the-art RNA-seq protocols. Some

of the longest TEs harbor an intact promoter and ORF sequences, and are therefore able to be transcribed and to retrotranspose under conditions that cause the removal of their repression, such as hypomethylation in cancer (see Section 3.1). Therefore, being able to resolve the quantification of these TEs can be crucial to properly study the contribution of transposons in such pathological conditions.

Since the early years of the NGS era, multi-reads have been handled in different ways, each with its own advantages and drawbacks: i) ignoring multi-reads by selecting unique alignments only. This option may lead to underestimating the expression levels of TEs and their derivates, as well as the overall expression level of a sample, but assigns reads with the highest confidence; ii) reporting the best alignment for each multi-mapping read based on the alignment quality score calculated by the mapping algorithm. Here, the results may vary based on how mismatches and gaps between the reads and the reference genome are weighted, making it difficult to provide the exact genomic location with high confidence; iii) keeping multi-reads, counting them once for each mapped feature. This prevents discarding potentially relevant loci from the downstream analysis. However, genomic features characterized by a high number of multi-reads, as well as the total library size, will be overestimated.

To avoid discarding relevant biological signals from multi-mapping or ambiguous reads, multi-mapping reads should be either assigned to a unique genomic feature or re-distributed across the multi-mapped regions. To accomplish the assignation to a unique genomic feature, available methods implement algorithms to assign, according to different criteria (see below), the genomic feature that is the source of transcription for those reads. Whenever this is not possible, the reads can be assigned computing a probability, they will be proportionally re-distributed across the mapped genomic features according to how likely they are to be the source of transcription (often based on the level of transcription of the genomic features, see below). This approach offers a more precise estimation of expression and reads coverage across genes, and some of the methods implementing it are discussed below.

In 2008, Mortazavi et al. [138] depicted one of the earliest efforts in this direction, in which multi-reads are recovered by distributing them across the aligned genes, proportionally to the amount of unique alignments on a given gene. This method resulted in an increase of expression levels estimates by more than 30% compared to discarding multi reads, for several mouse genes.

The importance to use multi-reads in gene expression profiling of cancer, has been more recently considered by Robert and Watson [139] with a survey on 12 common methods for gene-level expression quantification from RNA-seq data. The expression levels of hundreds of genes are underestimated by one or more of those methods; interestingly, many of these genes are implicated in human diseases. The quantification of such genes is proposed via multi-map groups (MMGs) of genes that multi-reads map to, and by this approach, MMGs are differentially expressed between normal and lung tumour mouse cells, while the methods based on unique counts failed to produce this result [139]. By avoiding quantifying the expression of individual ambiguous genes, Robert and Watson could retrieve important data that otherwise would have been missed, but, on the other hand, the information on the transcripts is not considered in the analysis. This technical gap was filled by the multi-mapper resolution tool (MMR), developed by Kahles et al. in 2016 [140]. In contrast to the previous methods, MMR returns an expression estimate for each individual gene or transcript, and it does not proportionally distribute multi-reads across the aligned features. Rather, MMR assumes that the reads coverage should be uniform within a local region, thus selecting the alignment that leads to the smoothest coverage signal across a window of a fixed length.

Recently, pseudo-alignment algorithms emerged as an alternative to aligning RNA-seq reads to a reference genome, by directly inferring the transcript from which the read originates [141,142]. The ambiguity of highly overlapping transcripts in the human genome is circumvented by probabilistically distributing the reads count across a given transcriptome, avoiding the generation of multi-reads in the first place. Tools based on pseudo-alignment have become valuable in transcriptomics, providing a fast and reliable method for transcript-level quantification.

Besides RNA-seq, several NGS methods are designed to meet specific needs in transcriptome analysis. Among these, cap analysis of gene expression (CAGE) is an high-throughput technology for sequencing the 5′ end of transcripts into short reads (tags) [143]. CAGE has been proven valuable for the discovery of novel transcription start sites (TSS) of either novel genes or alternative transcript isoforms of known genes [144]. Faulkner et al. [145] showed a method to recover short multi-reads produced by tag-based NGS technologies such as CAGE, in which a score is given to tag-TSS associations according to the amount of individual tags associated to the same TSS; multi-mapping tags are proportionally assigned to the mapped TSS according to the calculated scores. With this method, it has been demonstrated that up to 30% of transcripts initiate from within TEs [36], and that some of them are associated with enhancer regions in stem cells, regulating their pluripotency [46].

Therefore, rescuing multi-mapping CAGE tags, or multi-reads in other NGS technologies complementary to RNA-seq, has been fundamental in clarification of the extent to which TEs influence the transcriptional output of mammalian cells in both physiological and pathological contexts.

4.2. Current Computational Methods for TEs Transcriptome Analysis

General-purpose computational methods, such as the aforementioned ones, help with the recovering of ambiguous reads for their inclusion in downstream analyses, including those originating from TEs. However, some contexts of analysis require complementary specialized tools designed for TEs to survey the overall contribution of the various TE categories to the transcriptional output of a certain tissue, or to be able to properly distribute the RNA-seq signal among active TE instances and TEs expressed as part of other transcripts.

Several TE-centric tools have been developed to (i) identify and quantify expressed TEs from transcriptomic datasets that can be classified based on their capability of quantifying TE expression at the subfamily level (counting a subfamily as an individual entity) or at genomic instance level (to quantify the expression of individual elements), and (ii) discern TEs that are actively transcribed as individual transcriptional units from those that are co-expressed within other transcripts (Table 1).

Table 1. Computational tools and pipelines for transposable elements (TEs) transcriptome analysis.

Name	Resolution	TE Specificity	Detection of Active Transcription	Method Description	Reference
REdiscoverTE	Subfamily	All	Yes (Intergenic TEs are classified as autonomously transcribed)	Pseudo-alignment on a transcriptome of cDNA and individual genomic loci.	[95]
L1EM	Locus-level	LINE1	Yes	Categorizes L1 loci by the presence of promoter and polyA tail; EM-based quantification.	[146]
LIONS	Locus-level	TEs initiating transcripts	No	Identify and quantify TE-initiated transcripts based on read coverage on *de-novo* reconstructed exons and around TEs.	[147]
RepEnrich	Subfamily	All	No	Non-spliced alignment on a pseudo-genome of repeats sequences.	[148]
SalmonTE	Subfamily	All	No	Pseudo-alignment on TE consensus sequences.	[149]
SQuIRE	Locus-level	All	No	Spliced alignment followed by EM-based locus-level quantification.	[150]
TEcandidates	Locus-level	All	No	Alignment of *de novo* assembled contigs of TE-derived reads to the reference genome.	[151]
Telescope	Locus-level	All	No	Reassignment of multi-reads to the most probable source of transcript.	[152]
TEtools	Subfamily	All	No	Reference-free alignment on a provided set of TE sequences.	[153]
TEtranscripts	Subfamily	All	No	EM-based re-distribution of pre-aligned multi-reads.	[154]
TeXP	Subfamily	All	Yes	Removes *noise* derived from non-autonomous transcription of TEs.	[155]

From left to right: name of the software, resolution of expression estimation (e.g., TE (sub)family or locus-level), specificity of the software towards a particular category of TEs, ability of the software to discern autonomous from passive TE transcription, brief description of the method, reference of the associated publication. All the software listed in this table, including the source code, are freely available.

Criscione et al. [146] published RepEnrich in 2014. They rescued most multi-reads by assigning them proportionally to the subfamilies on which they align, and showed that many TEs subfamilies

are expressed in a tissue-specific manner, and significantly enriched in cancer [148]. Recently, Jung et al. [156] used TEtranscripts to improve the expression estimate of L1Hs in cancer, potentially active in the human genome. By quantifying L1Hs somatic insertions and their overall expression in whole-genome and RNA sequencing data from matched TCGA gastrointestinal cancer samples, they found that L1 insertions count and expression are significantly higher in cancer tissues compared to normal, and that L1 insertions causes abnormal mRNA splicing and gene expression [156].

TEtranscripts does not discern potentially autonomously transcribed TEs from pervasively transcribed ones. To do that, Navarro et al. [155] recently released TeXP method that removes the noise due to pervasive transcription from the RNA-seq signal mapping on evolutionarily young subfamilies. [155]. They applied this method in several RNA-seq datasets from cancer and healthy human cell lines and tissues, and found a greater amount of autonomous transcription for transposons in the human germline and in tumor cell lines.

A different approach to quantify the expression of TEs at class, superfamily or subfamily level is to align RNA-seq reads on a custom transcriptome of TEs sequences, rather than a reference genome. TEtools [153] is a pipeline that works in this way, enabling the analysis of a TE transcriptome by providing the sequences of TE instances and computing a class-superfamily-subfamily level count and a differential expression analysis. A recent work by Cebrià-Costa et al. used TEtools to perform a differential expression analysis of TEs in an epigenetic study on the function of histone 3 lysine 4 oxidation by LOXL2 in breast cancer cells, and to rule out the possibility that the overexpression of TEs were responsible for DNA damage response in LOXL2 KD cells [157].

As aforementioned, pseudo-alignment can quantify transcripts including both unique and ambiguous reads, avoiding the generation of multi-reads. Recently, TE-centric pipelines based on pseudo-alignment have been released as SalmonTE [149] and REdiscoverTE [95], that both leverage on Salmon's pseudo-alignment algorithm. Kong et al. illustrate REdiscoverTE using over five million genomic repetitive elements annotated by RepeatMasker [158] together with cDNA transcript sequences as well as the sequences of introns containing repetitive elements. They show that including all genomic repeats instances in the reference transcriptome allows taking in account the sequence diversity within TE subfamilies. This includes eventual genomic TE loci that significantly deviate from the Repbase consensus sequence, and results in a more accurate quantification of TE hierarchies. Further, the inclusion of intronic sequences containing repetitive elements allows mapping reads on TEs transcribed within unannotated alternative exons or retained introns. By applying this pipeline on 7750 TCGA cancer samples, Kong and colleagues [95] described the TE expression landscape in cancer, differentiated between the TEs co-expressed within host genes and intergenic TEs, and found the latter more expressed and more correlated with DNA demethylation, DNA damage and immune response in cancer [95].

Measuring the expression enrichment of TEs in RNA-seq data when comparing different cell types, developmental stages or pathological conditions can provide important evidences on the regulatory network in which TEs are involved. However, to deeply investigate TEs involvement in a specific mechanism or phenotype, it is crucial to study TEs expression at the individual genomic instance resolution. Indeed, for example, a different function would be expected for evolutionarily old TEs in respect to the youngest ones that own a promoter and are able to retrotranspose in the genome. For this purpose, Yang et al. published SQuIRE in 2019, the first bioinformatics tool designed for locus-specific quantification of interspersed repeats [150], based on the spliced alignment of a reference genome of RNA-seq data. By applying this method, they show a differential expression of individual TE instances across different tissues of healthy mouse, as well as of TEs differentially expressed in a *D. Melanogaster* model of amyotrophic lateral sclerosis, highlighting the structure of the transcripts containing such TEs, that would not have been possible without a locus-level resolution.

Besides SQuIRE, other tools reports the expression estimates of TEs at genomic instance level [146,151] by L1EM tool [146] that has been developed to quantify the expression of autonomously transcribed L1 elements at locus level. As reported by L1EM analysis, full-length L1 loci of the L1Hs

subfamily are highly expressed in stem and cancer cells, while being less expressed in differentiated tissue samples.

Bioinformatics analyses in TE-centric studies may not be limited to the expression of TEs instances. As we reviewed, TEs influence the transcription of coding and non-coding RNAs in several ways [36]. Jang et al. characterized the landscape of TE onco-exaptation across RNA-seq data from TCGA tumors and normal samples, which they reanalyzed using a pipeline for transcript assembly and integrated with data from the FANTOM5 consortium for the annotation of TE-derived transcription start sites [116]. This analysis revealed the prevalence of TE usage as novel regulatory sequences in cancer and its importance for oncogene activation and tumorigenesis. In this context, a recent tool, LIONS [147], is specifically designed to detect and quantify transcripts initiated from within TEs. This tool is able to estimate expression levels of both TEs and exons, and to compute a specific metric to discern TE-initiation from TE-exonization events based on read coverage. Finally, if more than one experimental group is being processed, LIONS performs a differential analysis between them.

Alternative approaches for an accurate quantification of TEs expression could also use data generated by new technologies, although less available than RNA-seq. For example, Deininger et al. developed a pipeline based on RNA-seq and 5′ RACE coupled with PACbio sequencing of 1200 base pair-long reads to estimate the expression of L1 RNAs expressed as independent transcriptional units [159]. In particular, they show that a large part of the total expression of full-length L1 elements derives by the transcription of a relatively small number of L1 loci. Indeed, this method anticipates the potential of long read sequencing in identifying the TEs contributing to the majority of expression and new insertions in several cancer conditions. Indeed, recent advancements on long-read sequencing, that obtain and map tens of thousands of base-pair long reads, should allow to identify the TEs expressed and contributing to new insertions in cancer conditions, and may signal a new era for the analysis of TEs in transcription regulation, other than for genomics as a whole [160,161].

Despite the limitations of NGS technologies for studying interspersed repetitive elements, recent efforts in bioinformatic research have undoubtedly reached the goal of increasing the level of confidence by which the expression levels of such elements are estimated, and enabled the discovery of several transcriptional regulatory networks in which they are involved in physiological and pathological conditions. Nonetheless, further efforts are still required to improve bioinformatic practices and increase the awareness of the biological relevance of the once called “junk DNA”.

5. Conclusions

In the current review, we summarized the latest findings on TEs, highlighting that, beyond their ability of being “jumping elements”, they contribute to the establishment of a vast regulatory network that, controlling genome plasticity, magnifies the cell type specific transcriptional complexity, both in health and diseases.

Although TEs are finely transcriptionally regulated in order to avoid the negative effects of their transposition, TEs activation physiologically takes part within the concerted spatio-temporal establishment of the cellular transcriptional programs. Indeed, we find that TEs can mediate multi-layered regulatory functions in cellular identity establishment in embryogenesis and development (see subheading 1) and that using the host signaling pathways, TEs are key players in the regulation of adult tissue plasticity and adaptation to environment, as occurs in the innate and adaptive immune response (see Section 2). Further, we also highlighted how TEs can represent evolutionary instruments that create novel functions that can be positively selected to promote cancer fitness and tumorigenesis (see Section 3). Importantly, we showed that many of these findings have been achieved through the employment of NGS sequencing technologies and bioinformatic tools that are in continuous development and improvement to overcome the limitations of TEs study that render their unambiguous annotation and analysis still puzzling (see Section 4).

Concluding, TE mediated regulatory networks represent a prolific source of still hidden players that could explain complex phenomena such as the establishment and progression of multifactorial

diseases. Transposons represent a new window for novel therapeutic opportunities and for deriving targetable molecules for personalized based therapies. For these purposes, it is necessary to develop computational and experimental methods to identify and characterize more systematically TEs at their genomic instances, in order to improve our knowledge about their implications in genome plasticity and functions in health and disease.

Author Contributions: F.M. performed literature searches, drafted the literature review and the figures, revised and finalized the manuscript. E.G., B.P., and R.V. performed literature searches, drafted the literature review and the figures. V.R. revised the manuscripts and the figures. B.B. conceptualized, revised, and finalized the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by the following grants to B.B.: Fondazione Regionale per la Ricerca Biomedica (FRRB CP2_12/2018) and Fondazione Cariplo (grant nr 2019-3416.) This work has been supported by the following grant to F.M.: Fondazione Cariplo (Bando Giovani, grant nr 2018-0321). This work has been supported by the following grant to V.R.: Fondazione Cariplo (Bando Giovani, grant nr 2019-1788).

Acknowledgments: Erica Gasparotto is a PhD student within the European School of Molecular Medicine (SEMM). Rebecca Vadalà is a PhD student within PhD Program in Translational and Molecular Medicine (DIMET).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

LINE	Long interspersed element
L1	Long interspersed element 1
SINE	Short interspersed element
LTR	Long terminal repeat
TE	Transposable element
ERV	Endogenous retrovirus
NGS	Next generation sequence
IFN	Interferon
NK	Natural killer
DC	Dendritic cell
NF-κB	Nuclear factor kappa B
AP-1	Activator protein-1
IRF	Interferon response factor
PRR	Pattern recognition receptor
PAMP	Pathogens associated molecular patterns
RLR	retinoic acid-inducible gene I-like receptors
RIG-I	Retinoic acid inducible gene-I
MDA5	Melanoma differentiation-associated gene 5
TLR	Toll like receptor
cGAS	GMP-AMP synthase
AID	Activation-induced cytidine deaminase
ss	Single strand
ds	Double strand
lnc	Long non coding
CRC	Colon rectal cancer
AGS	Aicardi-Goutières Syndrome
TREX1	DNA sensing pathways 3'repair exonuclease
ADAR1	Adenosine deaminase acting on RNA
DNMTi	DNA methyltransferase inhibitors
AML	Acute myeloid leukemia
CDK4/6	Cyclin dependent kinase 4/6
SLE	Systemic lupus erythematosus
ALS	Amyothropic latelar sclerosis
CIC	Cancer initiating cell
MSI	Microsatellite instable

MSS	Microsatellite stable
HL	Hodgkin Lymphoma
GAGE6	protooncogene G antigen 6

References

1. De Koning, A.P.; Gu, W.; Castoe, T.A.; Batzer, M.A.; Pollock, D.D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* **2011**, *7*, e1002384. [CrossRef] [PubMed]
2. Lander, E.S.; Linton, L.M.; Birren, B.; Nusbaum, C.; Zody, M.C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860–921. [CrossRef] [PubMed]
3. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory activities of transposable elements: From conflicts to benefits. *Nat Rev. Genet.* **2017**, *18*, 71–86. [CrossRef] [PubMed]
4. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* **2008**, *9*, 397–405. [CrossRef] [PubMed]
5. Percharde, M.; Sultana, T.; Ramalho-Santos, M. What Doesn't Kill You Makes You Stronger: Transposons as Dual Players in Chromatin Regulation and Genomic Variation. *Bioessays* **2020**, *42*, e1900232. [CrossRef] [PubMed]
6. Kazazian, H.H., Jr.; Moran, J.V. The impact of L1 retrotransposons on the human genome. *Nat. Genet.* **1998**, *19*, 19–24. [CrossRef]
7. Viollet, S.; Monot, C.; Cristofari, G. L1 retrotransposition: The snap-velcro model and its consequences. *Mob. Genet. Elements* **2014**, *4*, e28907. [CrossRef]
8. Okada, N.; Hamada, M.; Ogiwara, I.; Ohshima, K. SINEs and LINEs share common 3′ sequences: A review. *Gene* **1997**, *205*, 229–243. [CrossRef]
9. Malik, H.S.; Henikoff, S.; Eickbush, T.H. Poised for contagion: Evolutionary origins of the infectious abilities of invertebrate retroviruses. *Genome Res.* **2000**, *10*, 1307–1318. [CrossRef]
10. Canapa, A.; Barucca, M.; Biscotti, M.A.; Forconi, M.; Olmo, E. Transposons, Genome Size, and Evolutionary Insights in Animals. *Cytogenet Genome Res.* **2015**, *147*, 217–239. [CrossRef]
11. Kazazian, H.H., Jr. Mobile elements: Drivers of genome evolution. *Science* **2004**, *303*, 1626–1632. [CrossRef] [PubMed]
12. Belancio, V.P.; Hedges, D.J.; Deininger, P. Mammalian non-LTR retrotransposons: For better or worse, in sickness and in health. *Genome Res.* **2008**, *18*, 343–358. [CrossRef]
13. Muotri, A.R.; Chu, V.T.; Marchetto, M.C.; Deng, W.; Moran, J.V.; Gage, F.H. Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **2005**, *435*, 903–910. [CrossRef]
14. Beck, C.R.; Collier, P.; Macfarlane, C.; Malig, M.; Kidd, J.M.; Eichler, E.E.; Badge, R.M.; Moran, J.V. LINE-1 retrotransposition activity in human genomes. *Cell* **2010**, *141*, 1159–1170. [CrossRef]
15. Ewing, A.D.; Kazazian, H.H., Jr. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res.* **2010**, *20*, 1262–1270. [CrossRef] [PubMed]
16. Iskow, R.C.; McCabe, M.T.; Mills, R.E.; Torene, S.; Pittard, W.S.; Neuwald, A.F.; Van Meir, E.G.; Vertino, P.M.; Devine, S.E. Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* **2010**, *141*, 1253–1261. [CrossRef] [PubMed]
17. Baillie, J.K.; Barnett, M.W.; Upton, K.R.; Gerhardt, D.J.; Richmond, T.A.; De Sapio, F.; Brennan, P.M.; Rizzu, P.; Smith, S.; Fell, M.; et al. Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* **2011**, *479*, 534–537. [CrossRef]
18. Perrat, P.N.; DasGupta, S.; Wang, J.; Theurkauf, W.; Weng, Z.; Rosbash, M.; Waddell, S. Transposition-driven genomic heterogeneity in the Drosophila brain. *Science* **2013**, *340*, 91–95. [CrossRef]
19. Upton, K.R.; Gerhardt, D.J.; Jesuadian, J.S.; Richardson, S.R.; Sanchez-Luque, F.J.; Bodea, G.O.; Ewing, A.D.; Salvador-Palomeque, C.; van der Knaap, M.S.; Brennan, P.M.; et al. Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* **2015**, *161*, 228–239. [CrossRef]
20. Coufal, N.G.; Garcia-Perez, J.L.; Peng, G.E.; Yeo, G.W.; Mu, Y.; Lovci, M.T.; Morell, M.; O'Shea, K.S.; Moran, J.V.; Gage, F.H. L1 retrotransposition in human neural progenitor cells. *Nature* **2009**, *460*, 1127–1131. [CrossRef]

21. Reilly, M.T.; Faulkner, G.J.; Dubnau, J.; Ponomarev, I.; Gage, F.H. The role of transposable elements in health and diseases of the central nervous system. *J. Neurosci* **2013**, *33*, 17577–17586. [CrossRef] [PubMed]
22. Saleh, A.; Macia, A.; Muotri, A.R. Transposable Elements, Inflammation, and Neurological Disease. *Front. Neurol* **2019**, *10*, 894. [CrossRef] [PubMed]
23. Payer, L.M.; Burns, K.H. Transposable elements in human genetic disease. *Nat. Rev. Genet.* **2019**, *20*, 760–772. [CrossRef] [PubMed]
24. Deniz, O.; Frost, J.M.; Branco, M.R. Regulation of transposable elements by DNA modifications. *Nat. Rev. Genet.* **2019**, *20*, 417–431. [CrossRef] [PubMed]
25. Mc, C.B. The origin and behavior of mutable loci in maize. *Proc. Natl Acad Sci USA* **1950**, *36*, 344–355. [CrossRef]
26. McClintock, B. Controlling elements and the gene. *Cold Spring Harb Symp Quant. Biol* **1956**, *21*, 197–216. [CrossRef]
27. Sundaram, V.; Cheng, Y.; Ma, Z.; Li, D.; Xing, X.; Edge, P.; Snyder, M.P.; Wang, T. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* **2014**, *24*, 1963–1976. [CrossRef]
28. Bodega, B.; Orlando, V. Repetitive elements dynamics in cell identity programming, maintenance and disease. *Curr. Opin. Cell Biol.* **2014**, *31*, 67–73. [CrossRef]
29. Pennisi, E. Genomics. ENCODE project writes eulogy for junk DNA. *Science* **2012**, *337*, 1159–1161. [CrossRef]
30. Bourque, G.; Leong, B.; Vega, V.B.; Chen, X.; Lee, Y.L.; Srinivasan, K.G.; Chew, J.L.; Ruan, Y.; Wei, C.L.; Ng, H.H.; et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* **2008**, *18*, 1752–1762. [CrossRef]
31. Imbeault, M.; Helleboid, P.Y.; Trono, D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **2017**, *543*, 550–554. [CrossRef] [PubMed]
32. Morgan, H.D.; Sutherland, H.G.; Martin, D.I.; Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* **1999**, *23*, 314–318. [CrossRef] [PubMed]
33. Ferrari, R.; de Llobet Cucalon, L.I.; Di Vona, C.; Le Dilly, F.; Vidal, E.; Lioutas, A.; Oliete, J.Q.; Jochem, L.; Cutts, E.; Dieci, G.; et al. TFIIIC Binding to Alu Elements Controls Gene Expression via Chromatin Looping and Histone Acetylation. *Mol. Cell* **2020**, *77*, 475–487 e411. [CrossRef]
34. Schmidt, D.; Schwalie, P.C.; Wilson, M.D.; Ballester, B.; Goncalves, A.; Kutter, C.; Brown, G.D.; Marshall, A.; Flicek, P.; Odom, D.T. Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **2012**, *148*, 335–348. [CrossRef] [PubMed]
35. Zhang, Y.; Li, T.; Preissl, S.; Amaral, M.L.; Grinstein, J.D.; Farah, E.N.; Destici, E.; Qiu, Y.; Hu, R.; Lee, A.Y.; et al. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat. Genet.* **2019**, *51*, 1380–1388. [CrossRef]
36. Faulkner, G.J.; Kimura, Y.; Daub, C.O.; Wani, S.; Plessy, C.; Irvine, K.M.; Schroder, K.; Cloonan, N.; Steptoe, A.L.; Lassmann, T.; et al. The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* **2009**, *41*, 563–571. [CrossRef]
37. Rodriguez-Terrones, D.; Hartleben, G.; Gaume, X.; Eid, A.; Guthmann, M.; Iturbide, A.; Torres-Padilla, M.E. A distinct metabolic state arises during the emergence of 2-cell-like cells. *EMBO Rep.* **2020**, *21*, e48354. [CrossRef]
38. Lu, J.Y.; Shao, W.; Chang, L.; Yin, Y.; Li, T.; Zhang, H.; Hong, Y.; Percharde, M.; Guo, L.; Wu, Z.; et al. Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Rep.* **2020**, *30*, 3296–3311 e3295. [CrossRef]
39. Attig, J.; Agostini, F.; Gooding, C.; Chakrabarti, A.M.; Singh, A.; Haberman, N.; Zagalak, J.A.; Emmett, W.; Smith, C.W.J.; Luscombe, N.M.; et al. Heteromeric RNP Assembly at LINEs Controls Lineage-Specific RNA Processing. *Cell* **2018**, *174*, 1067–1081 e1017. [CrossRef]
40. Nekrutenko, A.; Li, W.H. Transposable elements are found in a large number of human protein-coding genes. *Trends Genet.* **2001**, *17*, 619–621. [CrossRef]
41. Perepelitsa-Belancio, V.; Deininger, P. RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat. Genet.* **2003**, *35*, 363–366. [CrossRef] [PubMed]
42. Roy-Engel, A.M.; El-Sawy, M.; Farooq, L.; Odom, G.L.; Perepelitsa-Belancio, V.; Bruch, H.; Oyeniran, O.O.; Deininger, P.L. Human retroelements may introduce intragenic polyadenylation signals. *Cytogenet Genome Res.* **2005**, *110*, 365–371. [CrossRef] [PubMed]

43. Gong, C.; Maquat, L.E. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3′ UTRs via Alu elements. *Nature* **2011**, *470*, 284–288. [CrossRef] [PubMed]
44. Kapusta, A.; Kronenberg, Z.; Lynch, V.J.; Zhuo, X.; Ramsay, L.; Bourque, G.; Yandell, M.; Feschotte, C. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* **2013**, *9*, e1003470. [CrossRef] [PubMed]
45. Kelley, D.; Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* **2012**, *13*, R107. [CrossRef]
46. Fort, A.; Hashimoto, K.; Yamada, D.; Salimullah, M.; Keya, C.A.; Saxena, A.; Bonetti, A.; Voineagu, I.; Bertin, N.; Kratz, A.; et al. Deep transcriptome profiling of mammalian stem cells supports a regulatory role for retrotransposons in pluripotency maintenance. *Nat. Genet.* **2014**, *46*, 558–566. [CrossRef] [PubMed]
47. Lu, X.; Sachs, F.; Ramsay, L.; Jacques, P.E.; Goke, J.; Bourque, G.; Ng, H.H. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat. Struct. Mol. Biol.* **2014**, *21*, 423–425. [CrossRef]
48. Hall, L.L.; Carone, D.M.; Gomez, A.V.; Kolpa, H.J.; Byron, M.; Mehta, N.; Fackelmayer, F.O.; Lawrence, J.B. Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* **2014**, *156*, 907–919. [CrossRef]
49. Jachowicz, J.W.; Bing, X.; Pontabry, J.; Boskovic, A.; Rando, O.J.; Torres-Padilla, M.E. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* **2017**, *49*, 1502–1510. [CrossRef]
50. Fadloun, A.; Le Gras, S.; Jost, B.; Ziegler-Birling, C.; Takahashi, H.; Gorab, E.; Carninci, P.; Torres-Padilla, M.E. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct Mol. Biol.* **2013**, *20*, 332–338. [CrossRef]
51. Percharde, M.; Lin, C.J.; Yin, Y.; Guan, J.; Peixoto, G.A.; Bulut-Karslioglu, A.; Biechele, S.; Huang, B.; Shen, X.; Ramalho-Santos, M. A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* **2018**, *174*, 391–405 e319. [CrossRef] [PubMed]
52. Hoebe, K.; Janssen, E.; Beutler, B. The interface between innate and adaptive immunity. *Nat. Immunol.* **2004**, *5*, 971–974. [CrossRef]
53. Broecker, F.; Moelling, K. Evolution of Immune Systems From Viruses and Transposable Elements. *Front. Microbiol.* **2019**, *10*, 51. [CrossRef] [PubMed]
54. Grandi, N.; Tramontano, E. Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Responses. *Front. Immunol.* **2018**, *9*, 2039. [CrossRef] [PubMed]
55. Kassiotis, G.; Stoye, J.P. Immune responses to endogenous retroelements: Taking the bad with the good. *Nat. Rev. Immunol.* **2016**, *16*, 207–219. [CrossRef] [PubMed]
56. Apostolou, E.; Thanos, D. Virus Infection Induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression. *Cell* **2008**, *134*, 85–96. [CrossRef]
57. Thomson, S.J.; Goh, F.G.; Banks, H.; Krausgruber, T.; Kotenko, S.V.; Foxwell, B.M.; Udalova, I.A. The role of transposable elements in the regulation of IFN-lambda1 gene expression. *Proc.. Natl. Acad. Sci. USA* **2009**, *106*, 11564–11569. [CrossRef]
58. Chuong, E.B.; Elde, N.C.; Feschotte, C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* **2016**, *351*, 1083–1087. [CrossRef]
59. Lee, H.C.; Chathuranga, K.; Lee, J.S. Intracellular sensing of viral genomes and viral evasion. *Exp. Mol. Med.* **2019**, *51*, 1–13. [CrossRef]
60. Hurst, T.P.; Magiorkinis, G. Activation of the innate immune response by endogenous retroviruses. *J. Gen. Virol.* **2015**, *96*, 1207–1218. [CrossRef]
61. Rolland, A.; Jouvin-Marche, E.; Viret, C.; Faure, M.; Perron, H.; Marche, P.N. The envelope protein of a human endogenous retrovirus-W family activates innate immunity through CD14/TLR4 and promotes Th1-like responses. *J. Immunol.* **2006**, *176*, 7636–7644. [CrossRef] [PubMed]
62. Heil, F.; Hemmi, H.; Hochrein, H.; Ampenberger, F.; Kirschning, C.; Akira, S.; Lipford, G.; Wagner, H.; Bauer, S. Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science* **2004**, *303*, 1526–1529. [CrossRef] [PubMed]
63. Chiappinelli, K.B.; Strissel, P.L.; Desrichard, A.; Li, H.; Henke, C.; Akman, B.; Hein, A.; Rote, N.S.; Cope, L.M.; Snyder, A.; et al. Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* **2015**, *162*, 974–986. [CrossRef] [PubMed]

64. Brisse, M.; Ly, H. Comparative Structure and Function Analysis of the RIG-I-Like Receptors: RIG-I and MDA5. *Front. Immunol.* **2019**, *10*, 1586. [CrossRef] [PubMed]
65. Zhao, K.; Du, J.; Peng, Y.; Li, P.; Wang, S.; Wang, Y.; Hou, J.; Kang, J.; Zheng, W.; Hua, S.; et al. LINE1 contributes to autoimmunity through both RIG-I- and MDA5-mediated RNA sensing pathways. *J. Autoimmun* **2018**, *90*, 105–115. [CrossRef]
66. Williams, B.R. PKR; a sentinel kinase for cellular stress. *Oncogene* **1999**, *18*, 6112–6120. [CrossRef]
67. Chu, W.M.; Ballard, R.; Carpick, B.W.; Williams, B.R.; Schmid, C.W. Potential Alu function: Regulation of the activity of double-stranded RNA-activated kinase PKR. *Mol. Cell Biol.* **1998**, *18*, 58–68. [CrossRef]
68. Takaoka, A.; Wang, Z.; Choi, M.K.; Yanai, H.; Negishi, H.; Ban, T.; Lu, Y.; Miyagishi, M.; Kodama, T.; Honda, K.; et al. DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response. *Nature* **2007**, *448*, 501–505. [CrossRef]
69. Hornung, V.; Ablasser, A.; Charrel-Dennis, M.; Bauernfeind, F.; Horvath, G.; Caffrey, D.R.; Latz, E.; Fitzgerald, K.A. AIM2 recognizes cytosolic dsDNA and forms a caspase-1-activating inflammasome with ASC. *Nature* **2009**, *458*, 514–518. [CrossRef]
70. Unterholzner, L.; Keating, S.E.; Baran, M.; Horan, K.A.; Jensen, S.B.; Sharma, S.; Sirois, C.M.; Jin, T.; Latz, E.; Xiao, T.S.; et al. IFI16 is an innate immune sensor for intracellular DNA. *Nat. Immunol.* **2010**, *11*, 997–1004. [CrossRef]
71. Zeng, M.; Hu, Z.; Shi, X.; Li, X.; Zhan, X.; Li, X.D.; Wang, J.; Choi, J.H.; Wang, K.W.; Purrington, T.; et al. MAVS, cGAS, and endogenous retroviruses in T-independent B cell responses. *Science* **2014**, *346*, 1486–1492. [CrossRef] [PubMed]
72. Metzner, M.; Jack, H.M.; Wabl, M. LINE-1 retroelements complexed and inhibited by activation induced cytidine deaminase. *PLoS ONE* **2012**, *7*, e49358. [CrossRef] [PubMed]
73. Manghera, M.; Ferguson-Parry, J.; Lin, R.; Douville, R.N. NF-kappaB and IRF1 Induce Endogenous Retrovirus K Expression via Interferon-Stimulated Response Elements in Its 5′ Long Terminal Repeat. *J. Virol.* **2016**, *90*, 9338–9349. [CrossRef] [PubMed]
74. Manghera, M.; Douville, R.N. Endogenous retrovirus-K promoter: A landing strip for inflammatory transcription factors? *Retrovirology* **2013**, *10*, 16. [CrossRef] [PubMed]
75. Li, W.; Lee, M.H.; Henderson, L.; Tyagi, R.; Bachani, M.; Steiner, J.; Campanac, E.; Hoffman, D.A.; von Geldern, G.; Johnson, K.; et al. Human endogenous retrovirus-K contributes to motor neuron disease. *Sci. Transl. Med.* **2015**, *7*. [CrossRef] [PubMed]
76. Hung, T.; Pratt, G.A.; Sundararaman, B.; Townsend, M.J.; Chaivorapol, C.; Bhangale, T.; Graham, R.R.; Ortmann, W.; Criswell, L.A.; Yeo, G.W.; et al. The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science* **2015**, *350*, 455–459. [CrossRef] [PubMed]
77. Reed, J.H.; Gordon, T.P. Autoimmunity: Ro60-associated RNA takes its toll on disease pathogenesis. *Nat. Rev. Rheumatol.* **2016**, *12*, 136–138. [CrossRef]
78. Liddicoat, B.J.; Piskol, R.; Chalk, A.M.; Ramaswami, G.; Higuchi, M.; Hartner, J.C.; Li, J.B.; Seeburg, P.H.; Walkley, C.R. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science* **2015**, *349*, 1115–1120. [CrossRef]
79. Thomas, C.A.; Tejwani, L.; Trujillo, C.A.; Negraes, P.D.; Herai, R.H.; Mesci, P.; Macia, A.; Crow, Y.J.; Muotri, A.R. Modeling of TREX1-Dependent Autoimmune Disease using Human Stem Cells Highlights L1 Accumulation as a Source of Neuroinflammation. *Cell Stem Cell* **2017**, *21*, 319–331 e318. [CrossRef]
80. De Cecco, M.; Ito, T.; Petrashen, A.P.; Elias, A.E.; Skvir, N.J.; Criscione, S.W.; Caligiana, A.; Brocculi, G.; Adney, E.M.; Boeke, J.D.; et al. L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **2019**, *566*, 73–78. [CrossRef]
81. Shankaran, V.; Ikeda, H.; Bruce, A.T.; White, J.M.; Swanson, P.E.; Old, L.J.; Schreiber, R.D. IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity. *Nature* **2001**, *410*, 1107–1111. [CrossRef] [PubMed]
82. Dunn, G.P.; Koebel, C.M.; Schreiber, R.D. Interferons, immunity and cancer immunoediting. *Nat. Rev. Immunol.* **2006**, *6*, 836–848. [CrossRef] [PubMed]
83. Koebel, C.M.; Vermi, W.; Swann, J.B.; Zerafa, N.; Rodig, S.J.; Old, L.J.; Smyth, M.J.; Schreiber, R.D. Adaptive immunity maintains occult cancer in an equilibrium state. *Nature* **2007**, *450*, 903–907. [CrossRef] [PubMed]
84. Schreiber, R.D.; Old, L.J.; Smyth, M.J. Cancer immunoediting: Integrating immunity's roles in cancer suppression and promotion. *Science* **2011**, *331*, 1565–1570. [CrossRef]

85. Jerby-Arnon, L.; Shah, P.; Cuoco, M.S.; Rodman, C.; Su, M.J.; Melms, J.C.; Leeson, R.; Kanodia, A.; Mei, S.; Lin, J.R.; et al. A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. *Cell* **2018**, *175*, 984–997 e924. [CrossRef]
86. Thommen, D.S.; Schumacher, T.N. T Cell Dysfunction in Cancer. *Cancer Cell* **2018**, *33*, 547–562. [CrossRef]
87. Robbez-Masson, L.; Tie, C.H.C.; Rowe, H.M. Cancer cells, on your histone marks, get SETDB1, silence retrotransposons, and go! *J. Cell Biol.* **2017**, *216*, 3429–3431. [CrossRef]
88. Goel, S.; DeCristo, M.J.; Watt, A.C.; BrinJones, H.; Sceneay, J.; Li, B.B.; Khan, N.; Ubellacker, J.M.; Xie, S.; Metzger-Filho, O.; et al. CDK4/6 inhibition triggers anti-tumour immunity. *Nature* **2017**, *548*, 471–475. [CrossRef]
89. Roulois, D.; Loo Yau, H.; Singhania, R.; Wang, Y.; Danesh, A.; Shen, S.Y.; Han, H.; Liang, G.; Jones, P.A.; Pugh, T.J.; et al. DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell* **2015**, *162*, 961–973. [CrossRef]
90. Cuellar, T.L.; Herzner, A.M.; Zhang, X.; Goyal, Y.; Watanabe, C.; Friedman, B.A.; Janakiraman, V.; Durinck, S.; Stinson, J.; Arnott, D.; et al. Silencing of retrotransposons by SETDB1 inhibits the interferon response in acute myeloid leukemia. *J. Cell Biol.* **2017**, *216*, 3535–3549. [CrossRef]
91. Canadas, I.; Thummalapalli, R.; Kim, J.W.; Kitajima, S.; Jenkins, R.W.; Christensen, C.L.; Campisi, M.; Kuang, Y.; Zhang, Y.; Gjini, E.; et al. Tumor innate immunity primed by specific interferon-stimulated endogenous retroviruses. *Nat. Med.* **2018**, *24*, 1143–1150. [CrossRef] [PubMed]
92. Smith, C.C.; Beckermann, K.E.; Bortone, D.S.; De Cubas, A.A.; Bixby, L.M.; Lee, S.J.; Panda, A.; Ganesan, S.; Bhanot, G.; Wallen, E.M.; et al. Endogenous retroviral signatures predict immunotherapy response in clear cell renal cell carcinoma. *J. Clin. Investig.* **2018**, *128*, 4804–4820. [CrossRef] [PubMed]
93. Bradner, J.E.; Hnisz, D.; Young, R.A. Transcriptional Addiction in Cancer. *Cell* **2017**, *168*, 629–643. [CrossRef]
94. Solovyov, A.; Vabret, N.; Arora, K.S.; Snyder, A.; Funt, S.A.; Bajorin, D.F.; Rosenberg, J.E.; Bhardwaj, N.; Ting, D.T.; Greenbaum, B.D. Global Cancer Transcriptome Quantifies Repeat Element Polarization between Immunotherapy Responsive and T Cell Suppressive Classes. *Cell Rep.* **2018**, *23*, 512–521. [CrossRef] [PubMed]
95. Kong, Y.; Rose, C.M.; Cass, A.A.; Williams, A.G.; Darwish, M.; Lianoglou, S.; Haverty, P.M.; Tong, A.J.; Blanchette, C.; Albert, M.L.; et al. Transposable element expression in tumors is associated with immune infiltration and increased antigenicity. *Nat. Commun.* **2019**, *10*, 5228. [CrossRef] [PubMed]
96. Munoz-Lopez, M.; Macia, A.; Garcia-Canadas, M.; Badge, R.M.; Garcia-Perez, J.L. An epi [c] genetic battle: LINE-1 retrotransposons and intragenomic conflict in humans. *Mob. Genet. Elements* **2011**, *1*, 122–127. [CrossRef]
97. Scott, E.C.; Gardner, E.J.; Masood, A.; Chuang, N.T.; Vertino, P.M.; Devine, S.E. A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res.* **2016**, *26*, 745–755. [CrossRef]
98. Krug, B.; De Jay, N.; Harutyunyan, A.S.; Deshmukh, S.; Marchione, D.M.; Guilhamon, P.; Bertrand, K.C.; Mikael, L.G.; McConechy, M.K.; Chen, C.C.L.; et al. Pervasive H3K27 Acetylation Leads to ERV Expression and a Therapeutic Vulnerability in H3K27M Gliomas. *Cancer Cell* **2019**, *35*, 782–797 e788. [CrossRef]
99. Jiang, J.C.; Upton, K.R. Human transposons are an abundant supply of transcription factor binding sites and promoter activities in breast cancer cell lines. *Mob. DNA* **2019**, *10*, 16. [CrossRef]
100. Choi, S.H.; Worswick, S.; Byun, H.M.; Shear, T.; Soussa, J.C.; Wolff, E.M.; Douer, D.; Garcia-Manero, G.; Liang, G.; Yang, A.S. Changes in DNA methylation of tandem DNA repeats are different from interspersed repeats in cancer. *Int. J. Cancer* **2009**, *125*, 723–729. [CrossRef]
101. Alves, P.M.; Levy, N.; Stevenson, B.J.; Bouzourene, H.; Theiler, G.; Bricard, G.; Viatte, S.; Ayyoub, M.; Vuilleumier, H.; Givel, J.C.; et al. Identification of tumor-associated antigens by large-scale analysis of genes expressed in human colorectal cancer. *Cancer Immun.* **2008**, *8*, 11.
102. 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*, 4172–4180. [CrossRef] [PubMed]
103. Stengel, S.; Fiebig, U.; Kurth, R.; Denner, J. Regulation of human endogenous retrovirus-K expression in melanomas by CpG methylation. *Genes Chromosomes Cancer* **2010**, *49*, 401–411. [CrossRef] [PubMed]
104. Siebenthall, K.T.; Miller, C.P.; Vierstra, J.D.; Mathieu, J.; Tretiakova, M.; Reynolds, A.; Sandstrom, R.; Rynes, E.; Haugen, E.; Johnson, A.; et al. Integrated epigenomic profiling reveals endogenous retrovirus reactivation in renal cell carcinoma. *EBioMedicine* **2019**, *41*, 427–442. [CrossRef] [PubMed]

105. Li, M.; Radvanyi, L.; Yin, B.; Rycaj, K.; Li, J.; Chivukula, R.; Lin, K.; Lu, Y.; Shen, J.; Chang, D.Z.; 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*, 5892–5911. [CrossRef]
106. Yandim, C.; Karakulah, G. Dysregulated expression of repetitive DNA in ER+/HER2- breast cancer. *Cancer Genet.* **2019**, *239*, 36–45. [CrossRef]
107. Wang-Johanning, F.; Liu, J.; Rycaj, K.; Huang, M.; Tsai, K.; Rosen, D.G.; Chen, D.T.; Lu, D.W.; Barnhart, K.F.; Johanning, G.L. Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer. *Int J. Cancer* **2007**, *120*, 81–90. [CrossRef]
108. Clayton, E.A.; Wang, L.; Rishishwar, L.; Wang, J.; McDonald, J.F.; Jordan, I.K. Patterns of Transposable Element Expression and Insertion in Cancer. *Front. Mol. Biosci.* **2016**, *3*, 76. [CrossRef]
109. Desai, N.; Sajed, D.; Arora, K.S.; Solovyov, A.; Rajurkar, M.; Bledsoe, J.R.; Sil, S.; Amri, R.; Tai, E.; MacKenzie, O.C.; et al. Diverse repetitive element RNA expression defines epigenetic and immunologic features of colon cancer. *JCI Insight* **2017**, *2*, e91078. [CrossRef]
110. Wolff, E.M.; Byun, H.M.; Han, H.F.; Sharma, S.; Nichols, P.W.; Siegmund, K.D.; Yang, A.S.; Jones, P.A.; Liang, G. Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. *PLoS Genet.* **2010**, *6*, e1000917. [CrossRef]
111. Anwar, S.L.; Wulaningsih, W.; Lehmann, U. Transposable Elements in Human Cancer: Causes and Consequences of Deregulation. *Int J. Mol. Sci.* **2017**, *18*, 974. [CrossRef] [PubMed]
112. Oricchio, E.; Sciamanna, I.; Beraldi, R.; Tolstonog, G.V.; Schumann, G.G.; Spadafora, C. Distinct roles for LINE-1 and HERV-K retroelements in cell proliferation, differentiation and tumor progression. *Oncogene* **2007**, *26*, 4226–4233. [CrossRef] [PubMed]
113. Nagai, Y.; Sunami, E.; Yamamoto, Y.; Hata, K.; Okada, S.; Murono, K.; Yasuda, K.; Otani, K.; Nishikawa, T.; Tanaka, T.; et al. LINE-1 hypomethylation status of circulating cell-free DNA in plasma as a biomarker for colorectal cancer. *Oncotarget* **2017**, *8*, 11906–11916. [CrossRef] [PubMed]
114. Strissel, P.L.; Ruebner, M.; Thiel, F.; Wachter, D.; Ekici, A.B.; Wolf, F.; Thieme, F.; Ruprecht, K.; Beckmann, M.W.; Strick, R. Reactivation of codogenic endogenous retroviral (ERV) envelope genes in human endometrial carcinoma and prestages: Emergence of new molecular targets. *Oncotarget* **2012**, *3*, 1204–1219. [CrossRef]
115. Babaian, A.; Romanish, M.T.; Gagnier, L.; Kuo, L.Y.; Karimi, M.M.; Steidl, C.; Mager, D.L. Onco-exaptation of an endogenous retroviral LTR drives IRF5 expression in Hodgkin lymphoma. *Oncogene* **2016**, *35*, 2542–2546. [CrossRef]
116. Jang, H.S.; Shah, N.M.; Du, A.Y.; Dailey, Z.Z.; Pehrsson, E.C.; Godoy, P.M.; Zhang, D.; Li, D.; Xing, X.; Kim, S.; et al. Transposable elements drive widespread expression of oncogenes in human cancers. *Nat. Genet.* **2019**, *51*, 611–617. [CrossRef]
117. Cruickshanks, H.A.; Tufarelli, C. Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter. *Genomics* **2009**, *94*, 397–406. [CrossRef]
118. Cruickshanks, H.A.; Vafadar-Isfahani, N.; Dunican, D.S.; Lee, A.; Sproul, D.; Lund, J.N.; Meehan, R.R.; Tufarelli, C. Expression of a large LINE-1-driven antisense RNA is linked to epigenetic silencing of the metastasis suppressor gene TFPI-2 in cancer. *Nucleic Acids Res.* **2013**, *41*, 6857–6869. [CrossRef]
119. Scarfo, I.; Pellegrino, E.; Mereu, E.; Kwee, I.; Agnelli, L.; Bergaggio, E.; Garaffo, G.; Vitale, N.; Caputo, M.; Machiorlatti, R.; et al. Identification of a new subclass of ALK-negative ALCL expressing aberrant levels of ERBB4 transcripts. *Blood* **2016**, *127*, 221–232. [CrossRef]
120. Lock, F.E.; Rebollo, R.; Miceli-Royer, K.; Gagnier, L.; Kuah, S.; Babaian, A.; Sistiaga-Poveda, M.; Lai, C.B.; Nemirovsky, O.; Serrano, I.; et al. Distinct isoform of FABP7 revealed by screening for retroelement-activated genes in diffuse large B-cell lymphoma. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E3534–E3543. [CrossRef]
121. Miki, Y.; Nishisho, I.; Horii, A.; Miyoshi, Y.; Utsunomiya, J.; Kinzler, K.W.; Vogelstein, B.; Nakamura, Y. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* **1992**, *52*, 643–645. [PubMed]
122. Goodier, J.L. Restricting retrotransposons: A review. *Mob. DNA* **2016**, *7*, 16. [CrossRef] [PubMed]
123. Aschacher, T.; Wolf, B.; Enzmann, F.; Kienzl, P.; Messner, B.; Sampl, S.; Svoboda, M.; Mechtcheriakova, D.; Holzmann, K.; Bergmann, M. LINE-1 induces hTERT and ensures telomere maintenance in tumour cell lines. *Oncogene* **2016**, *35*, 94–104. [CrossRef] [PubMed]
124. Lv, J.; Zhao, Z. Binding of LINE-1 RNA to PSF transcriptionally promotes GAGE6 and regulates cell proliferation and tumor formation in vitro. *Exp. Ther. Med.* **2017**, *14*, 1685–1691. [CrossRef]

125. Colombo, A.R.; Zubair, A.; Thiagarajan, D.; Nuzhdin, S.; Triche, T.J.; Ramsingh, G. Suppression of Transposable Elements in Leukemic Stem Cells. *Sci. Rep.* **2017**, *7*, 7029. [CrossRef]
126. Zhang, M.; Liang, J.Q.; Zheng, S. Expressional activation and functional roles of human endogenous retroviruses in cancers. *Rev. Med. Virol.* **2019**, *29*, e2025. [CrossRef]
127. Li, H.; Jiang, X.; Niu, X. Long Non-Coding RNA Reprogramming (ROR) Promotes Cell Proliferation in Colorectal Cancer via Affecting P53. *Med. Sci. Monit.* **2017**, *23*, 919–928. [CrossRef]
128. Li, L.; Feng, T.; Lian, Y.; Zhang, G.; Garen, A.; Song, X. Role of human noncoding RNAs in the control of tumorigenesis. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12956–12961. [CrossRef]
129. Wang, G.; Cui, Y.; Zhang, G.; Garen, A.; Song, X. Regulation of proto-oncogene transcription, cell proliferation, and tumorigenesis in mice by PSF protein and a VL30 noncoding RNA. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 16794–16798. [CrossRef]
130. Xu, L.; Elkahloun, A.G.; Candotti, F.; Grajkowski, A.; Beaucage, S.L.; Petricoin, E.F.; Calvert, V.; Juhl, H.; Mills, F.; Mason, K.; et al. A novel function of RNAs arising from the long terminal repeat of human endogenous retrovirus 9 in cell cycle arrest. *J. Virol.* **2013**, *87*, 25–36. [CrossRef]
131. Benatti, P.; Basile, V.; Merico, D.; Fantoni, L.I.; Tagliafico, E.; Imbriano, C. A balance between NF-Y and p53 governs the pro- and anti-apoptotic transcriptional response. *Nucleic Acids Res.* **2008**, *36*, 1415–1428. [CrossRef] [PubMed]
132. Cocucci, E.; Racchetti, G.; Meldolesi, J. Shedding microvesicles: Artefacts no more. *Trends Cell Biol.* **2009**, *19*, 43–51. [CrossRef] [PubMed]
133. Contreras-Galindo, R.; Kaplan, M.H.; Leissner, P.; Verjat, T.; Ferlenghi, I.; Bagnoli, F.; Giusti, F.; Dosik, M.H.; Hayes, D.F.; Gitlin, S.D.; et al. Human endogenous retrovirus K (HML-2) elements in the plasma of people with lymphoma and breast cancer. *J. Virol.* **2008**, *82*, 9329–9336. [CrossRef] [PubMed]
134. Seifarth, W.; Skladny, H.; Krieg-Schneider, F.; Reichert, A.; Hehlmann, R.; Leib-Mosch, C. Retrovirus-like particles released from the human breast cancer cell line T47-D display type B- and C-related endogenous retroviral sequences. *J. Virol.* **1995**, *69*, 6408–6416. [CrossRef] [PubMed]
135. Balaj, L.; Lessard, R.; Dai, L.; Cho, Y.J.; Pomeroy, S.L.; Breakefield, X.O.; Skog, J. Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nat. Commun.* **2011**, *2*, 180. [CrossRef] [PubMed]
136. Mangeney, M.; Renard, M.; Schlecht-Louf, G.; Bouallaga, I.; Heidmann, O.; Letzelter, C.; Richaud, A.; Ducos, B.; Heidmann, T. Placental syncytins: Genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 20534–20539. [CrossRef] [PubMed]
137. Treangen, T.J.; Salzberg, S.L. Repetitive DNA and next-generation sequencing: Computational challenges and solutions. *Nat. Rev. Genet.* **2011**, *13*, 36–46. [CrossRef]
138. Mortazavi, A.; Williams, B.A.; McCue, K.; Schaeffer, L.; Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **2008**, *5*, 621–628. [CrossRef]
139. Robert, C.; Watson, M. Errors in RNA-Seq quantification affect genes of relevance to human disease. *Genome Biol.* **2015**, *16*, 177. [CrossRef]
140. Kahles, A.; Behr, J.; Ratsch, G. MMR: A tool for read multi-mapper resolution. *Bioinformatics* **2016**, *32*, 770–772. [CrossRef]
141. Bray, N.L.; Pimentel, H.; Melsted, P.; Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **2016**, *34*, 525–527. [CrossRef] [PubMed]
142. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **2017**, *14*, 417–419. [CrossRef] [PubMed]
143. Takahashi, H.; Lassmann, T.; Murata, M.; Carninci, P. 5′ end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat. Protoc.* **2012**, *7*, 542–561. [CrossRef] [PubMed]
144. Carninci, P.; Kasukawa, T.; Katayama, S.; Gough, J.; Frith, M.C.; Maeda, N.; Oyama, R.; Ravasi, T.; Lenhard, B.; Wells, C.; et al. The transcriptional landscape of the mammalian genome. *Science* **2005**, *309*, 1559–1563. [CrossRef] [PubMed]
145. Faulkner, G.J.; Forrest, A.R.; Chalk, A.M.; Schroder, K.; Hayashizaki, Y.; Carninci, P.; Hume, D.A.; Grimmond, S.M. A rescue strategy for multimapping short sequence tags refines surveys of transcriptional activity by CAGE. *Genomics* **2008**, *91*, 281–288. [CrossRef]

146. McKerrow, W.; Fenyo, D. L1EM: A tool for accurate locus specific LINE-1 RNA quantification. *Bioinformatics* **2020**, *36*, 1167–1173. [CrossRef]
147. Babaian, A.; Thompson, I.R.; Lever, J.; Gagnier, L.; Karimi, M.M.; Mager, D.L. LIONS: Analysis suite for detecting and quantifying transposable element initiated transcription from RNA-seq. *Bioinformatics* **2019**, *35*, 3839–3841. [CrossRef]
148. Criscione, S.W.; Zhang, Y.; Thompson, W.; Sedivy, J.M.; Neretti, N. Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genom.* **2014**, *15*, 583. [CrossRef]
149. Jeong, H.H.; Yalamanchili, H.K.; Guo, C.; Shulman, J.M.; Liu, Z. An ultra-fast and scalable quantification pipeline for transposable elements from next generation sequencing data. *Pac. Symp. Biocomput.* **2018**, *23*, 168–179.
150. Yang, W.R.; Ardeljan, D.; Pacyna, C.N.; Payer, L.M.; Burns, K.H. SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Res.* **2019**, *47*, e27. [CrossRef]
151. Valdebenito-Maturana, B.; Riadi, G. TEcandidates: Prediction of genomic origin of expressed transposable elements using RNA-seq data. *Bioinformatics* **2018**, *34*, 3915–3916. [CrossRef] [PubMed]
152. Bendall, M.L.; de Mulder, M.; Iniguez, L.P.; Lecanda-Sanchez, A.; Perez-Losada, M.; Ostrowski, M.A.; Jones, R.B.; Mulder, L.C.F.; Reyes-Teran, G.; Crandall, K.A.; et al. Telescope: Characterization of the retrotranscriptome by accurate estimation of transposable element expression. *PLoS Comput. Biol.* **2019**, *15*, e1006453. [CrossRef] [PubMed]
153. Lerat, E.; Fablet, M.; Modolo, L.; Lopez-Maestre, H.; Vieira, C. TEtools facilitates big data expression analysis of transposable elements and reveals an antagonism between their activity and that of piRNA genes. *Nucleic Acids Res.* **2017**, *45*, e17. [CrossRef]
154. Jin, Y.; Tam, O.H.; Paniagua, E.; Hammell, M. TEtranscripts: A package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **2015**, *31*, 3593–3599. [CrossRef] [PubMed]
155. Navarro, F.C.; Hoops, J.; Bellfy, L.; Cerveira, E.; Zhu, Q.; Zhang, C.; Lee, C.; Gerstein, M.B. TeXP: Deconvolving the effects of pervasive and autonomous transcription of transposable elements. *PLoS Comput. Biol.* **2019**, *15*, e1007293. [CrossRef]
156. Jung, H.; Choi, J.K.; Lee, E.A. Immune signatures correlate with L1 retrotransposition in gastrointestinal cancers. *Genome Res.* **2018**, *28*, 1136–1146. [CrossRef]
157. Cebria-Costa, J.P.; Pascual-Reguant, L.; Gonzalez-Perez, A.; Serra-Bardenys, G.; Querol, J.; Cosin, M.; Verde, G.; Cigliano, R.A.; Sanseverino, W.; Segura-Bayona, S.; et al. LOXL2-mediated H3K4 oxidation reduces chromatin accessibility in triple-negative breast cancer cells. *Oncogene* **2020**, *39*, 79–121. [CrossRef]
158. Smit, A.; Hubley, R.; Green, P. RepeatMasker Open-4.0. 2015.
159. Deininger, P.; Morales, M.E.; White, T.B.; Baddoo, M.; Hedges, D.J.; Servant, G.; Srivastav, S.; Smither, M.E.; Concha, M.; DeHaro, D.L.; et al. A comprehensive approach to expression of L1 loci. *Nucleic Acids Res.* **2017**, *45*, e31. [CrossRef]
160. Zhou, W.; Emery, S.B.; Flasch, D.A.; Wang, Y.; Kwan, K.Y.; Kidd, J.M.; Moran, J.V.; Mills, R.E. Identification and characterization of occult human-specific LINE-1 insertions using long-read sequencing technology. *Nucleic Acids Res.* **2020**, *48*, 1146–1163. [CrossRef]
161. Istace, B.; Friedrich, A.; d'Agata, L.; Faye, S.; Payen, E.; Beluche, O.; Caradec, C.; Davidas, S.; Cruaud, C.; Liti, G.; et al. de novo assembly and population genomic survey of natural yeast isolates with the Oxford Nanopore MinION sequencer. *Gigascience* **2017**, *6*, 1–13. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Chapter 5

5.1 Summary, conclusions and future perspectives in translational medicine

Summary

Although for decades transposable elements are considered as “junk DNA”, now it is established that they are key molecule in the epigenetic regulation of cellular identity. TEs comprise 40-45% of the human genome and among them, LINE1 is the most representative superfamilies constituting 17% of the human genome (Casa & Gabellini, 2012; Lander et al., 2001; Viollet et al., 2014). Many indications suggest that the deregulated expression of TEs, as occurs in several auto immune diseases and in cancer, could control and shape innate and adaptive immune responses, through RNA and DNA sensing pathways that can promote IFN type I response (Kassiotis & Stoye, 2016). In this project, we studied the involvement of the RNA transcribed from LINE1 elements in the regulation on T cell identity and functions. At first, we found that, among TEs, LINE1 elements are chromatin enriched specifically in quiescence naïve $CD4^+$ T cells. By RNA-sequencing with short and long reads of chromatin associated RNA and the employment of *de novo* reconstruction pipelines, we discovered that LINE1 belonging to L1M families, are spliced as exons of non-canonical transcript variant of genes that are important for T cell activation.

We further discovered that LINE1-containing transcripts are controlled under the control of the transcription factor IRF4 that is expressed specifically in CD4+ T cells. Since it has been described that LINE1 RNAs are in partnership with nucleolin, and Kap1 (Percharde et al., 2018), we investigated the function of this complex in T cells. We found that LINE1-containing transcripts are kept at chromatin by nucleolin; this complex interferes with the expression of the corresponding genes *in cis*, hampering the deposition of H3K36me3. Upon cellular activation, LINE1-containing transcripts' levels decrease in a mTORC1 dependent manner through a splicing-suppression mechanism promoted by PTBP1, favoring the expression of the canonical transcripts through GTF2F1. We discovered that, in quiescent naïve CD4+ T cells, LINE1-containing transcripts control cell-activation genes and modulating effector T cell functions. Importantly, we found that in dysfunctional/ exhausted T cells (both in vitro exhausted T cells and ex vivo TILs), the LINE1-containing transcripts re-accumulate. In particular, we found that the mechanism that subtend this accumulation is the same described for quiescent T cells. Indeed, it is known that IRF4 not only contributes to T cell effector functions but also has a role in promoting exhausted phenotype (Mahnke et al., 2016; Man et al., 2017). Notably, we discovered that the modulation of LINE1-containing transcripts restored the TILs effector response, highlighting LINE1 transcripts as novel powerful molecular target to boost anti-tumoral immune response.

Conclusions

Cancer is one of the leading causes of death worldwide (Falzone et al., 2018), great efforts have been made to try to find a possible cure, starting from chemo-, radio- and targeted therapy. Immunotherapies represents the most promising frontiers, (Binnewies et al., 2018; Hoelder et al., 2012), however many limits have been found ascribed to extrinsic and intrinsic factors, like patients epigenetic background and intratumoral cellular variability (Falzone et al., 2018). To increase efficacy pf immunotherapies, more knowledge is needed to understand the molecular mechanism that subtend tumor microenvironment.

This project finds out another important function for TEs, in particular LINE1 elements, in the regulation of T cells, establishing the engagement of LINE1 elements in the regulation of T cells quiescence and exhaustion.

A very interesting aspect that came out from this project is that LINE1-containing transcripts re-expression in TILs seems to be a common feature among different TILs, since we demonstrated that this re-expression occurs both in $CD4^+$ TILs and $CD8^+$ TILs, and we confirmed these data in our *in vitro* exhausted model. Moreover, we verified this characteristic among distinct patients and different types of cancers, in fact, we retrieved the re-accumulation of LINE1-containing transcripts in TILs derived from colorectal and lung cancer patients. This is a very strong point, since the modulation of LINE1 transcripts could be used a common attribute to enhance effector and cytotoxic functions.

In detail we found that silencing LINE1-containing transcripts in *ex vivo* TILs and also in *in vitro* exhausted T cells restores cytokines secretion and more importantly the killing ability of these cells. All these pieces of evidence suggest that LINE1 transcripts could be a possible target for tumor therapeutic strategy.

Future perspectives

TME characterization is crucial to define cancer phenotype and it is incisive also in dictating response to therapy (Hanahan & Weinberg, 2011). Recent evidence suggested that type and location of TILs in TME could have a prognostic and therapeutic value (Galon et al., 2006; Pagès et al., 2005). Furthermore, in this scenario not only the high complexity and heterogeneity of TME but also TEs can play an important role. For these reasons we want to apply single-cell omics to understand if TEs could represent a novel, hidden layer that determines heterogeneity of TILs, and also if LINE1 transcripts could constitute a novel hallmark to define the phenotype of exhausted cells. To this purpose, at first, we are planning to define at single cell level the signature of TEs transcripts in TILs. We are planning to perform a single-cell RNA seq with a full-length technology (Gao et al., 2017). We decided to avoid the 5' or 3' scRNA-seq approach which gives us only partially information about 5' or 3' of mRNAs and to use a full-length technology to obtain isoform-level resolution that allows us to capture all TEs RNA isoform / derived

exons that we would otherwise lose. With this strategy we aim to determine which TEs RNA are expressed in which TILs subsets to identify a possible transcriptional signature of TEs that could perhaps define specific subsets of TILs. In order to increase the resolution of the experiments, we will execute it in memory $CD4^+$ and $CD8^+$ isolated from CRC and Lung cancer and normal adjacent tissues of patients. In our lab we have built an expertise in the analysis of repeats in datasets of scRNA-seq and we have set up a pipeline, called iRescue, to quantify the expression of TEs considering the uncertainty generated by highly homologous interspersed repetitive elements. By this pipeline we were able to detects TEs subfamilies in scRNA-seq. Indeed, applying our pipeline to scRNA-seq datasets derived from human PBMCs, we are able to discriminate by TEs expression subsets of innate and adaptive immune system as by a standard analysis with gene expression.

We are planning to validate scRNA-seq results with Prime Flow RNA Assay, a new technology that permits, by flow cytometry, to analyze RNA transcripts and receptor markers. We have already set up this technology analyzing LINE1 RNA content in PBMCs derived from healthy donors, in exhausted T cells and in TILs.

Further, since our RNA FISH data, suggest that LINE1 transcripts are homogenously expressed among TILs, we hypothesized that LINE1 transcripts expression could precede the expression of inhibitory receptors, possibly representing a pre-dysfunctional state in which effector functions are reduced; scRNA-seq

experiments and Prime flow will help in testing this important hypothesis.

scRNA-seq will allow the identification of specific LINE1 transcripts expressed at level of peculiar subsets. To verify if these transcripts can be used as targetable molecules to enhance the immunological response, we will knock down these and we will test the checkpoint receptor expression (PD-1, TIM-3, LAG-3, CTLA4), cytokines secretion (IFN-γ, PerfA, GrzB/K, IL10), killing capacity and suppression assay. More importantly, we are planning to co-colture memory $CD4^+$ T cells and $CD8^+$ T cells with organoids in order to reconstitute the TME and to test immunological function of TILs knocked down for LINE1 transcripts in a more physiological context.

This project will be focused in understanding how the modulation of LINE1 RNAs has the potential of being an important mechanism in determining at transcriptional level TILs function in the tumor microenvironment.

In conclusion, if we are able to identify the nature of LINE RNAs specifically expressed in TILs, we can speculate to promote their degradation through antisense oligonucleotides (ASOs), to restore TILs effector functions, as a cancer therapeutic adjuvant. Antisense oligonucleotides are already used for therapeutic strategy: in the last decades at least 8 drugs based on ASO technology were approved by FDA (Food and drug administration) and EMA (European medicines agency) and many others are in clinical trials. The ASOs approved were used

to treat different pathology like Cytomegalovirus infection, Hereditary transthyretin amyloidosis (hATTR), spinal muscular atrophy (SMA) and Duchenne muscular dystrophy (DMD); the ones that are in clinical trials were used to treat different types of tumors (leukemia, lung, CRC, prostate), neurodegenerative disease, diabetes and cardiovascular diseases (Dhuri et al., 2020).

All these future studies want to corroborate what we found in this project to validate a novel class of molecules, LINE1 RNAs as specific tumor microenvironmental targets usable in the clinic.

5.2 Chapter 5 bibliography

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., Coussens, L. M., Gabrilovich, D. I., Ostrand-Rosenberg, S., Hedrick, C. C., Vonderheide, R. H., Pittet, M. J., Jain, R. K., Zou, W., Howcroft, T. K., Woodhouse, E. C., Weinberg, R. A., & Krummel, M. F. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nature Medicine*. https://doi.org/10.1038/s41591-018-0014-x

Dhuri, K., Bechtold, C., Quijano, E., Pham, H., Gupta, A., Vikram, A., & Bahal, R. (2020). Antisense oligonucleotides: An emerging area in drug discovery and development. In *Journal of Clinical Medicine*. https://doi.org/10.3390/JCM9062004

Falzone, L., Salomone, S., & Libra, M. (2018). Evolution of cancer pharmacological treatments at the turn of the third millennium. In *Frontiers in Pharmacology*. https://doi.org/10.3389/fphar.2018.01300

Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., Zinzindohoué, F., Bruneval, P., Cugnenc, P. H., Trajanoski, Z., Fridman, W. H., & Pagès, F. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science*. https://doi.org/10.1126/science.1129139

Gao, R., Kim, C., Sei, E., Foukakis, T., Crosetto, N., Chan, L. K., Srinivasan, M., Zhang, H., Meric-Bernstam, F., & Navin, N. (2017). Nanogrid single-nucleus RNA sequencing reveals phenotypic diversity in breast cancer. *Nature Communications*. https://doi.org/10.1038/s41467-017-00244-w

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. In *Cell*. https://doi.org/10.1016/j.cell.2011.02.013

Hoelder, S., Clarke, P. A., & Workman, P. (2012). Discovery of small molecule cancer drugs: Successes, challenges and opportunities. In *Molecular Oncology*. https://doi.org/10.1016/j.molonc.2012.02.004

Mahnke, J., Schumacher, V., Ahrens, S., Käding, N., Feldhoff, L. M., Huber, M., Rupp, J., Raczkowski, F., & Mittrücker, H.

W. (2016). Interferon Regulatory Factor 4 controls T H1 cell effector function and metabolism. *Scientific Reports*. https://doi.org/10.1038/srep35521

Man, K., Gabriel, S. S., Liao, Y., Gloury, R., Preston, S., Henstridge, D. C., Pellegrini, M., Zehn, D., Berberich-Siebelt, F., Febbraio, M. A., Shi, W., & Kallies, A. (2017). Transcription Factor IRF4 Promotes CD8+ T Cell Exhaustion and Limits the Development of Memory-like T Cells during Chronic Infection. *Immunity*. https://doi.org/10.1016/j.immuni.2017.11.021

Pagès, F., Berger, A., Camus, M., Sanchez-Cabo, F., Costes, A., Molidor, R., Mlecnik, B., Kirilovsky, A., Nilsson, M., Damotte, D., Meatchi, T., Bruneval, P., Cugnenc, P.-H., Trajanoski, Z., Fridman, W.-H., & Galon, J. (2005). Effector Memory T Cells, Early Metastasis, and Survival in Colorectal Cancer. *New England Journal of Medicine*. https://doi.org/10.1056/nejmoa051424