



**HAL**  
open science

# Deregulation of alternative promoter usage in T-acute lymphoblastic leukemia

José David Abad Flores

► **To cite this version:**

José David Abad Flores. Deregulation of alternative promoter usage in T-acute lymphoblastic leukemia. Quantitative Methods [q-bio.QM]. Aix-Marseille Université, 2021. English. NNT: 2021AIXM0394 . tel-04419417

**HAL Id: tel-04419417**

**<https://hal.science/tel-04419417v1>**

Submitted on 26 Jan 2024

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

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

# AIX-MARSEILLE UNIVERSITÉ

FACULTÉ DES SCIENCES

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA  
SANTÉ

THÉORIES ET APPROCHES DE LA COMPLEXITÉ  
GENOMIQUE (TAGC)

THÈSE DE DOCTORAT GÉNOMIQUE ET  
BIOINFORMATIQUE

---

Dérégulation de l'utilisation de promoteurs alternatifs dans la  
leucémie aiguë lymphoblastique de type T

---

Par

**JOSÉ DAVID ABAD FLORES**

Soutenue le 29/10/21 devant le jury:

Dr. Jean-Christophe Andrau	Président
Dr. Meritxell Alberich Jorda	Rapporteur
Dr. Angélica Santana	Rapporteur
Dr. Lydie Pradel	Examinatrice
Dr. Denis Puthier	Examineur
Dr. Salvatore Spicuglia	Directeur de thèse

# **AIX-MARSEILLE UNIVERSITY**

FACULTY OF SCIENCE

DOCTORAL SCHOOL OF LIFE SCIENCES AND HEALTH

THEORIES AND APPROACHES OF GENOMIC COMPLEXITY  
(TAGC)

DOCTORAL THESIS

GENOMICS AND BIOINFORMATICS

---

## **Deregulation of alternative promoter usage in T-acute lymphoblastic leukemia**

---

By

**JOSÉ DAVID ABAD FLORES**

Defended the 29/10/21 in front of the jury:

Dr. Jean-Christophe Andrau	President
Dr. Meritxell Alberich Jorda	Reporter
Dr. Angélica Santana	Reporter
Dr. Lydie Pradel	Examiner
Dr. Denis Puthier	Examiner
Dr. Salvatore Spicuglia	Supervisor

## Résumé

Une dimension importante de la complexité du génome est l'utilisation de promoteurs alternatifs pour conduire une régulation génique omniprésente d'une manière spécifique au type cellulaire et dans le développement humain. Les promoteurs alternatifs sont fréquemment dérégulés dans les maladies, y compris le cancer, ainsi le choix du promoteur pourrait être parmi les forces motrices inconnues derrière les changements transcriptionnels oncogéniques. La leucémie aiguë lymphoblastique de type T (LAL-T) est un cancer hématologique agressif résultant de la transformation maligne des pro-géniteurs des cellules T, dans une certaine mesure en raison de l'expression anormale de facteurs de transcription. Mon objectif de thèse est d'évaluer la pertinence de la dérégulation des promoteurs alternatifs dans la LAL-T. L'analyse intégrative des données épigénomiques et transcriptomiques des précurseurs normaux des lymphocytes T et des LAL-T primaires a conduit à l'identification d'un promoteur alternatif de l'*ATP2C1* comme étant fréquemment régulé à la hausse chez les patients LAL-T. L'*ATP2C1* code pour l'ATPase de type 2C transportant le calcium membre 1 (également connue sous le nom de *SPCA1*). La pompe *SPCA1* est située sur la membrane de l'appareil de Golgi (AG) où elle transporte les ions  $\text{Ca}^{2+}$  et  $\text{Mn}^{2+}$  du cytosol vers l'AG contribuant ainsi à la voie sécrétoire. L'expression de l'*ATP2C1* a été impliquée dans le stress oxydatif, la régulation du cycle cellulaire et la survie des cellules cancéreuses. J'ai découvert que l'utilisation du promoteur spécifique de LAL-T de l'*ATP2C1* est liée à une signalisation des cellules T activée. Les analyses de l'expression génique et les tests de gène rapporteur ont démontré que le promoteur alternatif *ATP2C1* répond intrinsèquement à l'activation des lymphocytes T. La délétion ou la répression médiée par CRISPR du promoteur alternatif a entraîné l'absence d'activation de l'*ATP2C1*. De plus, l'inactivation génétique du gène *ATP2C1* a déclenché une activation des lymphocytes T exacerbée. J'ai émis l'hypothèse que la dérégulation épigénétique du promoteur alternatif *ATP2C1* pourrait conférer une survie cellulaire augmentée aux cellules leucémiques en interférant avec la signalisation des cellules T.

## Abstract

An important dimension of genome complexity is the use of alternative promoters to drive pervasive gene regulation in a cell type-specific manner and human development. Alternative promoters are frequently deregulated in disease, including cancer, thus promoter choice might be among the unknown driving forces behind the oncogenic transcriptional changes. T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive hematological cancer resulting from the malignant transformation of T cell progenitors, to some extent due to abnormal expression of transcription factors. My thesis aimed to assess the relevance of alternative promoter deregulation in T-ALL. Integrative analysis of epigenomic and transcriptomic data in normal T-cell precursors and primary T-ALLs lead to the identification of an alternative promoter of *ATP2C1* as frequently up-regulated in T-ALL patients. *ATP2C1* encodes the secretory pathway  $\text{Ca}^{2+}$  ATPase type I pump (also known as SPCA1). The *ATP2C1* pump is located on the membrane of the Golgi apparatus (GA) where it transports  $\text{Ca}^{2+}$  and  $\text{Mn}^{2+}$  ions from the cytosol into the GA, thus contributing to the secretory pathway. *ATP2C1* expression has been involved in oxidative stress, cell cycle regulation, and cancer cell survival. I found that T-ALL specific promoter usage of *ATP2C1* is linked to an activated T cell signaling. Analyses of gene expression and reporter assays demonstrated that the alternative *ATP2C1* promoter intrinsically responds to T-cell activation. CRISPR-mediated deletion or repression of the alternative promoter resulted in the lack of activation of *ATP2C1*. Moreover, genetic inactivation of the *ATP2C1* gene triggered exacerbated T-cell activation. I hypothesized that the epigenetic deregulation of the *ATP2C1* alternative promoter might confer increased cell survival of leukemic cells by interfering with T-cell signaling.

# Acknowledgements

*I'm still writing those....*

# Table of Contents

<b>Chapter 1 Gene Regulation</b> .....	1
1.1 General introduction .....	1
1.2 The double helix .....	1
1.3 Structure of genes .....	2
1.4 Transcriptional regulation in mammals .....	3
1.5 DNA methylation .....	5
1.6 Chromatin structure .....	6
1.7 Chromatin remodelers .....	7
1.8 Histone modifications.....	9
1.8.1 Histone acetylation.....	10
1.8.2 Histone methylation.....	12
1.8.2.1 Lysine methylation .....	12
1.8.2.2 Arginine methylation.....	12
1.8.2.3 Demethylation.....	12
1.8.3 Histone phosphorylation.....	13
1.8.4 Other histone modifications.....	14
1.9 Transcription factors .....	15
1.10 Transcriptional regulatory elements .....	16
1.10.1 Promoters.....	17
1.10.2 Enhancers .....	18
1.10.3 Silencers.....	19
1.10.4 Insulators .....	20
1.11 LncRNAs.....	20
1.12 Epigenetic regulation.....	22
1.12.1 Gene regulation by histone modifications .....	22
1.12.2 The epigenetic signature of regulatory elements .....	22
1.13 Epigenetic de-regulation in Cancer .....	24
1.13.1 Super-enhancers and Broad H3K4me3 domains.....	25
1.13.2 Broad H3K4me3 domains.....	26
1.14 Epigenomic resources .....	28

<b>Chapter 2 Alternative promoters</b> .....	30
2.1 Definition of alternative promoters .....	30
2.2 Impact of alternative promoters in protein function and gene expression.....	30
2.2.1 The upstream untranslated region and its role in regulating gene expression .....	31
2.2.2 Binding sites for regulatory proteins.....	32
2.2.3 Secondary structure .....	32
2.2.4 Upstream open reading frames (uORF) .....	32
2.2.5 Binding sites for regulatory RNAs.....	32
2.3 Regulation of alternative promoter usage .....	32
2.3.1 Alternative promoter expression is tissue-specific .....	33
2.3.2 Regulation of alternative promoter usage by enhancer elements .....	34
2.3.3 Regulation of alternative promoter usage by DNA methylation .....	35
2.3.4 Regulation of alternative promoter usage by chromatin looping.....	35
2.4 Biological implications of alternative promoters .....	36
2.4.1 Alternative promoters in development and disease.....	36
2.4.2 Deregulation of alternative promoters in cancer.....	37
2.5 Examples of alternative promoter regulation in normal T-cell differentiation and leukemia ....	38
2.5.1 Regulation of SATB1 AP usage by different transcription factor response during Th2 differentiation .....	39
2.5.2 Alternative Promoter usage of Notch1 in T-cell development and leukemogenesis .....	40
2.5.3 Nfatc1 in thymocyte differentiation and leukemia development.....	41
2.5.4 Regulation of alternative promoter usage by the acquired neomorphic promoter, the case of LMO2 .....	42
2.6 Genome-wide approaches for the study of alternative promoters.....	43
<b>Chapter 3 T-cell acute Lymphoblastic Leukemia (T-ALL)</b> .....	45
3.1 T-cell differentiation.....	45
3.2 T-acute lymphoblastic leukemia.....	46
3.2.1 Genomic and transcriptomic classification of T-ALL .....	46
3.2.3 Oncogenic NOTCH1 and T-ALL .....	48
3.2.4 T-ALL and cell-cycle deregulation.....	48
3.2.5 Transcription factors as oncogenes in T-ALL.....	49
3.2.6 Transcription factors as tumor suppressors in T-ALL.....	51
3.2.7 T-ALL and alterations in epigenetic regulators.....	52
3.2.8 Deregulation of T-ALL related oncogenes by genomic insertion of enhancer sites.....	53



3.2.9 The role of non-coding RNAs in T-ALL .....	53
3.2.9.1 Micro-RNAs and T-ALL.....	53
3.2.9.2 Inc-RNAs and T-ALL.....	54
3.3 Ca <sup>2+</sup> signaling and T-Cell activation .....	55
3.3.1 T-cell activation .....	55
3.3.2 Calcium-NFAT transcriptional signaling in T-cell activation .....	56
<b>Chapter 4 Calcium pump families and Ca<sup>2+</sup> signaling .....</b>	<b>58</b>
4.1 General overview of Ca <sup>2+</sup> signaling.....	58
4.2 Calcium pumps families.....	59
4.2.1 The SERCA pump family .....	62
4.2.2 The PMCA pump family.....	64
4.2.3 The ATP2C (SPCA) pump family.....	65
4.2.3.1 ATP2C1 gene.....	66
4.2.3.2 ATP2C1 and Hailey-Hailey disease .....	67
4.2.3.3 ATP2C1-NOTCH1 and DNA damage response in a Hailey-Hailey Disease model .....	69
4.3 Ca <sup>2+</sup> signaling and Ca <sup>2+</sup> pumps in cancer .....	70
4.4 Ca <sup>2+</sup> pumps inhibitors .....	73
<b>Chapter 5 Deregulation of alternative promoter usage in T-acute lymphoblastic leukemia.....</b>	<b>76</b>
5.1 Objective.....	76
5.2 Contributions.....	76
5.3 Results Part I and manuscript.....	77
5.4 Materials and methods I .....	105
5.6 Results part II Functional study of <i>ATP2C1</i> in a T-ALL model.....	112
5.6.1 Goal .....	112
5.6.2 Contributions.....	112
5.6.3 Results II .....	112
5.6.4 Materials and methods II .....	120
<b>Chapter 6 Discussion and perspectives .....</b>	<b>123</b>
6.1 Summary of main results.....	123
6.2. Alternative promoter identification .....	124
6.3 <i>ATP2C1</i> regulation in leukemia .....	125
6.4 Impact of ATP2C1 P1 isoforms in T-ALL.....	127
6.5 Function of <i>ATP2C1</i> as a modulator of T cell activation.....	129

6.6. Long term perspectives .....	132
6.6.1 The link between Hailey-Hailey disease and immunodeficiencies.....	132
6.6.2 Inhibitors of ATP2C1 as immunomodulators .....	132
<b>Bibliography</b> .....	<b>135</b>
<b>ANNEX I</b> .....	<b>154</b>
<b>ANNEX II</b> .....	<b>156</b>

**INTRODUCTION**  
**CHAPTERS 1-4**

# Chapter 1 Gene Regulation

## 1.1 General introduction

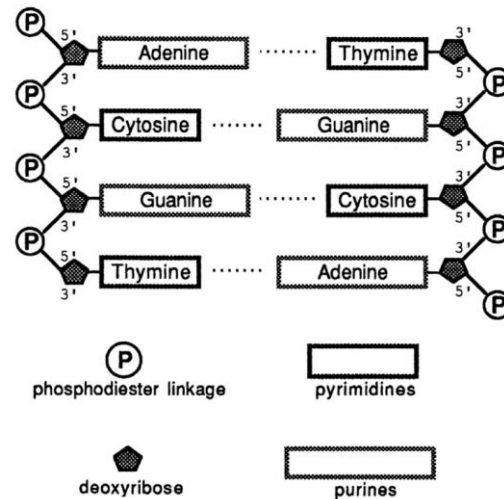
The human genome carries the genetic information in the form of DNA (deoxyribonucleic acid), and has an estimated size of 3.2 Mb, from a functional point of view the genomic sequences are distinguished into genes, pseudogenes, and noncoding DNA, approximately 2% of the genome codes for proteins, 0.5% are pseudogenes, however, most of the genome consists of introns and intergenic DNA, almost half of the intergenic DNA is made of transposons. The chemical structure of the genetic material, as well as the storage, processing and transfer of genetic information throughout generations, are similar in all living organisms (Hedge and Crowley 2019). The human genome reveals a high complexity in the spatial and temporal regulation of gene alternative transcripts, and their expression derived from a single locus, which can explain the low number of genes found in the human genome in contrast to the expected number, according to phenotypic entanglement. The posttranslational modification of proteins also adds a level of the gene product and function diversity (Hegde & Crowley, 2019).

Since the publication of the human genome sequence, we have gained new insights into genomics and how the sequence and structure participate in complex functions of human cells, and how genome architecture can be disrupted to produce disease. Comparative genomics has allowed the study of evolutionary conservation of gene and protein functions to determine our origins. Technological innovations have made easy the sequence analysis of different human genomes, highlighting the specific characteristics of each individual, allowing a better understanding of ourselves as human species and the traits we manifest, but also allowing us to reveal the genetic molecular mechanisms that lead to disease in humans and to generate therapeutic strategies based on these findings (Hegde & Crowley, 2019).

## 1.2 The double helix

The functional role of the human genome is to transfer, in a reliable manner, the genetic information from the parental cell to the daughter cell and from one generation to the next. This transfer is made in a semiconservative way, which means that one of the two strands of DNA from the parental cell remains intact throughout the cell division and serves as a template for copying the sequence.

The two DNA strands form a double helix by hydrogen bonding between the nitrogenous bases: purine bases (guanine and adenine) pairs with pyrimidine bases (cytosine and thymine) in the following manner; guanine (G) with cytosine (C) and adenine (A) with thymine (T) (Figure), the hydrogen bond formed between these pairs stabilize the double helix and allows the two complementary strands to remain together. The strands are oriented antiparallel to each other, which means they run in opposite directions, one has a 5' to 3' direction and the other is in a 3' to 5' direction (Figure 1.1). The double helix was described by Francis Crick and James Watson with the X-ray diffraction images produced by Rosalind Franklin.

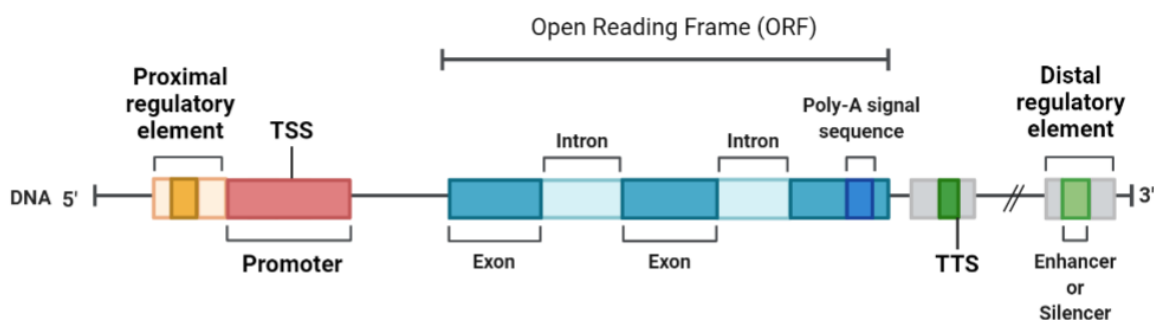


**Figure 1.1** The chemical structure of DNA. Taken from(Hegde & Crowley, 2019).

### 1.3 Structure of genes

A sequence coding for a single eukaryotic mRNA molecule is usually separated by non-coding sequences along the chromosomal DNA strand. The segments retained in the mature mRNA are called exons. During transcription, the exons are spliced together from a larger precursor RNA that also contains interspersed non-coding sequences called introns, the number of exons present in a single transcript depends on the gene and the organism (**Figure 1.2**). The non-coding sequences of the transcript are spliced out during mRNA maturation. Human genes tend to have small exons with a mean equal to 216bp, on the other hand, introns can be less than 100bp but can also exceed 10kb. Remarkably some introns carry significant information and even code for nested genes (Hegde & Crowley, 2019).

Individual exons may correspond to structural or functional domains of the proteins for which they code, for example, the heme-binding domain of globin. The origin of the exon/intron structure is thought to be ancient and to have established the divergence of eukaryotes and prokaryotes. In (Figure 1.2) we can observe a typical architecture of a gene, which includes the promoter region, the untranslated regions (5' UTR and 3' UTR), and of course the exonic and intronic sequences(Hegde & Crowley, 2019).



**Figure 1.2** The gene structure of eukaryotic organisms. Created with (BioRender.com).

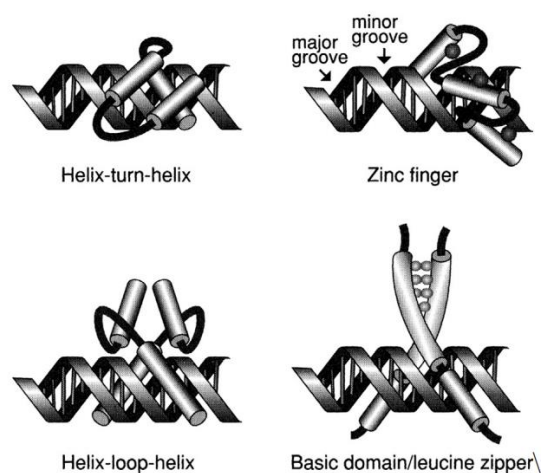
## 1.4 Transcriptional regulation in mammals

The expression of individual genes can be regulated at different levels. Multiple events take place before a gene sequence gets translated into a protein, including activation of the local DNA structure, initiation, and termination of transcription, processing of the primary transcript, transport of the mature transcript to the cytoplasm, and translation of the mRNA. All these steps can be a target of regulation and therefore are control points for altering gene expression. Several human genes show highly restricted tissue-specific expression patterns and this spatial and/or temporal restriction of gene expression can also be regulated at different levels (Hegde & Crowley, 2019).

Transcription initiation happens when the compact structure of DNA is relaxed and short sequence elements on the 5' end of the gene guide and active RNA polymerase, more details about these elements can be found in the subsection promoters of this chapter.

For the moment, we will define promoter as the region of DNA at the 5' end of the genes that bind the RNA polymerase, although the promoter definition can be broader and is also discussed with more detail in the subsection promoters.

There are three types of RNA polymerase I, II, and III. RNA polymerases I and III are dedicated to the transcription of genes coding for RNA molecules (rRNA and tRNA) which assists in the translation of coding genes. The minimal promoter sequence that is sufficient to begin any protein-coding gene transcription has an RNA polymerase II recognition signal as well as signals for general transcription factors (GTFs). The usual nomenclature of the transcription factors is TF followed by the Romanic number of the associated RNA polymerase. The GTFs such as TFIIB, TFIID, TFIIE, TFIIF, and TFIIH facilitate the binding and activation of RNA polymerase II into an activated transcriptional complex. TFs have some characteristic structural domains of DNA binding; these include zinc-finger motifs, helix-loop-helix motifs, helix-turn-helix motifs, and leucine zipper motifs (Figure 1.3) (Hegde & Crowley, 2019).

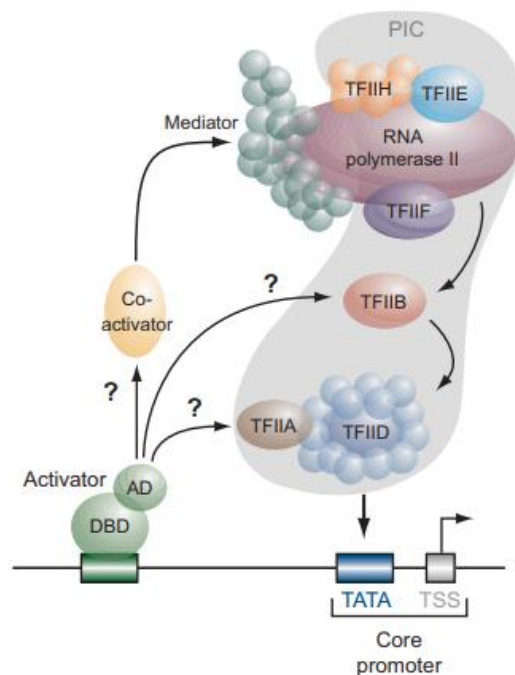


**Figure 1.3** The structure of DNA helix binding domains. Taken from (Hegde & Crowley, 2019).

GTFs assemble in the core promoter in an ordered fashion to form a transcription preinitiation complex (PIC), which directs the RNA polymerase II to the transcription start site (TSS). The first step in PIC assembly is binding of TFIID, a multi-subunit complex consisting of TATA box binding protein (TBP) and a set of tightly bound TBP associated factors (TAFs), then transcription proceeds to a series of steps including, promoter melting, clearance and escape before a fully functional RNA polymerase II elongation complex is formed. The current model of transcription regulation accounts for this as a cycle, in which complete PIC assembly is stimulated only once. After RNA polymerase II escapes from the promoter, a scaffold structure, composed of TFIID, TFII E, TFII H, and Mediator, remains on the core promoter, subsequent re-initiation of transcription only requires recruitment of RNA polymerase II-TFII F and TFII B, as seen in (Figure 1.4) (Hahn 2004).

The assembly of a PIC on the core promoter is sufficient to initiate transcription but at low levels, a process generally called basal transcription. The transcriptional activity must be stimulated by the second class of factors, denominated activators. Activators are sequence-specific DNA-binding proteins whose recognition sites are usually present in sequences upstream of the core promoter, in addition to a sequence-specific DNA-binding domain; an activator also has a separated activation domain that is required for the activator to stimulate transcription (Ptashne & Gann, 1997).

Activators function, at least in part, by increasing PIC formation through a mechanism thought to involve direct interactions with one or more components of the transcriptional machinery, termed the “target” (Orphanides et al., 1996). Finally, activators have also been proposed to function by recruiting activities that modify chromatin structure (de la Serna et al., 2005).



**Figure 1.4** The eukaryotic transcriptional machinery and factors involved in eukaryotic transcription by RNA polymerase II. Taken from (Maston et al., 2006).

The activators can be modulated by the third group of factors required for eukaryotic transcription: the coactivators (Lonard and O'Malley 2005). Generally, coactivators do not have intrinsic sequence-specific DNA binding sites; instead, they are recruited by protein-protein interactions with one or more DNA-bound activators. Coactivators work in many of the same ways as activators, by stimulating PIC assembly or modifying chromatin. The specific set of coactivators present in a cell can play a major role in determining the regulatory response (Lemon & Tjian, 2000).

## 1.5 DNA methylation

The DNA methylation is an epigenetic modification that consists in the formation of a covalent bond between a methyl group and the fifth carbon position of cytosine to form 5-methylcytosine (5mC), the enzymes in charge of this chemical reaction are the DNA methyltransferases (DNMTs) and they use S-adenosyl-methionine as the methyl donor. This modification occurs at CpG dinucleotides and is catalyzed by one of the following DNMTs: DNMT1, DNMT3a, and DNMT3b (Miranda & Jones, 2007).

DNA methylation is important for the formation of heterochromatin and in the long-term silencing of transcription. This epigenetic modification can be inherited throughout cellular divisions to maintain the silent state.

In the genome, we can discriminate two regions, according to the abundance of methylated sequences: CpG poor regions and CpG islands. CpG islands are sequences longer than 500bp that have GC content above 55%. The intergenic and intronic sequences of the genome are CpG poor regions. In healthy cells, methylation occurs in CpG poor regions whereas CpG islands are often hypomethylated, with a few exceptions like the inactive X chromosome (Takai & Jones, 2002).

During the carcinogenesis process, many CpG islands undergo hypermethylation, and the CpG poor regions become hypomethylated. This disruption of the DNA methylation patterns leads to changes in chromatin structure and silencing tumor suppressor genes as the main consequence (Jones & Baylin, 2007).

Exist in the genome two types of DNA methylation: de novo and maintenance methylation. De novo methylation is performed by DNMT3a and DNMT3b, this methylation is important for the formation of the methylation patterns in embryonic development and carcinogenesis (Okano et al., 1999).

DNA methylation in contrast to histone modifications is an excellent solution for long-term silencing due to its stability during cell division. The methylation of the promoter region blocks the reactivation of silent genes, even when the repressive histone marks are removed (McGarvey et al., 2007), this allows the daughter cell to harbor the same methylation patterns as the parental one, therefore DNA methylation plays a key role in some cellular process such as silencing of repetitive elements, X-inactivation, imprinting and development (Miranda & Jones, 2007).

The transposons are repetitive sequences that have been integrated into the mammalian genome over time, often produced after viral infections, and localized mostly in intergenic and intronic regions, the majority of transposons have long terminal repeat promoters, which



allows the transcription of the transposons, facilitating the translocation of these parasitic elements through all the genome, to keep the genome integrity these elements are silenced undergoing a methylation process (Robertson & Wolffe, 2000). However, in oncogenic states these CpG poor regions become hypomethylated, allowing the expansion of the transposons which in turn may play a role in tumorigenesis (Wilson et al., 2007).

CpG methylation is important for the formation and preservation of the X-inactivation and imprinted genes, which are non-mendelian forms of inheritance where one allele is methylated leading to monoallelic expression. Imprinting is important for the selection of the parental allele that will be expressed. Imprinted genes are methylated in the so-called imprinted control region (ICR), a process that takes place in the gonads, permitting the expression of the same allele in both parental and daughter cells (Jelinic & Shaw, 2007).

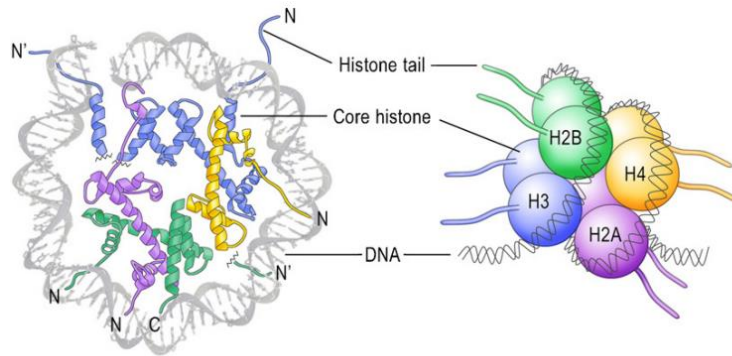
As expressed before the methylation of the promoter is associated with gene repression, which can be achieved by different mechanisms: alteration of transcription factor binding sites, prevention of transcriptional initiation by blocking the binding of the transcriptional machinery (Zhu et al., 2016), or by recruitment of co-regulators through proteins harboring a methyl-CpG-binding domain which in turn can compact the chromatin structure (Du et al., 2015).

## 1.6 Chromatin structure

The three billion base pairs that constitute the human genome is packed into 22 pairs of autosomes and one pair (XX or XY) of sex chromosomes. The chromosomal DNA can be divided into two regions according to the level of compaction: heterochromatin which represents the tightly packed regions, and euchromatin, the relaxed ones which are also the transcriptionally active DNA regions (Hedge and Crowley 2019).

Each of the 23 pairs of human chromosomes contains a single DNA duplex extending between the two telomeres; stretching the DNA of the human genome it would have a length greater than one meter, astonishing, compressing the DNA above 100,000 fold fit the chromosomes into the nucleus, this compaction process is achieved by coiling and folding the double helix into a series of progressively shorter and thicker structures. Proteins bind DNA to help direct and organize this folding, and the folded complex of DNA and protein is called chromatin (Hedge and Crowley 2019).

The fundamental unit of the chromatin is called the nucleosome and it is composed of an octamer of the four core histones (H3, H4, H2A, H2B) (Figure 1.5) around which 147 base pairs of DNA are wrapped. The core histones are globular except for their N-terminal tails which are unstructured (Kouzarides, 2007).



**Figure 1.5** Nucleosome representation of the four-nucleosome core histones. Adapted from (Gräff & Mansuy, 2008).

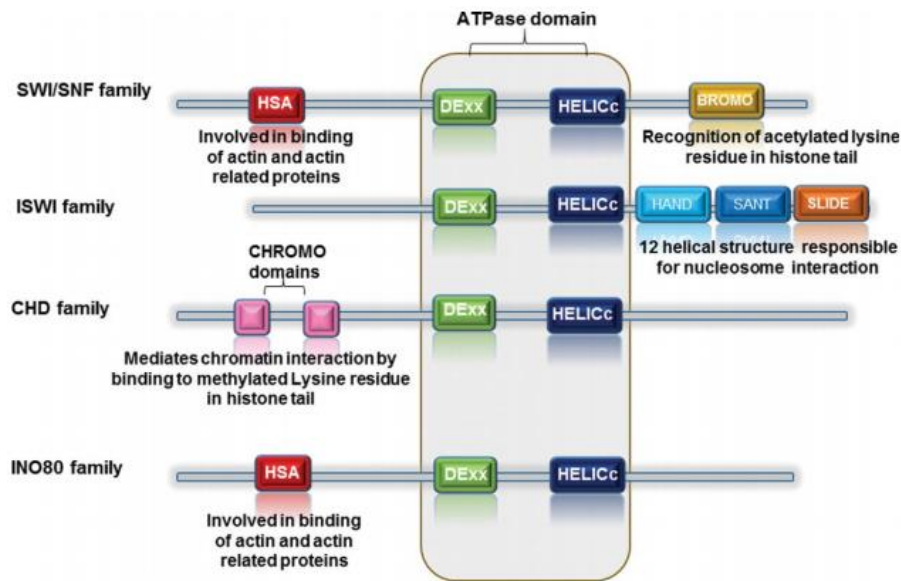
## 1.7 Chromatin remodelers

Chromatin is a highly dynamic structure that confers structural organization to the genome and regulates gene expression. The chromatin remodeling enzymes are the drivers of these dynamic chromatin structural changes, and they are categorized as writers, readers, and erasers, based on structure, ATP dependent chromatin remodelers consist of four families; SWI/SNF, ISWI, INO80, and CHD (Tyagi et al., 2016).

The plasticity of the chromatin is provided by (1) remodeling of chromosomes, (2) chemical modification of histones or incorporation of variants, (3) nonhistone DNA binding proteins, and (4) non-coding RNAs. The affinity of histones for DNA and DNA-associated proteins is governed by a combination of histone variants and post-transcriptional modifications (PTMs) of histones that further regulate the transcriptional activity and the accessibility of DNA for recombination, replication, and repair. The chromatin remodeling enzymes are in charge of performing these alterations (Strahl & Allis, 2000).

Chromatin remodelers catalyze a broad range of chromatin-changing reactions such as nucleosome sliding (sliding of an octamer across the DNA), changing the conformation of nucleosomal DNA, and histone variant exchange (Rippe et al., 2007). The remodeling enzymes are grouped in two categories according to their mode of action: (1) mediates histone post-translational modifications and (2) alters histone-DNA contact within the nucleosome coupled with ATP hydrolysis (Clapier & Cairns, 2009).

ATP-dependent chromatin remodelers are classified into four distinct families: switch/sucrose-non-fermenting (SWI/SNF), imitation switch (ISWI), chromo-domain-helicase-DNA binding (CHD), and inositol requiring 80 (INO80) (Figure1.5).



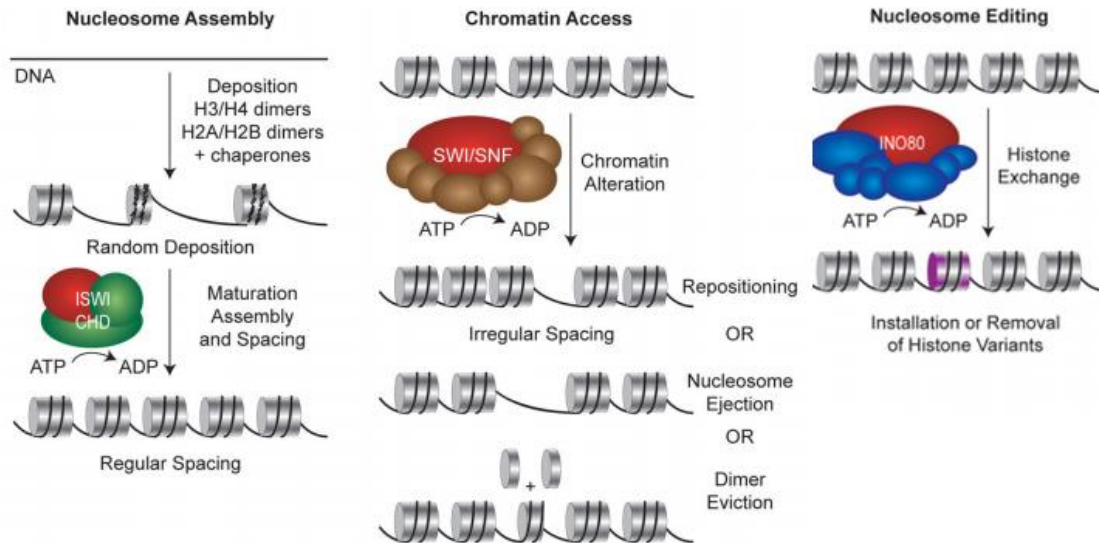
**Figure 1.6** The chromatin remodeler families. Adapted from (Tyagi et al., 2016).

The ATPase domain, present in all chromatin remodeler families, is split into two parts DExx and HELIC. The specific domains reside adjacent to the ATPase domain. SWI/SNF remodelers harbor bromodomains; ISWI remodelers - SANT-SLIDE modules; CHD remodelers - tandem chromodomains and members of INO80 family possess HSA (helicase SANT) domains (Figure 1.6). The specific domains have a role in remodeler recruitment to chromatin or interacting to specific histone modifications and/or they are involved in the regulation of the ATPase activity of the remodeler (Clapier & Cairns, 2009).

The remodelers are specialized in performing one of the following functions (1) nucleosome assembly and organization, (2) chromatin access, and (3) nucleosome editing (Clapier & Cairns, 2009).

**(1) Nucleosome assembly.** After replication, histone chaperones deliver histone complexes to nascent DNA behind the replisome, where assembly remodelers such as the ISWI and CHD subfamily help the initial DNA-histone complex to mature and next they form nucleosome arrays by spacing them. The assembly and spacing process also takes place during transcription at locations where nucleosomes have been ejected (Lusser et al., 2005).

**(2) Chromatin access.** Rendering the chromatin more accessible to proteins and RNA involves sliding nucleosomes along with the DNA, evicting nucleosome components (such as H2A–H2B dimers), or ejecting full nucleosome (Figure 1.7). These functions are primarily carried out by SWI/SNF subfamily. Access remodelers can expose binding sites for transcription activators or repressors at gene promoters or enhancers (Boeger et al., 2004).



**Figure 1.7** Action mechanism of ATP-dependent chromatin remodelers. Adapted from (Clapier & Cairns, 2009).

**(3)Nucleosome editing.** The INO80 subfamily conduct the replication-independent removal of a particular histone within a nucleosome and its replacement with either a canonical or a variant histone (Clapier & Cairns, 2009)

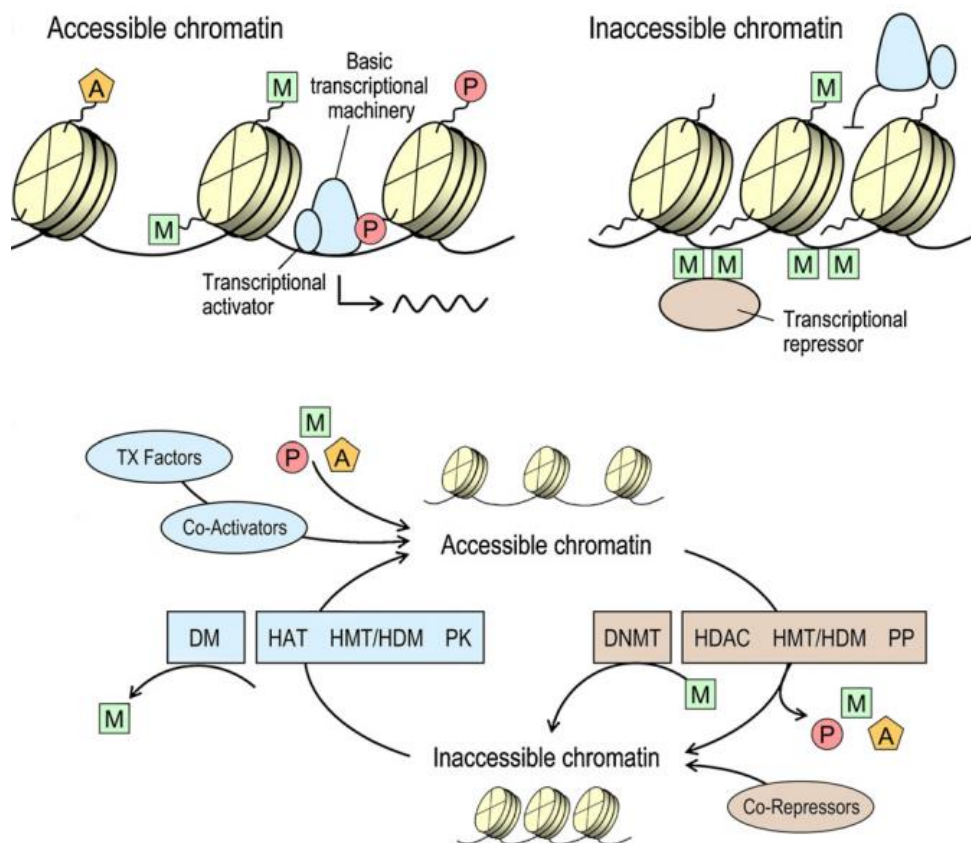
## 1.8 Histone modifications

The surface of the nucleosome is covered with a variety of modifications, they have been described in at least eight classes, and many sites for each class identified. These modifications can disrupt the chromatin structure or affect the recruitment of non-histone proteins to the chromatin, altering how DNA is packaged and controlling the recruitment of enzymatic complexes to manipulate DNA gene regulation, in this scenario histone modifications influence many fundamental biological processes (Figure 1.8)(Kouzarides, 2007).

A summary of the distinct types of histone modifications can be found in table 1, highlighting their associated functions.

**Table 1.1** Histone modification and associated function

Histone modification	Regulated functions
<b>Acetylation</b>	Transcription, Repair, Replication, Condensation
<b>Methylation (K)</b>	Transcription, Repair
<b>Methylation (R)</b>	Transcription
<b>Phosphorylation</b>	Transcription, Repair, Condensation
<b>Ubiquitylation</b>	Transcription, Repair
<b>Sumoylation</b>	Transcription
<b>ADP ribosylation</b>	Transcription
<b>Deimination</b>	Transcription
<b>Proline isomerization</b>	Transcription



**Figure 1.8** The transcriptional accessibility of chromatin. Taken from (Gräff & Mansuy, 2008).

### 1.8.1 Histone acetylation

Histone acetylation was discovered in 1964 by Allfrey and col. This chemical modification is performed by two main families of enzymes with opposite actions, the histone acetyltransferases (HATs) and histone deacetylases (HDACs). The HATs catalyze an enzymatic reaction in which an acetyl group is transferred to the  $\epsilon$ -amino group of lysine side

chains using acetyl coenzyme A (CoA) as cofactor thus neutralizing the positive charge of the lysine residues, and therefore weakening the interaction between histones and DNA (Allfrey et al., 1964).

Two types of HATs have been described, Type-A and Type-B, Type B is predominantly cytoplasmic and acetylates free histones, being the Type A of HATs a more diverse family, which can be subclassified into three different subgroups: GNAT, MYST, and CBP/p300. All these three enzymes together can acetylate multiple sites within the histone N-terminal tails and have the ability to neutralize positive charges and destabilize the electrostatic interaction that confers their functional role as transcriptional co-activators (Bannister and Kouzarides 2011).

Besides the tails, there are other additional sites for acetylation within the globular histone core, for example, H3K56. Like many other histone-modifying enzymes, the type A HATs are often associated with large protein complexes, the role of each protein component can be diverse but usually participates in controlling enzyme recruitment, activity, and substrate specificity (Yang & Seto, 2007).

On the contrary HDACs reverse lysine acetylation, restoring the positive charge of the lysine, this action stabilizes the chromatin structure, making them predominantly transcriptional repressors. There have been described at least 18 different HADACs in mammals belonging to four classes: class I to IV see table. Class I HADACs are found in the nucleus of all tissues, class II can be subdivided into two: IIa and IIb. Class IIa presents nuclear and cytoplasmic localization, switching between both sides in response to diverse signals. Class IIb localizes mainly in the cytoplasm and regulates signal transduction and motility. Class III of HADACs is also known as sirtuins (SIRT1-7), they are widely expressed in human tissues and participates in diverse biological functions such as oxidative stress, DNA repair, metabolism, and aging, little is known about HADAC11 the only member of class IV but it is expressed in the kidney, brain, testes, heart and skeletal muscle (Hull et al., 2016; Lakshmaiah et al., 2014). A summary of HADACs classification can be found in (Table 1.2).

Most of the time the HADACs forms multiple distinct complexes, even with other family members, for example, HADAC1 and HADAC2 are found together in the NuRD, Sin3a, and Co-REST complexes (Yang & Seto, 2008).

**Table 1.2 HADACs classification**

Group	Class	Name	Localization in cell	Localization in body
<b>Classical Zn dependent</b>	Class I	HADAC1 HADAC2 HADAC3 HADAC8	Nucleus	Ubiquitous
	Class IIa	HADAC4 HADAC5 HADAC7 HADAC9	Nucleus/Cytoplasm	Tissue-specific
	Class IIb	HADAC6 HADAC10	Cytoplasm	Tissue-specific
	Class IV	HADAC11	Nucleus/Cytoplasm	Tissue-specific
<b>NAD-dependent</b>	Class III	SIRT (1-7)	Nucleus/Cytoplasm	

This complexity is responsible for the difficulties in determining the specific activity or effect of each complex component, in some cases it is possible to determine which enzyme is required for a specific process, for example, the main participant in the control of embryonic stem cell differentiation is HADAC1 rather than HADAC2 (Dovey et al., 2010).

### 1.8.2 Histone methylation

The histone methylation usually targets the side chains of lysine and arginine residues, the lysine or arginine methylation does not alter the charge of the histone-like the acetylation or phosphorylation. Lysine residues may be mono, di or tri methylated, and arginine residues can be mono, symmetrically, or asymmetrically dimethylated (Bannister et Kouzarides 2011).

#### 1.8.2.1 Lysine methylation

The first histone lysine methyltransferase (HKMT) to be identified was SUV39H1, which targets H3K9 and since then diverse HKMTs have been described, most of them methylate the lysine residues within the N-terminal tails, the enzymatic activity of these HKMTs is accomplished by their SET domain by catalyzing the transfer of a methyl group from S-adenyl-methionine (SAM) to the  $\epsilon$ -amino group of a lysine residue (Bannister & Kouzarides, 2011).

#### 1.8.2.2 Arginine methylation

Two types of arginine methyltransferases have been identified the type-I and type-II, the type I specializes in mono- and asymmetric di-methylation of arginine, while type II generates mono- and symmetric di-methylation of arginine. Together they form a relatively large protein family with eleven members, and they are referred as to PRMTs, they transfer a methyl group from SAM to  $\omega$ -guanidino group of arginine, the most relevant enzymes for histone arginine methylation are PRMT1, 4, 5, and 6 (Bannister et Kouzarides 2011).

The histone methylation influences the recruitment and binding of different regulatory proteins to the chromatin (Morera et al 2016). Histone methylation can have either permissive or restrictive transcriptional character, depending on the position of the modified residue within the histone tail and/or the number of methyl groups added (Swygert and Peterson 2014), overall methylation of H3K4, H3K36 and H3K79 are associated with transcriptional activation while methylation of H3K9, H3K27 and H4K20 is associated with transcriptional repression (Kouzarides, 2007).

#### 1.8.2.3 Demethylation

Lysine methylation is a reversible process, the first lysine demethylase discovered was LSD1 (Shi Y et al 2004). At least six families of histone lysine demethylases have been described. The KDM1 family includes LSD1 (KDM1A) and LSD2 (KDM1B), both of which can demethylate H3K4me2/me1 but not H3K4me3 (Fang R et al 2010). The rest of the lysine demethylase families harbor the Jumonji (JmjC) domain, which has the potential to remove

the trimethyl mark, unlike the LSD family. The KDM2 family comprising KDM2 and KDM2B, also known as JHDM1A and JHDM1B respectively, targets H3K36me2/me1 and H3K4me3 (Cloos et al., 2008).

The KDM3 family includes KDM3A, KDM3B, and JMJD1C, with demethylase activities for H3K9me2/me1. The KDM5 family contains KDM5A, KDM5B, KDM5C, and KDM5D, all of which can demethylate H3K4me3/me2. KDM6 family includes UTX (KDM6A), JMJD3 (KDM6B), and UTY. UTX and JMJD3 are specific for H3K27me3/me2, while UTY has little catalytic activity (Z. Zhao & Shilatifard, 2019).

### 1.8.3 Histone phosphorylation

The phosphorylation of histones is as dynamic as the other modifications; in most cases, it can take place on serine, threonine, and tyrosine residues at the N-terminal tails. The main enzymes responsible for these chemical changes are called kinases and phosphatases that add or remove a phosphate group respectively (Oki et al., 2007).

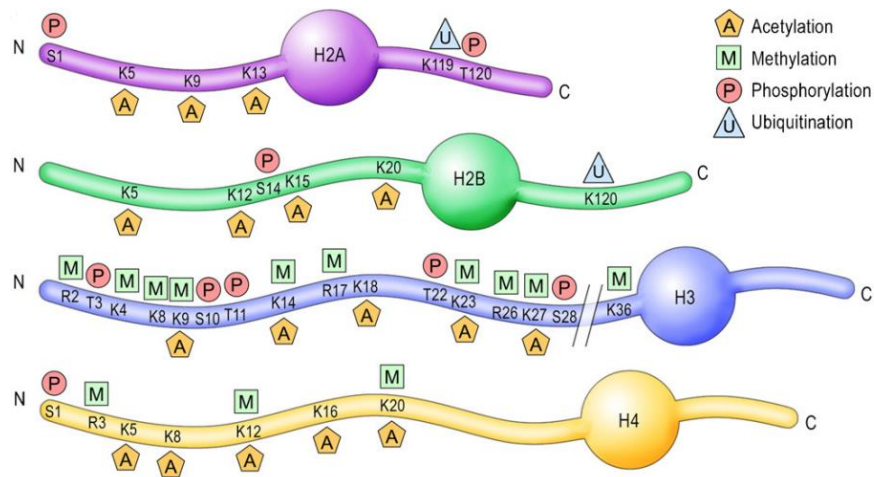
The histone kinases transfer a phosphate group from ATP to the hydroxyl group of the target amino-acid in the side chain, adding significantly more negative charge to the histone, affecting the chromatin structure, it is not well known how the kinases are recruited to their site of action on the chromatin, in few cases, for example, the mammalian MAPK1 enzyme has a DNA-binding domain with which is targeted towards the DNA and may be sufficient for specific recruitment (S. Hu et al., 2009).

There have been described cases where the phosphorylation occurs outside the N-terminal tails, for example, the phosphorylation of H3Y41 in the histone core region by the non-receptor tyrosine kinase (JAK2) (Dawson et al., 2009). Less is known about the role of histone phosphatases, it is speculated to have high activity within the nucleus. The phosphatase PP1 antagonizes Aurora B, a kinase that participates in mitosis (Sugiyama et al., 2002).

The best-known function of histone phosphorylation is the response to DNA damage, when phosphorylated histone H2A(X), demarcates large chromatin domains around the site of DNA breakage, however other studies have shown that histone phosphorylation participates in chromatin remodeling and other nuclear processes including gene regulation (Rosetto et al., 2012).

A schematic representation of the main histone modifications can be found in (Figure 1.9).





**Figure 1.9** Schematic representation of the core histones and their specific residue modification. Adapted from (Gräff and Mansuy 2008)

### 1.8.4 Other histone modifications

Other histone modifications have been characterized (Bannister et Kouzarides 2011).

**Deimination** is catalyzed by the peptidyl deiminase PAD14 which converts peptidyl arginines to citrulline, allowing the neutralization of the arginine positive charges (Cuthbert et al., 2004).

**$\beta$ -N-acetyl glucosamine** regulates a variety of proteins including the histones by modifying the lateral chain of serines and threonines with sugar residues of  $\beta$ -N-acetyl glucosamine (O-GlcNAc). In mammals, only the O-GlcNAc transferase catalyzes the transfer of the sugar residue from the donor (UDP- GlcNAc) to the target protein and only the  $\beta$ -N-acetylglucosaminidase is capable to remove it. For the moment this modification has only been observed in the histones H2A, H2B, and H4 (Sakabe et al., 2010).

**ADP ribosylation** process is performed by two enzymes with opposite functions; the ADP-ribose-polymerase and the poly-ADP-ribose-glycohydrolase. This modification is correlated with a relaxed chromatin state (Bannister et Kouzarides 2011).

**Ubiquitylation** is an important covalent modification. The ubiquitin is a polypeptide that is attached to the histone lysine by the action of three enzymes: E1 activating enzymes, E2 conjugating enzymes, and E3 ligating enzymes (Hershko et Ciechanover 1998). Two ubiquitylation sites have been well characterized in histone H2A and H2B. The H2AK119ub1 is implicated in gene repression, meanwhile, H2BK123ub1 plays a role in the transcription initiation and elongation processes. The ubiquitin can be removed by the action of the de-ubiquitin enzyme isopeptidase.

**Sumoylation** (small ubiquitin-like modifier) is a mechanism linked to ubiquitylation, using the same E1, E2, and E3 enzymes. Sumoylation can occurred on all four core histones and seems to function by antagonizing acetylation and ubiquitylation. Therefore, is hypothesized to have repressive functions (Shiio & Eisenman, 2003).

**Histone tail clipping.** Perhaps the most radical way to remove histone modifications is to remove the histone N-terminal tail in which they reside, a process referred to as tail clipping. In yeast, the proteolytic enzyme remains unknown, but the clipping process is involved in regulating transcription. The mouse enzyme was identified as Cathepsin L, which cleaves the N-terminus of H3 during ES cell differentiation (Duncan et al., 2008; Santos-Rosa et al., 2009).

**Histone proline isomerization.** Proline isomerases facilitate the interconversion between the cis and trans conformation of the peptidyl proline's peptide bond, which has the potential to stably affect peptide configuration. Proline isomerization is an important modification of the histone tail. It is, however, not a true covalent modification since the enzyme merely 'flips' the peptide bond by 180°, thereby generating chemical isomers rather than covalently modified products (Bannister et Kouzarides 2011).

## 1.9 Transcription factors

Transcription factors (TF) are cellular components that regulate gene expression, and their activity is critical to cellular function and responsiveness to the environment. The transcriptional regulatory apparatus has a main role in governing several biological processes, including cell cycle progression, maintenance of physiological and metabolic homeostatic state, differentiation, and development, this cellular transcriptional machinery must respond correctly to several internal and external stimuli (Vaquerizas et al., 2009).

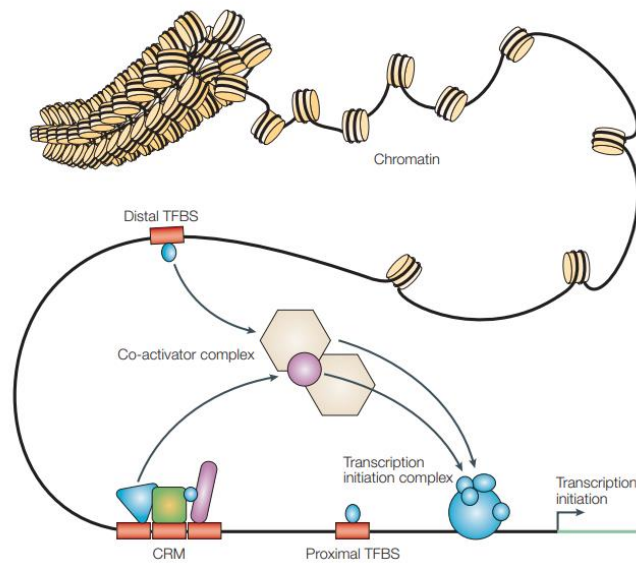
Several diseases originate from a disruption in the regulatory system, a high percentage of oncogenes are TFs, and 33% of developmental disorders are attributed to TFs miss function (Boyadjiev & Jabs, 2000; Furney et al., 2006). Alterations in the activity and specificity of TFs are believed to be major sources of evolutionary diversification (Lopez-Bigas et al., 2008).

Successful transcription by RNA polymerase in eukaryotes depends on different arrays of proteins target to the promoter sequence, these proteins include general transcription factors (GTFs), co-factors, histones, and chromatin remodeling proteins, in addition, some sequence-specific DNA binding TFs regulates transcription initiation of specific promoters (Lemon and Tjian 2000).

Mechanistically TFs are proteins capable of binding DNA in a sequence-specific manner and can regulate transcription (Figure 1.9) (Fulton et al., 2009; Vaquerizas et al., 2009). TFs have a high preference for specific binding sequences (Geertz et al 2012). TF DNA-binding specific sequences are often referred to as motifs, which, represents the set of related short sequences preferred by a given TF, and they can be used to scan longer sequences like promoters to identify potential binding sites (Lambert et al., 2018).

Most TF-binding sites are small ranging from 6 to 12 base pairs and they are flexible, so a typical human gene (>20 kb) will contain multiple potential binding sites for most TFs (Wunderlich and Mirny, 2009), therefore TFs must cooperate in a synergistic way to overcome this deficit in specificity. The majority of human TFs have to work together to achieve their function, but the details of their interactions and relationships remain enigmatic. The biochemical effects of TFs after binding DNA are also largely unmapped and known to be context-dependent. The resulting frustration in determining the relationship between TFs

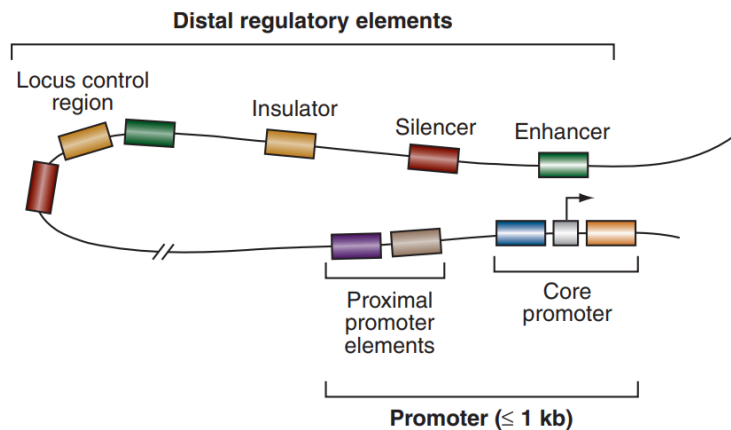
and gene sequences has been termed “the futility theorem” (Figure 1.10)(Wasserman & Sandelin, 2004).



**Figure 1.10** Schematic representation of transcriptional regulation. Transcription factors (TFs) bind to specific sites (TFBS), sets of TFs can act as Cis-regulatory modules (CRM). Adapted from (Wasserman & Sandelin, 2004).

## 1.10 Transcriptional regulatory elements

To understand the molecular mechanisms that control the specific expression patterns behind the biological process, such as development, proliferation, apoptosis, aging, or differentiation among others we need to define the so-called transcriptional regulatory elements. The cis-regulatory elements (CRE) are non-coding sequences of DNA that carries regulatory elements like transcription factors or histone modifications, according to their position from the TSS of the gene, they are subdivided into two groups, proximal regulatory elements: Promoters and distal regulatory elements: enhancers, silencers, insulators, and locus control region (Figure 1.11) (Maston et al., 2006).



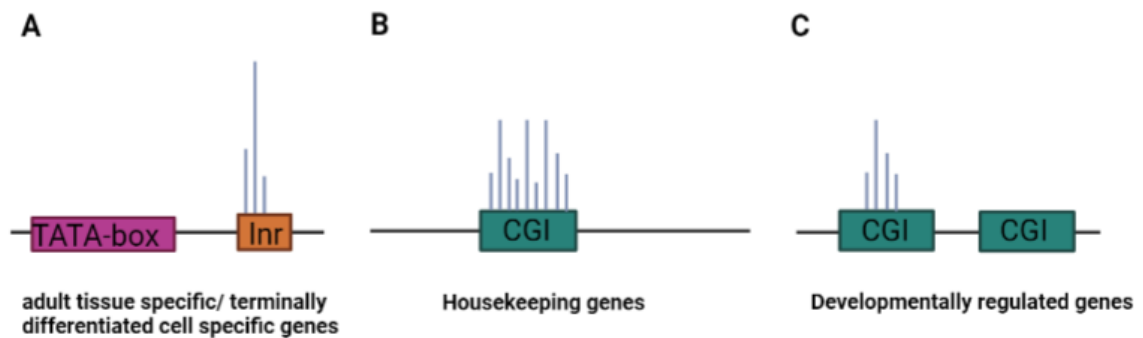
**Figure 1.11** Scheme of a gene regulatory region (taken from Maston et al., 2006).

### 1.10.1 Promoters

The promoter is the cis-regulatory element responsible for the initiation of gene transcription, is located upstream the 5' end of the transcription initiation site, and consists of two defined elements: the core promoter and the proximal promoter. The core promoter is the region at the beginning of a gene that anchors the basic transcriptional machinery and the transcription pre-initiation complex (PIC), establishes the position of the TSS and the direction of transcription, so is the minimal DNA region around the TSS that can induce transcription ( $\pm 50$ bp). Several core promoter motifs have been identified. The TATA box was the first described core promoter element that acts as a binding site for the TBP (Tata box binding protein) a subunit of the TFIID (Transcription factor IID).

Two types of core promoters have been described 'focused' and 'dispersed'. The 'focused' promoters have a single and well defined TSS (Figure 1.12a), and the 'dispersed' promoters have multiple closely-spaced TSS that are used with similar frequency (Figure 1.12b). Focused initiation occurs in promoters of highly cell-type-specific genes with restricted expression patterns, whereas dispersed initiation is mainly associated with housekeeping genes.

Three types of core promoters have been defined in metazoa based on different properties like initiation pattern, sequence composition and motifs, chromatin configuration, and gene function: (1) core promoters with sharp initiation patterns TATA-box, and Inr motifs (Figure 1.12a). These promoters have key regulatory elements near their TSS and are active in terminally differentiated cells in adult tissues and they acquire H3K4me3 and H3K27ac marks which are associated with active transcription. (2) Core promoters of housekeeping genes, present a well-defined nucleosome-depleted region (NDR), and they present H3K4me3 and H3K27ac marks (Figure 1.12b). In mammals, these promoters overlap individual CpG island (CGI) (3) Core promoters of key developmental transcription factors involved in patterning and morphogenesis. In mammals, they resemble housekeeping gene core promoters, which in embryonic stem cells however are distinctly bivalently marked with both H3K4me3 and the repressive modification H3K27me3 (Figure 1.12c). This possibly primes them for activation in the correct cell lineage and for silencing in all other cells, such promoters are associated with long individual CGIs or multiple CGIs (Haberle & Stark, 2018).



**Figure 1.12** Mammalian core promoter types. Adapted from (Haberle & Stark, 2018).

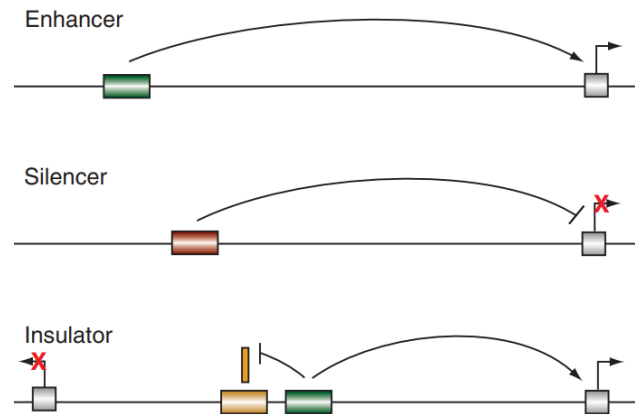
Because of the low activity of the core promoter, proximal promoter elements are required. The proximal promoter is the region upstream of the core promoter (up to a few hundred bases) that harbors multiple binding sites for activators and general transcription factors (GTFs) to increase the rate of transcription. Usually, the term promoter refers to the region of DNA encompassing the core and the proximal promoter elements although only a part of it is important for the regulation of transcription initiation (Andersson & Sandelin, 2020). More details of promoters and the mechanisms behind gene transcription regulation can be found in Chapter II Alternative promoters.

### 1.10.2 Enhancers

The enhancers are regulatory elements that are capable of regulating transcription in a spatial or temporal-specific manner, and they act independently of the distance and/or orientation relative to the promoter. Enhancers are composed of clusters of transcription factor binding sites (TFBSs) that work together to enhance transcription. The spatial organization and orientation of the TFBSs play a critical role in its regulatory activity, so the property of distance and orientation independence applies only to the enhancer cluster as a whole (Reményi et al., 2004).

Enhancers and proximal promoters are functionally similar, in the sense that they can bind the same activators in different genes, however contrasting with proximal promoters, enhancers are normally long-distance transcription regulatory elements, they can be located quite distal from the core promoter (Figure 1.13). They can be situated a hundred kilobase pairs upstream of the promoter, but they can also be located downstream of the promoter in an intronic region or even further from the 3' end of the gene (Lettice et al., 2003).

The distal control of transcription by enhancer elements can be explained by the DNA-looping model, where the enhancer and core promoter are brought together by looping out the DNA in question, many studies suggest that this model may be a general mechanism of enhancer function (Vilar & Saiz, 2005).



**Figure 1.13** Distal transcriptional regulatory elements. Taken from (Maston et al., 2006).

### 1.10.3 Silencers

Silencers are regulatory elements with a negative effect on gene transcription, which means that they have repressor activity, as enhancers, they function independently of distance and/or orientation from the promoter. They can be part of proximal promoters, of distal enhancers, or as an independent distal regulatory element, they can be located far from their target gene, in the intronic region or the 3' untranslated region, silencers can cooperate in binding DNA to have a synergistic effect (Harris et al., 2005; Sertil et al., 2003).

Silencers are binding sites for negative transcription factors called repressors. The repressors can also recruit negative cofactors called corepressors (Privalsky, 2004). An activator can switch to a repressor by differential cofactor recruitment (Ogbourne & Antalis, 1998; Perissi et al., 2004).

Many models of repression mechanisms have been described. repressors can act by blocking the binding of a nearby activator (Harris et al., 2005), by competition of the same binding site (L. Li et al., 2004), also the repressors can have the ability to impede the access of activators or GTFs by modifying the chromatin into a repressive state (Srinivasan & Atchison, 2004). Interestingly has also been observed that the repressor may block transcription by inhibition of PIC assembly (L. Chen & Widom, 2005).

Polycomb group response elements (PREs) are a class of silencers that also has anti-silencer activity depending on the factors that are bound to it, the switch depends on the transcription of non-coding RNAs within the same element, by transcribing these sequences somehow the access of repressive complexes to the DNA is blocked, suggesting that transcription of silencer elements may play a role in the regulation of silencer activity (Schmitt et al., 2005).

#### 1.10.4 Insulators

Insulators or boundary elements isolate genes so they cannot be affected by the transcription activity of neighboring genes, they defined domains to limit the action of transcriptional regulatory elements. Insulators have two principal properties: they can disturb enhancer-promoter communication (enhancer blocking activity) and they can arrest the spread of repressive chromatin (heterochromatin barrier activity) (Recillas-Targa et al., 2002). They usually have a length between 0.5 to 3 kb and they function in a position-dependent, orientation-independent manner (Maston et al., 2006).

The number of insulator elements in the human genome is enigmatic but now is believed that insulator elements may be less common than first envisaged, and they can only be found in regions with a high density of coding or regulatory information (Fourel et al., 2004).

The only protein known to mediate insulator activity in vertebrates is CTCF (CCCTC binding factor), which plays a role in different loci, including the chicken globin 5´HS4 and the mammalian H19/Igf2 ICR (Bell et al., 1999; Bell & Felsenfeld, 2000). The CTCF activity can be regulated by DNA methylation, post-translational modifications, and cofactors interaction (A. G. West & Fraser, 2005).

The precise mechanisms by which insulators carry out their activity are not well characterized. Two models have been proposed to explain insulator function (Capelson & Corces, 2004).

The first model links insulators with the transcriptional machinery. In this model, enhancer blocking activity can be explained by the inability of an insulator-bound activator to interact with its target promoter. Heterochromatin barrier activity is explained by the recruitment of gene activating factors or histone-modifying enzymes, that serve as an anchor site for a permissive chromatin state that, in turn, disrupts the spread of repressive chromatin (Defossez et al., 2005).

The second model associates insulators with the structural organization of chromatin. This model proposes a role for insulators in physically separating chromatin into independent structural domains. In this model, positioning an insulator between an enhancer and its target promoter, results in enhancer blocking activity due to physical obstruction between the two elements, preventing the communication among them. Likewise, flanking a gene with insulator elements provides heterochromatin barrier activity because of the creation of an independent expression domain (Capelson & Corces, 2004).

#### 1.11 LncRNAs

The encyclopedia of DNA elements (ENCODE), an international consortium, has shown that up to 80% of the human genome is transcribed, but only 1.5% of it is protein-coding sequences (Neph et al., 2012). Non-coding RNAs can be divided into two groups based on the length of the transcript: short and long noncoding RNAs. The first group includes well-

characterized RNAs such as tRNAs, snRNAs, snoRNAs, miRNAs, piRNAs among others. The second group includes rRNA and Long non-coding RNAs (lncRNAs).

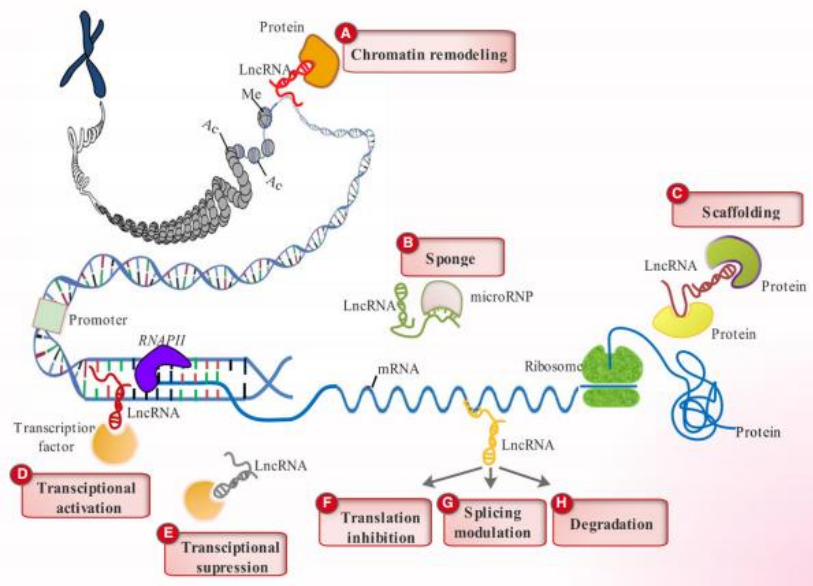
lncRNAs are transcripts of more than 200bp in length without a large open reading frame nor protein-coding potential. The complexity of different organisms is well correlated with the amount and diversity of non-coding transcripts. lncRNAs shared common features with mRNAs; as they are often capped, polyadenylated, and undergo splicing, however, they also have specific features: lncRNAs have fewer exons in their structure, 42% of lncRNAs have 2 exons in contrast with only 6% of mRNAs, lncRNAs are shorter, in the order of hundred than the mRNAs which are in the order of thousand base pairs, another distinguishing feature is that 95% of all multi exonic mRNA are spliced whereas only 25% of lncRNAs undergo alternative splicing. lncRNAs have lower expression levels, their expression is more tissue-specific, and they are mostly located in the nucleus (Derrien et al., 2012).

lncRNAs are less conserved than mRNAs, however, their sequence conservation is not always correlated with their function; we have two clear examples: XIST and MALAT1, the first one, presents a low degree of sequence conservation, nevertheless, their X chromosome silencing function is well conserved across mammals and MALAT1 although it has a high degree of sequence conservation between human and mice, knock-out experiments in human and mouse cell lines have shown different phenotypes. These examples highlight that lncRNA sequence conservation is not always a predictor of functionality and that functional conservation is more complex in lncRNAs (Diederichs, 2014; Galupa & Heard, 2015).

Several studies showed that lncRNAs play an important role in diverse cellular processes, mainly in the regulation of gene expression at the epigenetic, transcriptional (Ng et al., 2012), and post-transcriptional level (Joon et al., 2012). Some lncRNAs have been associated with the development and human diseases (Sparber et al., 2019).

Several mechanisms for lncRNAs gene expression regulation have been proposed: lncRNAs can recruit different protein components of the chromatin remodeling complex to change the chromatin structure, they can function as sponges by base-pairing with their complementary miRNAs and altering their effects, lncRNAs can act as a scaffold by providing anchor sites for proteins that function together in the same biological pathway, they activate transcription of certain genes by guiding transcription factors to their promoters, lncRNAs are capable of suppressing transcription by sequestering transcription factors and keeping them away from their promoters, they can modulate mRNA functioning through base pairing with them and therefore interfering with the translation or splicing process or tag them for mRNA degradation (Figure 1.14)(Salehi et al., 2017).





**Figure 1.14** Mechanisms for LncRNAs gene regulation. Adapted from (Salehi et al., 2017).

## 1.12 Epigenetic regulation

The biological processes that involved temporally and spatially precise patterns of gene expression such as development and differentiation take advantage of changes within the chromatin structure, they rely specifically on the covalent modifications of histones above mentioned. Diverse studies demonstrate that the enzymes responsible for these modifications function in a coordinated manner to control gene expression in the short term and the inheritance to their progeny in a long-term manner.

### 1.12.1 Gene regulation by histone modifications

Histone modifications have different functions. First, except for methylation, histone modifications result in a change in the net charge of nucleosomes, which could loosen inter- or intranucleosomal DNA-histone interactions. Second, histone modification can be recognized by other proteins (Seet et al., 2006). Individual histone modifications or modification patterns might be read by other proteins that influence chromatin dynamics and function (Jenuwein & Allis, 2001). Third, some modifications directly influence higher-order chromatin structure. For example, acetylation of H4 K16 inhibits the formation of compact 30 nm fibers (Shogren-Knaak et al., 2006). These mechanisms are not mutually exclusive.

### 1.12.2 The epigenetic signature of regulatory elements

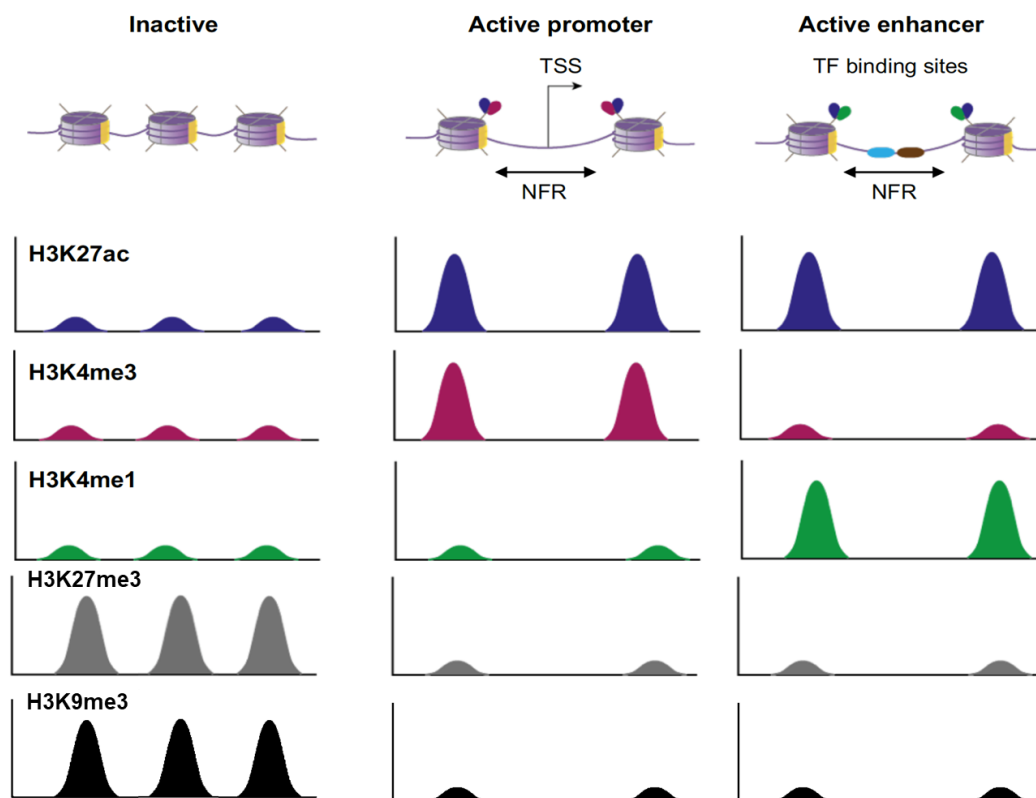
Gene transcription activation in eukaryotes is coordinated by multiple TFs and co-factors acting on regulatory DNA sequences, like promoters and enhancers. Deciphering the regulatory information encoded in the genome will require a thorough understanding of the relationships between the transcriptional activities of the.

Progress in different fields like epigenetics and chromatin biology suggests a histone code (Strahl et al., 2000) of increasing complexity with profound implications for chromatin as both

a receptive substrate and a predictive signal in a variety of biological processes (Margueron et al., 2005).

Occupancy of TFs at enhancers and promoter is associated with nucleosome-free regions (NFR), exhibiting high sensitivity to DNA nucleases such as the DNase I (Gross and Garrard 1988) Nucleosomes directly flanking TF binding regions are less mobile and tagged with specific histone modifications, including, but not limited to, H3K4me1 and H3K27ac.

H3K4me1 was the first histone modification linked to distal regulatory regions. The ENCODE project analysis of histone modifications associated the H3K4me1 with distal enhancer regions in contrast to H3K4me3 which is present at active promoters (Figure 1.15) (Heintzman et al., 2007).



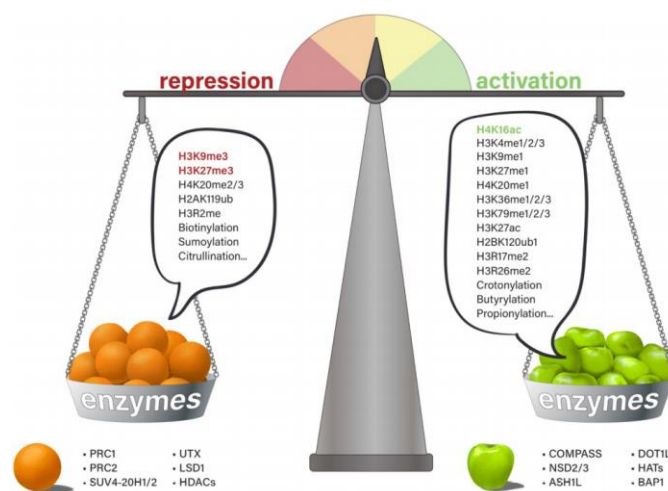
**Figure 1.15** The epigenetic signature of regulatory elements. Adapted from (Pundhir et al., 2015).

Inactive chromatin is enriched in both H3K27me3 and H3K9me3 histone marks whereas active enhancers and promoters share some features like nucleosome depletion and enrichment of histone acetylation H3K27ac. The high-resolution profiles of these markers and the dichotomy of enrichment for trimethylated H3K4 and monomethylated H3K4 at active promoters and enhancers (Figure 1.15) define chromatin signatures that can use novel regulatory elements in the human genome (Heintzman et al., 2007).

### 1.13 Epigenetic de-regulation in Cancer

The enzymes in charge of histone modifications contribute to chromatin compaction, nucleosome dynamics, and transcription, the implementation of these modifications can be initiated in response to intrinsic and external stimuli. Dysregulation of these processes can shift the balance of gene expression and is frequently observed in human cancers (Figure 1.16).

The genes encoding histone-modifying enzymes/complexes can be altered by the following mechanisms; gain or loss of function mutations, overexpression, and repression by promoter hypermethylation or chromosomal translocation, the alteration can also occur in the site of histone modification (Audia & Campbell, 2016).



**Figure 1.16** Balance representation between activation and repression histone modifications marks. Taken from (Z. Zhao & Shilatifard, 2019).

Among the frequently mutated targets in cancer, we found chromatin-bound proteins (Shen et al., 2013). The dysregulation of chromatin-associated proteins can act as drivers on specific types of cancer (Garraway and Lander 2013).

Transcription activation and repression are controlled by an array of histone modifiers and chromatin-bound proteins. A balance between specific modifications and modifiers is needed to execute the proper gene expression program during a specific cellular stage (Figure 1.15). The disruption of the balance can alter cell phenotype and trigger disease onset and progression (Z. Zhao & Shilatifard, 2019).

The methylation marks associated with transcriptional activation include H3K4, H3K36, and H3K79 and the ones associated with transcriptional repression occur on H3K9, H4K20, and H3K27 (Figure 1.15). The bivalent domains (BD) are large regions of H3K27 methylation harboring smaller regions of H3K4 methylation. BD can maintain pluripotency by silencing developmental genes in embryonic stem cells (ESCs) while keeping them poised for activation during differentiation stages (Bernstein et al., 2006). Altering the balance of these histone modifications can modify gene expression patterns and contribute to the pathogenesis of cancer (Zhao & Shilatifard 2019).

Great effort has been devoted to understanding the role of histone modifications and the enzymatic machinery involved in the implementation of these modifications during development and disease, especially for cancer. Interestingly, histone modifiers often reside within large multi-protein complexes for proper function, such as MLL/COMPASS, PRC2, and HDAC complexes (Zhao & Shilatifard 2019). Here we described some examples of the implication of the altered histone modifier machinery with cancer.

Histone H3K4 methylation at enhancers and promoters is performed by methyltransferases in the COMPASS family including SET1A, SET1B, and MLL1-4. The trimethylation of H3K4 at promoters is done preferentially by SET1A and SET1B. SET1A has been implicated in breast cancer metastasis, lung cancer, and colorectal cancer tumorigenesis through methylation of histones (Salz et al., 2015).

In acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL) patients, MLL1 is frequently mutated through translocation with other oncogenes, the resulting chimeric proteins lack the catalytic SET domain of MLL1 and drive leukemogenesis (Smith et al., 2011).

Histone H3K27ac is deposited by CBP/p300, the acetylation can function as a signal recognized by bromodomain (BRD)-containing proteins like BRD2, BRD3, and BRD4. Mutation, aberrant expression, and gene fusions have been found in these proteins and implicate their roles in cancer development and progression (Fujisawa & Filippakopoulos, 2017).

### 1.13.1 Super-enhancers and Broad H3K4me3 domains

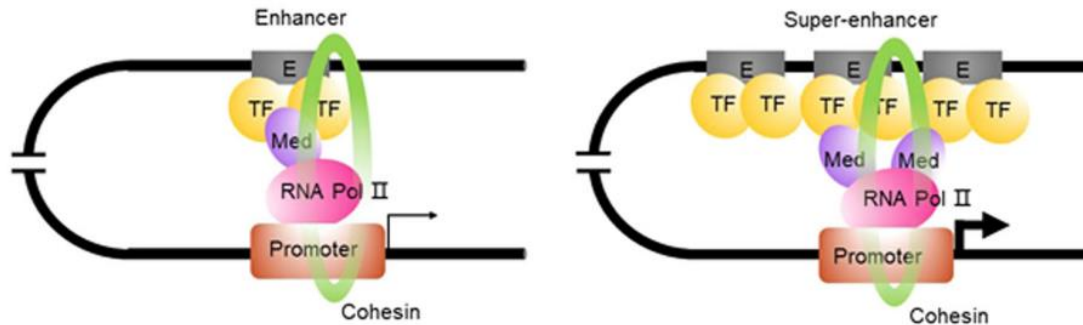
In addition to typical enhancers, described before, the genome is also comprised of large stretches of enhancers in close linear proximity, often spanning several kilobases in length called super-enhancers (Figure 1.17)(Whyte et al., 2013).

Enhancers and super-enhancers share similarities, both are occupied by the same components generally associated with enhancer activity, including transcription factors (TFs), co-activators such as the Mediator complex, chromatin regulators, and the RNA polymerase II (pol II) complex. However, super-enhancers harbor these factors on an average 10-fold higher density than typical enhancers(Denes et al., 2013). As a result, super-enhancers can drive higher levels of transcription of their target genes than typical enhancers (Figure 1.16).

Super-enhancers were first identified in mouse embryonic stem cells (ESC) based on the high-density occupancy of ESC-specific master transcription factors OCT4, SOX2, and NANOG at regulatory enhancers nearby genes that define ESC identity (Whyte et al., 2013). Super-enhancers were found to regulate the expression of key oncogenic drivers in many tumor samples, suggesting that cancer cells are often addicted to the super-enhancer-driven transcriptional programs (Bradner et al., 2017).

A key property of enhancers and super-enhancers is that they contain clustered binding sites for multiple TFs, allowing the coordinated binding of multiple TFs, including master TFs and transcriptional effectors of signaling pathways, ensuring integration of intrinsic and extrinsic environmental cues at these elements (J. Yan et al., 2013).

Occupancy of TFs at enhancers and super-enhancers are associated with regions of nucleosomal depletion and increased DNA accessibility (Gross & Garrard, 1988). The abovementioned regions have a distinct chromatin signature including, but not limited to, histone H3 lysine 4 monomethylation (H3K4me1), histone H3 lysine 27 acetylation (H3K27ac) (Thandapani 2019).



**Figure 1.17** Schematic representation of enhancer and super-enhancer. Adapted from (Jia et al., 2020).

Preferential enrichment of the H3K4me1 mark over H3K4me3 is used to differentiate enhancers from active promoters (Heintzman et al., 2007).

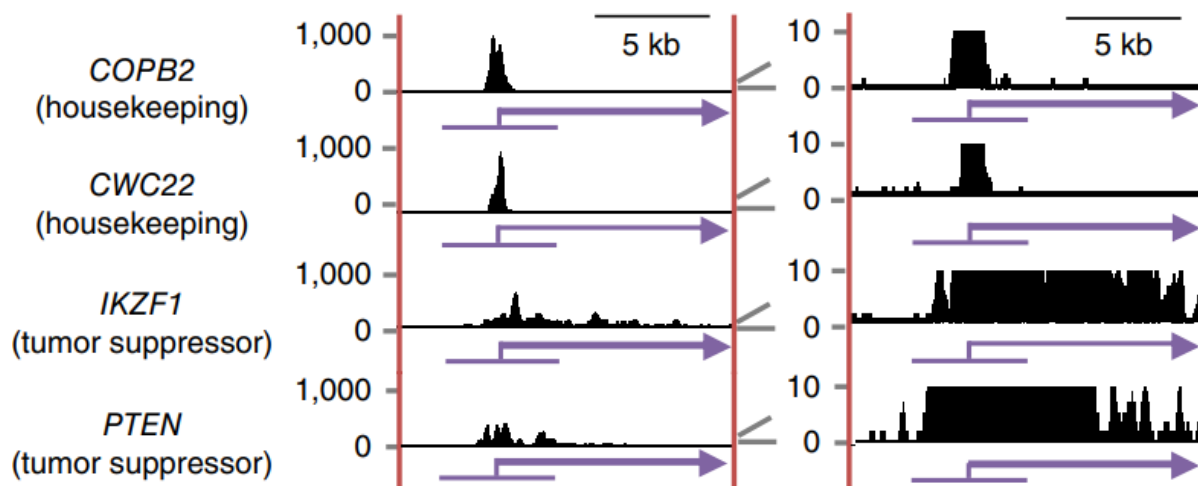
H3K27ac differentiates active enhancers from inactive/poised enhancers containing H3K4me1 (Creyghton et al., 2010). These two modifications in combination with nuclease hypersensitivity data or coactivator occupancy of Mediator 1 provide a robust readout of active enhancers and super-enhancers and have been used for their annotation in different studies. Binding of pioneering TF is the primary event that primes enhancer/ super-enhancer elements for activation by the subsequent recruitment of co-activator proteins such as histone modifiers, ATP-dependent chromatin remodelers, and Mediator complex (Calo & Wysocka, 2013).

Activated super-enhancer facilitate long-range physical interaction that brings the super-enhancer and promoter in close 3D proximity via looping out the intervening DNA sequence. The resultant chromatin loops enable interactions with the basal RNA pol II transcription machinery at target promoters to activate transcription (Levine et al., 2014). The exact mechanism of chromatin loop formation is still not clear, but it was previously shown that cohesin binding is enriched at promoters and enhancers that form chromatin loops (Kagey et al., 2010).

### 1.13.2 Broad H3K4me3 domains

The H3K4me3 mark is associated with open chromatin, and therefore found in active promoters and enhancers; the length of this mark varies in size and can be found in broader deposits even of several Kb, characteristic, that is shared with other active epigenetic marks, and in which the length-size of the mark is used to differentiates between enhancers (<1Kb long) and clusters of enhancers (also called super-enhancers) that usually spanned more than 10Kb (Hnisz et al., 2017).

Several groups of research (Benayoun et al., 2015; K. Chen et al., 2015; Koch et al., 2011; Pekowska et al., 2010; Suzuki et al., 2017; Zacarías-Cabeza et al., 2015) have revealed enrichment of H3K3me2/me3 mark several Kb away from the TSS into the gene body (Broad H3K4me3 domains). Pekowska and col. 2010 discovered that Broad H3K4me3 domains tag tissue-specific genes in T-cells. An example of a gene harboring a broad H3K4me3 domain can be found in (Figure 1.18).



**Figure 1.18** Examples of sharp H3K4me3 domains in two housekeeping genes (top) and broad H3K4me3 domains of two tumor suppressor genes (bottom). Taken from (Chen et al., 2015).

Broad H3K4me3 domains are more often associated with cell type-specific genes (Benayoun et al. 2014), Broad H3K4me3 are also found in the HOX gene clusters and T-cell receptor loci (Bernstein et al., 2005; Zacarias-Cabeza et al. 2015). Moreover, Chen et al 2015, found the loss of Broad H3K4me3 domains tagging a set of tumor suppressor genes in cancer cells otherwise present in normal cells. High levels of RNA pol II recruitment and pausing have been associated with Broad H3K4me3 and thus helping with transcriptional consistency, (Benayoun et al., 2014), conferring to those genes harboring H3K4me3 resistance to environmental perturbations (Chen et al., 2015). H3K4me3 broad domains seem to be more implicated in defining a specialized chromatin structure for fine-tune cell-specific gene regulation rather than up-regulation of gene expression (Benayoun et al., 2015).

In a study in which I collaborate during my Ph.D. and recently published (**ANNEX I**), we analyzed epigenomic data of human T-cell precursors and T-ALL samples and we found that oncogenes were significantly associated with gain of Broad H3K4me3 domains in leukemic samples including the major T-ALL oncogenes highlighting the importance of these domains in the oncogenic events leading to T-cell transformation (Belhocine et al., 2021)

## 1.14 Epigenomic resources

The regulatory elements are usually identified by functional genomics approaches or sequence conservation, they often have cell or tissue specificity. As mentioned before regulatory elements like promoters and enhancers regulate gene expression usually by TF binding. TFs bind to specific DNA sequences (motifs), which can be identified using chromatin immunoprecipitation followed by sequencing (ChIP-seq) assays. They bind to DNA in nucleosome-free regions (NFR), these regions can be identified using DNase I hypersensitivity assays. DNase I footprinting can also help to identify high-resolution TF-binding sites within the larger DNase I hypersensitive sites (DHSs) (Neph et al., 2012).

Histone modification can also be identified by ChIP-seq assays and although most sequence-specific TFs and some chromatin marks lead to highly localized ChIPseq signals (hundreds of nucleotides), other marks (such as H3K9me3 and H3K36me3) are associated with large genomic domains that can cover up to a few megabases. Thus, epigenetic changes can alter TF accessibility in different cellular states and may modify the activity of regulatory elements, resulting in cell-type specificity of their associated genes (Khurana et al., 2016).

Distal regulatory elements may regulate gene expression by interacting with promoters in a three-dimensional (3D) structure of the genome manner. Linking the distal elements to their target protein-coding genes in the 3D chromatin structure. Multiple approaches have been used to link cis-regulatory regions to their target genes. For example, chromosome conformation capture (3C) technology has demonstrated that regulatory sequences can control transcription by looping to and physically contacting target coding genes that are located tens or hundreds of kilobases away (Hughes et al., 2014). The 3C technology probes one-versus-one contacts in the 3D space of the genome. Further variations of the 3C technology have since been developed that probe one-versus-all (4C), many-versus-many (5C), and all-versus-all (HiC) contacts (de Laat & Dekker, 2012).

Several large-scale efforts such as ENCODE (Dunham et al., 2012) and the National Institutes of Health (NIH) Roadmap Epigenomics Consortium (Chadwick, 2012) BLUEPRINT Consortium (Martens & Stunnenberg, 2013) were launched to create a comprehensive map of regulatory regions, in table 1.3 we summarize the different sources of non-coding element annotations. Gene expression studies of different tissues can point out regions that are associated with tissue-specific expression. The Genotype-Tissue Expression (GTEx) project has provided an atlas of gene expression across multiple tissues and many individuals, which can be used to identify potential regulatory regions (The GTEx Consortium, 2013).

**Table1.3** Non-coding annotations

<b>Annotation</b>	<b>Resource</b>
<b>Transcription start sites</b>	GENCODE FANTOM
<b>Transcription factor-binding sites and motifs</b>	ENCODE Roadmap epigenomics JASPAR Transfac CIS-BP
<b>DHS (regions of open chromatin)</b>	ENCODE Roadmap epigenomics
<b>Histone marks</b>	ENCODE Roadmap epigenomics
<b>Histone marks of cell types from the blood</b>	BLUEPRINT
<b>Integrated chromatin states (including enhancers)</b>	ENCODE Roadmap epigenomics (derived from methods such as ChromHMM and Segway) FANTOM
<b>Enhancer–promoter linkages</b>	ENCODE Roadmap epigenomics FunSeq2
<b>Transcription factors– target gene linkages</b>	ENCODE (derived from CHIP-seq) ENCODE (derived from DHS) Roadmap epigenomics Remap (derived from CHIP-seq)
<b>Topologically associated domains from HiC</b>	ENCODE
<b>Various types of ncRNAs</b>	GENCODE miRBase snoRNABase GtRNAdb MiTranscriptom



# Chapter 2 Alternative promoters

## 2.1 Definition of alternative promoters

The main mechanism for transcript and protein diversity is the alternative splicing, in this process a set of exon combinations are used to generate multiple transcripts from a single gene that encode different protein isoforms with different functions. Alternative promoter usage adds a level of complexity to the isoform diversity generation (Pajares et al., 2007).

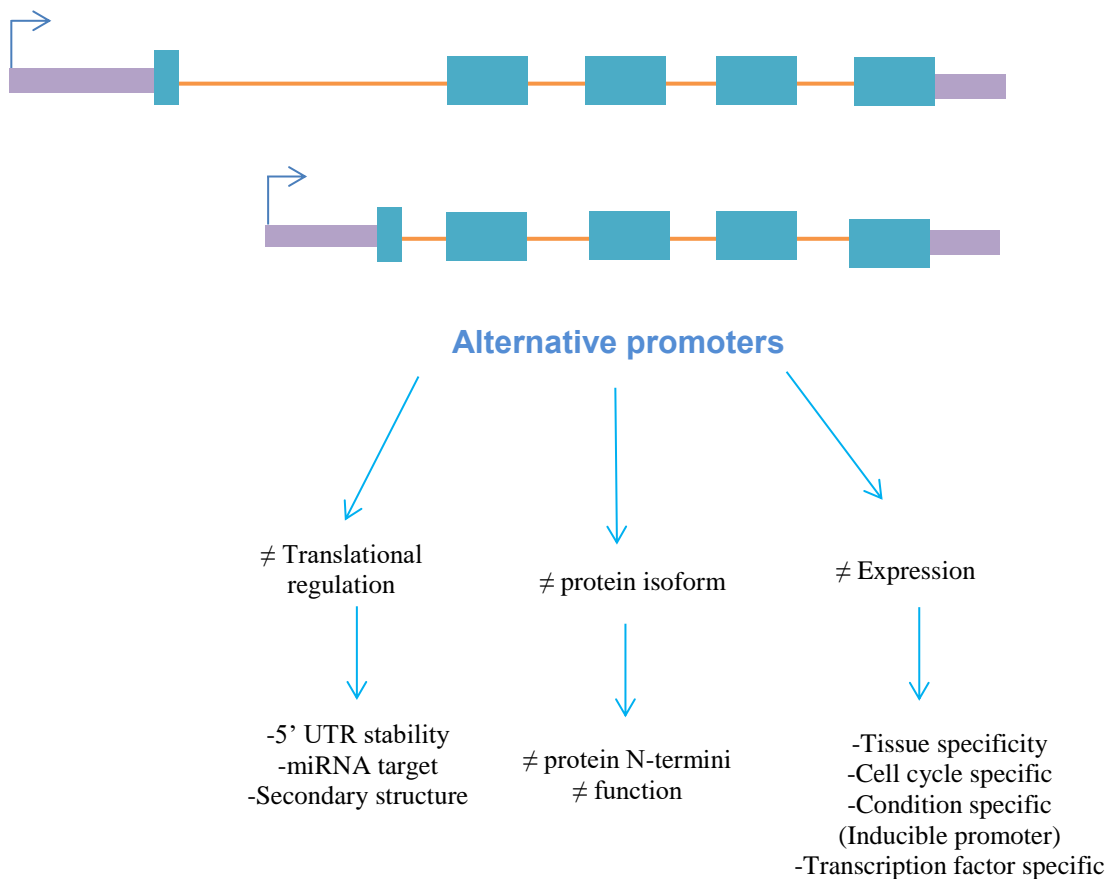
An alternative promoter (AP) is an alternative region within the same gene body from which transcription can initiate and therefore originates a variety of transcript isoforms of the gene without the need to undergo alternative splicing to generate isoform diversity (Davuluri et al., 2008).

Genome-wide analyses indicate that more than 60% of the human genes use alternative splicing and 30-50% of the human genes have multiple AP that can span up to thousands of kilobases (Kimura et al., 2006).

The existence of multiple transcripts for a single gene that differs in their 5' termini reflects the presence of APs. During the biogenesis of mRNAs, regulation of transcription initiation represents the first layer in the control of gene expression (Davuluri et al., 2009).

## 2.2 Impact of alternative promoters in protein function and gene expression

The use of APs leads to transcripts differing in their first exons or the 5' UTR. The use of alternative first exons leads to transcripts with different open reading frames (ORFs) and diversifies the repertoire of encoding proteins, giving rise to protein isoforms with alternative N-termini (Figure 2.1) (Davuluri et al., 2009).



**Figure 2.1** Impact of APs on gene expression and protein function.

Transcripts sharing the same coding region but a different 5'UTR can be subject to differential translational regulation, the mechanisms of translational regulation by 5' UTR are briefly described in the following sections.

### 2.2.1 The upstream untranslated region and its role in regulating gene expression

The region that spanned from the TSS to the initiation codon for protein translation in the mRNA is called the 5' untranslated region (UTR) and is important for the control of gene expression because influences the stability of the mRNA and the efficiency of translation. Existing reports underline these regions as templates for post-transcriptional regulation (Hinnebusch et al., 2016a).

Different transcripts with different UTRs from the same gene can have tissue specificity, therefore, controlling the expression of a protein in different developmental, physiological, and pathological states.

For several years the attention of the mRNA architecture has only been focused on the coding section, most recently this paradigm has shifted towards the untranslated regions; the 5'UTR and 3'UTR of the mRNA, where it is possible to find motifs necessary for mRNA regulatory aspects. Briefly, we can describe the regulatory motifs present in UTRs as follows:

### 2.2.2 Binding sites for regulatory proteins

Proteins can bind to unique RNA sites for regulation purposes, they can either bind to short RNA sequences or specific secondary RNA structures, and then modulate mRNA function, some examples include the iron response elements which are structural hairpin motifs involved in iron metabolism; and the AU rich sequence motif usually present in 3'UTRs which binds proteins for the stabilization or destabilization of mRNA (Hinnebusch et al., 2016b).

### 2.2.3 Secondary structure

The secondary structure within the 5' UTR besides acting as a scaffold for protein binding, can regulate translational efficiency in two more levels. Some structural motifs act as internal ribosome entry and induce cap-independent translation, however, most of the 5' UTR secondary structures represses cap-dependent translation by inhibiting binding or scanning of the translational machinery (Leppek et al., 2018).

### 2.2.4 Upstream open reading frames (uORF)

Short reading frames within the 5'UTRs, in humans usually 17 residues, can block the access of ribosomes to the correct start codon by providing premature start codons or generating ribosome stalling.

### 2.2.5 Binding sites for regulatory RNAs

Some examples of regulatory RNAs are small interfering RNAs (siRNAs) and microRNAs (microRNAs) which target mRNA by complementary base-pairing and affect mRNA stability or translational efficiency.

The before mentioned mechanisms have also gained importance by the fact that oncogenes, tumor suppressors, and genes associated with gene proliferation tend to have atypically long and complex 5'UTRs that contain more regulatory elements (Hughes 2006).

The synthesis of unstable or inefficient mRNA might appear as a waste of cellular energy but provides new possibilities to regulate protein expression. Elements anchored in UTRs are useful for changing the stability or efficiency of mRNA translation, giving a quicker mechanism for altering protein expression without the necessity of novel transcription (Hughes, 2006).

## 2.3 Regulation of alternative promoter usage

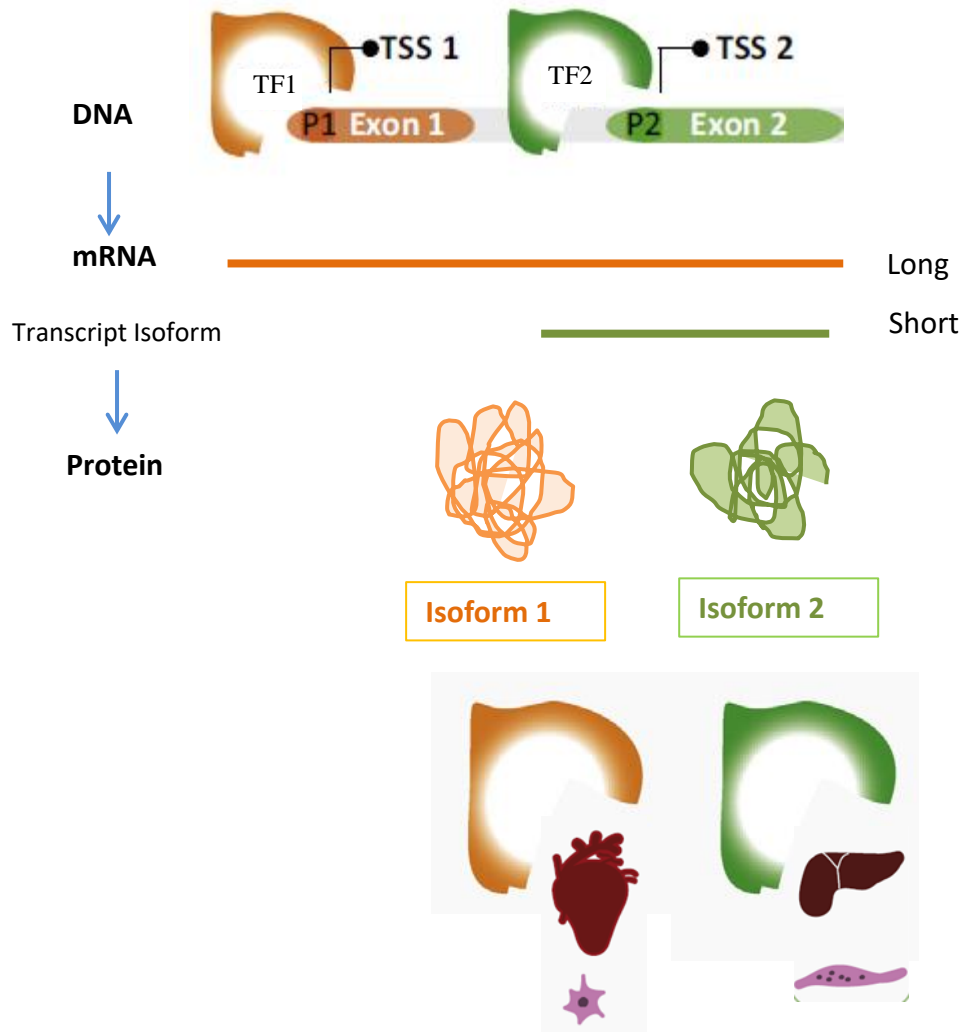
The molecular mechanisms responsible for the choice of APs and TSSs can be by alteration of the chromatin state and regulation mediated by cell and tissue-specific transcription factors. Understanding the biological importance of APs and tissue-specific TSSs requires learning how the choice of specific TSS is made and which transcription factor and regulatory networks are involved.

Promoters and enhancers are a major control hub for gene regulation that integrate information from a multitude of signaling pathways, through binding of signal responsive activators and repressors but also by epigenetic and post-transcriptional mechanisms.

### 2.3.1 Alternative promoter expression is tissue-specific

APs have a critical role in gene regulation during the determination of cell function and fate, APs increase the transcriptional repertoire of a gene allowing regulation and expression of different transcript isoforms in various tissues and developmental stages (Figure 2.2) (de Klerk & 't Hoen, 2015).

Transcriptome-wide studies suggest that TSS use is highly tissue-specific and that the number of alternative TSSs differs by tissue type, with the hippocampus accounting for a larger number of TSSs than any other tissue. The FANTOM Consortium is leading most of the research in the field of promoters and TSSs. In a survey that includes approximately 200 human primary cell types, 150 human tissues, and 250 human cancer cell lines, it was shown that on average there are four TSSs per gene. An interesting finding from this large TSS survey is that most genes are regulated in a tissue-specific manner and only a small percentage can be considered as true housekeeping. The use of alternative tissue-specific TSSs seems to be regulated by the presence of enhancer regions more than alternative core promoters. Half of all detected CpG islands promoters and more than 90% of all promoters lacking both CpG islands and a TATA box exhibit cell-type restricted expression due to the presence of proximal enhancers (FANTOM Consortium 2014).



**Figure 2.2** APs of the same gene can have different transcripts that code for different proteins with tissue-specific expression.

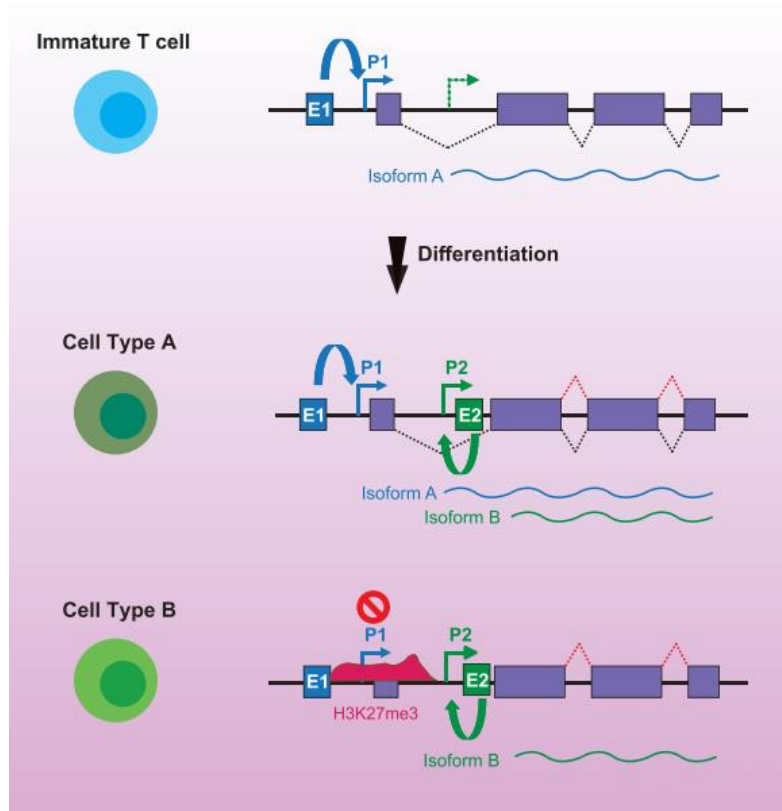
### 2.3.2 Regulation of alternative promoter usage by enhancer elements

The regulation of AP usage can be achieved by many mechanisms, including the regulation by distal regulatory elements like enhancers. Maqbool and col. Analyze the stage-specific transcriptional programs in the development and activation of T cells, they analyzed a wide epigenomic and transcriptional data set of mouse T cell differentiation, revealing the role of multiple enhancers and PRC2 in controlling AP choice in T cell development.

They propose a model of AP usage guided by enhancers, they study five *loci*, three with pre-existing or remaining after activation (*Runx3*, *S1pr1*, and *Nfatc1*) and two with activation-induced (*Lef1* and *mir181*), promoter-enhancer long-distance interactions respectively, highlighting the functional dynamics of enhancers in various contexts or cell types (Maqbool et al., 2020).

In the *Runx3* locus, AP usage in CD4 and CD8 subtypes is directed by putative enhancers number one and two towards the distal promoter P1, whereas putative enhancer number three controls P2, evidencing pre-existing contacts in the choice of APs.

They also observed that polycomb repressive complex 2 (PRC2) participates in AP choice repressing one specific TSS and leaving the other intact for activation (Figure 2.3) influencing gene isoforms expression.



**Figure 2.3** AP usage during differentiation can be mediated by enhancer elements and epigenetic marks. Taken from (Maqbool et al., 2020)

### 2.3.3 Regulation of alternative promoter usage by DNA methylation

One of the regulatory mechanisms of AP usage described by Maunakea and col. is the methylation of APs. This mechanism is one of the drivers of different isoforms expression in the same cell type but in different regions of the brain, highlighting the role of intragenic methylation in the control of context-specific expression of APs (Maunakea et al., 2010).

They found *SHANK3*, a gene related to the autism and 22q deletion syndrome to have at least two intragenic evolutionary conserved and differentially methylated CGI regions with AP properties according to H3K4me3 ChIP-seq data and CAGE-tag data, and later found out to be involved in the tissue-specific expression of the gene across species (Maunakea et al., 2010).

### 2.3.4 Regulation of alternative promoter usage by chromatin looping

One study revealed a mechanism behind the transcription of the gene *ZEB2* controlling the epithelial-mesenchymal transition (EMT) in triple-negative breast cancer (TNBC) cells after the stimulation with TNF $\alpha$ , which in turn activates at least two AP-1 signaling pathways including PI3K-Akt and MAPK/ERK, facilitating cancer invasion. The study demonstrates the induction of *ZEB2* under the stimulation of TNF $\alpha$  and its role in EMT providing a possible molecular mechanism for TNBC aggressiveness. The human gene *ZEB2* is formed by two promoters follow by different non-coding first exons separated by 2.2Kb and spliced to a

common exon 2. Transcript from exon 1a was found in all breast tumors analyzed while exon 1b shows a more restricted expression (Qiao et al., 2015).

The coordinate regulation of multiple genes through chromatin looping has been established (Kadauke et al., 2009). AP-1 regulates the expression of both *ZEB2* transcripts, differing in the 5' untranslated region but coding for the same protein, however, only one transcript has the binding site for AP-1, Qiao, and col. Give evidence that supports chromatin looping between the 1a and 1b promoters of the *ZEB2* gene as a transcriptional regulatory mechanism controlled by AP-1 signaling (Qiao et al., 2015).

## 2.4 Biological implications of alternative promoters

Many studies focusing on single genes have shown that the choice of a specific TSS has critical roles during development, and aberrations in AP and TSS use lead to various diseases including cancer, neuropsychiatric disorders, and developmental disorders. Whereas some disorders are caused by epigenetic changes or genetic aberrations in the promoter region, others are caused by genetic changes in distal elements affecting long-range transcriptional regulation (de Klerk and Hoen 2015).

### 2.4.1 Alternative promoters in development and disease

Transcriptional regulation plays a central role in producing different cell types from the same genomic content. Throughout embryonic development, cells make and respond to cell fate decisions by turning on new transcription programs required to generate progressively more specialized cell types. Similar events drive the differentiation of specialized cells from proliferating precursors in the adult stem cell lineages that maintain and repair many tissues throughout the life span. Understanding how cell-type-specific transcription is achieved forms the very basis of understanding differentiation and development in multicellular organisms.

Tissue and stage-specific transcription programs are established by the intricate interplay among promoter-proximal and distal DNA elements, and protein complexes that interact with them. Much recent work has focused on the role of the stage or tissue-specific transcriptional activators and repressors acting upon distal enhancer elements to control the time and place of expression of developmental genes (de Klerk and Hoen 2015).

However, evidence has emerged that alternative core promoter motifs and their recognition factors can play roles in cell-type-specific transcription programs in certain tissues, as illustrated by the tissue-specific pattern of transcripts derived from APs of several genes, a compendium of genes harboring APs and the respective tissue-specific expression patterns can be found in **ANNEX 2**.

As mentioned before the use of APs plays a fundamental role in tissue-specific and differential regulation of a gene in various human tissues. To determine the cellular fate, genes require guidance cues that enable them to express precise isoforms in the right cell types at appropriate times. Such cues are partly provided by the use of APs and chromatin state of the corresponding genomic regions, which are altered in disease settings (Davuluri et al., 2008).

The availability of mammalian genomes and the advent of high-throughput molecular technologies have improved our ability to further investigate and increase our understanding of normal molecular processes that are disrupted in a disease state. Genetic aberrations such as deletions, insertions, and polymorphisms in APs that are associated with the disease have been reported for many genes (Davuluri et al., 2008) (**ANNEX 2**).

Developmentally regulated genes in the human brain are spatially and temporally regulated, in some cases, through differential transcription from APs, causing different isoforms expression in the human brain. Disruption of these genes is implicated in several neuropsychiatric disorders, including Parkinson's disease, schizophrenia, bipolar disorder, and autism, one specific example is the case of *NRG1* and Schizophrenia (Liu et al., 2005). Another example of deregulation of AP usage causing disease outside cancer and neuropsychiatric disorders is the case of *GNAS* and pseudohypoparathyroidism type IB (PHPIB) (Tan et al., 2007).

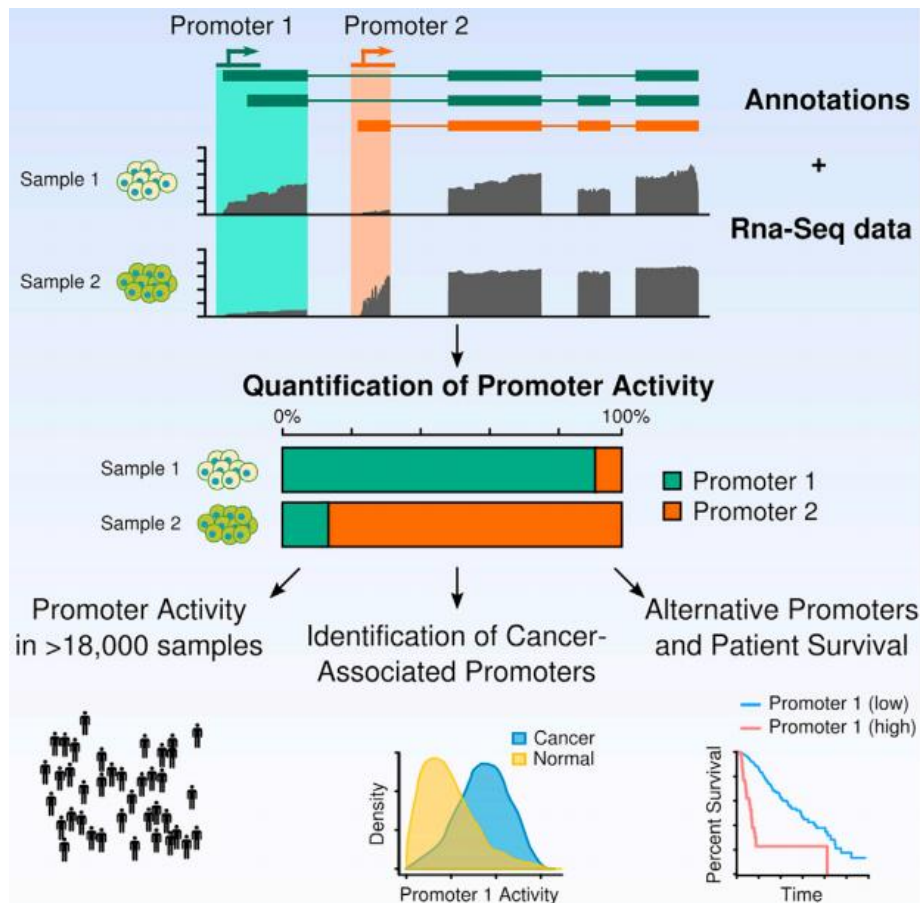
#### 2.4.2 Deregulation of alternative promoters in cancer

Expression of inappropriate alternative UTR might contribute to the development of diseases, a reduced expression of *BRCA1*, a tumor suppressor gene, is believed to be a hallmark in the development of some sporadic cases of breast and ovarian cancers. *BRCA1* has two separate promoters that generate different 5' UTRs. In some breast cancers, downregulation of *BRCA1* is achieved by shifting the expression from a shorter 5'UTR with efficient translation to a longer 5' UTR that contains secondary structures and uORFs that inhibit translation (Sobczak & Krzyzosiak, 2002).

Another example is the oncogene *Mdm2* overexpressed in some choriocarcinomas due to activation of an AP that allows the expression of a short 5'UTR without the inhibitory uORF embedded in the longer constitutively expressed 5'UTR (Brown et al., 1999).

Demircioglu and col. Analyzed 18,468 RNA-seq samples, providing the largest survey of active promoters in human tissues and cancers demonstrate that APs are frequently required for the regulation of tissue-specific and cancer-related isoforms. The scale of the data generated permitted for the first time analysis of patient-to-patient promoter usage and associate the difference in AP usage with patient survival (Figure 2.4). They conclude that for genes with independent regulated APs, the promoter activity is a better predictor of patient survival than gene expression, and possibly contributing to cell transformation in cancer. (Demircioglu et al., 2019).





**Figure 2.4.** Deregulation of APs is associated with patient survival. Taken from (Demircioğlu et al., 2019).

This study highlights the extensive role of APs in isoform expression throughout specific contexts and isoform diversity regulation and underscores how patient-to-patient variation in promoter usage is a link to pathological traits in cancer. They construct a comprehensive catalog of active promoters over 42 cancer types and tissues, that will be a helpful resource to the comprehension of gene regulation networks and non-coding mutations in cancer. Cancer-specific promoters could be relevant for the development of novel diagnostic approaches, drugs, and therapies, and they will allow the accurate design of genome-wide functional screens (Demircioğlu et al., 2019).

An interesting example of isoforms generated by the same gene but with opposite protein product function is the *BCL2L1* locus, in which the expression of the isoform BCL-xS promotes apoptosis while the expression of BCL-xL inhibits apoptosis and is important in the context of B-cell lymphoma (Warren et al., 2019).

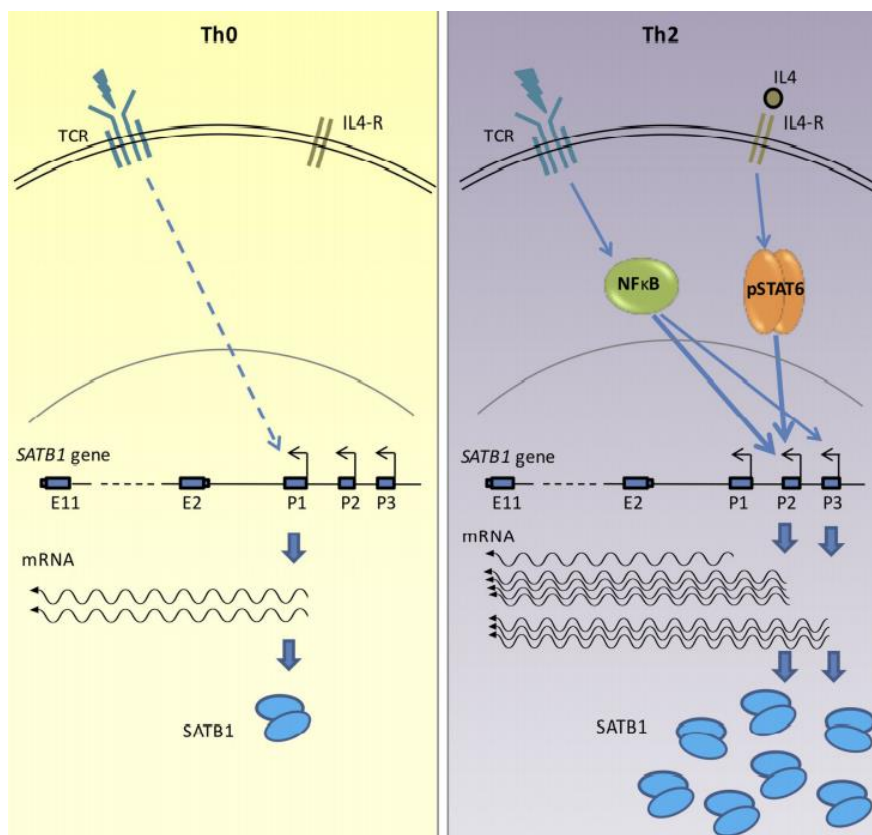
## 2.5 Examples of alternative promoter regulation in normal T-cell differentiation and leukemia

In the following section, some characteristic examples of AP regulation in normal T-cell differentiation and leukemia are described.

### 2.5.1 Regulation of SATB1 AP usage by different transcription factor response during Th2 differentiation

Khare SP and col. performed an analysis of publically available T cell transcriptome and identified a large regulatory region at the *SATB1* gene locus which encodes for multiple transcripts that differ in the TSS corresponding to APs and studied *SATB1* expression during T-helper cell differentiation (Khare et al., 2019).

*SATB1* stands for special AT-rich binding protein 1 and participates in the development of T cells in the thymus and the periphery. Khare SP and col. found that *SATB1* expression is regulated by three APs: P1-proximal, P2 middle, and P3 distal during peripheral differentiation of CD4+ T cells. The helper T cells use P2 and P3 while T-reg cells rely on the P1 promoter, highlighting the relevance of proinflammatory cytokines in promoter switching (Khare et al., 2019).



**Figure 2.5** *SATB1* expression regulations by APs responding to different cytokines stimuli. Taken from (Khare et al., 2019)

*SATB1* expression is orchestrated via an intricate regulatory network of NFkB signaling and cytokine signaling. The P1 promoter is preferentially expressed in the naïve CD4+ T cells and Th0 on the contrary P2 and P3 promoters are used by the Th2 cells. P2 promoter is positively regulated in Th2 by STAT6 TF which is downstream of cytokine signaling and binds to *SATB1* P2 promoter. P2 and P3 are also regulated by NFkB TF which is downstream of the TCR signaling pathway. P1 promoter is used by Treg cells and Th0 cells and poorly correlates with protein expression (Figure 2.5)

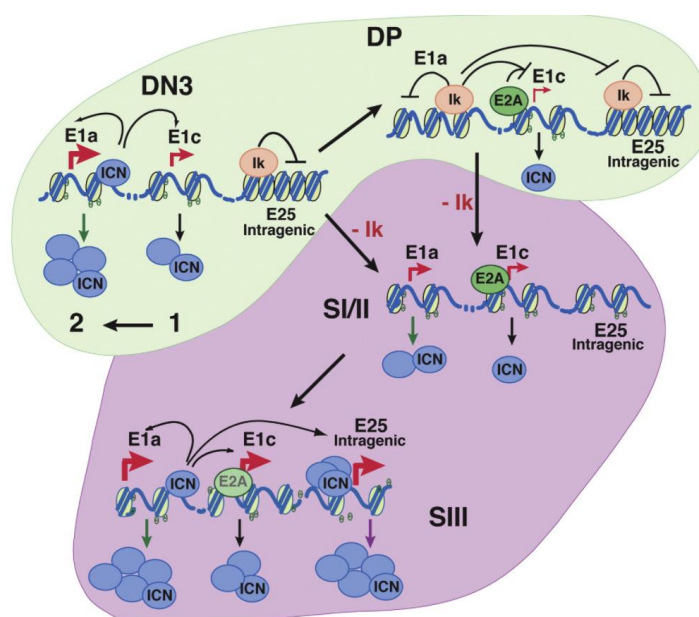
Some genes are regulated through AP usage during immune cell activation. Many of these genes originate transcripts without any change in the coding sequence of DNA, suggesting that the change in AP usage is driven under different TFs repertoire in different biological or physiological contexts (Khare et al., 2019).

## 2.5.2 Alternative Promoter usage of Notch1 in T-cell development and leukemogenesis

Another example of APs regulating a gene during T cell differentiation was established by Gomez del Arco and col. Whom demonstrated that the *Notch1* locus contains a combination of promoters and enhancers that interplay forming a feed-forward loop that regulates NOTCH signaling at specific stages of T cell development and leukemic transformation. Ikaros is a key regulator of this process acting at the epigenetic level to limit recruitment of TFs. The activity of alternative *Notch1* promoters was increased in Notch1 and Ikaros knock-out leukemic cells as well as in other leukemic models where *Notch1* promoter remained intact. Two alternative Notch1 promoter regions were mapped at the 5' and intragenic locations alongside the canonical promoter. Transcriptional analysis of this locus during T cell development revealed differentially AP usage throughout the process (Gómez-del Arco et al., 2010).

The three sets of *Notch1* promoters acquired more permissive chromatin before NOTCH activation during leukemogenesis mediated by loss of Ikaros. The permissive chromatin acquisition was far more evident in the intragenic promoters in pre-leukemic thymocytes. Both APs produce proteins with ligand-independent activation of NOTCH signaling properties, the upstream APs by altering the receptor processing during the secretion process and the intragenic by the lack of the ligand-binding and cleavage domains.

Gomez del Arco and col. proposed the presence of a feedback loop in Notch signaling controlled by a network of epigenetic and transcriptional regulators and Notch receptors with a different activity based on ligand dependency; local chromatin surrounding the Notch1 locus allows access to the transcription machinery and enhancer proteins. The presence of Ikaros at the proximity of all three sets of Notch1 APs prevents the access of transcription machinery by the generation of restrictive chromatin. Ikaros might prevent the activity of positive chromatin regulators such as MLL and HATs by association with negative chromatin remodeling factors, for example, Mi-2b and HDACs. Increases access of transcription machinery to both ligand-dependent and independent Notch1 promoters is generated by the loss of Ikaros. This promotes an increase in ICN that works as a potent transcriptional enhancer from at least two regions of the Notch1 locus (Figure 2.6) (Gomez del Arco et al., 2010).

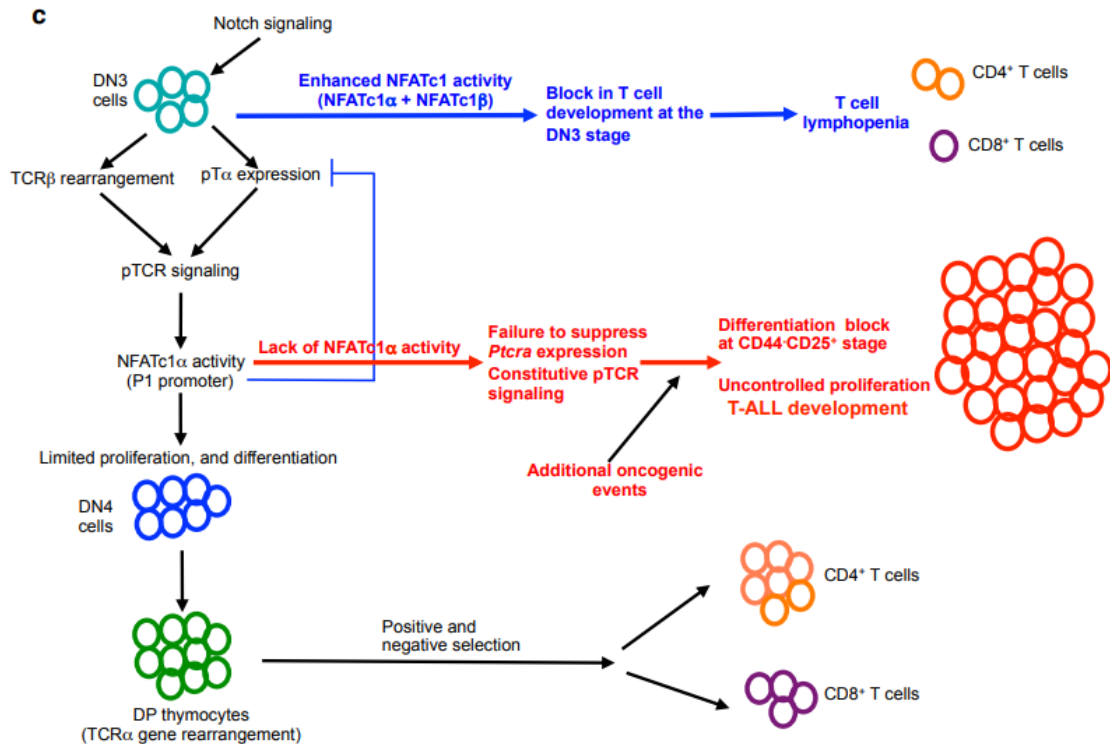


**Figure 2.6** Regulation of Notch1 signaling during T cell development and leukemogenesis. At the DN3 stage, ligand-dependent signaling supported by the canonical E1c promoter (1) activates the alternative E1a promoter to produce more ICN in the feed-forward loop (2). Both E1a and E1c promoters are in a permissive chromatin environment (broadly spaced nucleosomes). The intragenic promoter E25 is not transcriptionally active or in poised chromatin due to the action of negative regulatory factors such as Ikaros and possibly E2A (tightly spaced nucleosomes). During the transition from DN to DP, increased Ikaros binding in the vicinity of all three Notch1 promoters restricts chromatin and attenuates transcription. Only transcription from the canonical promoter (E1c) remains active albeit at a basal level. Following the loss of Ikaros at stages I and II (SI/II), the chromatin of all three Notch1 promoters becomes more accessible, and basal transcription increases at both the E1a and E1c promoters. This triggers both feed-forward and feed-back loops that further increase Notch1 signaling and Notch1 transcription respectively. Importantly, the increase in chromatin accessibility at all of the Notch1 promoters facilitates recruitment of ICN binding to its target sites, corroborating with a progressive ICN accumulation to seal transition to an aggressive leukemic state. This leukemic state is demarcated by a Notch1 locus with three fully active transcriptional promoters supporting both ligand-dependent and ligand-independent phases in Notch1 signaling. Taken from (Gomez del Arco et al., 2010).

### 2.5.3 *Nfatc1* in thymocyte differentiation and leukemia development

Klein-Hessling and col. Characterized a differential expression pattern of *Nfatc1* in a specific stage manner during T cell development by activating distal P1 or proximal P2 promoters, pre-TCR-negative thymocytes express only *Nfatcb* isoform derived from activation of P2 and pre-TCR positive thymocytes express both P1 (*Nfatc1a*) and P2 (*Nfatcb*) derived isoforms, they showed that induction of the P1 promoter in pre-TCR negative thymocytes besides the already activated P2 promoter blocks thymocyte development causing severe T-cell lymphopenia, they also demonstrate that the activity of *Nfatc1* suppresses the B lineage fate of immature thymocytes and assures their differentiation into T-cells. A threshold level of *Nfatc1* activity is needed in TCR-positive thymocytes to prevent Notch3-induced T-ALL. *Nfatc1* is crucial for the T-cell fate of thymocytes (Klein-Hessling et al., 2016).

The switch from *Nfatc1b* only to both *Nfatc1a* and *b* at the pTCR-positive DN3 stage is necessary for T-lineage commitment in the Thymus. They characterize a new enhancer element E2 that regulates the P1 isoform. They observed that E2 has an open chromatin conformation at the DN3 stage but is the TF occupancy that induces P1 activity in pTCR-positive cells. They showed that levels of *Nfatc1* increase systematically from DN1 to DN3 stage and proposed a model where binding of *Nfatc1* to the E2 element auto-regulates its expression by activating P1 transcript expression, increasing, even more, the *Nfatc1* levels to a threshold necessary for T-cell commitment. Any change that alters the threshold of *Nfatc1* activity, either reducing or increasing the activity will disturb the delicate balance towards unwanted phenotypes either T-cell lymphopenia or cooperating with other oncogenic events towards the development of leukemia (Figure 2.7) (Klein-Hessling et al., 2016).



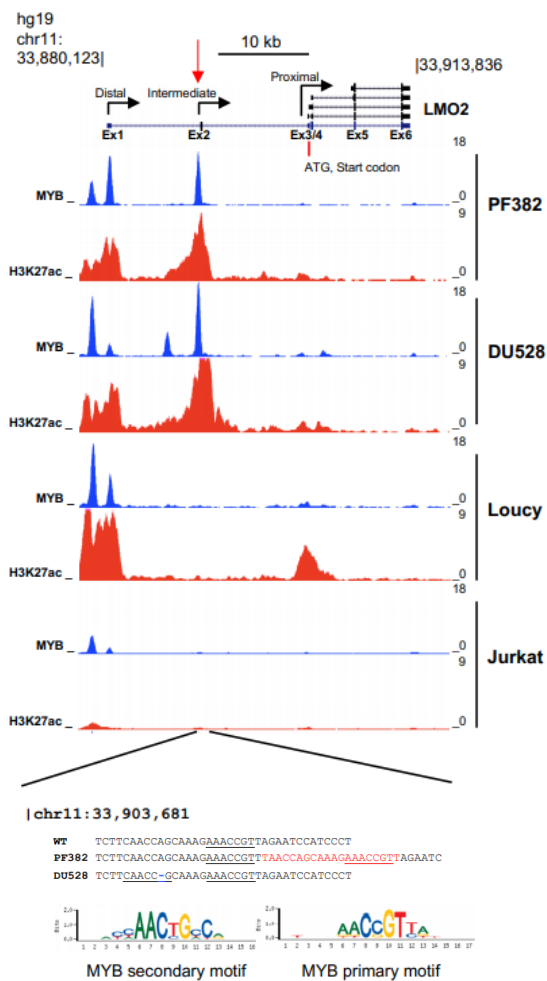
**Figure 2.7** Unbalanced *Nfatc1* activity mediated by an AP during T-cell differentiation can lead to lymphopenia or T-ALL. Taken from (Klein-Hessling et al., 2016).

### 2.5.4 Regulation of alternative promoter usage by the acquired neomorphic promoter, the case of LMO2

LIM-domain-only-protein-2 (LMO2) is important for the formation of multimeric transcriptional complexes that include TAL1, LDB1, GATA, RUNX, ETS1, and MYB, is crucial in T-cells development, it has been found that in mice *Lmo2* is silenced after the early T cell progenitor (ETP) stage, and when overexpressed leads to T-ALL in transgenic models. In human thymi, LMO2 is also downregulated after T lineage commitment and LMO2 overexpression in humans is associated with leukemic alterations in T-cells but not in other lineages (Rahman et al., 2017).

Reported mechanisms for dysregulation of LMO2 expression include chromosomal translocations such as t(11;14)(p13;q11) and t(7;11)(q35;p13), deletions of an upstream negative regulatory region, and retroviral insertional mutagenesis in LMO2 locus as a result of gene therapy furthermore 50% of T-ALL cases overexpresses LMO2, only 10% of the cases present a cytogenetic lesion (Rahman et al., 2017).

Rahman and col. described a mechanism where somatic mutation of a non-coding region (intron) of LMO2 activates the oncogene expression in T-ALL. They analyzed cell lines and T-ALL patients harboring heterozygous mutations where the majority presented putative de novo MYB, ETS1, and RUNX1 consensus binding sites. CAGE analysis in mutant cell lines identified the usage of an intermediate promoter site (Figure 2.8) with monoallelic overexpression of LMO2. The dysregulation of LMO2 expression was supported by CRISPR-Cas9 experiments, LMO2 down-regulation was observed when disruption of the mutant allele in PF-382 cell was performed (Rahman et al., 2017).



**Figure 2.8** ChIP-Seq tracks at the LMO2 locus for MYB and H3K27ac in PF-382, DU.528, and Loucy cell lines, somatic mutation creates a *de novo* MYB binding site. Modified from (Rahman et al., 2017).

## 2.6 Genome-wide approaches for the study of alternative promoters

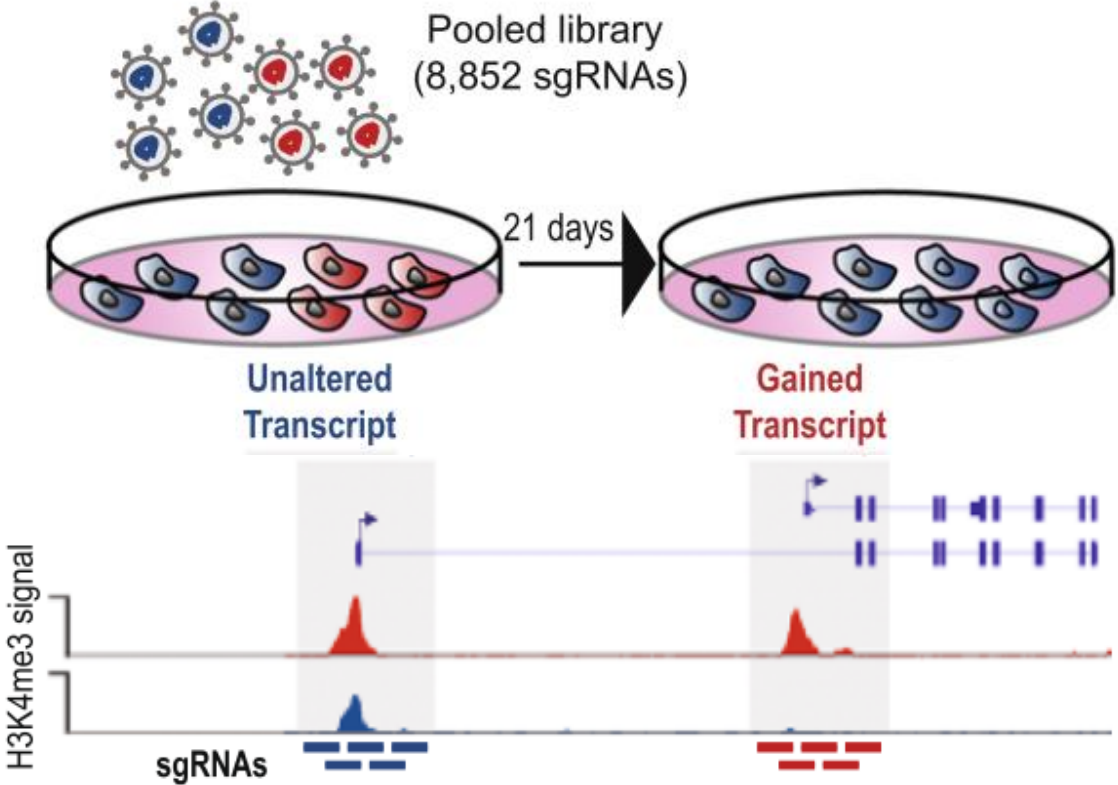
The use of APs and transcription start sites (TSSs) in protein-coding transcripts was established before the development of transcriptome-wide approaches, through studies based on a method called cap analysis of gene expression (CAGE). CAGE still represents the basic technology for the detection of TSSs. Several high throughput CAGE methods such as DeepCAGE have been developed. These transcriptome-wide studies suggest that TSS is highly tissue specific and that the number of alternative TSSs differs by tissue type (de Hoon and Hayashizaki 2008).

Genome-wide studies of promoters using the H3K4me3 histone modification, an epigenetic mark at active promoters, or CAGE (cap analysis of gene expression) tag sequencing of the 5' end of transcripts have found that TSSs are frequently differentially used in cancer (Kaczkowski et al., 2016; Muratani et al., 2014; Takahashi et al., 2012).

Because any change in a cell's identity and function will be reflected in a change in gene expression, transcriptome profiling by RNA sequencing (RNA-seq) is one of the most widely studied large-scale molecular phenotypes in cancer, as the study performed by Demircioglu and col. 2019 and described in the subsection AP deregulation in cancer. Analysis of gene

expression in cancer has uncovered fundamental insights into tumor biology (Hoadley et al., 2018), enabled stratification of cancer types (Cancer Genome Atlas Research Network, 2012), predicted clinical outcome (Gerstung et al., 2015), and guided treatment decisions (Cancer Genome Atlas Research Network, 2011), forming a cornerstone of data-driven precision oncology.

Davies R and col. Adapted the CRISPR interference (CRISPRi) technology to target specific promoters inside a gene, generating specific transcripts loss-of-function genetic screen (Figure 2.9), and use it to test 820 isoforms that are gained in gastric cancer (GC), they identified a subset of GC associated isoforms and found that some isoforms from the same gene have opposite functions, specifically *ZFH3* tumor suppressor gene, expresses an isoform with a contradictory oncogenic role that is associated with poor patient outcome(Davies et al., 2021).



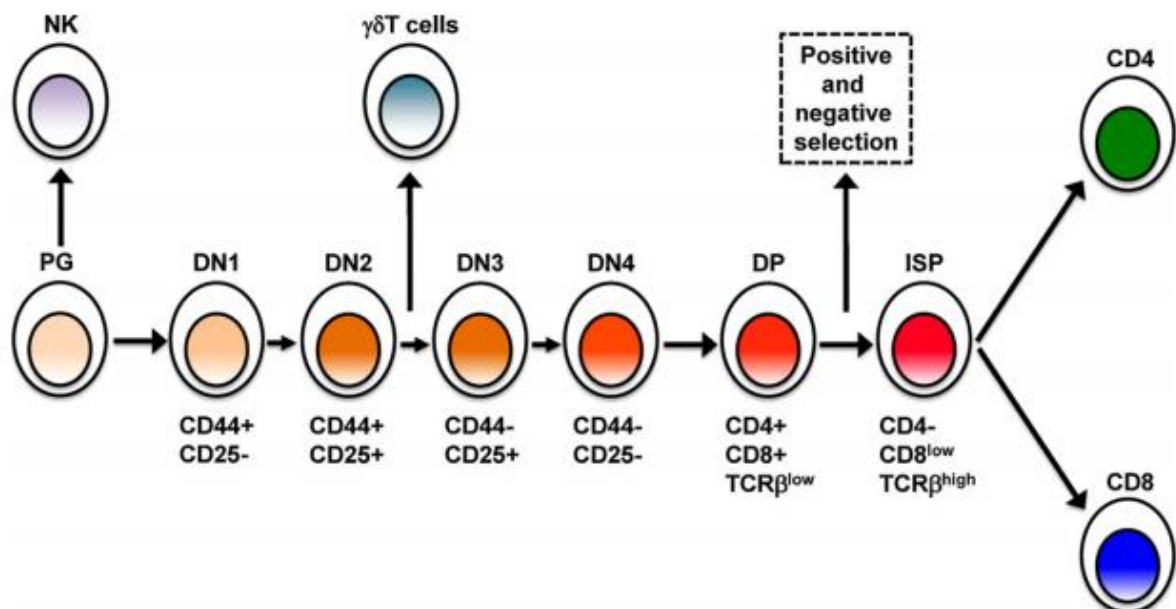
**Figure 2.9.** Scheme describing an isoform-specific CRISPRi screening. Modified from (Davies et al., 2021)

# Chapter 3 T-cell acute Lymphoblastic Leukemia (T-ALL)

To understand the molecular mechanisms disturbed in T-ALL, which is the main model of my thesis work, first I present a brief description of normal T-cell differentiation, and then we focused on the mechanisms originating T-ALL.

## 3.1 T-cell differentiation

T-cells develop in the thymus, where maturation of lymphocyte precursors into functional T-cells occurred. The differentiation of progenitor cells depends on bidirectional signals between the developing thymocytes and the thymic epithelial cells, thymic CD4<sup>-</sup>CD8<sup>-</sup> DN cells give rise to the CD4<sup>+</sup> CD8<sup>+</sup> double-positive (DP) population, which differentiate to mature CD8<sup>+</sup> CD4<sup>-</sup> or CD8<sup>-</sup>CD4<sup>+</sup> single-positive (SP) cells. The DN population can be subdivided by cell surface expression of CD25 and CD44. CD44<sup>+</sup> CD25<sup>-</sup> (DN1) cells differentiate to become CD44<sup>+</sup> CD25<sup>+</sup> (DN2) cells, which then differentiate to become CD44<sup>-</sup> CD25<sup>+</sup> (DN3). The DN3 population gives rise to the CD44<sup>-</sup>CD25<sup>-</sup> (DN4) subset, which undergoes a phase of rapid proliferation before differentiation into the DP population, in general via a cycling immature CD8<sup>+</sup> intermediate single positive (ISP) cell (D'Acquisto & Crompton, 2011) (Figure 3.1).



**Figure 3.1** Schematic representation of T-cell differentiation in the thymus. Taken from (D'Acquisto and Crompton, 2011)

Maturation from the DP population to the mature SP T-cell populations involves the positive selection of the  $\alpha\beta$ TCR repertoire to ensure appropriate MHC restriction and negative selection of potentially self-reactive clones. TCR repertoire selection is dependent on



interactions between the TCR on the developing thymocyte and its MHC/peptide ligand on the thymic epithelial cells. Strength and duration of TCR signal are thought to broadly determine the DP cell's fate with the strongest signals leading to negative selection and apoptosis, (in the case of TCR recognizing self-antigens), intermediate signals leading to positive selection, and, weaker signals or lack of TCR signaling leading to DP cell death by neglect. Transcription factors secreted from the thymic epithelial cells play a critical role in TCR repertoire selection and differentiation from DP to SP cell (D'Acquisto and Crompton, 2011).

### **3.2 T-acute lymphoblastic leukemia**

T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive type of blood cancer that accounts for about 15% of pediatric and 25% of adult ALL cases and is considered a paradigm for the multistep nature of cancer initiation and progression (Kimura and Mulligan 2020). Genetic and epigenetic reprogramming events, which transform T-cell precursors into malignant T-ALL lymphoblasts, have been extensively characterized over the past decade, for the identification of the different types of T-ALL.

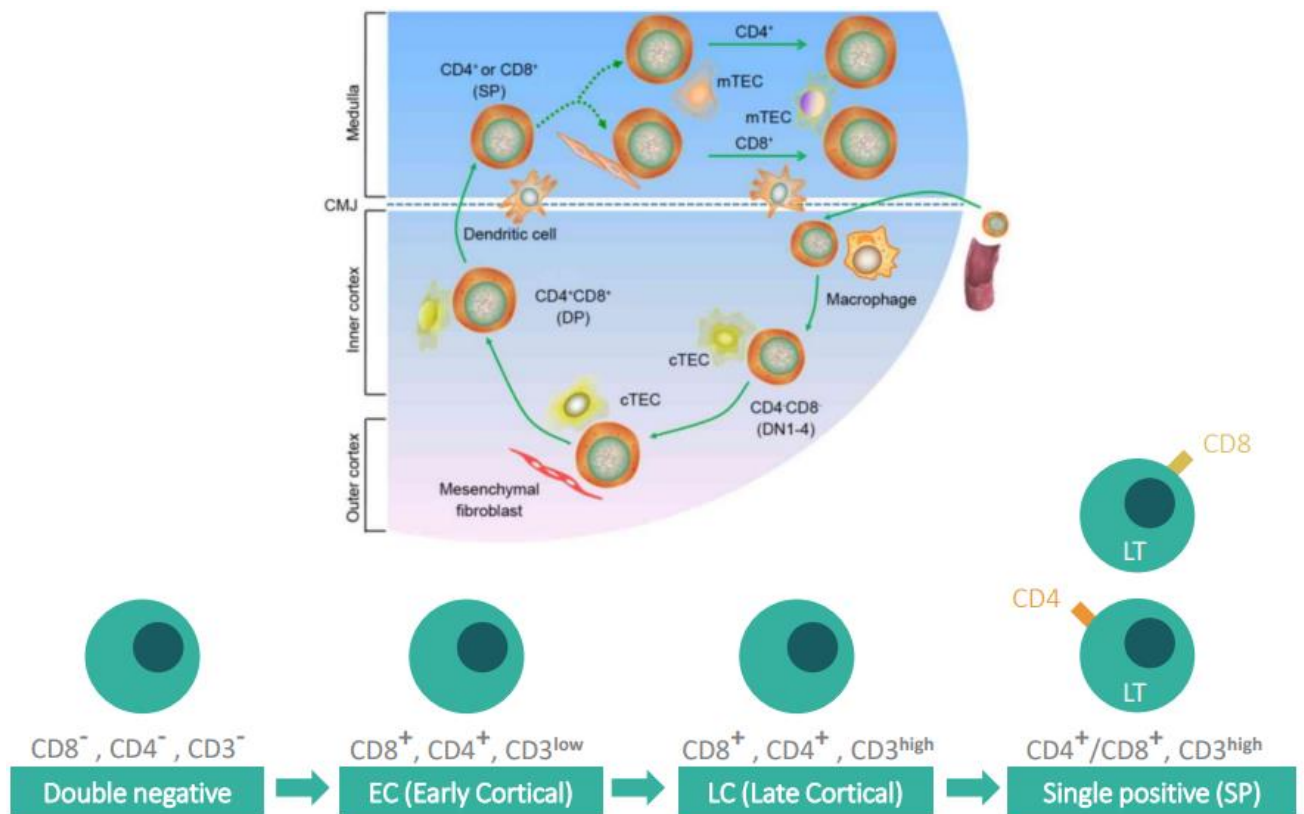
Clinically, patients with T-ALL typically present with elevated white cell counts in their blood and hematopoietic failure, with neutropenia, anemia, and thrombocytopenia. In addition, they frequently show mediastinal thymic masses and meningeal infiltration of the central nervous system at diagnosis.

T-ALL is biologically and genetically heterogeneous with gene expression signatures that identify different clinical-biological groups associated with T cell arrest at different stages of thymocyte development (Belver and Ferrando 2016). Oncogenic Notch signaling resulting from activating mutations in NOTCH1 is a major driver of T-ALL transformation. Aberrant expression of transcription factor oncogenes as a result of chromosomal translocations and other chromosomal rearrangements is common in T-ALL (Belver and Ferrando 2016).

Recurrent mutations and deletions in T-ALL frequently involve cell cycle regulators, but also transcription factors, tumor suppressors, epigenetic factors, and negative regulators of NOTCH1, Janus kinase (JAK)–signal transducer and activator of transcription (STAT), PI3K, and MAPK signaling (Belver & Ferrando, 2016).

#### **3.2.1 Genomic and transcriptomic classification of T-ALL**

Currently, different clinically relevant biological groups of T-ALL are recognized, and these are associated with unique gene expression signatures and with immunophenotypes that reflect thymocyte developmental arrest at different stages of development (Figure 3.2).



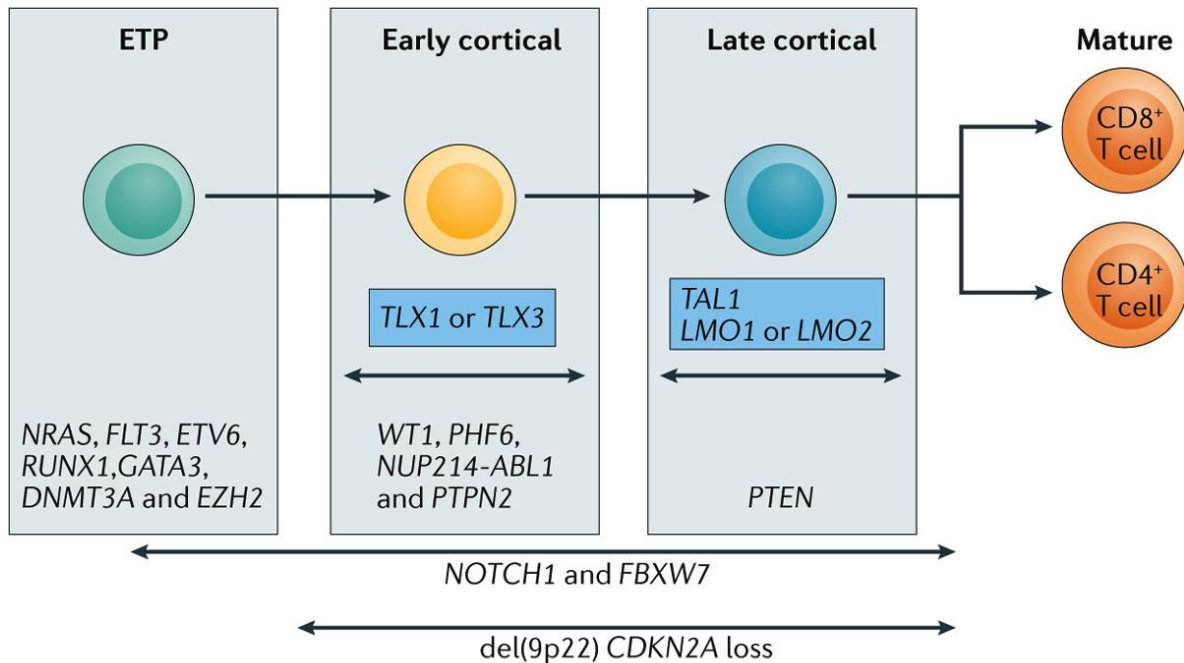
**Figure 3.2** Immunophenotype of human T-cell development. Adapted from (F. Yan et al., 2017).

Early T-lineage progenitor (ETP) leukemias show a block at the earliest stages of T cell differentiation ( $CD4^-CD8^-$  cells) and a transcriptional program related to that of early T-lineage progenitor cells but are also very closely associated with the hematopoietic stem cells and myeloid progenitors (Coustan-Smith et al., 2009). Genetically, early immature ETP T-ALL has a lower prevalence of NOTCH1 mutations, rarely has CDKN2A deletions, and is characteristically associated with mutations in genes encoding signaling factors (for example, NRAS and fms related tyrosine kinase 3 (FLT3) or epigenetic regulators (for example, isocitrate dehydrogenase 1 (IDH1), IDH2 and DNA methyltransferase 3A (DNMT3A) that are frequently mutated in myeloid leukemias (Zhang et al., 2012).

ETP T-ALL is also associated with mutations that disrupt the activity of important transcription factors governing hematopoietic and T cell development, for example, runt-related transcription factor 1 (RUNX1), GATA binding protein 3 (GATA3), and ETS variant 6 (ETV6) (Van Vlierberghe, P. et al. 2013). Although ETP T-ALL accounts for approximately 10% of pediatric T-ALL cases, its incidence increases with age, accounting for approximately 40–50% of adult T-ALLs (Van Vlierberghe et al., 2011).

T-ALLs with a  $CD1a^+$ ,  $CD4^+$ , and  $CD8^+$  immunophenotype, which corresponds to the early stages of cortical thymocyte maturation, show a particularly favorable prognosis. These leukemias are typically associated with activation of the TLX1, TLX3, NKX2-1, and NKX2-2 homeobox genes; they have the highest prevalence of NOTCH1 mutations and almost universally harbor deletions of the CDKN2A locus (Belver and Ferrando 2016) (Figure 3.2 and 3.3).

The third group of T-ALLs with a more mature late cortical thymocyte immunophenotype ( $CD4^+$ ,  $CD8^+$ , and  $CD3^+$ ) typically show activation of the TAL1 oncogene (Belver and Ferrando) (Figure 3.2) (Figure 3.3).



**Figure 3.3** Molecular subgroups of T-ALL associated with oncogenic programs and gene expression signatures. Taken from (Belver and Ferrando 2016).

### 3.2.3 Oncogenic NOTCH1 and T-ALL

NOTCH1 is a class I transmembrane glycoprotein that functions as a ligand-activated transcription factor, enabling the direct transduction of extracellular signals at the cell membrane into transcriptional changes in specific target genes in the nucleus (Dumortier A et al 2005). In the hematopoietic system, activation of the NOTCH1 receptor in the thymus is crucial for early T cell fate specification and thymocyte development (Radtke F et al.1999). Aberrant constitutively active NOTCH1 signaling was first identified in rare T-ALLs harboring the t(7;9)(q34;q34.3) chromosomal translocation, which leads to the expression of a truncated and constitutively active form of NOTCH (Ellisen L W et al 1991).

However, in most T-ALLs NOTCH1 is not activated by chromosomal translocations, but occurs as a result of activating mutations that disrupt specific domains responsible for controlling the initiation and termination of NOTCH1 signaling (Weng, A P et al 2004). In addition, 8–30% of T-ALLs harbor mutations in the F-box and WD repeat domain containing 7 (FBXW7), which results in the impaired degradation of activated NOTCH1 (O’Neil J et al 2007).

### 3.2.4 T-ALL and cell-cycle deregulation

The loss of cell cycle control is a hallmark of cancer and has a prominent role in the pathogenesis of T-ALL. As mentioned above, the tumor suppressors p16<sup>INK4A</sup> and p14<sup>ARF</sup> encoded by the CDKN2A locus in the short arm of chromosome 9 are lost via chromosomal deletions in most (>70%) T-ALLs (Ferrando A A et al 2002).

The p16<sup>INK4A</sup> tumor suppressor inhibits G1 to S phase cell cycle progression by directly binding to and inactivating cyclin D–CDK4 and cyclin D–CDK6 complexes, whereas p14<sup>ARF</sup> facilitates cell cycle arrest and apoptosis in response to cellular stress *via* inhibition of MDM2, a negative regulator of p53 (Kamijo, T. et al.1998)

Aside from the loss of CDKN2A, 15% of T-ALLs show chromosomal deletions in 13q14.2, encompassing the retinoblastoma 1 (RB1) locus, which encodes a master regulator of cell cycle progression (Mullighan, C G et al 2007). Moreover, approximately 12% of cases harbor 12p13.2 deletions involving the CDKN1B gene, which encodes p27<sup>KIP1</sup>, an inhibitor of cyclin E–CDK2 and cyclin D–CDK4 complexes (Remke M et al 2009). The t(12;14)(p13;q11) and t(7;12)(q34;p13) translocations, which are present in approximately 3% of T-ALLs, result in aberrant cell cycle progression by driving aberrantly high levels of cyclin D2 (CCND2) expression (Clappier E et al 2006), further highlighting the prominent role of cell cycle deregulation in T cell transformation.

### 3.2.5 Transcription factors as oncogenes in T-ALL

Approximately 60% of T-ALLs are characterized by the aberrant expression of a class II bHLH transcription factor (TAL1, TAL2, LYL1, or BHLHB1) and an LMO protein (LMO1 or LMO2). TAL1 aberrant expression can be driven by interchromosomal or intrachromosomal rearrangements that place *TAL1* under the regulatory sequences controlling the expression of the T cell receptor- $\alpha$  (TCRA) and TCRD genes or the SCL/TAL1 interrupting locus (STIL), a gene that is a neighbor of TAL1.

In addition, heterozygous somatic mutations in a precise site upstream of the TAL1 locus have been found in a fraction of T-ALL tumors. These intergenic mutations create a 5' enhancer containing new MYB transcription factor binding sites, thus driving monoallelic aberrant expression of TAL1 (Mansour M R et al 2014; Navarro et al., 2015). Several genes activated by TAL1 have been highlighted as potential mediators of its oncogenic activity, including NKX3-1 (Kusy S et al 2010)

Globally, TAL1 forms autoregulatory loops with GATA3 and RUNX1 and directly activates MYB in a positive feed-forward regulatory loop that strengthens and stabilizes this transcriptional network (Sanda, T et al 2012). Genes encoding TAL1-related to class II bHLH factors such as LYL1, TAL2, and BHLHB1 are also aberrantly expressed in rare cases of T-ALL that harbor chromosomal translocations that place them near enhancers in the loci encoding the T cell receptor (TCR) (Belver L and Ferrando A 2016).

LMO1 and LMO2 are aberrantly expressed at high levels in approximately 10% of T-ALLs as a result of the t(11;14)(p15;q11) and t(11;14)(p13;q11) chromosomal translocations, respectively (McGuire E A et al 1989). In addition, aberrant LMO2 expression can be found in up to 5% of T-ALLs driven by small chromosomal deletions in the vicinity of the LMO2 locus (Van Vlierberghe, P. et al. 2006). LMO proteins do not interact directly with DNA, but instead, form transcriptional complexes with TAL1 and other bHLH factors.

This finding, together with the frequent coexpression of LMO1 and LMO2 with TAL1 and/or LYL1 in T-ALL (Ferrando, A. A. et al. 2002) supports the hypothesis that bHLH and LMO oncogenes have a common and cooperative role in T-ALL.

The oncogenic activity of LMO2 may underlie the development of T-ALL that has been observed in patients with X-linked severe combined immunodeficiency syndrome undergoing retrovirus-based gene therapy. In these trials, re-expression of interleukin-2 receptor- $\gamma$  (IL2RG) chain in hematopoietic progenitors using  $\gamma$ -retroviral vectors rescued immunodeficiency but was followed by the development of T-ALL with high levels of LMO2 expression and retroviral integrations upstream of the LMO2 gene (Hacein-Bey-Abina, S. et al. 2008).

HOX factors encoded by homeotic genes are developmentally important transcriptional regulators responsible for the control of embryonic body patterning along the anterior-posterior (head-tail) axis (Garcia-Fernandez 2005). Deregulated expression of HOXA9 and HOXA10 is involved in the pathogenesis of approximately 3% of T-ALLs, which harbor chromosomal translocations and inversions that move the cluster of HOXA paralogues to the vicinity of the TCRB and TCRG loci (Speleman, F. et al. 2005). In addition, the expression of HOXA genes is a characteristic of early immature ETP T-ALLs and T-ALLs that harbor translocations and chromosomal rearrangements driving the expression of fusion oncogenes such as KMT2A–MLLT1 (also known as MLL–ENL) (Rubnitz, J. E. et al. 1996) suggesting that HOXA dysregulation is generally pathogenic in T-ALL.

However, the most characteristic T-ALL HOX oncogenic factors are the TLX genes. TLX1, the founding member of this family (which also includes TLX2 and TLX3), was first identified in the translocation breakpoint of leukemias harboring the t(10;14)(q24;q11) rearrangement. This chromosomal translocation, present in approximately 5–10% of pediatric T-ALLs and 30% of adult T-ALLs, places TLX1 under the control of strong enhancers in the TCRA and TCRD loci, thus inducing high levels of TLX1 expression in T-ALL lymphoblasts (Hatano, M. et al. 1991).

The direct pathogenic role of TLX1 in T cell transformation has been firmly established in transgenic mice, in which forced expression of TLX1 induced the development of clonal T cell leukemias that had a transcriptional program similar to that of TLX1+ human tumors (De Keersmaecker, K. et al. 2010).

TLX3 is also involved in the pathogenesis of T-ALL, in this case as a result of the t(5;14)(q35;q32) translocation present in 20–25% of pediatric T-ALLs and 5% of adult T-ALLs. Unlike most oncogenic translocations involving T-ALL transcription factor oncogenes, this rearrangement does not involve the TCR loci but instead places the TLX3 oncogene under the control of strong T cell regulatory elements in the BCL11B locus (Bernard et al., 2001).

Notably, tumors with high expression of TLX1 or TLX3 show a marked overlap in their transcriptional signatures, and analyses of the direct targets of these transcription factors suggest that they directly regulate a broadly overlapping set of genes. This finding is further supported by a set of mutations common to leukemias expressing TLX1 and TLX3, but which are rarely or less frequently present in other T-ALLs, including loss-of-function mutations in the BCL11B, protein tyrosine phosphatase non-receptor type 2 (PTPN2), Wilms tumor 1 (WT1) and PHD finger protein 6 (PHF6) tumor suppressors and the NUP214–ABL1 fusion oncogene (Graux, C. et al. 2004).

NKX2-1 and NKX2-2 are two highly related HOX genes and can be targeted by translocations to the TCR loci and chromosomal inversions, mechanisms that result in their aberrant expression in approximately 5% of pediatric T-ALLs. NKX2-1- and NKX2-2-rearranged leukemias share a gene expression signature related to that of T-ALLs with high TLX1 expression and show a similar arrest at the early cortical stage of thymocyte development (Homminga et al., 2011).

Cieslak and col. Performed found that HOXA5-9 transcription factors repress the T-cell receptor enhancer ( $E\alpha$ ) at the early stages of T-cell differentiation. Decommission is required for TCRA locus activation and enforced  $\alpha\beta$  T-lineage differentiation. The HOXA mediated repression of  $E\alpha$  is paralleled by the ectopic expression of homeodomain-related oncogenes in T-ALL (Cieslak et al., 2020).

MYC is a transcription factor and master regulator of cell growth and proliferation that is broadly involved in the pathogenesis of human cancer (Stine, Z. E. Et al 2015). In early T cell

development, MYC plays an important part in the control of cell growth downstream of NOTCH1 and pre-TCR signaling (Dose, M. et al 2006).

Although MYC overexpression is directly linked with T cell transformation in only approximately 1% of T-ALLs through its translocation to the TCRA and TCRD loci, MYC has been broadly highlighted as an important mediator of NOTCH1 induced transformation, and the identification of a long-range distal NOTCH1 controlled MYC enhancer (N-Me) has formally established a direct role for NOTCH1 in MYC expression (Herranz, D. et al. 2014).

Like NOTCH1, MYC is targeted for proteasomal degradation by FBXW7, and therefore, FBXW7 mutations found in T-ALL increase the levels of both NOTCH1 and MYC proteins. Mechanistically, MYC drives cell growth and proliferation in T-ALL, but it also has an important role as a driver of leukemia-initiating activity (King, B. et al. 2013).

MYB encodes an oncogenic leucine zipper transcription factor that is activated in rare cases of T-ALL harboring a t(6;7)(q23;q34) chromosomal translocation (Clappier, E. et al. 2007). Interestingly, patients with MYB-translocated T-ALL are frequently young children under 2 years of age and characteristically show a marked increase in the expression of proliferation and mitosis genes. Moreover, focal duplications of the MYB locus driving increased MYB expression are found in approximately 10% of T-ALLs in both children and adults (Lahortiga, I. et al. 2007).

### 3.2.6 Transcription factors as tumor suppressors in T-ALL

ETV6 is an ETS family transcriptional repressor strictly required for the development of hematopoietic stem cells. ETV6 mutations account for approximately 25% of cases of ETP T-ALL. They are characteristically amino-terminal or carboxy-terminal nonsense and frameshift mutations, resulting in truncated, dominant-negative forms of ETV6 that can abrogate the transrepressor activity of wild type ETV6 (Van Vlierberghe, P. et al 2011).

RUNX1 is a tumor suppressor and master regulator transcription factor with prominent roles in hematopoietic development. Somatic mutations in RUNX1 are found in approximately 5% of T-ALLs, typically in the immature ETP group (Zhang, J. et al. 2012). In addition, germline heterozygous mutations in RUNX1 are found in kindreds affected with platelet disorder, familial, with associated myeloid malignancy (FPDMM; OMIM ID 601399), a leukemia predisposition syndrome characterized by a moderate decrease in platelet numbers and an increased risk of acute myeloid leukemia (AML). Interestingly, FPDMM families also show an increased occurrence of T-ALL (Owen, C. J. et al. 2008).

GATA3, an important transcriptional regulator of T cell differentiation with a crucial role in the development of early T cell progenitors, is also recurrently mutated in immature ETP T-ALL. GATA3 mutations found in T-ALL are typically heterozygous point mutations in the zinc finger DNA-binding protein domain and recurrently involve R276, a residue needed for binding of GATA3 to DNA (Zahirieh, A. et al. 2005).

BCL11B is a Kruppel-like C2H2-type zinc finger transcription factor first identified as a tumor suppressor gene in mouse thymic lymphomas induced by ionizing radiation. Bcl11b inactivation in mouse T cell progenitors results in early arrest at the DN2–DN3 differentiation stage. In humans, BCL11B mutations are present in approximately 10% of T-ALLs (Gutierrez, A. et al. 2011). However, the specific genes and pathways controlled by BCL11B in T cell transformation remain to be elucidated

LEF1 is a member of the lymphoid enhancer factor/T cell factor (LEF/TCF) family of transcription factors, which are key mediators of WNT signaling. Monoallelic or biallelic deletions involving the LEF1 locus or mutations in LEF1 are present in approximately 10–15% of T-ALLs. Notably, T-ALLs with LEF1 inactivation show high levels of MYC expression and a characteristic differentiation arrest at the early cortical thymocyte stage of differentiation that resembles that of TLX1+ tumors (Gutierrez, A. et al. 2010).

Deletions and mutations in the WT1 tumor suppressor gene are present in approximately 10% of T-ALLs. WT1 mutations found in T-ALL are predominantly heterozygous frameshift mutations resulting in truncation of the C-terminal zinc finger domains of this transcription factor and are frequently associated with oncogenic expression of the TLX1, TLX3, and HOXA oncogenes. In the hematopoietic system, WT1 has been implicated in the maintenance of primitive stem cell quiescence and in promoting the differentiation of more mature progenitors (Ellisen, L. W. Et al 2001).

### 3.2.7 T-ALL and alterations in epigenetic regulators

PHF6 is a plant homeodomain (PHD) containing factor inactivated by mutations and deletions in approximately 16% of pediatric T-ALLs and 38% of adult T-ALLs, as well as in approximately 3% of AMLs. Germline PHF6 mutations are pathogenic in Börjeson–Forssman–Lehmann syndrome (BFLS; OMIM ID 301900), a rare X-linked disorder, and one case of T-ALL has been reported in a patient with BFLS (Chao, M. M. et al. 2010).

Interestingly, PHF6 is located on Xq26, and PHF6 mutations are almost exclusively found in male patients with T-ALL. The higher frequency of PHF6 mutations in males may result from an increased mutation rate in X chromosome genes in cells from males. PHF6 is primarily a nucleolar protein and has been implicated in ribosome biogenesis and splicing. However, a role for PHF6 in chromatin remodeling and transcription regulation is supported by its interaction with the NurD nucleosome repositioning and histone deacetylation complex and with multiple subunits of the PAF1 transcriptional elongation machinery (Todd and Picketts 2012).

Polycomb repressive complex 2 (PRC2) is a major epigenetic regulator involved in transcriptional repression through its writing of the histone H3 lysine 27 trimethyl (H3K27me3) epigenetic mark. In up to 25% of T-ALLs PRC2 function is recurrently disrupted by the loss-of-function mutations and deletions in the enhancer of zeste 2 (EZH2), embryonic ectoderm development (EED), and suppressor of zeste 12 (SUZ12) genes, which encode core components of this complex. Notably, activation of gene expression by NOTCH1 is coupled with the loss of H3K27me3 in NOTCH1 target genes, and PRC2 disrupting mutations found in T-ALL are frequently associated with activating mutations in NOTCH1. The model that emerges from these results indicates that loss of PRC2 activity may amplify the oncogenic effects of NOTCH1 mutations by priming NOTCH1 target genes for transcriptional activation (Ntziachristos, P. et al. 2012).

KDM6A (also known as UTX), a second H3K27me3 histone demethylase, is recurrently targeted by loss-of-function mutations in approximately 5–15% of T-ALLs and functions as a bona fide tumor suppressor. The contrasting roles of KDM6A (tumor suppressor) and KDM6B (required for NOTCH1-induced transformation) illustrate that epigenetic regulators with similar enzymatic activities can have opposing roles in the same disease. This contrast probably reflects the association of these factors with different transcriptional complexes (Ntziachristos, P. et al. 2014).

### 3.2.8 Deregulation of T-ALL related oncogenes by genomic insertion of enhancer sites

An insertion of 7.5 Kb upstream of the TAL1 TSS has been identified (Mansour et al., 2014; Navarro et al., 2015). The locus is enriched in H3K27ac and the insertion removal downregulates or abolishes the expression of TAL1, Mansour, and col. Demonstrate that the insertion often acts as an MYB site, and with less frequency as a TAL1, GATA3, or RUNX1 site (Mansour et al., 2014).

A model has been proposed where MYB binds to the insertion site and recruits other cofactors to up-regulate *TAL1* expression. The insertion deregulates the repressed locus by creating an MYB site that recruits MYB and major TF (Mansour et al., 2014).

Other cases of *de novo* enhancer sites deregulating gene expression have been described in the literature for the *LMO1* and *LMO2* loci (Rahman et al., 2017; Abraham et al., 2017; Li et al., 2017) (see Chapter II)

### 3.2.9 The role of non-coding RNAs in T-ALL

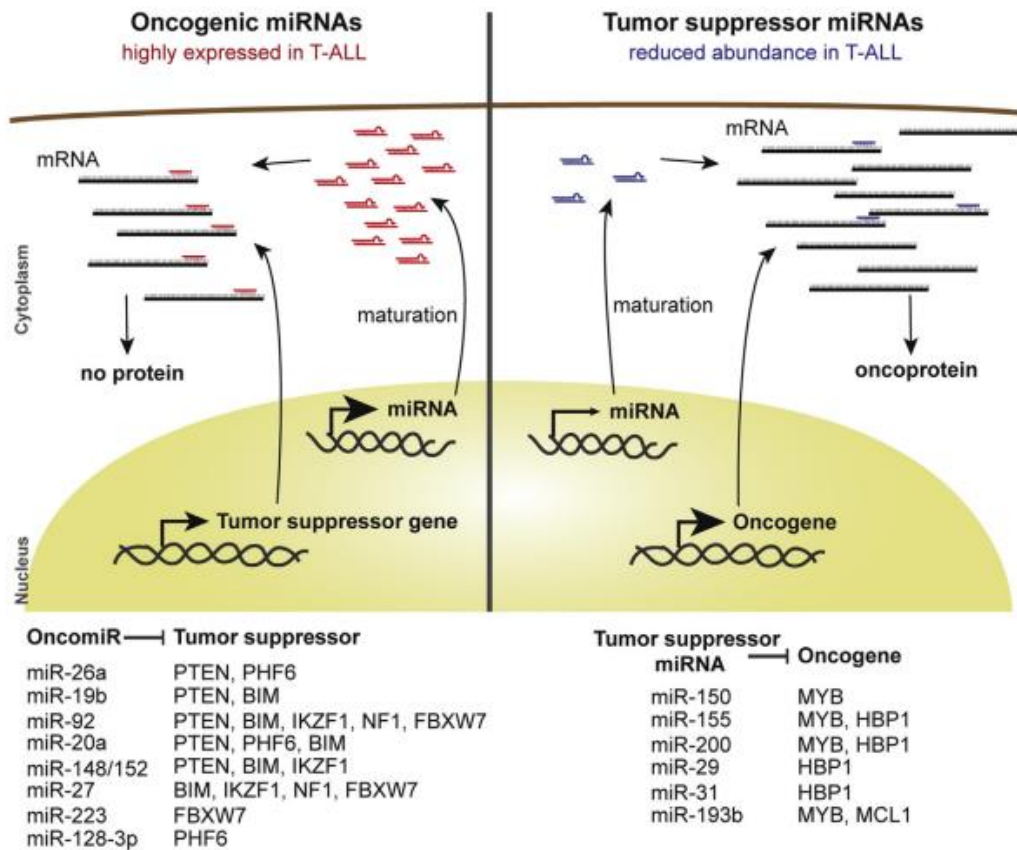
The protein-coding genes account for 2% of the entire genome, suggesting that the human transcriptome is predominantly composed of non-coding RNAs, diverse classes of RNAs have been identified in the genome such as microRNAs, long non-coding RNAs, and circular RNAs.

#### 3.2.9.1 Micro-RNAs and T-ALL

Micro-RNAs (miRNAs) are small non-coding RNAs, associated with normal development and cancer. Mavrakis et al identified in T-ALL a set of miRNAs (miR-19b, miR-20a, miR-26a, miR-92, and miR-223) that cooperate to suppress a network of tumor suppressor genes including PHF6, PTEN, BIM, and FBXW7, and they were able to accelerate leukemia onset in a NOTCH1 induced murine model of T-ALL, thus acting as oncomiRs (Figure 3.4) (Durinck K et al 2015).

Sanghvi and co-workers also identified a set of miRNAs that act as tumor suppressor miRNAs in T-ALL (miR-29, miR-31, miR-150, miR-155, and miR-200) by post-transcriptional activation of the MYB and HBP1 oncogenes.



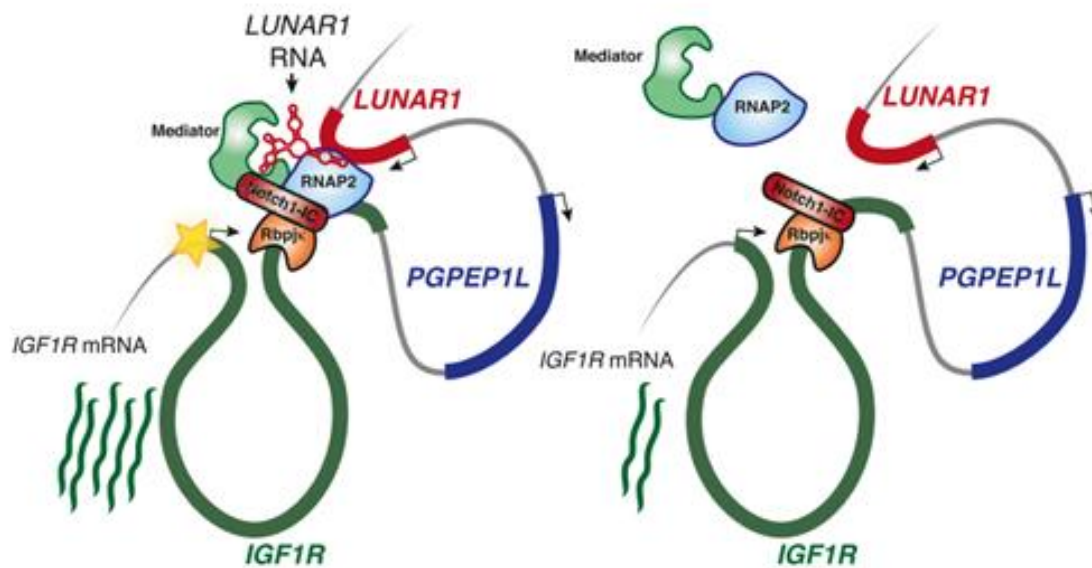


**Figure 3.4** MicroRNAs (miRNAs) implicated in T-ALL disease biology. Taken from (Durinck et al., 2015).

### 3.2.9.2 lnc-RNAs and T-ALL

Long non-coding RNAs are transcripts with a length of at least 200bp that lack protein-coding potential. They can be positionally located as antisense, intronic, intergenic, or overlapping transcripts with protein-coding genes, they have not evolutionary conservation, they have been involved in diverse functional mechanisms in gene expression regulation, mainly with association with chromatin modifier enzymes, serving as scaffolds among multiple proteins, acting as a guide to target chromatin remodelers to their sites of regulation, inducing conformational changes and thus activating or inactivating the protein complex (Durinck K et al 2015).

Trimarchi and col. Identified a set of lncRNAs in T-ALL with aberrant NOTCH1 signaling and they characterized LUNAR1 as an oncogenic lncRNA. LUNAR1 localizes in the nucleus and is overexpressed in T-ALL cases that have to activate NOTCH1 mutations. LUNAR1 is located in cis to the IGF1R gene and activates its expression through interaction between LUNAR1 and an intronic enhancer element located in the IGF1R locus as discovered by chromosome conformation capture analysis (Hi-C). In vitro knockdown of LUNAR1 significantly decreased leukemic cell growth by down-regulation of the IGF1R signaling (Figure 3.5). Furthermore, in vivo oncogenic capacity of LUNAR1 was supported by xenograft assays and later supported by other studies in which LUNAR1 was identified as the top candidate of NOTCH1 regulated lncRNAs in T-ALL and normal T cell development (Trimarchi et al., 2014).



**Figure 3.5** Model for *cis*-regulation of gene expression by *LUNAR1*. Taken from (Trimarchi et al., 2014).

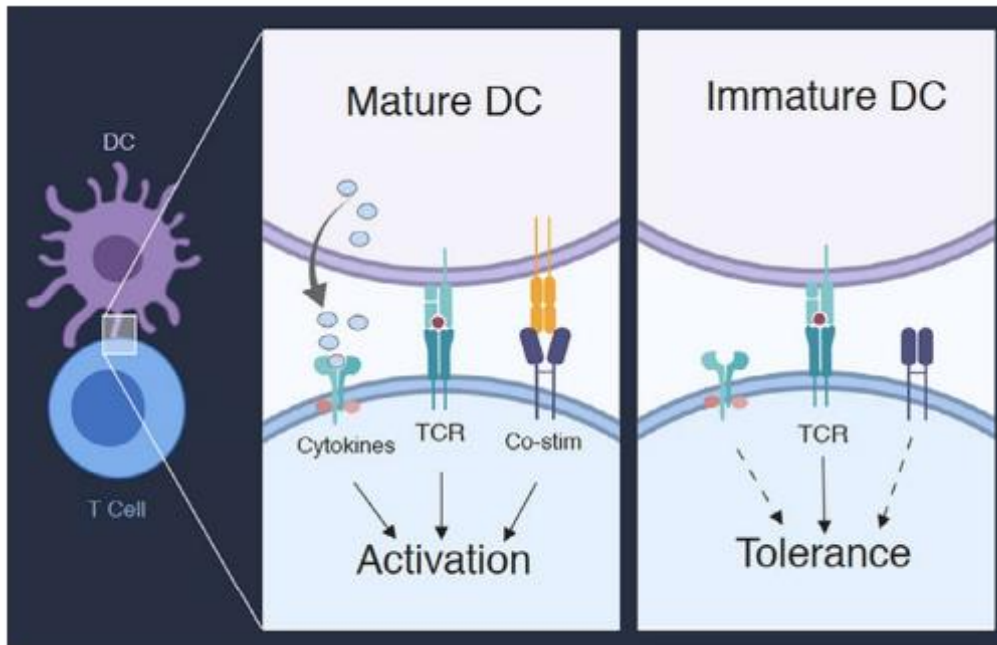
### 3.3 $Ca^{2+}$ signaling and T-Cell activation

T-cell activation is important for T-cell function, the genetic and cellular alterations that define cancer provide the immune system with the means to generate T-cell responses that recognize and eradicate cancer cells. However, elimination, of cancer by T-cells is complex, since it manages the delicate balance between the recognition of non-self and the prevention of autoimmunity. In the denominated ‘Cancer-Immunity Cycle’ (Chen and Mellman, 2013). The reason why I consider it relevant to discuss the biology of T-cell activation at the end of the T-ALL Chapter. Besides the current understanding of the multiple factors that govern T-cell activation have been used to create some immunotherapy strategies, like the immune check-point inhibitors (Saibil & Ohashi, 2020). In the next section, a brief introduction of T-cell activation is described.

#### 3.3.1 T-cell activation

The immune system has the challenge to respond to a universe of pathogens while limiting autoreactivity, for T-cells this process begins in the thymus where “central tolerance” or “thymic selection” takes place (Daley SR et al 2017). In the thymus, the immature T-cells undergo a process of T-cell receptor (TCR) rearrangement in their variable, diversity, and joining segments (Davis MM 1990). For a T-cell to mature and leave the thymus, the rearranged T-cell receptor must recognize peptide antigens presented by the major histocompatibility complex (MHC) of the antigen-presenting cells (APC), a process called “positive selection” on the contrary T-cells expressing T-cell receptors with high affinity to the self-peptide MHC complexes undergo “negative selection” meaning cell death. Negative selection is a mechanism preventing strong autoreactive T cells to leave the thymus and cause autoimmunity (Saibil and Ohashi 2020).

The stimulation of a TCR in a mature T-cell with an antigen-MHC complex can lead to different outcomes. T-cell activation is induced when delivered with the appropriate additional co-stimulatory signals and T-cell tolerance when T-cells display a hypo-responsive state called “anergy” this state is characterized by a lack of proliferation and low production of interleukin-2 (IL-2) (Schwartz 2003) (Figure 3.6).



**Figure 3.6** The maturation status of dendritic cells (DCs) determines T-cell activation and tolerance. Taken from (Saibil and Ohashi 2020).

The functional state of dendritic cells (DCs) is determinant for the decision between T cell activation and tolerance. The current model of T cell activation is that immature DCs are tolerogenic and induce T cell tolerance through deletion or anergy and that mature DCs are activating and induce a robust immune response against the antigens they present (Osorio F et al 2015).

### 3.3.2 Calcium-NFAT transcriptional signaling in T-cell activation

$Ca^{2+}$  signaling activates the nuclear factor of activated T cells (NFAT), a family of TFs composed of four members (NFATC1-C4). NFAT regulates gene transcription during T-cell activation and differentiation, cardiac valve development, differentiation of skeletal muscle fibers among others. They are implicated in biological processes such as transplant rejection, osteoporosis, myocardial hypertrophy, allergy, and autoimmune disease (Horsley et al., 2002).

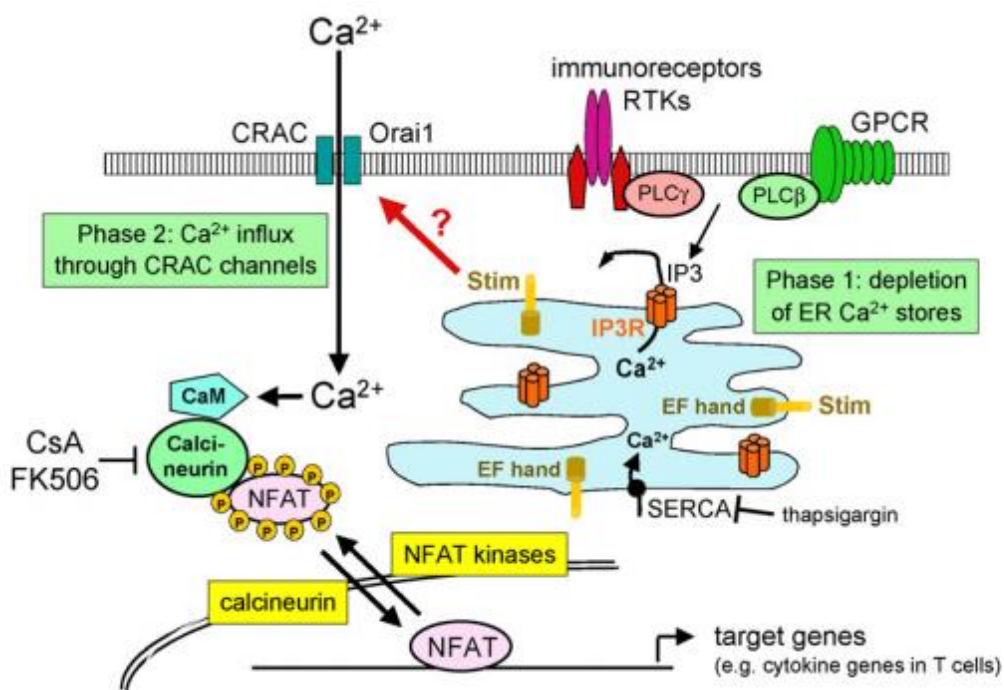
NFAT proteins are phosphorylated and reside in the cytoplasm, when cells are stimulated, they are dephosphorylated by calcineurin, a calmodulin-dependent serine/threonine phosphatase, and translocate to the nucleus (Crabtree GR et al 2002).

The stimulation of the immune cells including B-cells, T-cells, mast cells, and NK cells, through their respective immune receptors, activate similar downstream signaling pathways and TFs (Gwack Y et al 2007). The immune receptors are coupled to tyrosine kinases and the Src and ZAP70/Syk families whose activation induces tyrosine phosphorylation and activation of  $PLC\gamma$  with the consequent generation of  $InsP3$  and diacylglycerol as second messengers.  $InsP3$  mediated depletion of ER  $Ca^{2+}$  stores resulting in  $Ca^{2+}$  release-activated

Ca<sup>2+</sup>-channels (CRAC) opening and Ca<sup>2+</sup> influx across the plasma membrane and induce activation of NFAT TF. Diacylglycerol binds to Ras-GRP and protein kinase C, thus activating MAP kinase and IKK (I $\kappa$ B kinase) pathways, which lead to activation of the AP-1 (Fos-Jun) and NF $\kappa$ B transcription factors respectively (Figure 3.7) (Macian F et al 2001; Karin M et al 2005).

NFAT, AP-1, and NF $\kappa$ B act in concert with secondary transcription factors to induce the transcription of a large number of genes that regulate lymphocyte proliferation and differentiation. NFAT, AP-1, and NF $\kappa$ B were shown to be optimally activated in response to different patterns of Ca<sup>2+</sup> signaling in Jurkat T cells (Dolmetsch RE et al 1997). Transient high Ca<sup>2+</sup> peaks induced sustained activation of JNK and NF $\kappa$ B, but not NFAT, whereas prolonged low increases in Ca<sup>2+</sup>, which were insufficient to activate JNK or NF $\kappa$ B, sufficed to induce NFAT (Dolmetsch RE et al 1997).

The long term response of immune cells to sustained Ca<sup>2+</sup> signaling involves transcriptional programs that include proliferation, differentiation, and acquisition of effector functions by naive T and B lymphocytes, following the first encounter with antigen and transcription of cytokine, chemokine, and other activation associated genes by differentiated effector T cells upon second exposure with the antigen (Gwack et al., 2007).



**Figure 3.7** Schematic view of NFAT activation cycle. Taken from (Gwack et al., 2007).

# Chapter 4 Calcium pump families and Ca<sup>2+</sup> signaling

## 4.1 General overview of Ca<sup>2+</sup> signaling

Basal Ca<sup>2+</sup> concentrations are tightly controlled within a narrow submicromolar range by an array of Ca<sup>2+</sup> channels and pumps that are susceptible to dysregulation in cancer. Transient changes in cytosolic Ca<sup>2+</sup> induce downstream signaling events which regulate a wide range of cellular functions (Berridge et al., 2003). Ca<sup>2+</sup> signaling is required for every stage of the eukaryotic cell cycle, including activation and expression of transcriptional factors and cyclin-dependent kinases which are necessary for cell cycle progression (Hogan et al., 2003; Roderick and Cook 2008) as well as centrosome duplication and separation (Fukasawa, 2007; Matsumoto and Maller 2002). Crosstalk with other signaling mechanisms such as the Ras pathway regulates cell cycle transition and cell proliferation (Cook and Lockyer, 2006; Cullen and Lockyer, 2002). Dynamic regulation of Ca<sup>2+</sup> signaling is achieved by the cooperation of various cellular components including receptors, channels, transporters, buffering proteins, and downstream effectors (Berridge et al., 2003). Thus, inappropriate activation of Ca<sup>2+</sup> influx channels or downregulation of Ca<sup>2+</sup> efflux and sequestration mechanisms could increase basal Ca<sup>2+</sup> to augment Ca<sup>2+</sup> signaling and tumor cell proliferation (Monteith et al., 2007).

The participation of Ca<sup>2+</sup> in many different cell processes demands an efficient and precise control of Ca<sup>2+</sup> homeostasis. Specialized proteins that bind Ca<sup>2+</sup> with high specificity and affinity in the intracellular compartment perform this task; the free Ca<sup>2+</sup> intracellular concentration oscillates around 100 nM. These proteins belong to two broad groups: those that are embedded to membranes and move Ca<sup>2+</sup> across them, serving only as Ca<sup>2+</sup> buffers, and those that are not membrane-bound but in addition to Ca<sup>2+</sup> buffer properties, also participates in the transduction of the message by targeting proteins or enzymes. These proteins are called Ca<sup>2+</sup> sensors (Brini et al., 2012).

The mechanism on how Ca<sup>2+</sup> sensors decode the Ca<sup>2+</sup> signal has been clarified only for a few proteins, being calmodulin the best-studied example, it is general hypothesized that the decoding process involves a conformational change of the sensor protein upon binding Ca<sup>2+</sup> and upon contacting the target enzyme (Brini & Carafoli, 2009).

The processing of the Ca<sup>2+</sup> message by specialized proteins indicates that, as a rule, Ca<sup>2+</sup> does not transduce signals per se: the participation of a sensor protein that binds it and then contacts the targets of the message appears to be important, however, some proteins can bind and decode the Ca<sup>2+</sup> signal without the need of intermediate sensors, such as the case of protein kinase C (Oancea and Meyer 1998). Some proteins can have both Ca<sup>2+</sup> binding and Ca<sup>2+</sup> decoding functions, like calcineurin, which can respond to the decoded Ca<sup>2+</sup> message of calmodulin through its calmodulin-binding domain but also contains a separate calmodulin-like subunit that binds Ca<sup>2+</sup> (Aramburu et al., 2001). A more complex example is

the case of calpain that contains a covalently bound calmodulin-like domain within the catalytic subunit, and an additional smaller calmodulin-like subunit (Bertipaglia & Carafoli, 2007).

The translation of the  $\text{Ca}^{2+}$  message requires the regulation in a spatiotemporal manner of the sensor proteins. This regulation is accomplished by the action of membrane  $\text{Ca}^{2+}$  protein transporters and channels that move  $\text{Ca}^{2+}$  in both senses across the membrane boundaries, fulfilling the demands of the sensor proteins and, in turn, of the  $\text{Ca}^{2+}$  signaling functions. These membrane proteins are classified into different subclasses according to the transport mechanism, affinity for  $\text{Ca}^{2+}$ , and total transport capacity. The ATPases are high affinity, low-capacity systems; the exchangers ( $\text{Na}^+/\text{Ca}^{2+}$ ) have the opposite property, meanwhile, the channels are proteins that let  $\text{Ca}^{2+}$  flow passively across membranes when some gating mechanisms induce their opening (Carafoli et al., 2001). Voltage-gated (Catterall, 2000) and ligand-gated (Strigrow & Ehrlich, 1996) channels are the best-studied transport  $\text{Ca}^{2+}$  systems. Trp channels and store-operated channels were discovered later (Parekh & Putney, 2005; Venkatachalam & Montell, 2007).

Besides versatility,  $\text{Ca}^{2+}$  signaling has other properties, among them its auto-regulatory capacity at transcriptional and post-transcriptional levels. Autoregulation means that the expression and the activity of a variety of proteins that control  $\text{Ca}^{2+}$  signaling and cellular levels are regulated by the  $\text{Ca}^{2+}$  signal itself.  $\text{Ca}^{2+}$  can regulate gene transcription in an early or late response manner (Santella & Carafoli, 1997). The best characterized  $\text{Ca}^{2+}$  transcriptional regulation system is the CREB model. CREB is a transcription factor of the bZIP family that binds to the cAMP-responsive element (CRE) and to the  $\text{Ca}^{2+}$  responsive element (CARE) on DNA to be phosphorylated and activated by calmodulin kinases. The  $\text{Ca}^{2+}$  dependent phosphatase calcineurin then regulates the phosphorylation state of CREB in a complex process involving other phosphatases (Brini & Carafoli, 2009).

$\text{Ca}^{2+}$  can also regulate gene transcription without the intervention of protein kinases and phosphatases, but rather by the direct interaction with transcription factors. An interesting example of this type of transcriptional regulation is the one mediated by the downstream regulatory element antagonistic modulator (DREAM), a  $\text{Ca}^{2+}$  binding protein that acts as a gene repressor (Carrión et al., 1999). In the  $\text{Ca}^{2+}$  free form, it represses transcription by binding to specific DNA sites in the promoters of several genes. The binding of  $\text{Ca}^{2+}$  removes DREAM from the sites, restoring transcription. Among the genes regulated by DREAM we can find *NXC3* a plasma membrane  $\text{Na}/\text{Ca}$  exchanger (Gomez-Villafuertes et al., 2005).

An increase in  $\text{Ca}^{2+}$  basal concentration above the optimal values (100-200nM) can be tolerated for short periods but a sustained increase can lead to permanent activation of detrimental enzymes like proteases, phospholipases, and nucleases that can lead to cell discomfort and cell death. The negative outcome produced by the sustained increase of  $\text{Ca}^{2+}$  levels illustrates the ambivalent nature of  $\text{Ca}^{2+}$  signaling (Farber, 1981).

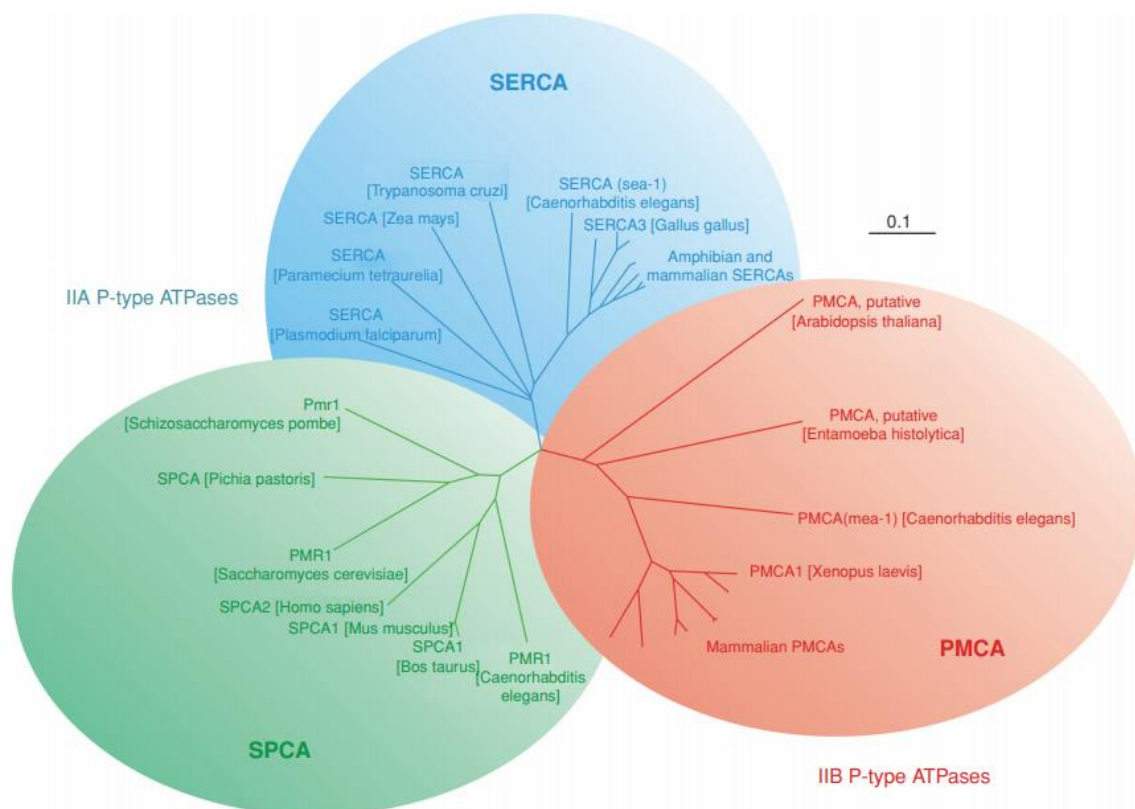
## 4.2 Calcium pumps families

In eukaryotes nine  $\text{Ca}^{2+}$  ATPase pumps from three multigenic families have been described so far: three endo(Sarco)plasmic reticulum (SERCA), four plasma membrane (PMCA), and two secretory pathway (SPCA)  $\text{Ca}^{2+}$ -ATPase pumps, these last two been the most recently

identified. The number of pump isoforms is further increased by alternative splicing. The three pump types share the basic features of catalytic activity but differ in tissue distribution, regulation, and role on the homeostasis of  $\text{Ca}^{2+}$  (Figure 4.1) (Brini et al., 2012).

$\text{Ca}^{2+}$  is important for the control of key cellular functions in all eukaryotic organisms, including fertilization, muscle contraction, secretion, for several phases of metabolism, gene transcription, and apoptosis to list some of them. The specificity of cellular  $\text{Ca}^{2+}$  signals depends on a complex array of ion channels, pumps, and exchangers that drive the fluxes of  $\text{Ca}^{2+}$  ions across the plasma membrane and the membrane of intracellular organelles, and on soluble proteins that bind  $\text{Ca}^{2+}$  in a reversible manner (Berridge et al., 2003).

Mammal  $\text{Ca}^{2+}$  pump types belong to the family of P-type ATPases, which conserve ATP energy in the form of a phosphorylated aspartic acid (P-type) (Pedersen & Carafoli, 1987) from a highly conserved sequence, **SDKTGT** (L/I/V/M) (T/I/S). The P-type ATPases integrated a superfamily of at least eight subfamilies with hundreds of members. The sequences have 15% of identity to each other but contain eight conserve core regions (Axelsen & Palmgren, 1998). The introduction of substrate specificity in the subfamilies as a result of sudden changes in sequence evolution. Five branches have been identified in the phylogenetic tree; those of the  $\text{Ca}^{2+}$  ATPases belong to the subfamilies IIA (SERCA and SPCA) and IIB (PMCA) (Figure 4.1)(Brini et al., 2012).

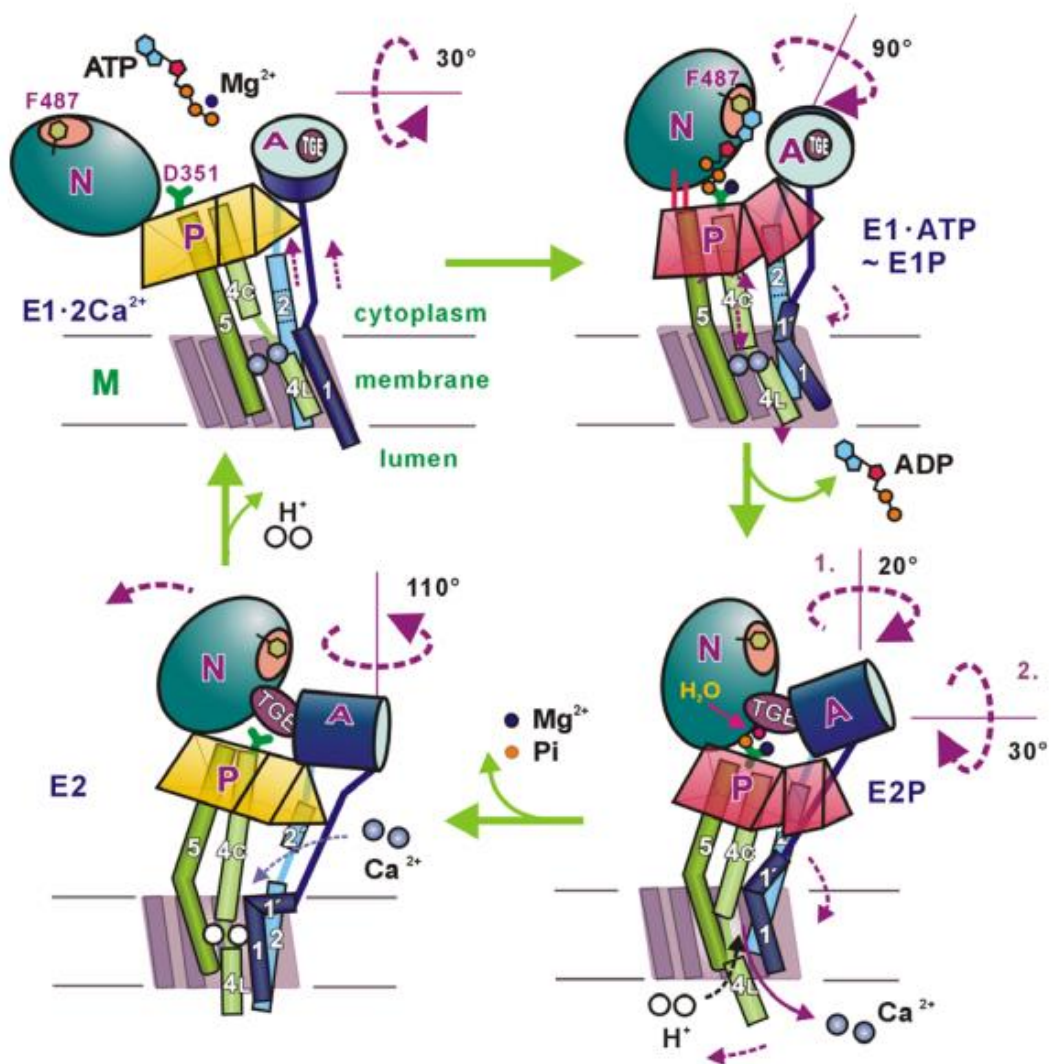


**Figure 4.1** Phylogenetic tree of the  $\text{Ca}^{2+}$ -ATPase pumps families. Adapted from (Brini et al., 2012)

The three  $\text{Ca}^{2+}$  pump types share properties like membrane topology and reaction mechanism and although only the 3D structure of a SERCA group member has been solved,

features of the 3D structure are believed to be shared with all the types of  $\text{Ca}^{2+}$  pumps (Toyoshima et al., 2000; Toyoshima & Nomura, 2002). However, the three pump types differ in regulation, the action of some inhibitors, and biology function (Brini et al., 2012).

In figure 4.2 a simplified scheme of the enzymatic reaction pertinent for all three types of  $\text{Ca}^{2+}$  pumps is shown.



**Figure 4.2** Cycle transport of  $\text{Ca}^{2+}$  ATPase-pumps. Adapted from(Toyoshima, 2008).

The two basic conformational states of the pumps, the E1 state in which the enzyme interacts with  $\text{Ca}^{2+}$  with high affinity at one side of the membrane, and the E2 state, in which the affinity for  $\text{Ca}^{2+}$  drops, leading to the release of the ion at the opposite site (de Meis & Vianna, 1979).

According to the 3D structure of the solved SERCA pump, there are two  $\text{Ca}^{2+}$  binding sites in the transmembrane helices (Clarke et al., 1989). These two sites are often created by residues with acidic properties located in the transmembrane domains 4, 5, 6, and 8. The PMCA and SPCA pumps however lack one essential acidic residue in transmembrane



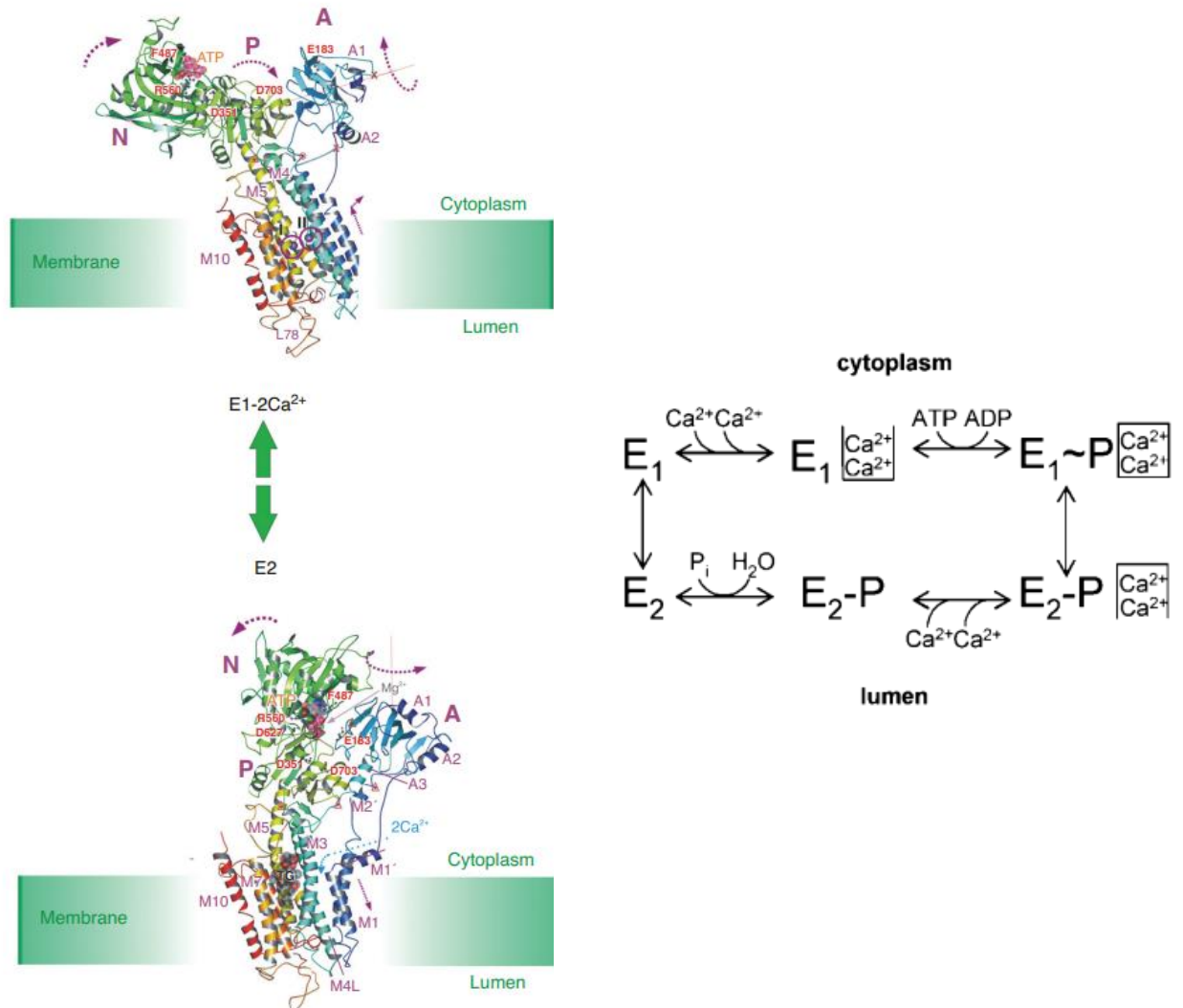
domain 5, therefore losing one  $\text{Ca}^{2+}$  binding site, altering the stoichiometry of the  $\text{Ca}^{2+}/\text{ATP}$  transport from 2.0 for the SERCA pump to 1.0 for the PMCA and SPCA pumps (Brini et al., 2012).

#### 4.2.1 The SERCA pump family

The first ATP-driven  $\text{Ca}^{2+}$  transport system was observed in a fraction of skeletal muscle identified as the vesicles of sarcoplasmic reticulum (SR) (HASSELBACH & MAKINOSE, 1961). The purified pump responsible for the process is a protein of about 110 kDa (MacLennan, 1970). The transport process, initially analyzed on the SR of skeletal and cardiac muscles, was also identified in the endoplasmic reticulum (ER) of all cells. The SERCA pump controlled the cytoplasmic  $\text{Ca}^{2+}$  levels in muscle cells, where its capacity to clear  $\text{Ca}^{2+}$  from the cytosol induced relaxation. Some SERCA pump specific inhibitors have been discovered; thapsigargin (Thastrup et al., 1990), cyclopiazonic acid (CPA) (Seidler et al., 1989) and 2,5-di(t-butyl) hydroquinone (tBHQ) (Brini et al., 2012; Oldershaw & Taylor, 1990).

The chemical structure of the SERCA pump is a single polypeptide chain folded into four major domains; a transmembrane (M) domain, integrated of ten transmembrane helices (M1-M10), and three cytosolic domains, the actuator (A) domain, the phosphorylation (P) domain and the nucleotide-binding (N) domain. The binding and translocation of  $\text{Ca}^{2+}$  induce a conformational change from a compact structure, involving the cytosolic portion domains, to a more open structure (Figure 4.3) (Toyoshima 2008).

$\text{Ca}^{2+}$  binds to two sites, (the site I and site II) in the M domain, the binding sites exist in a high-affinity state, which permits access of  $\text{Ca}^{2+}$  from the cytosolic side (E1 state), or in a low-affinity state, in which the sites face the luminal side (E2 state) and favor the release of  $\text{Ca}^{2+}$  to it. The binding of the two  $\text{Ca}^{2+}$  to the two sites occurs sequentially. The P-domain is located in the cytosolic portion of the pump and is composed of two regions: a short N-terminal region connected to M4 that holds the catalytic D351, and a longer C-terminal portion connected to M5. The A domain contains the highly conserved TGES sequence, which plays a key role in the hydrolysis of the phosphorylated D (Brini et al., 2012).



**Figure 4.3** The SERCA pump conformational states and Calcium transport cycle. Adapted from (Toyoshima, 2008)

Three human genes *ATP2A1-3* encode the three principal SERCA proteins. The number of pump variants is increased due to the tissue-dependent alternative splicing process. Most of the SERCA pumps are expressed in muscle either skeletal, smooth, or heart muscle, and therefore participates in the physiology of muscular contraction. Post-translational modifications of the SERCA pump have been described, like glycosylation, oxidation, s-nitrosylation, and phosphorylation, which can control their activity and highlights the role of the pump in some processes like apoptosis, protein translation, RE redox sensor (Brini et al., 2012).

There are two well-known SERCA disorders, Brody and Darier disease, which are phenotypically different. Brody disease is a rare autosomal recessive disorder affecting the skeletal muscle during exercise, characterized by painless cramps, slow muscle relaxation, and stiffness (Brody 1969). This disorder was linked to the reduced uptake of  $\text{Ca}^{2+}$  in the sarcoplasmic reticulum and recessive mutations in *ATP2A1*, the gene encoding the SERCA1 isoform (Odermatt et al., 1996).

Darier disease is an autosomal dominant genetic skin disorder characterized by keratotic papules because of the loss of desmosomal proteins at the cell-to-cell junctions that bind keratin filaments. There have been over 100 mutations in *ATP2A2*, the gene encoding SERCA2, reported in patients, distributed through the gene (Burge & Wilkinson, 1992).

#### 4.2.2 The PMCA pump family

In 1966 Schatzmann discovered a system of ATP-dependent  $\text{Ca}^{2+}$  expelled in erythrocytes (Schatzmann, 1966), later the same system was found in other tissues including excitable ones. The pump responsible was purified in 1979 using the calmodulin column (Niggli et al., 1979). The  $\text{Ca}^{2+}$  pump has a stoichiometry of 1:1 $\text{Ca}^{2+}$ /ATP, the PMCA pump has high  $\text{Ca}^{2+}$  affinity but low transport capacity (Brini et al., 2012). As with all the P-type pumps, the PMCA pump is inhibited by  $\text{La}^{3+}$  and vanadate. The PMCA pump was cloned in 1988 (Shull & Greeb, 1988; Verma et al., 1988), demonstrating the same topology organization as the SERCA pump: ten transmembrane domains and a cytosolic region composed of three domains (A, N, and P domains). The catalytic phosphorylation site (SDKTGLT) is conserved, among the differences concerning the other two  $\text{Ca}^{2+}$  pumps we found the first cytosolic loop and the  $\text{Ca}^{2+}$  binding sites, the site I does not exist in PMCA pump, The long C-terminal tail of the PMCA pump contains most of the regulatory domains, being the most important the calmodulin-binding domain (James et al., 1989).

The C-terminal tail of the pump is proposed to interact with two sites in the first and second cytosolic loops of the enzyme under non-activated conditions blocking access to the active site (Falchetto et al., 1991, 1992). Calmodulin interacts with its binding domain removing it from its anchor sites next to the active center, liberating the pump from auto-inhibition. Among the PMCA pump activators, we can list calmodulin and acidic phospholipids (Niggli et al., 1981). Acidic phospholipids bind to two sites in the pump molecule: one is the calmodulin-binding domain (BRODIN et al., 1992), the other is in the cytosolic loop that connects transmembrane domains 2 and 3 (Zvaritch et al., 1990).

The activation by acidic phospholipids may be physiologically relevant, as they integrated the membrane environment of the pump. Phosphatidylinositol bi phosphorylated (PIP<sub>2</sub>) is the most active acidic phospholipid, and its membrane concentration is regulated during  $\text{Ca}^{2+}$  signaling processes suggesting a possible PIP<sub>2</sub>-mediated reversible cycle of the PMCA activation model (Choquette et al., 1984).

Other activation mechanisms of the pump are the cleavage by calpain, and phosphorylation by protein kinase C (PKC) and PKA. The cleavage by calpain takes place immediately upstream of the C-terminal calmodulin-binding domain and activates the pump irreversibly, making it calmodulin insensitive (James et al., 1989).

The PMCA gene family is integrated by four basic gene products (*ATP2B1-4*). Alternative splicing of these four transcripts raises the number of isoforms to 30 (Strehler & Zacharias, 2001). The four main isoforms differ in tissue specificity and calmodulin affinity see table. PMCA 1 and 4 are ubiquitous and present poor calmodulin affinity, PMCA 2 and 3 have higher calmodulin sensitivity and their expression is tissue-restricted: PMCA2 is expressed mainly in the nervous system and in the mammary gland meanwhile, PMCA3 is expressed in the nervous system (Strehler & Zacharias, 2001).

**Table 4.1 Properties of PMCA family**

	PMCA1b	PMCA2b	PMCA3b	PMCA4b
<b>Tissue specificity</b>	Ubiquitous	Brain	Brain	Ubiquitous
<b>Kd CaM</b>	40-50nM	2-4nM	8nM	30-40nM

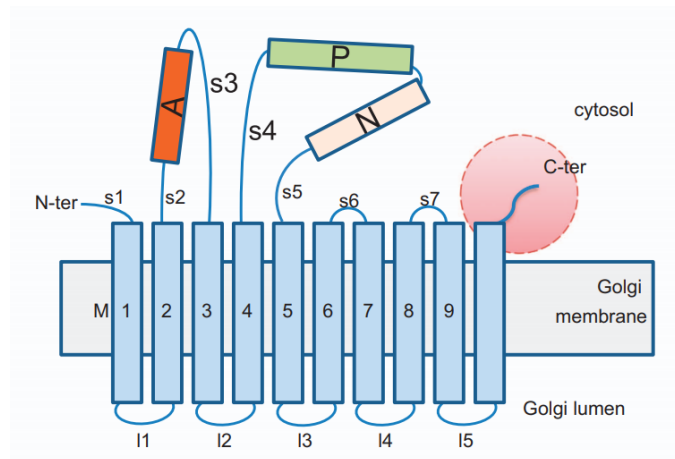
Among the PMCA2 properties that distinguish them from the other three PMCA pumps, we find its high resting activity, making them less responsive to calmodulin in comparison to the 4-to-6-fold increase activity of the other pumps in response to it, in summary, although PMCA2 binds calmodulin with high affinity, its activity only increases by 20% or 30%. This property could explain its main role in cells with special  $\text{Ca}^{2+}$  homeostasis demands like neurons (Elwess et al., 1997).

#### 4.2.3 The ATP2C (SPCA) pump family

A  $\text{Ca}^{2+}$  pump was identified in the Golgi membrane, this ATP-dependent  $\text{Ca}^{2+}$  uptake was detected into purified Golgi vesicles (Neville MC et al 1981). The  $\text{Ca}^{2+}$  affinity of the Golgi pump is higher than that of the SERCA pump and PMCA pump. The first member of the Golgi subfamily (Pmr1, plasma membrane ATPase-related pump) was identified in yeast (Rudolph et al., 1989). In humans, the first SPCA pump (SPCA1) was identified in a study that described mutations causing the acantholytic Hailey-Hailey skin disease see below for more details (Z. Hu et al., 2000). A second human SPCA pump (SPCA2) was later discovered (Vanoevelen et al., 2005; Xiang et al., 2005). One distinctive property of the SPCA pumps is the efficient transport of  $\text{Mn}^{2+}$ . This function is related to the presence of many  $\text{Mn}^{2+}$  requiring enzymes in the lumen of the Golgi compartment, like the glycosyltransferases. N-acetyl glucosamine transferases, mannosyl transferases, glucuronyl transferases, and fucosyl transferases are also Golgi lumen enzymes that require  $\text{Mn}^{2+}$  (van Baelen et al., 2001).

The transport of  $\text{Mn}^{2+}$  by the pump is important for the regulation of the  $\text{Mn}^{2+}$  concentration in the cytosol. Oxidative damage in yeast cells lacking superoxide dismutase (SOD) is suppressed in cells lacking functional Pmr1: they accumulate  $\text{Mn}^{2+}$  in the cytoplasm, which replaces SOD as a scavenger of ROS (Lapinskas et al., 1995). The SPCA pump has also an important role in the process of  $\text{Mn}^{2+}$  detoxification. The SPCA pump  $\text{Ca}^{2+}$  uptake into the Golgi vesicles is near 70% in keratinocytes (Callewaert et al., 2003).

The SPCA pumps are predicted to be structured in 10 transmembrane helices and three cytosolic domains (A, P, and N) as identified in the SERCA pump (Figure 4.5), as the other two pump families discussed before, SPCA pumps also have the critical consensus sites like the sequence (SDKTGTLT) surrounding the catalytic site. The SPCA pumps are shorter than the SERCA and, PMCA pumps due to the shorter luminal loops that connect some of the transmembrane domains, and to the absence of a long cytoplasmic tail (Brini et al., 2012).



**Figure 4.5** Schematic representation of the SPCA pump domains. Adapted from (Micaroni et al., 2016).

The human SPCA2 is 32 residues longer in the NH<sub>2</sub>-terminal domain than the SPCA1 (Xiang et al., 2005), also differs in the COOH terminal portion which contains putative protein binding motifs like PDZ binding motif that can be implicated in some functions like intracellular trafficking or directs the location of the SPCA2 pump (Heilker et al., 1996).

The SPCA pump only has one Ca<sup>2+</sup> transport site, corresponding to SERCA site II, consistent with a probable 1 :1 Ca<sup>2+</sup>/ATP transport stoichiometry. The site appears to be formed by oxygens of residue E329 in transmembrane domain 4 and N796 and D800 in transmembrane domain 6 (Wei et al., 2000). Studies on the Pmr1 yeast enzyme discovered that the Mn<sup>2+</sup> selectivity of the pump is instead defined by residue Q783 in transmembrane domain 6 and by conformation-sensitive packing interactions between Q783 and V335 in the transmembrane domain (D. Mandal et al., 2000; D. M. Mandal et al., 2000).

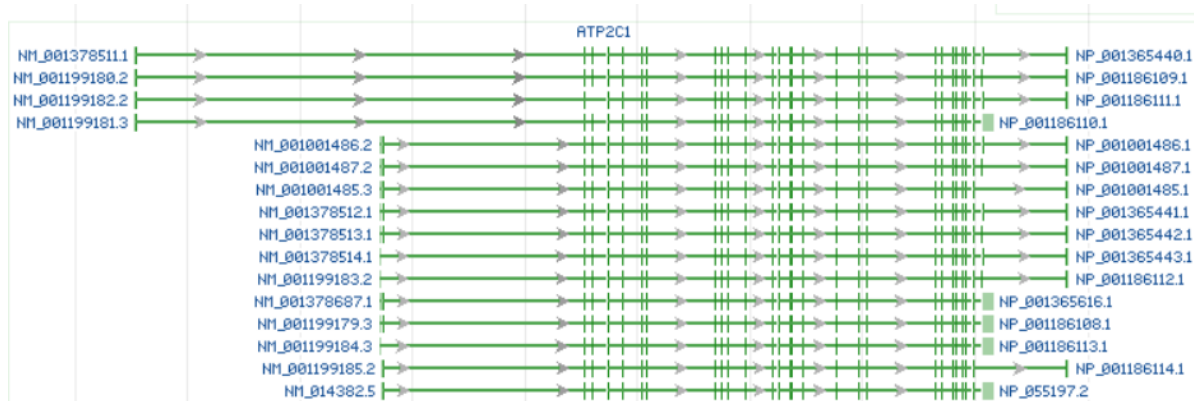
The high Ca<sup>2+</sup> affinity of the SPCA pumps has been associated with a slower E1-P(Ca<sup>2+</sup>)/E2P transition in the reaction cycle. This high Ca<sup>2+</sup> affinity of the SPCA pumps ensures that the Golgi compartment will be constantly refilled with Ca<sup>2+</sup>, which is required inside the vesicles for the optimal activity of several important enzymes, most notably the endoproteases that perform the proteolytic processing of prohormones (Oda, 1992).

Differently from SERCA and PMCA pumps which are electrogenic proton exchangers (H<sup>+</sup>), the SPCA1 pump does not have this function. Protons are essential for many important functions in the lumen of the Golgi vesicles and must thus be kept inside. Both SPCAs transport Mn<sup>2+</sup> with an affinity that is as high as that for Ca<sup>2+</sup>. Mn<sup>2+</sup> is required for optimal activity of a Golgi-located casein kinase (D. W. West & Clegg, 1984).

#### 4.2.3.1 *ATP2C1* gene

The *ATP2C1* gene is located on chromosome 3q22.1, spans 166 kb, and consists of 28 exons. In addition to the four distinct splice isoforms SPCA1 1a–d, corresponding to *ATP2C1* 1a–d produced by alternative processing of the *ATP2C1* pre-mRNA reported in most literature (Nellen RG et al 2017), other splice isoforms (SPCA1 1e–f and SPCA1 2a–d, corresponding to *ATP2C1* 1e–f and *ATP2C1* 2a–d) are also recorded in NCBI (<https://www.ncbi.nlm.nih.gov/gene/27032>) and UniProt

(<https://www.uniprot.org/uniprot/P98194>) Of these splice isoforms, SPCA1 1a (NP\_055197.2) is the “canonical” sequence containing 919 amino acids. SPCA1 2d (NP\_001365440.1) is the longest isoform containing 983 amino acids. The isoforms 2a-d are the only ones derived from the most distal promoter P1 of the gene. SPCA1 1c is an aberrant Ca<sup>2+</sup> pump with limited functional capacity due to the absence of exon 27, which results in the disruption of transmembrane 10 (M10) (van Baelen et al., 2004).



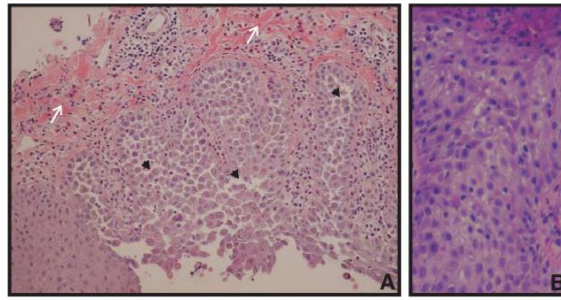
**Figure 4.6** ATP2C1 transcript isoforms taken from NCBI.

#### 4.2.3.2 ATP2C1 and Hailey-Hailey disease

The brothers Hugh Edward Hailey and William Howard Hailey, both dermatologists, described a skin disorder that they referred to as familial benign chronic pemphigus, and later, it received the name of Hailey-Hailey disease (HHD) in honor to them (Hailey and Hailey 1939). The disorder affects females and males and is inherited in an autosomal dominant manner, the symptoms appear usually in the third or fourth decade of life. The disorder is characterized by blisters and itchy erosions mainly located at sites of sweating and friction like the groin and the axillar regions, with pain and an unpleasant smell of the skin following macerations. The HHD can be exacerbated by several external factors like friction, sweating, infections, and ultraviolet exposure (Brini et al., 2009). The incidence of HHD is estimated to be 1/50,000 (Foggia & Hovnanian, 2004).

Inactivating mutations in one allele of the gene *ATP2C1* were discovered in patients with the disease (Hu et al., 2000). Although the disorder is often benign some squamous cell carcinomas may develop from the skin lesions. In mice, the disruption of the *ATP2C1* gene causes several squamous cell carcinomas instead of the acantholytic skin disorder of humans (Okunade et al., 2007).

Histologically, the skin of the patients with the disease presents a loss of adhesion between suprabasal keratinocytes (acantholysis) and abnormal keratinization of the epidermis (dyskeratosis) (Figure 4.7) (Metze et al., 1996). Over 100 additional mutations of the *ATP2C1* gene were later identified distributed within the gene (Micaroni et al., 2016).

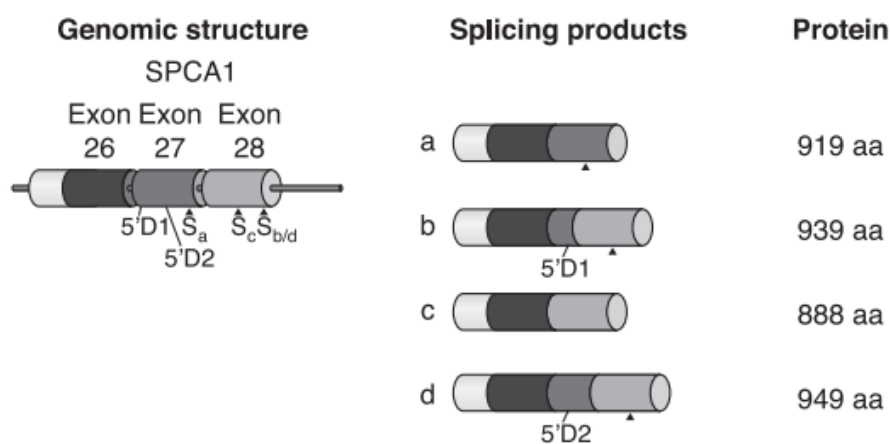


**Figure 4.7** (A) Histology of Hailey-Hailey's disease's skin (acantholysis indicated by arrowheads; dyskeratosis indicated by white arrows) (B) Histology of normal skin. Taken from (Brini and Carafoli 2009).

In a review made by Micaroni and col., from 166 *ATP2C1* mutations causing HHD, approximately 55% lead to a premature termination codon (PTC), arguing the possibility of *ATP2C1* haploinsufficiency as the prevalent mechanism for the dominant inheritance of the disease. Approximately 14% of the mutations were nonsense, 36% were insertion/deletion, 20% were splice-site mutations and 30% were missense mutations. Many of these mutations predict the absence or reduction of the mutated *ATP2C1* product via nonsense-mediated mRNA decay (Micaroni et al., 2016).

Mutations are located through all the *ATP2C1* without apparent clustering, presenting allelic heterogeneity, and distributed all over the encoded sequence and affect all the protein domains, as well as in the intron splice sites generating alternative splicing and/or truncated proteins (Micaroni et al., 2016).

Interestingly they found only one mutation in the most 3' region (exons 27-28) of the gene where differential splicing generates four different isoforms (SPCA1a, SPCA1b, SPCA1c, and SPCA1d), (Figure 4.8) none of which arises from the most distal promoter P1. This mutation located in exon 27 was a nonsense mutation that generates a stop codon (S887X) on the last transmembrane domain (M10) (H. Li et al., 2003).



**Figure 4.8** SPCA1 main splicing isoforms. Adapted from (Brini and Carafoli 2009)

The four splice variants before mentioned differ in their C-terminal tail sequence (Figure 4.8) and this is likely to be important for the functionality of the pump. Indeed, the C-terminal tail could have a role in mediating interactions with cytoplasmic effectors for intracellular

signaling or for targeting the single isoforms to specific sub-organellar localization. This would in turn make them unique not only in triggering the  $\text{Ca}^{2+}$  influx into the Golgi apparatus but also in mediating different cytoplasmic signaling, thus orchestrating membrane trafficking at different levels along the secretory pathway. However, these speculations need to be experimentally proven (Micaroni M et al 2016).

Missense mutations generate single amino acid substitution. The missense mutations occurring along with the *ATP2C1* gene sequence target mainly exons (12, 13, 18, 21, and 23). These exons encode for M4, P, and M5/M6 domains, which are critical for  $\text{Ca}^{2+}/\text{Mn}^{2+}$  binding and can compromise the stability or structure of the protein, causing a severe effect on SPCA1 functionality, without a reduction in its levels (Micaroni et al., 2016).

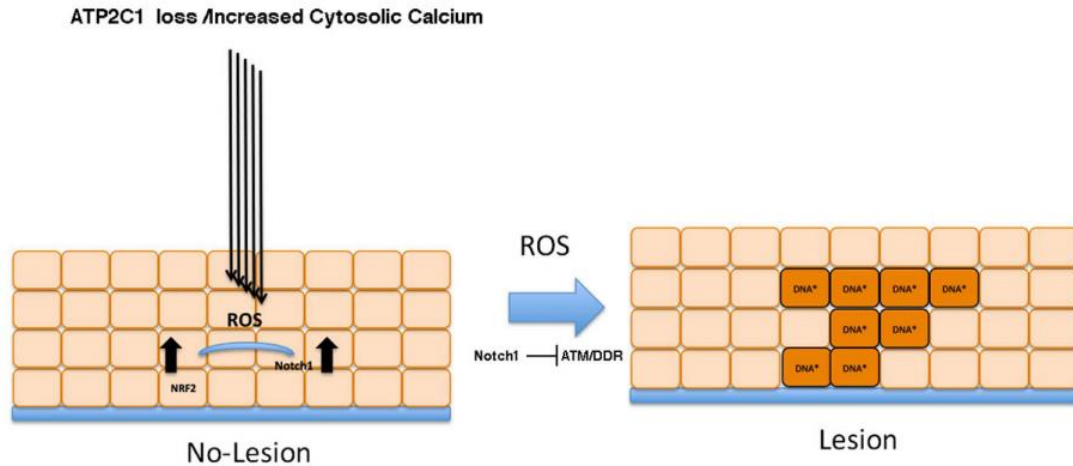
#### ***4.2.3.3 ATP2C1-NOTCH1 and DNA damage response in a Hailey-Hailey Disease model***

Cialfi and col. identified *ATP2C1* as a crucial regulator of epidermal homeostasis by oxidative stress regulation, they use siRNA technology to inactivate the *ATP2C1* gene and demonstrate an increase in oxidative stress and the up-regulation of NOTCH1 in a model of cultured human keratinocytes, they also observed a decrease of the DNA damage response (DDR) in skin lesions of patients with HHD using RNA-seq experiments, which could indicate that an *ATP2C1*/NOTCH1 axis could be critical for keratinocyte function, suggesting a possible model for HHD pathology (Cialfi et al., 2016).

DNA damage is crucial for MYC mediated replication and stress-induced keratinocyte differentiation, notably DNA damage induced by genotoxic agents induces keratinocyte differentiation (Freije et al., 2014). The loss of *ATP2C1* functions in a manner consistent with DNA damage-induced differentiation. This process is accompanied by increased Notch1 activation. Notch signaling regulates keratinocyte growth and differentiation (Rangarajan et al., 2001). Human cells expressing Notch1 show inactivation of the protein kinase ataxia-telangiectasia mutated ATM (an activator of the DDR which mobilizes and orchestrates one of the most extensive signaling networks in response to the induction of DNA damage) and other DDR components (Vermezovic et al., 2015). It has been shown that downregulation of the DDR partially constitutes a mechanism associated with astrocyte differentiation (Schneider et al., 2012).

Consistent with this model, it has been shown that the DDR is down-regulated upon the initiation of epidermal differentiation. The loss of *ATP2C1* may allow Notch1 activation to trigger the differentiation response. This response would be increased by Notch1 mediated inhibition of the DNA repair/ATM pathway in cells that accumulate irreparable DNA damage (Figure 4.9). As HHD is a skin-specific disease, one keratinocyte specific function of *ATP2C1* might be to protect the epidermal cells from temporally inappropriate activation of Notch1 (Cialfi et al., 2016).





**Figure 4.9** Proposed model of the mechanism to Hailey-Hailey disease by downregulation of the DNA damage response. From (Cialfi et al., 2016).

### 4.3 Ca<sup>2+</sup> signaling and Ca<sup>2+</sup> pumps in cancer

As mentioned before the intracellular calcium ions (Ca<sup>2+</sup>) works as a second messenger to regulate gene transcription, cell proliferation, migration, and death. Evidence has emerged that intracellular Ca<sup>2+</sup> homeostasis is altered in cancer cells and the alteration is implicated in tumor initiation, angiogenesis, progression, and metastasis.

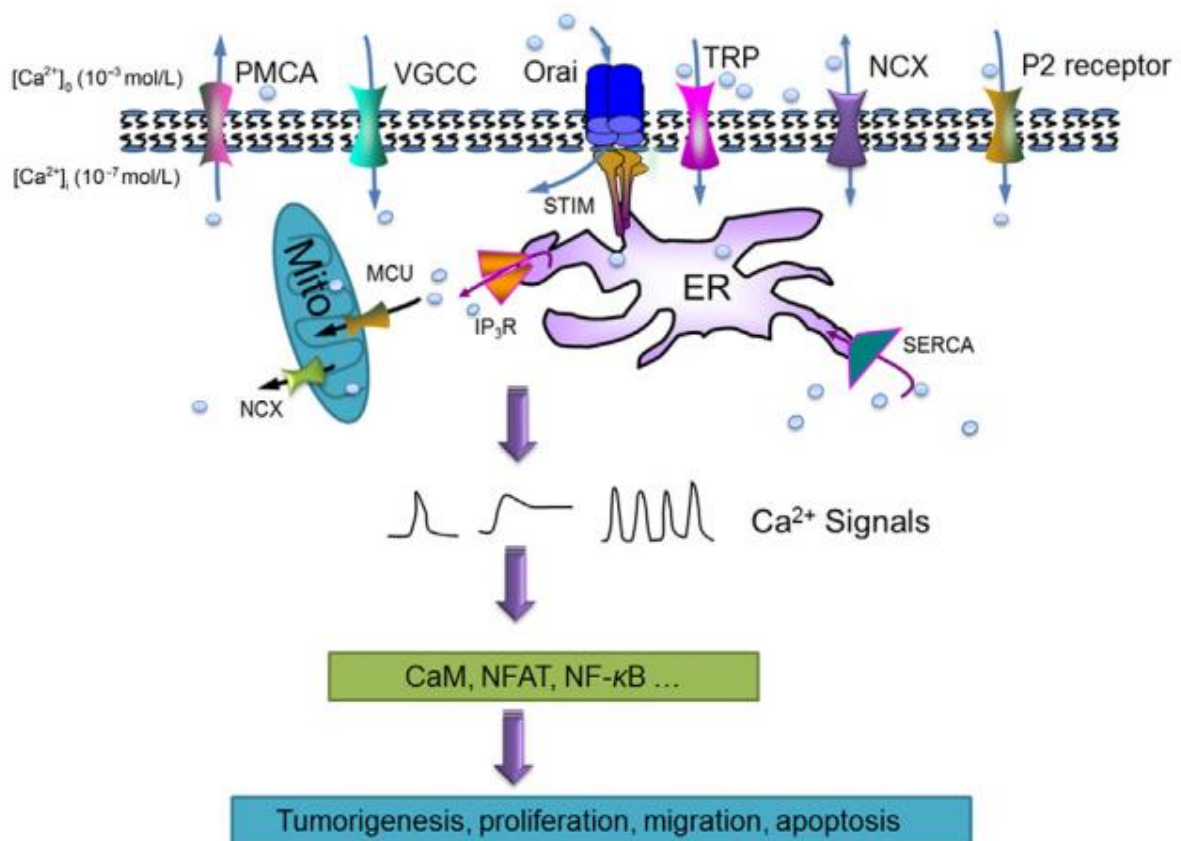
Zhu H and col. Showed that intracellular Ca<sup>2+</sup> oscillations provide essential proliferation signals for esophageal cancer cells. The frequency, amplitude, and duration of the intracellular Ca<sup>2+</sup> oscillations compose the specific Ca<sup>2+</sup> codes for selective activation of transcription factors for gene transcription, cell proliferation, and migration (Berridge, 1997; Parekh, 2011). The decoding of the oscillatory Ca<sup>2+</sup> signals is achieved by intracellular downstream effectors, including calmodulin (CaM), nuclear factor of activated T cells (NFAT), nuclear factor-κB (NFκB), calmodulin-dependent protein kinase II (CaMKII), and calpain, which differ in their on and off rates for Ca<sup>2+</sup> and activate different cellular process (Cullen, 2003; Q. Hu et al., 1999; Smedler & Uhlén, 2014).

Different Ca<sup>2+</sup> regulated kinases and enzymes often localize at different compartments within the cell. Therefore the size, kinetics, and spatial profile of a cytoplasmic Ca<sup>2+</sup> signal are all important in determining which Ca<sup>2+</sup> dependent response will be activated, when and for how long. Intracellular Ca<sup>2+</sup> oscillations may reduce the effective Ca<sup>2+</sup> threshold for signaling transduction, increasing signal detection at low levels of stimulation (Dolmetsch et al., 1998).

Disturbances in the Ca<sup>2+</sup> signaling contribute to the development of cancer phenotypes because tumors tend to remodel their Ca<sup>2+</sup> signaling network to proliferate at high rates, to increase cell motility and invasion, to escape death, to fool immune attack, or to have neovascularization (Kondratskyi et al., 2013).

Tumorigenic pathways have been associated with altered expression levels or abnormal activation of Ca<sup>2+</sup> channels, transporters, or Ca<sup>2+</sup>-ATPases pumps. Correction of these altered Ca<sup>2+</sup> signals could provide potential cancer therapies (Cui et al., 2017).

The intracellular  $\text{Ca}^{2+}$  homeostasis is governed by a network composed of various  $\text{Ca}^{2+}$  channels and transporters: IP<sub>3</sub>R receptor mediating  $\text{Ca}^{2+}$  release from endoplasmic/sarcoplasmic reticulum (ER/SR); SERCA pumps or SPCA pumps;  $\text{Ca}^{2+}$  channels or transporters allowing  $\text{Ca}^{2+}$  influx across the plasma membrane (PM) from extracellular  $\text{Ca}^{2+}$  reservoir, such as voltage-gated  $\text{Ca}^{2+}$  channel (VGCC), transient receptor potential channel (TRP),  $\text{Ca}^{2+}$  release-activated  $\text{Ca}^{2+}$  channel (CRAC),  $\text{Na}^+/\text{Ca}^{2+}$  exchanger (NCX) and purinergic receptor; mitochondrial  $\text{Ca}^{2+}$  uniporter (MCU) regulating mitochondrial  $\text{Ca}^{2+}$  uptakes (Figure 4.10). Alterations in the activity or expression of any of them can lead to disruption of  $\text{Ca}^{2+}$  signaling and carcinogenesis (Cui et al., 2017).



**Figure 4.10** Scheme of  $\text{Ca}^{2+}$  channels/transporters/pumps involved in tumorigenesis. From (Cui et al., 2017)

Mutations and altered expression levels of SERCA isoforms have been identified in various cancers, such as cancers of the colon, gastric, lung, myeloid leukemia, and choroid plexus tumors (Dang & Rao, 2016) (Table 4.2). Overexpression of SERCA2 was found in colorectal cancer cells, which could drive proliferation and migration (Fan et al., 2014).

SERCA3 was reported to diminish during the multistage process of colon tumorigenesis after initial increased expression during cell differentiation (Brouland et al., 2005). SERCA3 was also found to be downregulated in B lymphocytes after the infection of Epstein Barr virus, a

human gamma herpesvirus involved in various malignancies including Burkitt's and other lymphomas (Dellis et al., 2009).

Altered expression of SPCA isoforms occurs in different types of cancer including breast, colon, and prostate (Monteith et al., 2012). SPCA1 is highly expressed in basal-like breast cancers and has a low expression in the luminal subtypes (Grice et al., 2010).

In a study Jenkins and col. analyzed the effect of the tumor microenvironment (low O<sub>2</sub>, cell density, 3D organization) in human colon cancer cell line HCT116. They found that SPCA1 and SPCA2 are up-regulated by hypoxia (3%O<sub>2</sub>) but only SPCA2 is up-regulated by high cellular density conditions. They also provide evidence that SPCA2 is involved in maintaining Mn<sup>2+</sup> in the Golgi in live cells. The up-regulation of SPCA2 under hypoxia is correlated with ROS generation, emphasizing its role in cancer cell survival. Increased SPCA2 in cells grown at high density and under hypoxia points to its role in cell cycle progression and tumor growth (Jenkins et al., 2016).

Mn<sup>2+</sup> is a cofactor of a variety of enzymes such as oxidases, kinases, N-glycosylases, DNA/RNA-polymerases, and the mitochondrial superoxide dismutase (MnSOD) (Culotta et al., 2005). Mn<sup>2+</sup> displays antioxidant effects, can scavenge free radicals, and regulate reactive oxygen species (ROS) generation under hypoxia (Coassin et al., 1992). Recent reports also linked Mn<sup>2+</sup> homeostasis with the control and progression of the cell cycle: exposure to MnCl<sub>2</sub> led to the arrest of A549 cells in G<sub>0</sub>/G<sub>1</sub> phases (P. Zhao et al., 2008). Notably, MnSOD can induce cell cycle arrest and display anti-proliferative function (Bernard et al., 2001).

Grice and col. demonstrate that inhibition of SPCA1 in MDA-MB-231 results in pronounced changes in cell proliferation and morphology in three-dimensional culture, without alterations or changes in global calcium signaling or cytoplasmic Ca<sup>2+</sup> levels. Instead, the effects of SPCA1 inhibition reside in altered regulation of calcium-dependent enzymes located in the secretory pathway, such as pro-protein convertases. Inhibition of SPCA1 produced an alteration in the processing of insulin-like growth factor receptor (IGF1R), which significantly reduced the level of functional IGF1R $\beta$  and accumulation of the inactive trans-Golgi network pro-IGF1R form. This observation reveals that some calcium transporters can regulate the processing of proteins important in tumor progression without major alterations in cytosolic calcium signaling, suggesting inhibition of SPCA1 as an alternative strategy to direct inhibitors of IGF1R and reduced the processing of other proprotein convertase substrates important in basal breast cancers or other cancer types (Grice et al., 2010).

Feng and col. identified up-regulation of SPCA2 in breast cancer-derived cells and human breast tumors. Knockdown of SPCA2 resulted in reduced growth as well as decreased colony formation of MCF7 cells in soft agar and attenuated tumor formation in xenografted mice. Overexpression of SPCA2 may confer increased proliferation and colony formation capacity in soft agar assay in MCF10A cells, a nonmalignant mammary epithelial cell line (Feng et al., 2010).

Knock-down of SPCA2 and low Ca<sup>2+</sup> conditions can reduce the ERK1/2 pathway activity, which may result in decreased proliferation in breast cancer cells. SPCA2 appears to be a constitutive Ca<sup>2+</sup> entry pathway, which in turn promotes the proliferative potential of cancer cells (Feng et al., 2010).

**Table 4.2 Ca<sup>2+</sup> pumps and associated cancer**

Ca <sup>2+</sup> ATPase pump	Cancer type	Change
<b>SERCA2</b>	Colon	Increased
<b>SERCA3</b>	Gastric, lung, choroid plexus tumors, and myeloid leukemia	Decrease
<b>SPCA1</b>	Breast cancer	Increased
<b>SPCA2</b>	Breast cancer	Increased
<b>PMCA1</b>	Oral cancer	Decreased
<b>PMCA2</b>	Breast cancer	mRNA elevated
<b>PMCA4</b>	Colon cancer	Decreased

PMCA2, the isoform predominantly expressed in mammary epithelia during lactation, is highly expressed in certain numbers of breast cancer cell lines (Lee et al., 2005). In breast cancer cell lines, PMCA2 is expressed 100-fold more than non-tumorigenic lines. Therefore, PMCA2 can keep low cytosolic Ca<sup>2+</sup> levels and avoid apoptosis by preventing increased uptake of Ca<sup>2+</sup> into mitochondria.

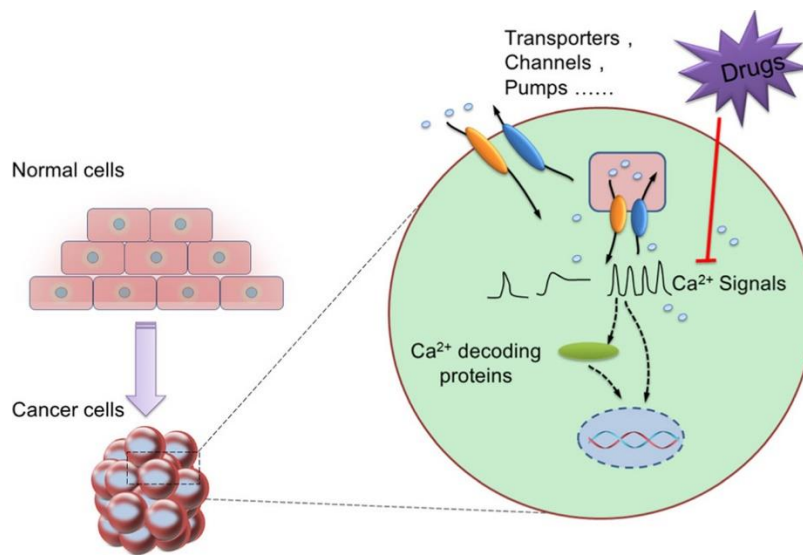
PMCA4 and PMCA1 are down-regulated in the colon or oral squamous cell carcinoma, respectively, which increase cytosolic Ca<sup>2+</sup> to enhance cell proliferation (Aung et al., 2007; Saito et al., 2006). These observations suggest that different cancer cells may develop different ways to satisfy the needs for intracellular Ca<sup>2+</sup> signaling; meaning that either up-regulation or down-regulation of Ca<sup>2+</sup> ATPases is used to promote a particular type of cancer and to escape from normal cellular control leading to carcinogenesis (Cui et al., 2017).

#### 4.4 Ca<sup>2+</sup> pumps inhibitors

The sustained high cytoplasm Ca<sup>2+</sup> levels are toxic for cells by activating cell death signaling (McConkey & Orrenius, 1997). Ca<sup>2+</sup> ATPases can be easily targeted by shutting down these pumps to generate such toxic cytosolic Ca<sup>2+</sup> concentrations for either apoptosis or necrosis. A PMCA selective inhibitor [Pt (O, O´ -acac) (γ-acac) (DMS)] is used to rapidly induce apoptosis in MCF-7 cells (Muscella et al., 2011).

Thapsigargin (TG) is a selective inhibitor of the SERCA pump used to inhibit Ca<sup>2+</sup> uptake into ER and deplete ER Ca<sup>2+</sup> stores. The application of TG as a chemotherapeutic agent has been proposed in prostate and other cancers. However, the main problem of TG as a therapeutic strategy is its non-selectivity (Denmeade et al., 2012).

Research of TG as a chemotherapy drug has been focused on tumor targeting. G202 has been developed as an analog to TG conjugated to prostate-specific membrane antigen (PSMA) targeting peptide which is a type II membrane carboxypeptidase and is overexpressed in prostate cancer cells and most tumor endothelial cells, but not in normal tissue epithelium. Once it reaches the tumor, it binds with PSMA and subsequently, PSMA can cleave the peptide and release an active cytotoxic analog of TG. G202 later termed mipsagargin, inhibits tumor progression in prostate, breast, and bladder cancers, while minimizing the toxicity effects on the host animals. G202 is currently in phase II clinical trial for prostate cancer and progressive glioblastoma (Denmeade et al., 2012).



**Figure 4.11** Model for drug targeting Ca<sup>2+</sup> signaling in cancer. Adapted from (Cui et al., 2017)

# RESULTS PART I

# Chapter 5 Deregulation of alternative promoter usage in T-acute lymphoblastic leukemia

## 5.1 Objective

To assess the implications of alternative promoter deregulation in T-ALL. Specifically of *ATP2C1*, which contains an AP found to be significantly deregulated in T-ALL with the pipeline developed in our laboratory.

## 5.2 Contributions

Bioinformatic contribution.

The bioinformatic pipeline used to identify the genes with AP deregulated in T-ALL was developed by Dr. Quentin Ferré, a former Ph.D. of our lab in collaboration with Dr. Denis Puthier.

Experimental contribution

To carry out this project, my supervisor and I routinely discussed and conceptualized the experimental designs for the validation of the best candidate gene (*ATP2C1*) found by the bioinformatics pipeline.

I performed the analysis of RNA-seq data and Chip-seq data for *ATP2C1*, to identify the samples with differentially *ATP2C1* AP usage. I performed the search and collection of data for the *in-silico* analysis (GSEA and GO terms of TF). I performed all experimental works except for the presented in Figure 6 D-E, the epigenetic inactivation of P1 by CRISPRi, which was performed by Iris Manosalva (biologic engineer working in our laboratory)

Manuscript contribution: I contributed to the writing of the manuscript and editing figures

## 5.3 Results Part I and manuscript

### An alternative promoter of *ATP2C1* is ectopically activated in a subset of T-acute lymphoblastic Leukemia with an activated T-cell phenotype

**Authors:** José David Abad Flores<sup>1-2</sup>, Quentin Ferré<sup>1-2</sup>, Iris Manosalva<sup>1-2</sup>, Agata Cieslak<sup>3</sup>, Guillaume Charbonnier<sup>1-2</sup>, Denis Puthier<sup>1-2</sup>, Vahid Asnafi<sup>3</sup>, Salvatore Spicuglia<sup>1-2</sup>

#### Affiliations

<sup>1</sup>Aix-Marseille University, Inserm, Theories and Approaches of Genomic Complexity (TAGC), UMR1090, 13288 Marseille, France

<sup>2</sup>Equipe Labellisée Ligue Contre le Cancer

<sup>3</sup>Université de Paris (Descartes), Institut Necker-Enfants Malades (INEM), Institut national de la santé et de la recherche médicale (Inserm) U1151, and Laboratory of Onco-Hematology, Assistance Publique-Hôpitaux de Paris, Hôpital Necker Enfants-Malades, 75015 Paris, France

#### Summary

An important dimension of genome complexity is the use of alternative promoters to drive pervasive gene regulation in a cell type-specific manner and during human development. Alternative promoters are frequently deregulated in disease, including cancer, thus promoter choice might be among the unknown driving forces behind the oncogenic transcriptional changes. Here we assessed the relevance of alternative promoter deregulation in T-cell acute lymphoblastic leukemia (T-ALL). Integrative analysis of epigenomic and transcriptomic data in normal T-cell precursors and primary T-ALLs lead to the identification of an alternative promoter of *ATP2C1* as frequently up-regulated in T-ALL patients. We found that T-ALL's specific promoter usage of *ATP2C1* is linked to an activated T-cell signaling phenotype. Analyses of isoform-specific gene expression and reporter assays demonstrated that the alternative *ATP2C1* promoter intrinsically responds to T-cell activation. CRISPR-mediated deletion or repression of the alternative promoter resulted in the lack of activation of *ATP2C1*. We suggest that epigenetic deregulation of the *ATP2C1* alternative promoter might contribute to cell survival of a subset of T-ALL with an activated phenotype.



## Results

### Identification of alternative promoter usage in T-ALL

To identify alternative promoter usage in T-ALL, we compared the H3K4me3 signal between a series of 11 primary T-ALL and the 5 main thymic subpopulations, including immature double negative CD34<sup>+</sup> (tCD34: CD34<sup>+</sup>/CD3<sup>-</sup>/CD4<sup>-</sup>/CD8<sup>-</sup>), early cortical (EC: TCR $\alpha\beta$ <sup>-</sup>/CD3<sup>-</sup>/CD4<sup>+</sup>/CD8<sup>+</sup>), late cortical (LC: TCR $\alpha\beta$ <sup>+</sup>/CD3<sup>low</sup>/CD4<sup>+</sup>/CD8<sup>+</sup>), single positive CD4 (SP4: TCR $\alpha\beta$ <sup>+</sup>/CD3<sup>+</sup>/CD4<sup>+</sup>/CD8<sup>-</sup>), single positive CD8 (SP8: TCR $\alpha\beta$ <sup>+</sup>/CD3<sup>+</sup>/CD4<sup>-</sup>/CD8<sup>+</sup>), along with 11 primary T-ALL samples from treatment-naive individuals, using previously published datasets (Cieslak et al.; Belhocine et al.; **Supplemental Table 1**).

We developed a bioinformatic pipeline to identify alternative promoters that are more frequently active in T-ALL as compared to T cell precursors (**Fig. 1A**; see Materials and Methods for details). As promoter activity can lead to multiple related Transcription Start sites (TSSs) {Lenhard, 2012 #5902}, we first combined the TSS that overlapped the same H3K4me3 peak based on ENSEMBL transcripts belonging to the same gene to define single promoter regions. We next selected 1970 genes harboring at least 2 active promoters and for which an H3K4me3 peak overlap in at least one sample. The H3K4me3 coverage was then quantified at each promoter. We designed a statistical procedure to identify genes with alternative promoters preferentially active in T-ALL. We identified 8 genes with an AP significantly active in T-ALL, but not in normal thymocytes (**Supplemental Table 2**).

To assess the relevance of AP usage for these 8 genes, we compared the expression of the different isoforms between a large series of T-ALL and the normal thymic populations using a previously published RNA-seq dataset (Bond et al. 2017; Cieslak et al. 2020) (**Supplemental Table 2**). We found that only the *ATP2C1* gene harbor AP-associated transcripts that were significantly upregulated in T-ALL. *ATP2C1* harbor 3 transcript isoforms associated with the most upstream promoter (hereafter P1) and 12 transcript isoforms associated with the internal promoter (hereafter P2)(**Supplemental Figure 1A**). P1-associated transcripts are predicted to

encode for a protein harboring 36 additional AA from the N-Terminal part (**Supplemental Figure 1A**).

As shown in figure 1B, the P2 promoter appeared to be active (high H3K4me3 signal) in all thymocyte populations and T-ALL samples. In contrast, the P1 promoter appeared to be active in more than half of the T-ALL samples (7 out of 11), while no signal was detected for the majority of normal thymocyte subpopulations, except for the most immature CD34+ thymocytes for which a weak signal of H3K4me3 was observed. Global analysis of *ATP2C1* isoform expression based on RNA-seq data revealed that all thymic populations and roughly half of the T-ALL samples (23 out of 41) express only the P2-associated transcripts, while the other half of the T-ALL samples express both the P1- and P2-associated transcripts (**Fig. 1C-D**). We concluded that the P2-associated isoforms are similarly expressed in normal and leukemic samples while the P1-associated isoforms are specifically up-regulated in a subset of T-ALL samples.

To obtain a more complete view of the epigenomic landscape of the *ATP2C1* locus, we analyzed ChIP-seq data of 6 histone modifications from each thymic subpopulation and 5 T-ALL samples generated within the BLUEPRINT consortia (Cieslak et al., 2020; Martens & Stunnenberg, 2013). ChIP-seq data were integrated into 11 chromatin states using the ChromHMM tool (**Supplemental Figure 1B**). Visual inspection of the chromatin states associated with the P1 and P2 promoters revealed a distinct epigenetic dynamic at both promoters. As expected, the P2 promoter was associated with the active promoter and TSS chromatin states. The P1 promoter was associated with the active promoter and TSS chromatin states in the T-ALL samples, but not in the normal T cell precursors. However, in the normal T cell precursors, the P1 promoter was not associated with repressive chromatin marks associated with Polycomb (H3K27me3) or heterochromatin (H3K9me3). This suggested that P1 is not actively repressed in the thymocyte populations.

*ATP2C1* gene encodes for the secretory pathway  $\text{Ca}^{2+}$  ATPase type I pump (also known as SPCA1). The *ATP2C1* pump is located on the membrane of the Golgi apparatus (GA) where it transports  $\text{Ca}^{2+}$  and  $\text{Mn}^{2+}$  ions from the cytosol into the GA, thus contributing to the secretory pathway (Brini et al., 2012). *ATP2C1* expression has been involved in oxidative stress, cell cycle regulation, and cancer cell survival

(Baron et al., 2010; Dang D et al., 2016; Jenkins et al., 2016). We hypothesized that alternative promoter usage might contribute to altering the function and regulation of *ATP2C1* in T-ALL, therefore we explored the potential mechanisms leading to the ectopic activation of *ATP2C1* P1 promoter.

### **Alternative promoter usage of *ATP2C1* is associated with the “immune state” of the T-ALL samples**

To gain insights into the potential mechanisms of AP usage of *ATP2C1* in T-ALL, we analyzed the functional enrichment associated with a set of 120 T-ALL samples expressing high or low levels of P1-associated transcripts. To this goal, we computed the proportion of P1- over P2- associated transcripts and selected the first (P1-high) and fourth (P1-low) quartiles of the ranked samples (**Supplemental Figure 2B**). We then performed a comparative Gene Set Enrichment Analyses (GSEA) with these two groups (**Fig. 2A-B and Supplemental Figures 3A-B**). We observed that T-ALL samples with a high level of P1 transcripts were associated with an active inflammatory response (Including type I and II interferon, IL2-STAT5, and TNFa signaling) as well as hypoxia and protein maturation and secretion, while T-ALL samples with a low level of P1 transcripts were associated with Hedgehog signaling and cell cycle regulation (including E2F targets and cell cycle checkpoint). Consistent results were obtained by analyzing an independent RNA-seq dataset of 31 T-ALL samples (Kalender Atak et al., 2013) (**Supplemental figures 4A-B**).

To gain insight into the potential regulation of P1 and P2 promoters we investigated the binding of transcription factors (TF) using ENCODE data (Sloan et al., 2016; **Supplemental Table 3**). Consistent with the GSEA analyses, we found that the P1 promoter was preferentially bound by TFs associated with the inflammatory and immune response (**Fig. 2C-D**), such as STAT, ATF/JUN, and NFATC family of TFs but also with TFs with known oncogenic roles in T-ALL, such as TAL1, MYC, and MYB (Kimura & Mullighan, 2020). On the other hand, the P2 promoter was bound by transcription factors associated with cell cycle regulation, such as E2F6 and PHF8, also consistent with the GSEA analyses.

Taken together, these results suggest that the *ATP2C1* P1 promoter might be ectopically activated concerning the inflammatory or immune state of the T-ALL

samples or in response to the hypoxic conditions. In contrast, the *ATP2C1* P2 transcripts might be naturally regulated through the cell cycle in both normal and leukemic conditions.

### ***ATP2C1* P1 is an inducible promoter**

To gain insight into the regulation of *ATP2C1* promoter usage we analyzed the H3K4me3 and H3K27ac profiles of *ATP2C1* promoters in seven T-ALL cell lines (**Fig. 3A**), using available datasets (Supplemental Table 1). As expected, all cell lines displayed high levels of H3K4me3 and H3K27ac around the P2 promoter, consistent with the ubiquitous activity of this promoter. In contrast, only three T-ALL cell lines (CCRF-CEM, LOUCY, and RPMI-8402) displayed a high level of H3K4me3 and H3K27ac at the P1 promoter, consistent with the activation of this promoter in a subset of T-ALL only. In addition, the Jurkat cell line displayed moderated levels of H3K4me3 and H3K27ac at the P1 promoter. We then designed primer sets to assess the expression of all *ATP2C1* transcripts (hereafter, total *ATP2C1* transcripts), as well as, P1- and P2-associated transcripts (**Supplemental Figure 5A**). We analyzed the relative expression of *ATP2C1* transcripts in three T-ALL cell lines with high (CCRF-CEM), intermediate (Jurkat), or null (SIL-ALL) levels of H3K4me3 at the P1 promoter. The three cell lines expressed similar levels of total *ATP2C1* and P2-associated transcripts. However, P1-associated transcripts were detected at the highest level in the CCRF-CEM cell line, while the Jurkat cell line displayed a low level (8 fold less than CCRF-CEM) and no signal was detected in the SIL-ALL cell line (**Fig. 3B**), consistent with the profile of histone modifications.

To gain insight into the potential mechanisms of P1 deregulation in T-ALL, we searched for available ChIP-seq from TFs performed in any of the studied T-ALL cell lines, using the ReMap database (Gheorghe et al., 2019) (**Fig. 3C-D**). Consistent with the more general analyses using ENCODE data, we found that the TAL1 oncogene was bound to the P1 promoter in the cell lines with an active P1. In addition, we observed that several lymphoid-specific factors playing major roles in T-cell differentiation, including TCF12(HEB), RUNX1, TCF3(E2A), and GATA3 ((Naito et al., 2011) were specifically bound to the P1 promoter in the P1-active cell lines.

The functional gene enrichments associated with T-ALLs expressing the P1 transcripts, as well as the nature of TFs associated with the P1 promoter, suggested that the P1 promoter might be induced by stress conditions, such as hypoxia, as well as T cell activation signaling. Thus, we assessed the relative expression of *ATP2C1* isoforms after 7 days of culture in hypoxia conditions (1% O<sub>2</sub>; **Fig. 4A and Supplemental Figure 5B**), after 4 hours of PMA/Ionomycin treatment (**Fig. 4B**), or after high-density culture (**Supplemental Figure 5C** “). The total *ATP2C1*-transcripts were significantly induced by hypoxia and PMA/Ionomycin stimulation, but not high-density culture, in all three cell lines (CCRF-CEM, Jurkat and Sil-All displayed a 2.7-, 2.6- and 1.9-fold increase, respectively, in hypoxic conditions and a 3.3-, 2.6- and 1.8- fold increase, respectively, after PMA/Ionomycin stimulation). However, P1-associated transcripts were more induced than total *ATP2C1*-transcripts (CCRF-CEM, Jurkat and Sil-All displayed a 2.8-, 6.1-, 4.1- fold increase, respectively, in hypoxic conditions and a 9.4-, 6.3- and 79.1- fold increase, respectively, after PMA/Ionomycin stimulation), while the P2-associated transcripts were either not or weakly induced. Note that P1-transcripts induction was also observed in the SIL-ALL cell line which did not express the P1 transcripts and did not display detectable levels of H3K4me3 nor H3K27ac at the P1 promoter, suggesting that even in cells where the P1 is fully silenced it is possible to induce the associated transcripts after cell stimulation. Overall, these results indicate that the P1 promoter is highly inducible, efficiently responding to hypoxia and T-cell stimulation.

To further validate the intrinsic activation of the P1 promoter, we cloned the P1 and P2 promoters upstream of the luciferase reporter gene (**Fig. 5A**) and assessed their promoter activity in unstimulated CCRF-CEM and Jurkat cells or after 4 hours of PMA/Ionomycin stimulation. In both cell lines, the P2 promoter has a basal promoter activity (34 and 26 folds in CCRF-CEM and Jurkat, respectively, as compared to the empty vector), while weakly induced by the PMA/Ionomycin treatment (1.8 folds in both cell lines). In contrast, the P1 promoter displayed a low basal activity (6 and 4 folds in CCRF-CEM and Jurkat, respectively, as compared to the empty vector), but was highly stimulated by the PMA/Ionomycin treatment (22 and 12 folds in CCRF-CEM and Jurkat, respectively, as compared to the unstimulated cells). We observed that the P1 promoter contains at least two binding sites (STAT3 and NFATC4) potentially associated with T cell activation signaling (**Fig. 5B**). Consistent with this

observation, concomitant mutation of STAT3 and NFATC4 sites within the P1 promoter strongly reduced the promoter activation mediated by the PMA/Ionomycin stimulation. Therefore, while P2 harbors an intrinsic promoter activity, P1 is a preferentially inducible promoter.

### **P1 is required for the induction of *ATP2C1***

To directly address the role of the P1 promoter, we engineered the CCRF-CEM cell line to delete the P1 promoter. We obtained three CCRF-CEM clones harboring a 2.4 kb homozygous deletion including the P1 promoter (**Fig. 6A**). The P1 deletion in CCRF-CEM cells resulted in undetectable levels of P1-associated transcripts, as expected, as well as a global reduction of 3.5 to 5.8 of total *ATP2C1* gene expression (**Fig. 6B**), suggesting that, in this cell line, P1 activity contribute to around half of the *ATP2C1* expression. Importantly, P1 deletion did not consistently impact P2-associated transcripts, indicating that P1 does not regulate or interfere with P2 activity. Next, we analyzed the response of *ATP2C1* transcripts to the PMA/Ionomycin stimulation. Strikingly, *ATP2C1* was no longer induced in the absence of the P1 promoter (**Fig. 6C**).

To assess whether the lack of *ATP2C1* induction was not due to an artifact of the clonal selection process, we performed CRISPRi-mediated silencing of the P1 promoter using at CCRF-CEM cell line constitutively expressing the dCas9-KRAP-MeCP2 double repressor complex (Adamson et al., 2016). Lentiviral expression of two gRNAs located at different positions within P1 (**Fig. 6D**) resulted in consistent repression of P1-transcripts, while the P2- transcripts were not affected (**Fig. 6E**). Thus, consistent with the P1 genetic deletion, epigenetic silencing of P1 resulted in a strong impairment of *ATP2C1* induction (**Fig. 6F**). Taken together, we concluded that P1 is a PMA/Ionomycin responsive element and absolutely required for the induction of the *ATP2C1* gene.

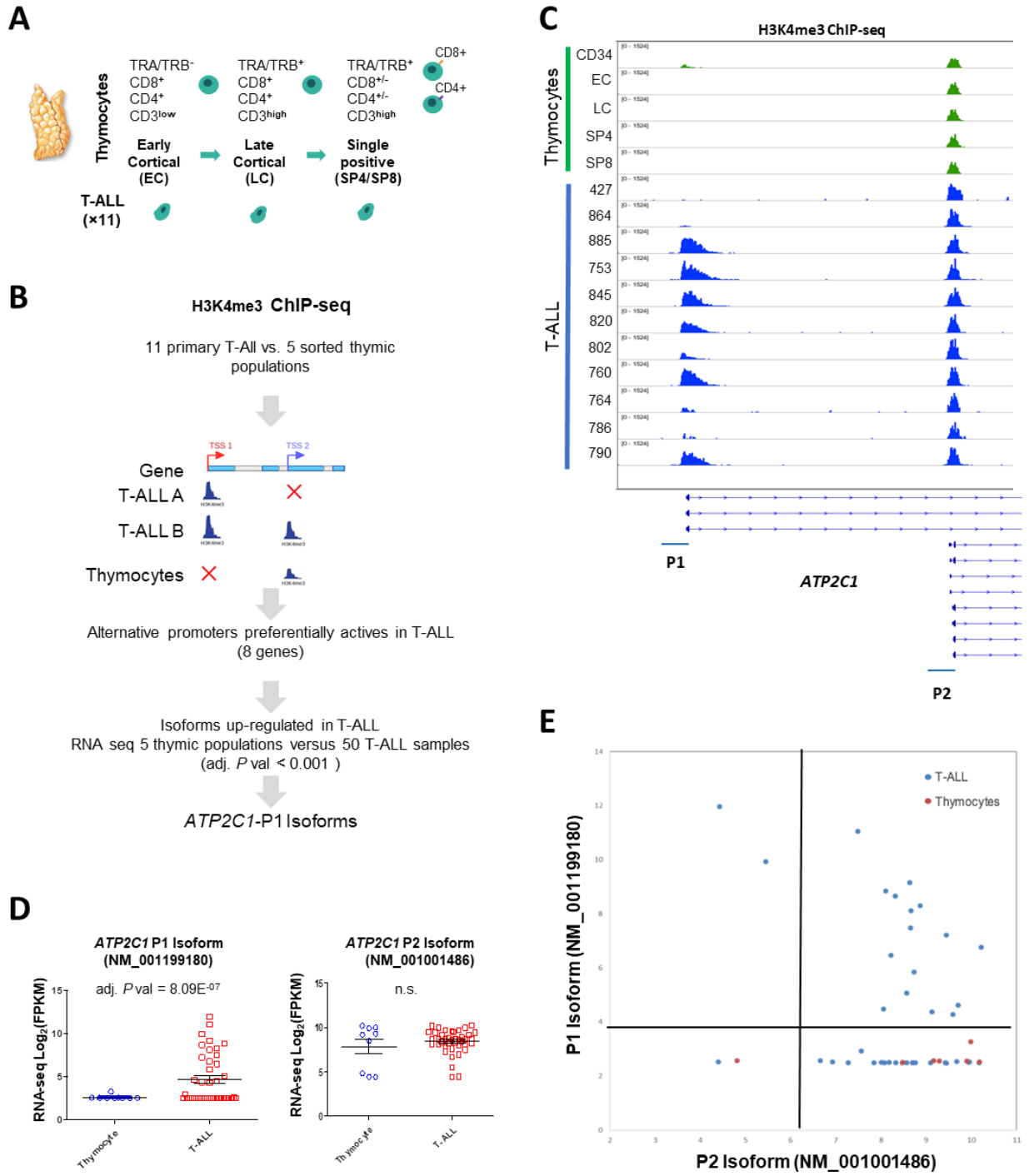
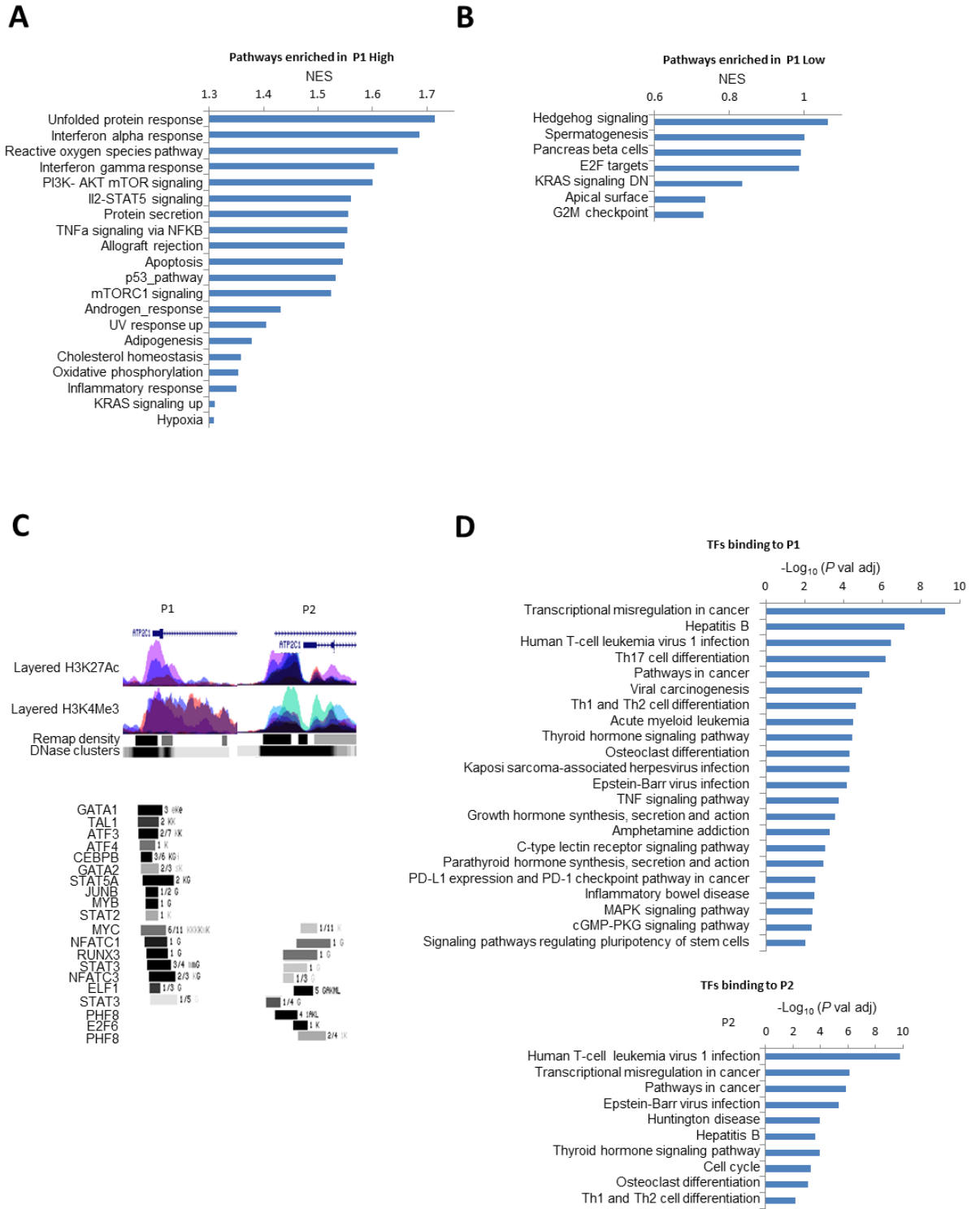
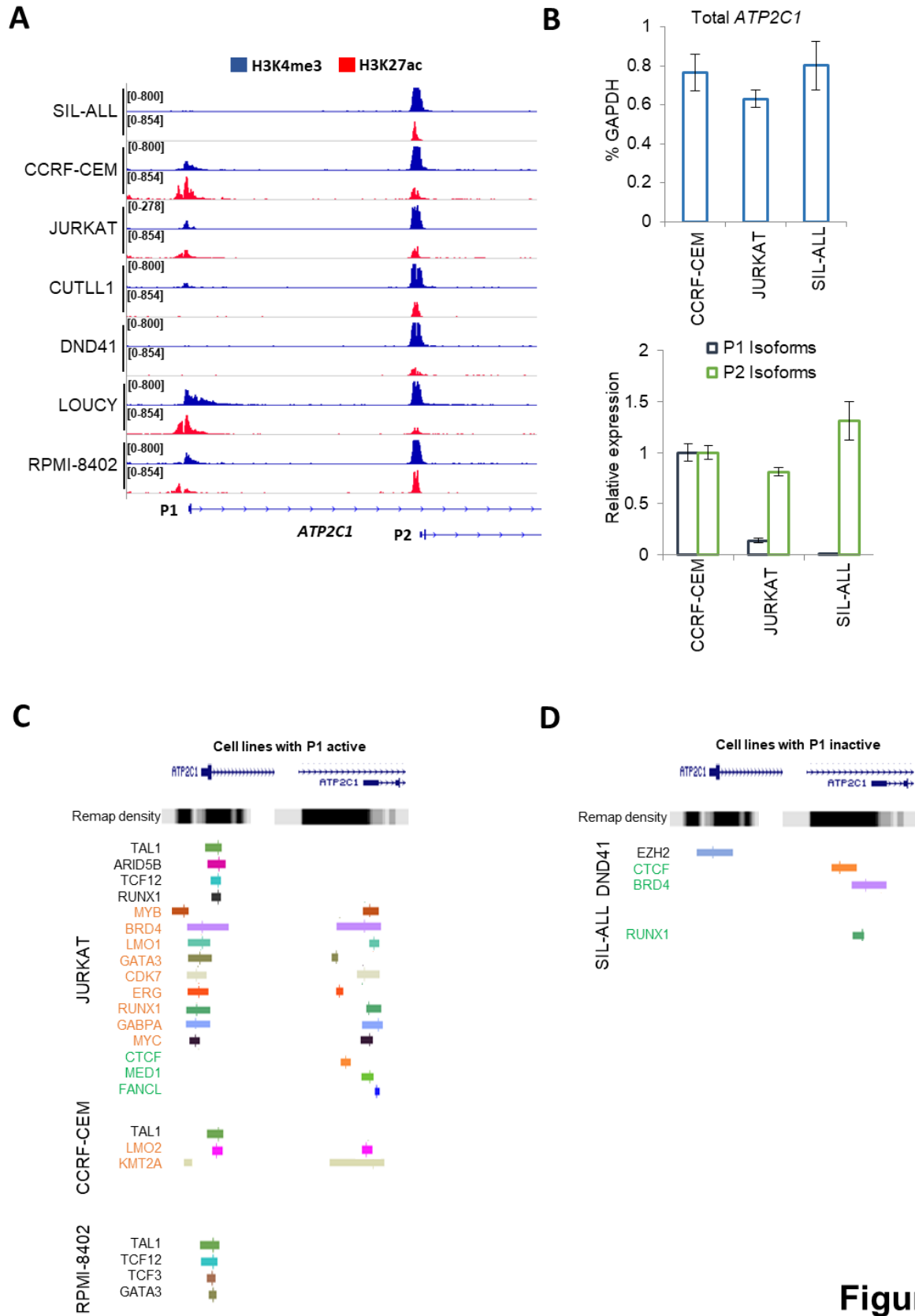


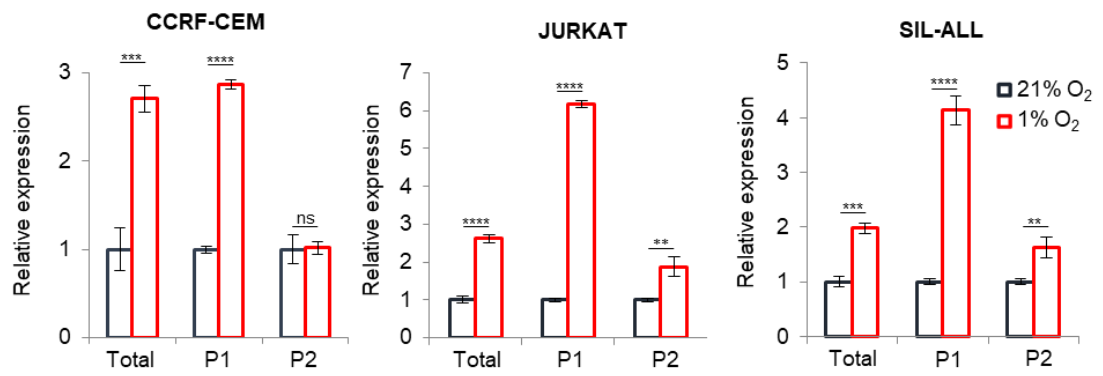
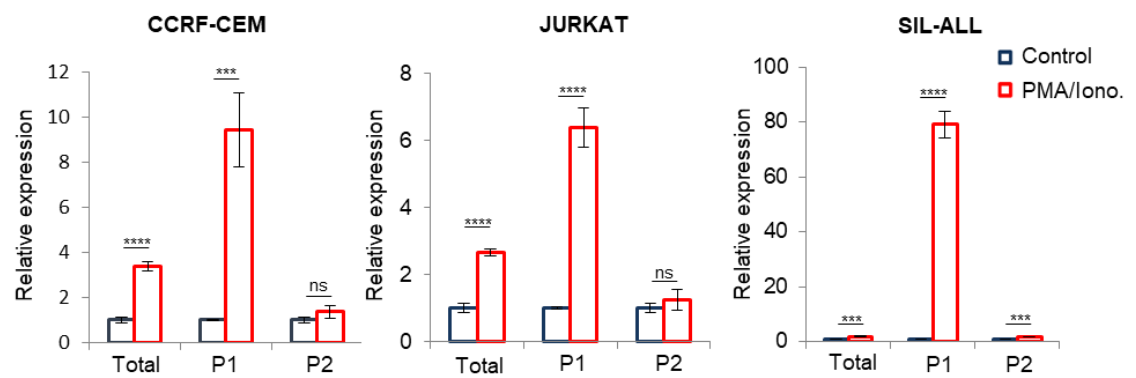
Figure 1

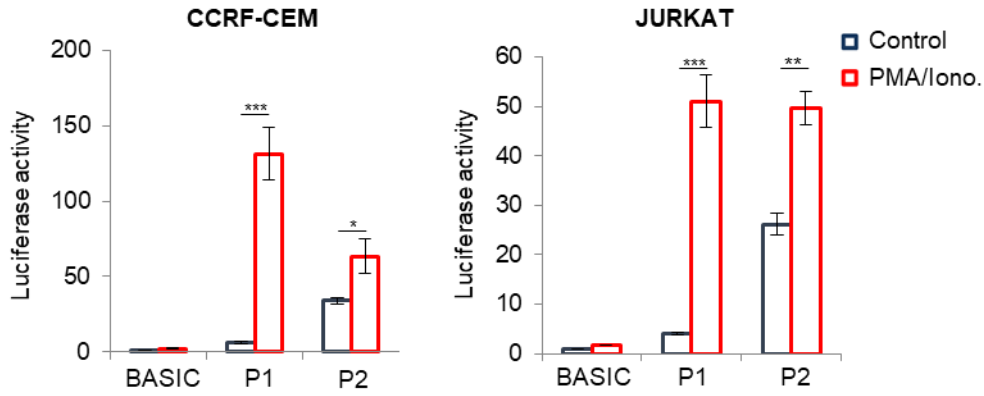
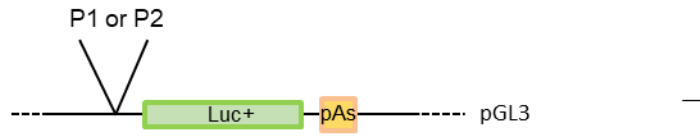
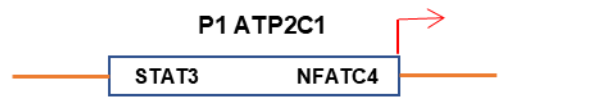


**Figure. 2**





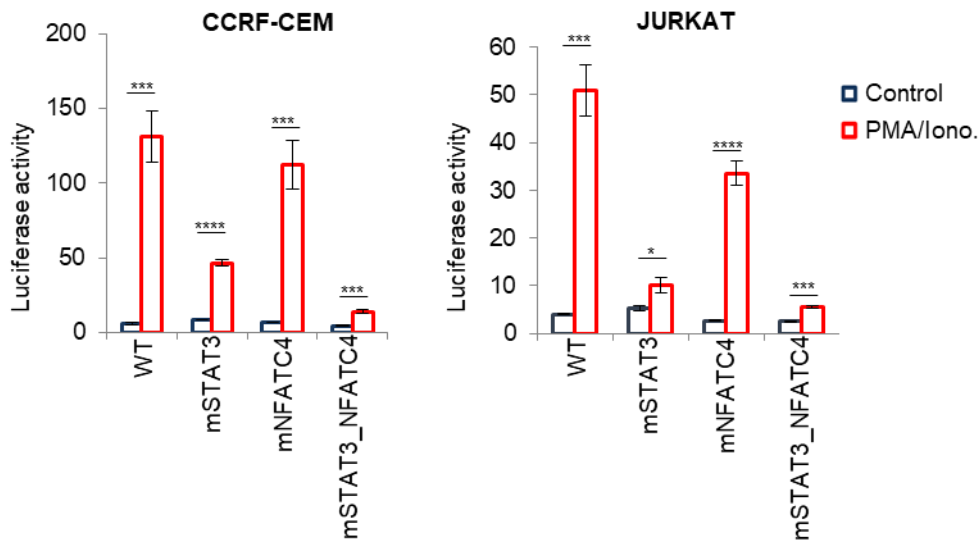
**A****B****Figure 4**

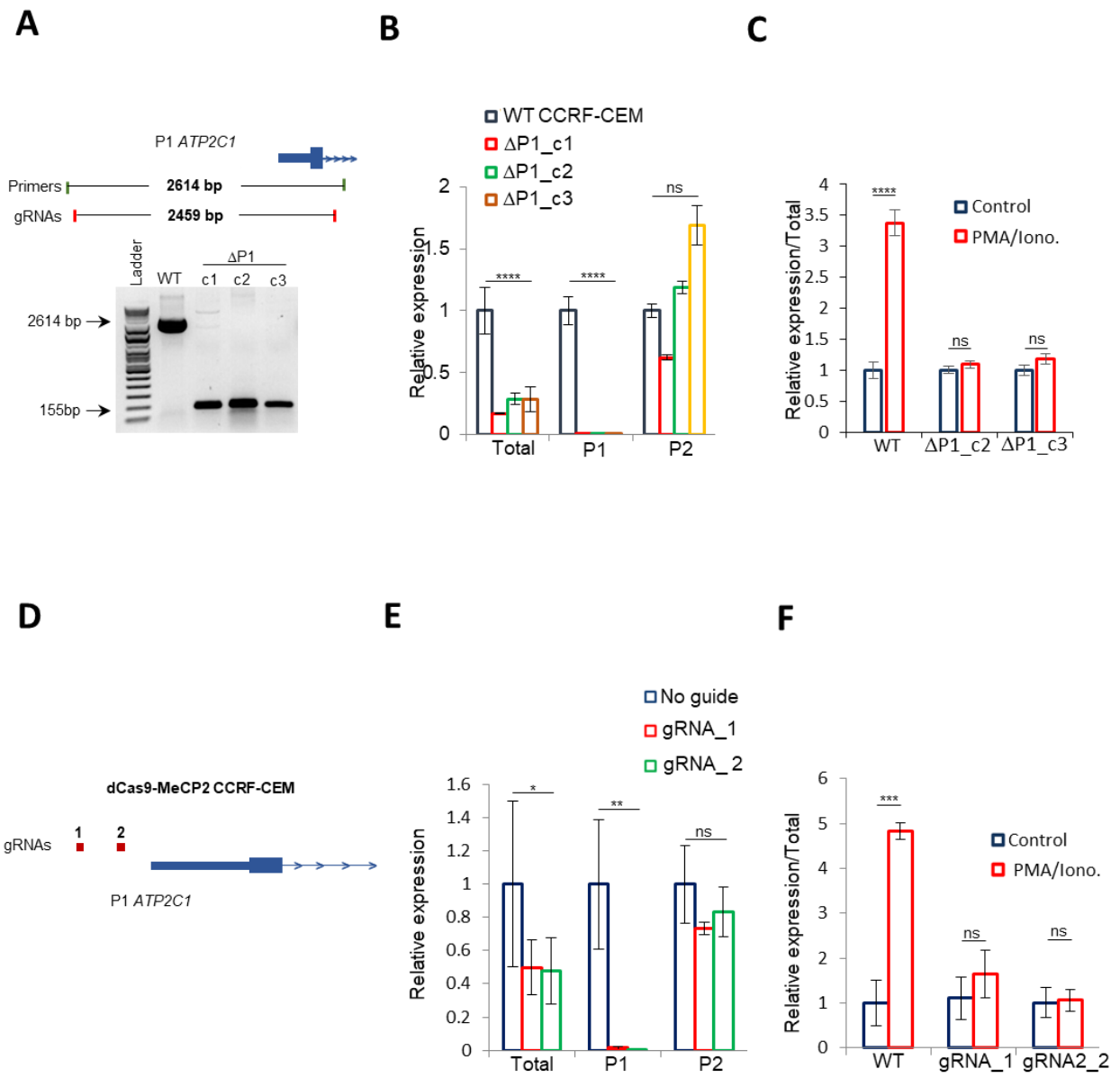
**A****B****STAT Motif**

ATTTTTTTTGGAAAGTC  
 ATTTTTTTTGTCCGTC

**NFATC4 Motif**

TATGGAAGT  
 TATAACAAGT

**Figure 5**



**Figure. 6**

## Figures Legends

### Figure 1. An alternative promoter of *ATP2C1* is frequently deregulated in T-ALL

A. Schematic representation of the major stages of human thymopoiesis and T-ALL blasts.

B. Experimental approach for the identification of alternative promoter usage in T-ALL.

C. The genomic tracks show H3K4me3 ChIP-seq data from 5 normal thymic subpopulations and 11 T-ALL samples. The *ATP2C1* transcripts are visualized. The P1 and P2 promoters are indicated. The P2 promoter is active in all the samples (thymocytes and T-ALL) as indicated by the presence of the H3K4me3 peak in the P2 region. The P1 promoter is only active in a subset of the T-ALL samples.

D and E. Dot (D) and scatter (E) plots representing the relative expression of P1- and P2-associated isoforms (NM\_001199180 and NM\_001001486, respectively) based on RNA-seq data from thymic subpopulations and 41 T-ALL samples. Statistical significance was assessed by DE-seq.

### Figure 2. Expression of *ATP2C1* P1 isoforms and TFs binding to P1 is associated with inflammatory pathways response and misregulation in cancer

A and B. GSEA analyses of gene expression profiles based on RNA-seq from T-ALL samples expressing high or low levels of P1-isoforms. (A) Representative gene sets enriched in the high P1 isoforms expression subgroup and (B) representative gene sets enriched in the low P1 isoforms expression subgroup.

C. The genomic track shows a representative subset of TFs binding to the P1 and P2 region of *ATP2C1*. The data was obtained from ENCODE database and visualized with the UCSC genome browser track.

D. Graph showing the most enriched KEGG pathways associated with TFs binding to P1 (upper panel) or P2 (lower panel) promoters. Data plotted are the adjusted *P*-value (-log<sub>10</sub> scale) of the enrichment.

### Figure 3. *ATP2C1* P1 promoter activity is T-ALL cell line dependent

A. The genomic tracks show H3K4me3 and H3K27ac ChIP-seq signal from different T-ALL cell lines around the *ATP2C1* P1 and P2 promoters, P2 is active in all cell lines as indicated by the presence of H3K4me3 and H3K27ac peaks while P1 is only active in a subset of T-ALL cell lines (CCRF-CEM, JURKAT, LOUCY, and RPMI-8402).

B. RT-qPCR analysis of the overall expression of *ATP2C1* (total) in three selected cell-lines with different P1 activity (CCRF-CEM, JURKAT, and SIL-ALL) was performed using a set of primers common to both isoforms (relative localization of the different set of primers used for the qPCR experiments is illustrated in the (SUPPFIG 5A), Values represent the percentage of total *ATP2C1* expression levels as compared to the *GAPDH* housekeeping gene (upper panel). RT-qPCR analysis of the expression of P1- and P2-associated transcripts in the three

selected cell lines (CCRF-CEM, JURKAT, and SIL-ALL) was performed using a specific set of primers. Values represent the relative expression levels as compared to the expression in the CCRF-CEM cell line and normalized by the *GAPDH* housekeeping gene (lower panel).

C and D. The genomic track show the binding of TFs to the P1 and P2 promoters in the cell lines with an active P1 promoter (C) and the cell lines with an inactive P1 promoter (D). TFs ChIP-seq data were obtained from ReMap 2020 (Chèneby et al., 2020) database and visualized with UCSC genome browser tracks.

#### **Figure 4. *ATP2C1* P1 isoforms are preferentially induced by hypoxia and PMA/Ionomycin stimulation**

A. RT-qPCR analysis of the overall expression of *ATP2C1* (total) and the expression of the P1- and P2-associated transcripts in hypoxic conditions and normal oxygen conditions in CCRF-CEM (left panel), Jurkat (middle panel), and Sil-ALL (right panel) cell lines. Values represent the relative expression levels as compared to the normal oxygen conditions and normalized by the *B2M* housekeeping gene. Significance was assessed by a two-sided Student t-test. Error bars represent the s.d. of three independent replicates. \*\*\*\*  $P \leq 0.0001$ , \*\*\*  $P \leq 0.001$ , \*\*  $P \leq 0.01$ , ns  $> 0.05$ .

B. RT-qPCR analysis of the overall expression of *ATP2C1* (total), and the expression of the of P1- and P2-associated transcripts in non-stimulated and stimulated with PMA/Ionomycin conditions in CCRF-CEM (left panel), Jurkat (middle panel) and Sil-ALL (right panel) cell lines. Values represent the relative expression levels as compared to the non-stimulated condition and normalized by the *B2M* housekeeping gene. Significance was assessed by a two-sided Student t-test. Error bars represent the s.d. of three independent replicates. \*\*\*\*  $P \leq 0.0001$ , \*\*\*  $P \leq 0.001$ , ns  $> 0.05$ .

#### **Figure 5. Luciferase assay of *ATP2C1* P1 and P2 activity in wildtype and P1 mutants of STAT3 and NFATC4 TFBS**

A. A simplified representation of the pGL3 vector containing the P1 or P2 *ATP2C1* promoters is shown in the upper panel. Promoter activity quantification by luciferase assay of P1 and P2 promoters in non-stimulated and stimulated with PMA/Ionomycin conditions is shown for CCRF-CEM (left panel) and Jurkat (right panel) cell lines. Values represent the relative luciferase activity as compared to the pGL3 basic vector and normalized by the Renilla control vector. Significance was assessed by a two-sided Student t-test. Error bars represent the s.d. of three independent replicates. \*\*\*  $P \leq 0.001$ , \*\*  $P \leq 0.01$ , \*  $P \leq 0.05$

B. Relative localization of the two TFBSs mutated in the *ATP2C1* P1 promoter (STAT3 and NFATC4) and their respective sequences taken from JASPAR 2020 (Fornes et al., 2020) are shown in the upper panel. Promoter activity quantification by luciferase assay in non-stimulated and stimulated with PMA/Ionomycin conditions for the P1 *ATP2C1* WT, the P1 single mutated (mSTAT3 or mNFATC4), and the P1 double mutated (mSTAT3 and NFATC4) is shown for CCRF-CEM (left panel) and Jurkat (right panel) cell lines. Values represent the relative luciferase activity as compared to the pGL3 basic vector and normalized by the Renilla control vector. Significance was assessed by a two-sided Student t-test. Error bars represent the s.d. of three independent replicates. \*\*\*\*  $P \leq 0.0001$ , \*\*\*  $P \leq 0.001$ , \*  $P \leq 0.05$ .

**Figure 6. The *ATP2C1* P1 knock-out by CRISPR-Cas9 and CRISPRi prevent the induction of P1 *ATP2C1* by PMA/Ionomycin stimulation**

A. Schematic representation of the gRNAs used for the deletion of P1 *ATP2C1* region by CRISPR-Cas9 technology in the CCRF-CEM cell line and the primers used for the detection of clones harboring a homozygous deletion of P1 *ATP2C1* ( $\Delta$ P1) (upper panel). Agarose gel electrophoresis of three clones (c1-c3) harboring a homozygous deletion of P1 *ATP2C1* ( $\Delta$ P1) is shown in the lower panel. PCR products were amplified from the genomic DNA of each clone with a set of primers designed for deletion detection. The expected amplicon size is 155bp for  $\Delta$ P1 and 2614 bp for the WT genomes.

B. RT-qPCR analysis of the overall expression of *ATP2C1* (total), and the expression of the P1- and P2-associated transcripts in the WT and the three  $\Delta$ P1 CCRF-CEM clones (c1-c3). Values represent the relative expression levels as compared to the WT and normalized by the *B2M* housekeeping gene. Significance was assessed by a two-sided Student t-test. Error bars represent the s.d. of three independent replicates. \*\*\*\*  $P \leq 0.0001$ , ns  $> 0.05$ .

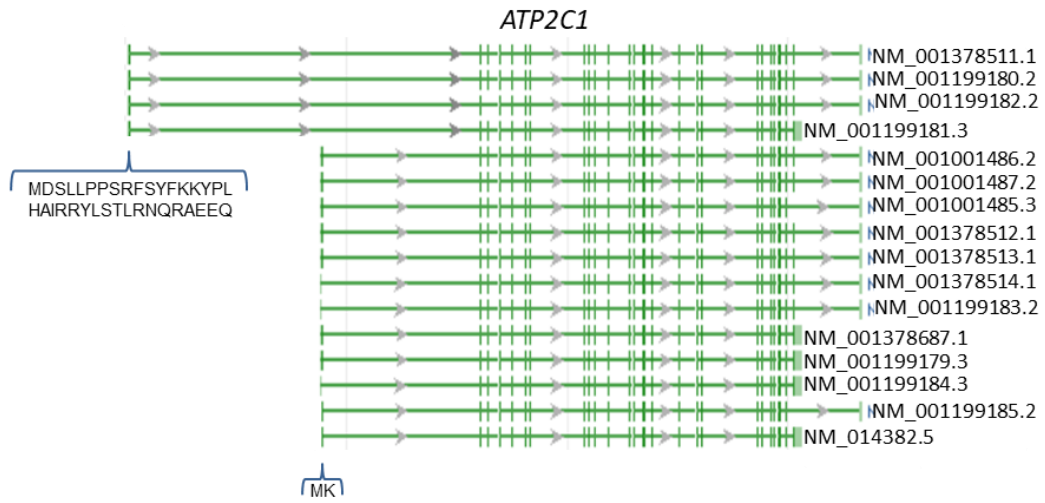
C. RT-qPCR analysis of the overall expression of *ATP2C1* (total) in non-stimulated and PMA/Ionomycin stimulated conditions in the WT and two  $\Delta$ P1 CCRF-CEM clones (c2 and c3). Values represent the relative expression levels as compared to the non-stimulated condition and normalized by the *B2M* housekeeping gene. Significance was assessed by a two-sided Student t-test. Error bars represent the s.d. of three independent replicates. \*\*\*\*  $P \leq 0.0001$ , ns  $> 0.05$ .

D. Schematic representation of the gRNAs used for the silencing of P1 *ATP2C1* by CRISPRi technology using a CCRF-CEM cell line expressing dCas9-KRAB-MeCP2.

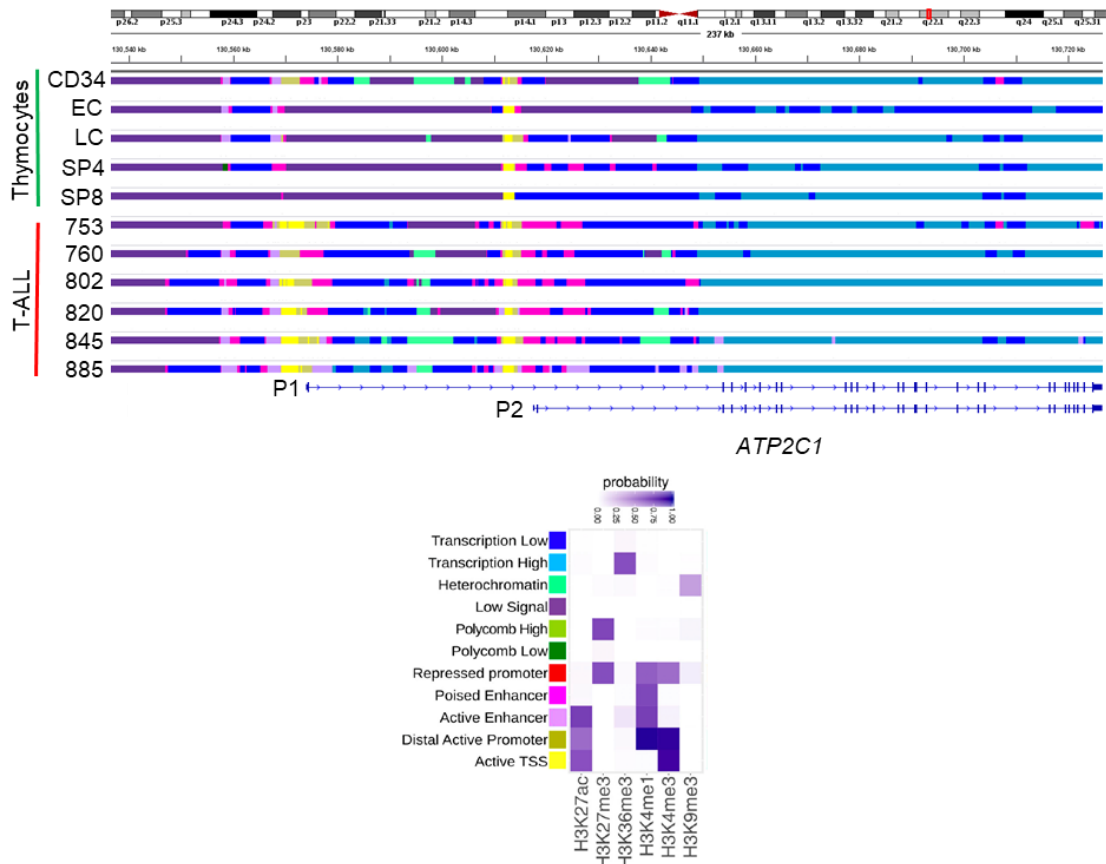
E. RT-qPCR analysis of the expression of the P1- and P2-associated transcripts in the non-infected CCRF-CEM and the CCRF-CEM expressing the P1 targeted gRNAs. Values represent the relative expression levels as compared to the non-infected CCRF-CEM cell line and normalized by the *B2M* housekeeping gene. Significance was assessed by a two-sided Student t-test. Error bars represent the s.d. of three independent replicates. \*\*  $P \leq 0.01$ , \*  $P \leq 0.05$ , ns  $> 0.05$ .

F. RT-qPCR analysis of the overall expression of *ATP2C1* (total) in non-stimulated and PMA/Ionomycin stimulated conditions in non-infected CCRF-CEM and the CCRF-CEM expressing the P1 targeted gRNAs. Values represent the relative expression levels as compared to the non-stimulated condition and normalized by the *B2M* housekeeping gene. Significance was assessed by a two-sided Student t-test. Error bars represent the s.d. of three independent replicates. \*\*\*  $P \leq 0.001$ , ns  $> 0.05$ .

**A**

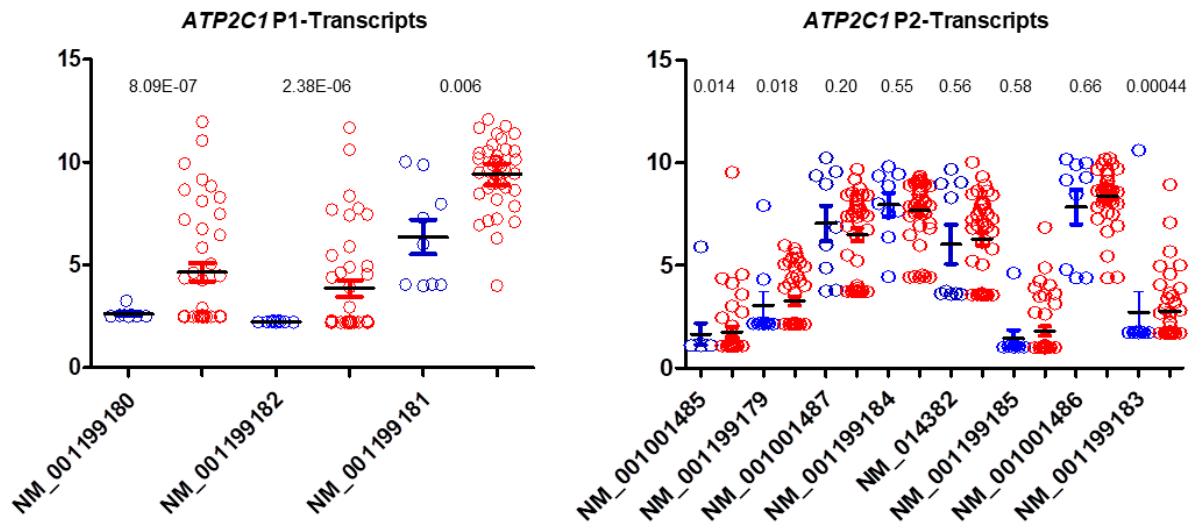
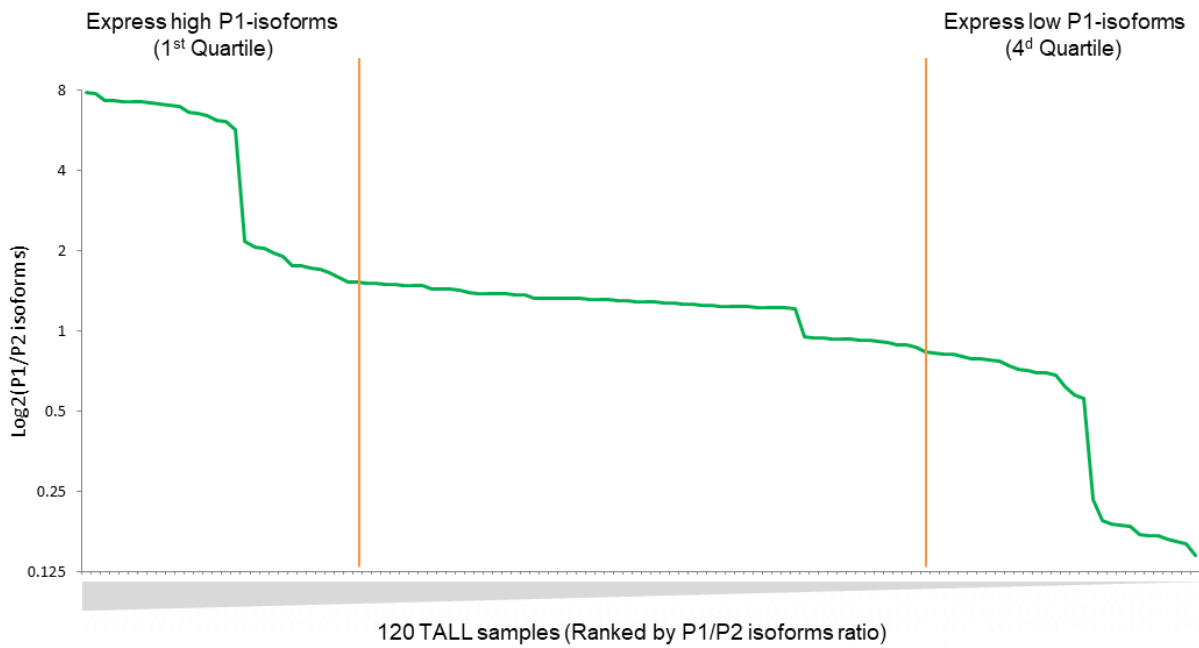


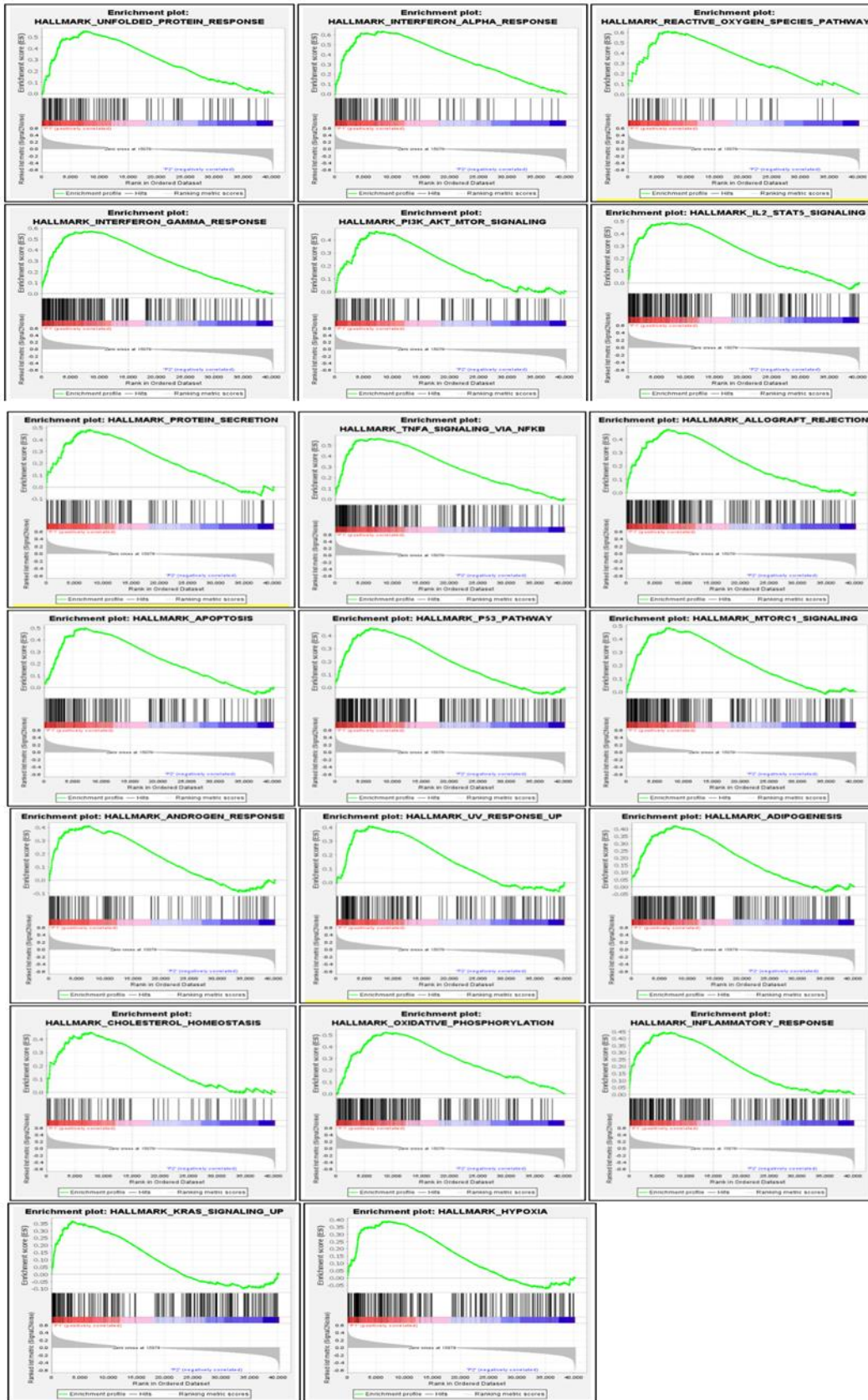
**B**



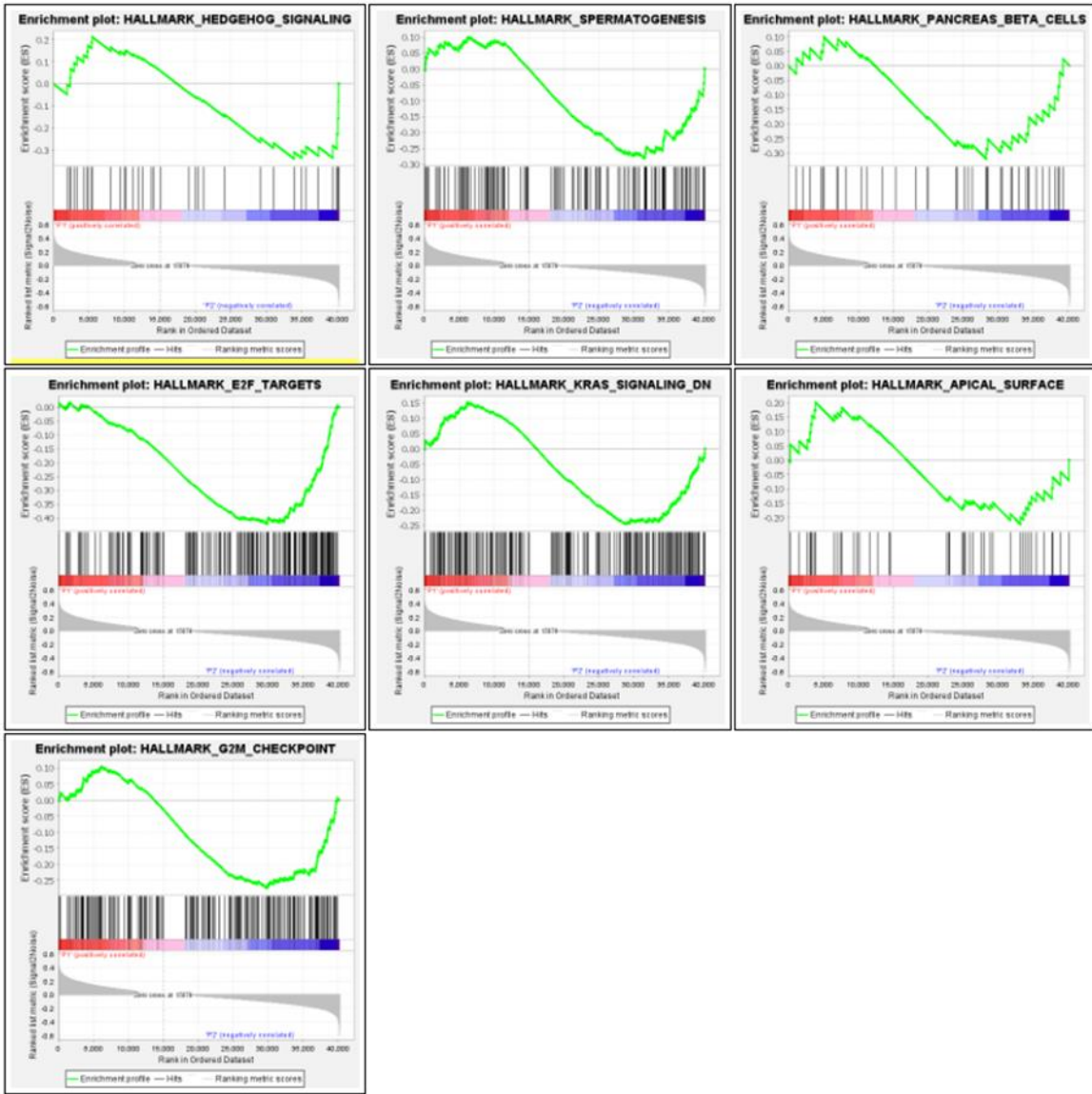
**SUPPLEMENTAL FIGURE 1**



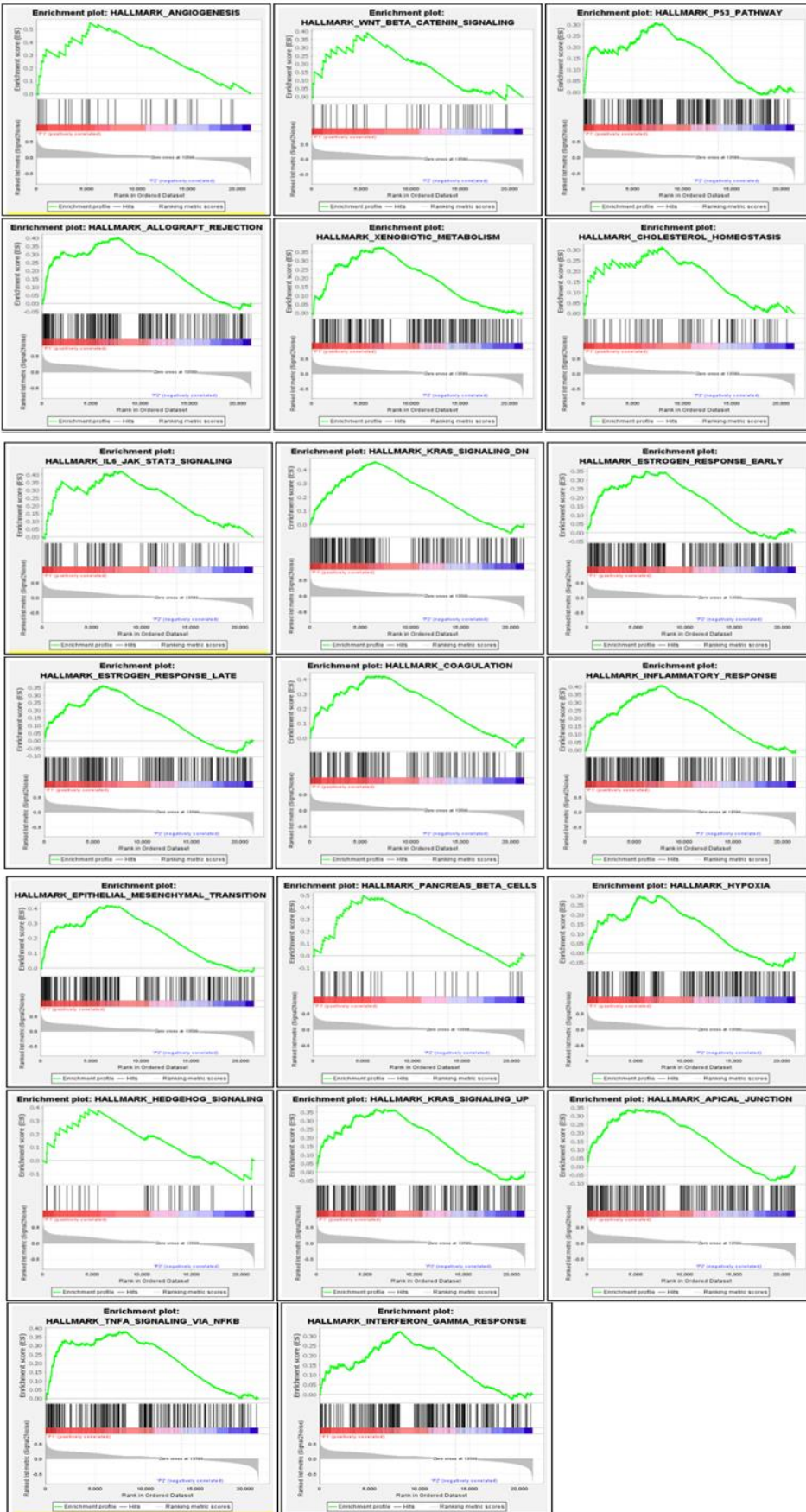
**A****B****SUPPLEMENTAL FIGURE 2**



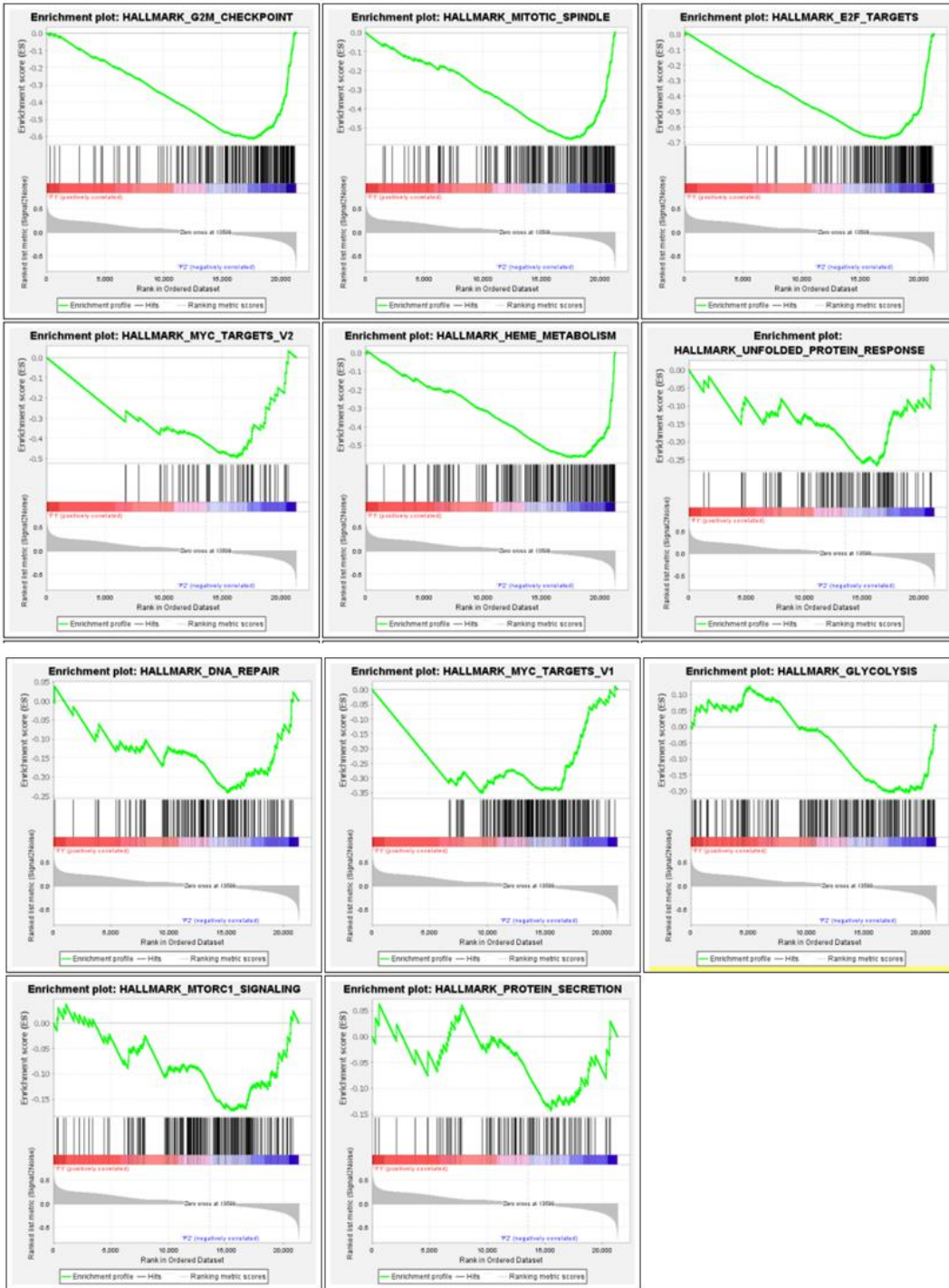
**SUPPLEMENTAL FIGURE 3A**



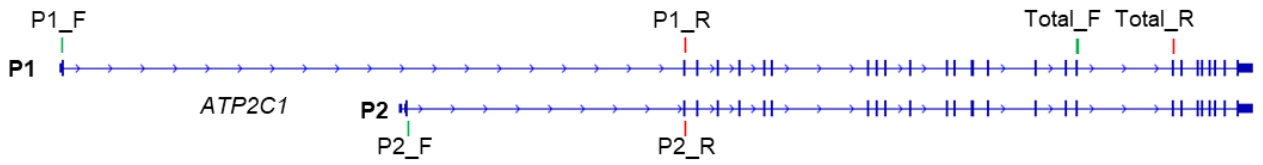
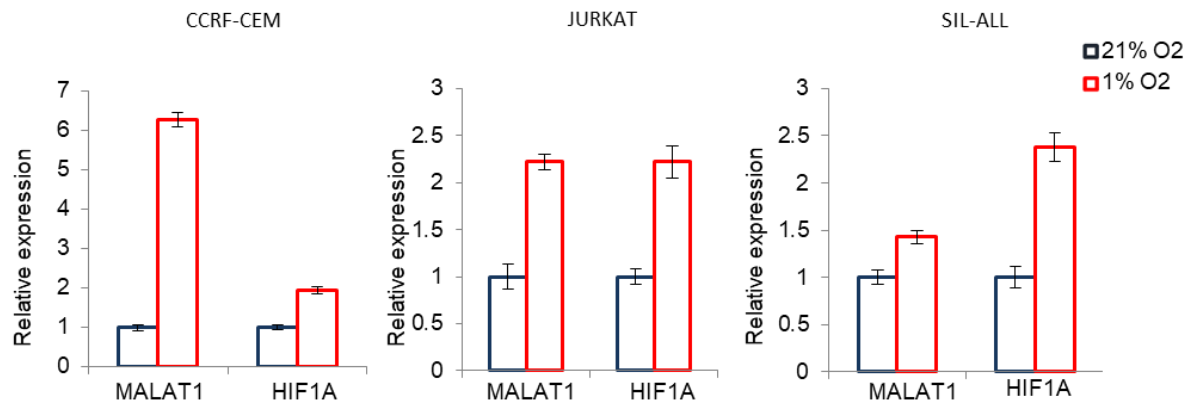
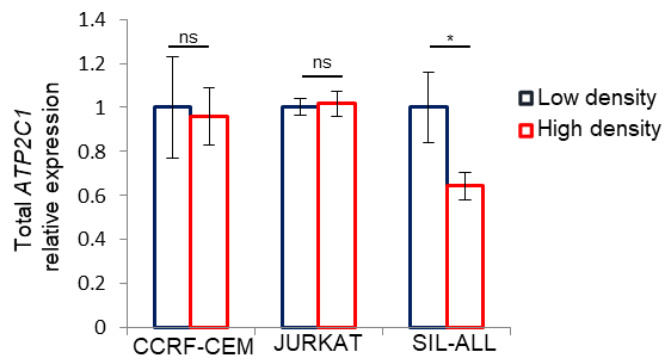
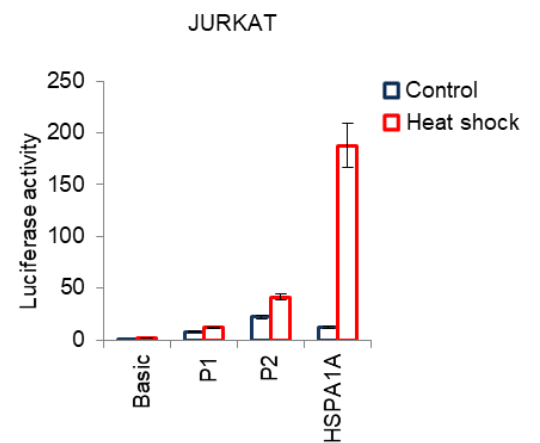
**SUPPLEMENTAL FIGURE 3B**



**SUPPLEMENTAL FIGURE 4A**



**SUPPLEMENTAL FIGURE 4B**

**A****qPCR primers****B****C****D****SUPPLEMENTAL FIGURE 5**



**SUPPLEMENTAL FIGURE 6**

## Legends supplemental figures

### **SUPPLEMENTAL FIGURE 1 ATP2C1 diverse isoforms expression**

A. Representation of all the *ATP2C1* isoforms described in NCBI gene database (<https://www.ncbi.nlm.nih.gov/gene/>), the transcript identifier as well as the amino acid differences between P1 and P2 isoforms in the 5' region of the gene are shown.

B. Dot plots representing the relative expression of P1-associated isoforms (NM\_001199180, NM\_001199181, and NM\_001199182) (left panel) and P2-associated isoforms (NM\_001001485, NM\_001199179, NM\_001001487, NM\_001199184, NM\_014382, NM\_001199185, NM\_001001486, and NM\_001199183) (right panel) based on RNA-seq data from thymic subpopulations and 41 T-ALL samples. Statistical significance was assessed by DE-seq.

### **SUPPLEMENTAL FIGURE 2 Classification of T-ALL samples**

A. Graphical representation of the ranked 120 T-ALL samples following the Log<sub>2</sub> (P1/P2 isoforms) ratio calculation, the first and fourth quartiles represent the P1 high and P1 low subset, respectively.

### **SUPPLEMENTAL FIGURE 3 Enrichment plots from Gene Set Enrichment Analysis (GSEA) for the 120 RNA-seq dataset**

A. Enrichment plots from the top 20 gene sets enriched in P1 high subset (first quartile) of 120 ranked T-ALL samples processed in this study.

B. Enrichment plots of the 7 gene sets enriched in P1 low subset (fourth quartile) of 120 ranked T-ALL samples processed in this study.

### **SUPPLEMENTAL FIGURE 4 Enrichment plots from Gene Set Enrichment Analysis (GSEA) for the ATAK et al. dataset**

A. Enrichment plots from the top 20 gene sets enriched in P1 high subset (11 T-ALL samples) of 31 T-ALL samples using RNA-seq data published by Atak et al, 2013.

B. Enrichment plots of the 11 gene sets enriched in P1 low subset (20 T-ALL samples) of 31 T-ALL samples using RNA-seq data published by Atak et al, 2013.

### **SUPPLEMENTAL FIGURE 5 RT-qPCR primers, hypoxia positive controls, and heat shock**

A. Schematic representation of the relative localization of the RT-qPCR primers designed for the assessment of the relative expression of total *ATP2C1*, and the P1 and P2 isoforms.

B. RT-qPCR analysis of the expression of *MALAT1* and *HIF1A* in hypoxic conditions and normal oxygen conditions in CCRF-CEM (left panel), Jurkat (middle panel), and Sil-ALL (right panel) cell lines. Values represent the relative expression levels as compared to the normal oxygen conditions and normalized by the *B2M* housekeeping gene. Significance was assessed by a two-sided Student t-test. Error bars represent the s.d. of three independent replicates.



C. RT-qPCR analysis of the overall expression of *ATP2C1* (total) in low and high-density conditions in CCRF-CEM, Jurkat, and Sil-ALL cell lines. Values represent the relative expression levels as compared to the low-density conditions and normalized by the *B2M* housekeeping gene. Significance was assessed by a two-sided Student t-test. Error bars represent the s.d. of three independent replicates. \*  $P \leq 0.05$ , ns  $> 0.05$ .

D. Promoter activity quantification by luciferase assay of P1, P2, and *HSPA1A* promoters after heat shock is shown for Jurkat cell line. Values represent the relative luciferase activity as compared to the pGL3 basic vector and normalized by the Renilla control vector. Significance was assessed by a two-sided Student t-test. Error bars represent the s.d. of three independent replicates.

### SUPPLEMENTAL FIGURE 6 Transcription factor binding sites (TFBSs) of P1 and P2 *ATP2C1*

A. The genomic track shows a representative subset of TFBSs present in the P1 (A) and P2 (B) regions of *ATP2C1*. The data was obtained from JASPAR 2020 database and visualized with the UCSC genome browser track.

**Supplemental Table S1:** List of sequenced samples and data resources

Sample	Source or reference	Link	Accession number
ChIP-seq H3K4me3 T-ALL 864	Blueprint	<a href="http://dcc.blueprint-epigenome.eu">http://dcc.blueprint-epigenome.eu</a>	S01PSEH1
ChIP-seq H3K4me3 T-ALL 885	Blueprint	<a href="http://dcc.blueprint-epigenome.eu">http://dcc.blueprint-epigenome.eu</a>	S01PTCH1
ChIP-seq H3K4me3 T-ALL 845	Blueprint	<a href="http://dcc.blueprint-epigenome.eu">http://dcc.blueprint-epigenome.eu</a>	S01PV8H1
ChIP-seq H3K4me3 T-ALL 820	Blueprint	<a href="http://dcc.blueprint-epigenome.eu">http://dcc.blueprint-epigenome.eu</a>	T10C T-ALL
ChIP-seq H3K4me3 T-ALL 802	Blueprint	<a href="http://dcc.blueprint-epigenome.eu">http://dcc.blueprint-epigenome.eu</a>	S01S1JH1
ChIP-seq H3K4me3 T-ALL 760	Blueprint	<a href="http://dcc.blueprint-epigenome.eu">http://dcc.blueprint-epigenome.eu</a>	T12C T-ALL
ChIP-seq H3K4me3 T-ALL 790	(Belhocine et al 2021)	<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	GSE151297
ChIP-seq H3K4me3 T-ALL 764	(Belhocine et al 2021)	<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	GSE151297
ChIP-seq H3K4me3 T-ALL 427	(Belhocine et al 2021)	<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	GSE151297
ChIP-seq H3K4me3 TALL 786	This study		
ChIP-seq H3K4me3 TALL 753	This study		
ChIP-seq H3K4me3 RPMI-8402	(Belhocine et al 2021)	<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	GSE151297
ChIP-seq H3K4me3 SIL-ALL	(Belhocine et al 2021)	<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	GSE151297
ChIP-seq H3K4me3 CCRF-CEM	(Belhocine et al 2021)	<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	GSE151297
ChIP-seq H3K4me3 Jurkat		<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	GSM1464990
ChIP-seq H3K4me3 CUTLL1	(Wang et al. 2011)	<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	GSM732911
ChIP-seq H3K4me3 DND41	ENCODE (Bernstein et al. 2012)	<a href="https://www.encodeproject.org">https://www.encodeproject.org</a>	ENCSR000ARA
ChIP-seq H3K27ac RPMI-8402			
ChIP-seq H3K27ac SIL-ALL			
ChIP-seq H3K27ac CCRF-CEM			

ChIP-seq H3K27ac Jurkat		<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	GSM1296384
ChIP-seq H3K27ac Loucy		<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	GSM1916974
ChIP-seq H3K27ac CUTLL1			
ChIP-seq H3K27ac DND41		<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	GSM1314136
RNA-seq normal thymocytes and 41 T-ALL series	(Cieslak et al. 2020) (Bond et al. 2017)	<a href="https://www.ncbi.nlm.nih.gov/geo/">https://www.ncbi.nlm.nih.gov/geo/</a>	GSE142522
RNA-seq 120 T-ALL series	This study		
RNA-seq T-ALL cell lines	(Atak et al. 2013)	European Genome-phenome Archive	EGAS00001000536
ChIP-seq H3K4me3 human thymocyte populations (EC, LC, SP4, SP8)	Blueprint (Cieslak et al. 2020)	<a href="http://dcc.blueprint-epigenome.eu">http://dcc.blueprint-epigenome.eu</a> Direct track hub: <a href="http://genome-euro.ucsc.edu/cgi-bin/hgTracks?hubUrl=https://raw.githubusercontent.com/guillaumecharbonnier/mw-cieslak2019/master/src/hub/hub.txt&amp;genome=hg19">http://genome-euro.ucsc.edu/cgi-bin/hgTracks?hubUrl=https://raw.githubusercontent.com/guillaumecharbonnier/mw-cieslak2019/master/src/hub/hub.txt&amp;genome=hg19</a>	

### Supplemental Table S2. AP Genes

tracking_id	gene_id	gene_short_n	locus	Chr.gene	start.gene	end.gene	length	ID	log2FoldC	pvalue.x	padj.x
NM_001195098	ANKRD28	ANKRD28	chr3:15708743-15901053	chr3	15708743	15901053	6382	NM_001195098	-2.36022	0.000169	0.001465
NM_001195099	ANKRD28	ANKRD28	chr3:15708743-15901053	chr3	15708743	15901053	6073	NM_001195099	0.21704	0.788721	0.857017
NM_015199	ANKRD28	ANKRD28	chr3:15708743-15901053	chr3	15708743	15901053	6338	NM_015199	-0.03767	0.918697	0.947682
NM_001199180	ATP2C1	ATP2C1	chr3:130569368-130745646	chr3	1.31E+08	1.31E+08	3682	NM_001199180	-4.42486	1.39E-08	8.09E-07
NM_001199182	ATP2C1	ATP2C1	chr3:130569368-130745646	chr3	1.31E+08	1.31E+08	3595	NM_001199182	-4.18625	5.23E-08	2.38E-06
NM_001199183	ATP2C1	ATP2C1	chr3:130569368-130745646	chr3	1.31E+08	1.31E+08	3555	NM_001199183	-2.84819	3.66E-05	0.000439
NM_001199181	ATP2C1	ATP2C1	chr3:130569368-130745646	chr3	1.31E+08	1.31E+08	5010	NM_001199181	-1.93399	0.001085	0.006289
NM_001001485	ATP2C1	ATP2C1	chr3:130569368-130745646	chr3	1.31E+08	1.31E+08	3692	NM_001001485	-1.98526	0.00323	0.01474
NM_001199179	ATP2C1	ATP2C1	chr3:130569368-130745646	chr3	1.31E+08	1.31E+08	5021	NM_001199179	-2.03837	0.004334	0.01839
NM_001001487	ATP2C1	ATP2C1	chr3:130569368-130745646	chr3	1.31E+08	1.31E+08	3464	NM_001001487	1.121889	0.108727	0.204514
NM_001199184	ATP2C1	ATP2C1	chr3:130569368-130745646	chr3	1.31E+08	1.31E+08	4883	NM_001199184	0.464865	0.418291	0.552225
NM_014382	ATP2C1	ATP2C1	chr3:130569368-130745646	chr3	1.31E+08	1.31E+08	4934	NM_014382	0.574429	0.431895	0.564888
NM_001199185	ATP2C1	ATP2C1	chr3:130569368-130745646	chr3	1.31E+08	1.31E+08	3511	NM_001199185	-0.5158	0.456204	0.587632
NM_001001486	ATP2C1	ATP2C1	chr3:130569368-130745646	chr3	1.31E+08	1.31E+08	3494	NM_001001486	0.327608	0.54588	0.666798
NM_004326	BCL9	BCL9	chr1:147013181-147098015	chr1	1.47E+08	1.47E+08	6276	NM_004326	0.536564	0.081675	0.166284
NM_182557	BCL9L	BCL9L	chr11:118754474-118781613	chr11	1.19E+08	1.19E+08	7738	NM_182557	0.099324	0.77886	0.850022
NM_012090	MACF1	MACF1	chr1:39547088-39952810	chr1	39547088	39952810	17871	NM_012090	-0.4243	0.097195	0.188651
NM_001256750	PEX5L	PEX5L	chr3:179512746-179831453	chr3	1.8E+08	1.8E+08	9198	NM_001256750	-3.23607	1.21E-05	0.000182
NM_016559	PEX5L	PEX5L	chr3:179512746-179831453	chr3	1.8E+08	1.8E+08	9204	NM_016559	-2.73416	6.57E-05	0.000693
NM_001256751	PEX5L	PEX5L	chr3:179512746-179831453	chr3	1.8E+08	1.8E+08	9132	NM_001256751	-1.96528	0.004246	0.018133
NM_001256752	PEX5L	PEX5L	chr3:179512746-179831453	chr3	1.8E+08	1.8E+08	9099	NM_001256752	-1.54971	0.033047	0.085501
NM_001256755	PEX5L	PEX5L	chr3:179512746-179831453	chr3	1.8E+08	1.8E+08	8739	NM_001256755	-1.30901	0.067429	0.144481
NM_001256754	PEX5L	PEX5L	chr3:179512746-179831453	chr3	1.8E+08	1.8E+08	8934	NM_001256754	-1.28335	0.09533	0.185982
NM_001256756	PEX5L	PEX5L	chr3:179512746-179831453	chr3	1.8E+08	1.8E+08	8829	NM_001256756	-1.06301	0.150747	0.260136
NM_001256753	PEX5L	PEX5L	chr3:179512746-179831453	chr3	1.8E+08	1.8E+08	9027	NM_001256753	-0.39646	0.522345	0.646477
NM_032627	SSBP4	SSBP4	chr19:18530145-18549111	chr19	18530145	18549111	1787	NM_032627	-2.14307	5.38E-05	0.000593
NM_001009998	SSBP4	SSBP4	chr19:18530145-18549111	chr19	18530145	18549111	1721	NM_001009998	-0.77901	0.05423	0.123459
NM_014975	MAST1	MAST1	chr19:12949258-12985766	chr19	12949258	12985766	4923	NM_014975	-3.87229	2.76E-12	5.92E-10
NM_138734	NRXN2	NRXN2	chr11:64373645-64490660	chr11	64373645	64490660	3535	NM_138734	-3.40047	3.78E-10	3.96E-08
NM_015080	NRXN2	NRXN2	chr11:64373645-64490660	chr11	64373645	64490660	6623	NM_015080	-0.88227	0.129481	0.232628
NM_138732	NRXN2	NRXN2	chr11:64373645-64490660	chr11	64373645	64490660	6413	NM_138732	0.139793	0.785522	0.854746

**Supplemental Table S3.** Transcription factors

<i>ATP2C1</i> P1 TFs					<i>ATP2C1</i> P2 TFs							
TAL1	ZNF592	NBN	GATAD2B		CTCF	NFRKB	POLR2A	MEIS2	SMAD5			
	HDAC1	ZNF687	MYC	IKZF1		NFATC3	GATA3	ARNT	SMC3			
	MLLT1	DPF2	NFIC	MXI1	ZNF384		RAD21	HMBOXSKIL	REST	CC2D1A		
	DPF2	MNT	ARID1B	E4F1		ZNF143	RCOR1	SMARCA5				
	CBFA2T3	SOX6	ZEB2	GATA1		TBX21	RFX5	TRIM22	STAT3			
	SMARCA4	ZSCAN29	TAL1			ATF7	ARID3A	RUNX3	RB1			
	CHAMP1	HDAC2	SMARCC2			KDM5A	SIN3A	MEF2A	WRNIP1			
	ATF3	GATAD2B	TRIM28			DACH1	NCOA1	STAT1	TFAP4	IKZF1		
	ATF7	NR2F6	GATA3	RCOR1	ZNF639		BRD9	MTA2	ZNF592	IKZF2		
	TRIM24	KDM1A	NCOA2			RELB	1	XRCC5	DPF2	RBFOX2		
	PHB2	FOXK2	MYC	ZNF282		ZNF263	KDM5B	L3MBTL2				
	TCF12	NCOR1	TAF9B	ATF4	1	PKNOX1		HDAC1	POLR2G	ZBED1		
	RELB	DACH1	EP400	TEAD4	FOXM1		GATAD2B	EP300	BCL3	2	HDAC2	
	SMARCE1	RFX5	CEBPB	HDAC3		MAX	SAP30	CLOCK	NFATC1			
	IKZF1	GATA2	ARID3A	KLF16		AGO2	NBN	PHF8	MYC	MGA		
	CUX1	NCOA1	HDAC1	CBFA2T2		MXI1	NFRKB	MTA1	ZFX	KDM4A		
	MYNN	SMAD1	CREM	CC2D1A		ZNF687	CHD1	YY1	MNT			
	ATF2	REST	ZNF24	PML	ARNT		CBX5	NFATC3	SREBF2			
	MTA2	MEIS2	STAT5A	NFIC		ZNF217	BCL11A	BACH1				
	ZBED1	EP300	JUN	TBX21	MBD2		EGR1	3	EBF1	CTBP1	TAF1	ASH2L
	ETV6	NFATC1	CREB3	SIX5		RB1	SAFB	SIN3B	EP400	PHF20		
	SKIL	MXI1	L3MBTL2	ZNF318		CBX1	U2AF1	TAF9B	C11orf30			
	BMI1	FOS	IKZF2	JUND	LEF1		ZBTB40	TBP	GABPB1			
	NBN	CHD1	BHLHE40	HCFC1		NRF1	PML	RNF2	HCFC1	SP1		
	RUNX3	E2F8	RNF2	NEUROD1		RBM39	ETV6	ZNF143	JUND			
	TBL1XR1	C11orf30	NR2F1			HNRNPL	PRDM10	SMAD5				
	MAX	CBX3	JUNB	WRNIP1		MEF2B	ELF1	TAF15	SMARCA4			
	MYB	STAT2	NRF1	MAFK	RLF		NR2C1	ELF4	GABPA	CREM	ELK1	
	SPI1	TCF7	STAT3	MAFF	HES1		ZNF639	POU2F2	RBBP5			
	CBX5	NR2F2	ZNF217	E2F1		FOSL2	CREB1	M	ETS1	CBX3	1	
	ZNF207	ELF4	GABPA	EED		HNF4A	E4F1	CBFB	MEIS2	SPI1	2	
	MITF	MEF2B	BACH1	BRD9	USF2		NFYB	HNRNPLL	UBTF	BHLHE40		
	TBP	RBFOX2	IRF1	HMBOX1		AGO1	HNRNPK	E2F6	PAX5			
	PAX5	RUNX1	ZNF184	NFATC3		EED	MGA	PHF8	RNF2	ZBTB7A		
	NFE2	ESRRA	ZNF592	BCL11A		IKZF1	GMEB1	EGR1	KDM5B			
	IRF4	1	STAT1	POLR2A	BCL3		HSF1	HDAC1	MXI1	RELB	POLR2A	
	ELF1	CBFB	NFRKB	TRIM22		HDAC2	MYC	IKZF1	IKZF1	SPI1		
	MTA3	SRF	ZNF316	EBF1								
	MLLT1	BATF	1	MEF2C	ZBTB2	HDAC6						
	MEF2A	POU2F2	SMAD5									
	TAF1	YY1	GATA3	SMARCA5								
	SMAD1	HSF1	SRF	SPI1								
	CTCF	ZNF592	FOXA1	GATA3								
	REST	SIN3A	MAX	CHD1	FOXA2							

## 5.4 Materials and Methods I

### Chip-seq

ChIP-seq of H3K4me3 from human samples was performed following the BLUEPRINT protocol (<http://dcc.blueprint-epigenome.eu/#/md/methods>) (Cieslak et al. 2020). ChIP-seq libraries were generated with the MicroPlex Library Preparation Kit (Diagenode), according to the manufacturer's instructions. ChIP samples were sequenced in house in single-end 75nt mode using the NextSeq® 500/550 (Illumina) according to the manufacturer's instructions and processed following the BLUEPRINT protocol.

### RNA-seq

The extraction of total RNA was performed using the RNeasy Plus Mini kit (Qiagen) according to the protocol recommended by the supplier. Total RNA was quantified using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific) and stored at  $-80^{\circ}\text{C}$  until needed. Extracted RNA was used for the RNA-seq library preparation, using the TruSeq RNA Library Prep Kit v2 (Illumina). Libraries were paired-end sequenced on the Illumina NextSeq 500 sequencer.

### RNA-seq data analysis

We used two datasets of RNA-seq samples, a first dataset consists of 41 adult T-ALLs and 5 thymic subpopulations which have been previously published (Bond et al. 2017; Cieslak et al. 2020), and a second dataset consists of 120 adult T-ALLs (unpublished). Normalization and differential expression analyses were performed using DESeq2 (Anders and Huber 2010). Control lists were determined based on a reference gene list computed as optimally matched for similar signal levels of expression using R script. The statistical comparisons were performed using the Wilcoxon test unless mentioned otherwise.

### Identification of alternative promoter usage in T-ALL

Human genome annotation was downloaded as a GTF (Gene Transfer Format) file from Ensembl (release 92, build GRCh38). Only conventional autosomes (1-22) and gonosomes were considered for further analysis. GTF file manipulations were performed using gtfk command line suite from the pygtf (v0.7.4.dec0+138e). Transcripts whose promoters overlapped the TSS from another gene were flagged using the 'divergent' subcommand of gtfk (with argument `-S/--no-strandness, --upstream 2500, --downstream 2500`). The TSS together with transcript-related information extracted from the GTF file and convert to bed format using the '5p\_3p\_coord' subcommand of gtfk. These TSS were then intersected with the merged peaks H3K4me3 peaks from T-ALL and thymic samples (command `mergeBed` from the bedtools suite). H3K4me3 peak coverage was computed from the bigwig file using the coverage subcommand of gtfk with `--stat` set to `sum`.

Peaks that did not overlap with at least one known RefSeq TSS in at least one of the studied samples were discarded. TSSs from the same gene overlapping with an H3K4me3 peak were pooled into a single promoter. H3K4me3 ChIP-seq coverage was computed by summing the number of mapped reads for each base pair of the peak. For each gene, a

contingency table of the H3K4me3 peak coverage was generated from the data; each line represents a different promoter for the gene, and each column the sample. Peak coverage is averaged across all samples of a given condition (cell lines, healthy thymic cells, leukemic cells). In the alternative local approach, the V score is computed for every possible pair of samples, without grouping by condition.

From this contingency table, we computed a Cramer's V-score (Cohen 1988), providing a likelihood of alternative promoter usage between T-ALL and thymocyte samples. Three successive criteria are applied: (I) We selected genes with a V greater than 0.2. In addition, (II) at least one significant ANOVA ratio when calculated on the coverage table between the two sets of samples was required. Finally, (III) to reduce the number of false positives, we also filter the genes based on the density of each of their H3K4me3 peaks (peak coverage divided by peak length): we require each gene to have at least two TSS where the mean peak density across all samples is higher than the median (roughly equal to  $\sigma$ ) of all densities across all samples. In the alternative local approach, we only require at least one sample to be higher than the median.

### **Cell culture**

Cell lines Jurkat (ACC-282), Sil-All (ACC-511), and CCRF-CEM (ATCC® CRM-CLL-119) were obtained from the ATCC (American type culture collection) and maintained in RPMI 1640 media (Thermo Fisher Scientific) supplemented with 10% FBS (Thermo Fischer Scientific) at 37°C and 5% CO<sub>2</sub>. Cells were passaged every three days at 3x10<sup>5</sup>cells/mL and routinely tested for mycoplasma contamination.

### **PMA/Ionomycin stimulation**

Phorbol 12-myristate 13-acetate (PMA; Sigma Aldrich, P8139) and ionomycin (Sigma Aldrich, I3909) were used to induce T-cell activation. Each cell line (Jurkat, Sil-All, and CCRF-CEM) was plated in 12-well plates (5x10<sup>6</sup>cells/well) in media containing 20 ng/mL of PMA and 2.5µM of ionomycin and incubate for 4 hours at 37°C and 5% CO<sub>2</sub>. After incubation cells were harvested and RNA extraction was performed. For each cell line, three independent stimulations were made and compared to a non-stimulated control.

### **Hypoxia**

Cells were cultured in flasks at the appropriate cellular concentration, a triplicate for each cell line was cultured under hypoxic conditions 1% O<sub>2</sub> at 37 °C for seven days using the InvivoO<sub>2</sub> hypoxic chamber (Baker Ruskin), media was changed every 2-3 days to maintain the appropriate cellular concentration, in parallel a triplicate control for each cell line was done in normal oxygen conditions 21% O<sub>2</sub> at 37 °C, after the seven days of incubation cells were harvested and RNA extraction was performed.

### **Gene expression analysis**

RNA was extracted using the RNeasyPlus Mini Kit (Qiagen) according to the manufacturer's instructions. 2.5 µg of total RNA was reverse transcribed using Master Mix SuperScript® VILO™ (Thermo Fisher Scientific, 11755-250). 1:5 cDNA dilutions were used for the RT-qPCR. RT-qPCR reactions were made using the PowerSYBR® Green PCR Master Mix (Thermo Fisher Scientific, 4367659) and the measurement was made using the Applied

Biosystems QuantStudio 6 Flex Real-Time PCR System. The relative expression was calculated using the relative standard curve method from the mean of three biological replicates using *B2M* as a housekeeping gene and the values were normalized by the value of the control (non-stimulated condition). Primer sets are listed in **Supplemental Table 4**

### **Luciferase reporter assay and mutagenesis**

The *ATP2C1* promoters P1 (762 bp, chr3:130,568,799-130,569,560; HG38 genome assembly) and P2 (1080 bp, chr3:130,612,082-130,613,161) were cloned into the pGL3 Basic vector (Promega, E1751) upstream the luciferase gene at the KpnI and HindIII restriction sites. Site-specific mutagenesis of the P1 *ATP2C1* was done using the Q5 site-directed Mutagenesis kit (NEB, E0554S) using a set of primers listed in **Supplemental Table 4**. For cell transfection 2x10<sup>6</sup> Jurkat or CCRF-CEM cells were mixed with 2 µg of each construct and 400 ng of Renilla vector using the Neo Transfection System (Thermo Fisher Scientific). For the Jurkat cell line the following conditions were used, pulse voltage 1,200V, pulse width 40 ms, and 1 pulse. The conditions for CCRF-CEM were as follows, pulse voltage 1230 V, pulse width 40 ms, and 1. After transfection cells were cultured in 2mL of media in 12 well plates. After 20 hours, half of the cell population was treated with PMA/Ionomycin as described in the section PMA/Ionomycin stimulation for 4 hours. Cells were harvested and luciferase reporter assay was performed using the Dual-Luciferase® Reporter Assay System (Promega, E1910) following manufacture instructions. Luciferase activity was measured using a Multimode Plate Reader VICTOR Nivo™ (Perkin Elmer). Data were normalized to Renilla values and represented as the fold change of relative light units over the basic pGL3 vector from non-stimulated cells. Experiments were performed in triplicate.

### **CRISPR-CAS9 genome editing**

For the deletion of *ATP2C1* P1 promoter region, target specific P1 *ATP2C1* ALT-R® crRNAs flanking a 2.4 Kb P1 region (chr3:130,848,608-130,851,066) were designed using the IDT Custom Alt-R® CRISPR-Cas9 guide RNA Tool (<https://www.idtdna.com>). ALT-R® crRNAs and common ALT-R® tracrRNA were chemically synthesized and obtained from Integrated DNA Technologies (IDT), P1 *ATP2C1* specific crRNAs sequences are listed in **Supplemental Table 4**. Each RNA was dissolved in a Nuclease-Free IDTE buffer 200 µM stock solution. Stock solutions were stored at -80 °C. To prepare the crRNA:tracrRNA duplex, 2.5 µL of each stock solution was mixed and annealed by heating 95 °C for 5 min, followed by gradual cooling until room temperature on the bench. For the RNP complex formation, 1.5 µL of 62µM Cas9 protein (Alt-R® S.p. Cas9 nuclease, v.3, IDT) was mixed with 1.5µL of crRNA:tracrRNA duplex previously prepared with 1µL of PBS and incubated at room temperature for 20min. The RNP complexes flanking the P1 region were transfected into the CCRF-CEM cell line using the Neo Transfection System (Thermo Fisher Scientific). The transfection of 5x10<sup>5</sup> with 0.9 µL of each complex was performed by triplicate with the following conditions, pulse voltage 1,230 V, pulse width 40 ms, and 1 pulse. Three days after transfection, cells were cultured in 3x96-well plates at limit dilution (0.5 cells/100 µL/well). Two primers were designed flanking the target region to identify the wild-type and the mutant alleles. After 2-4 weeks the clones were screened for homozygous deletion using the kit Phire Tissue Direct PCR Master Mix (Thermo Fisher Scientific, F170L).

## CRISPRi experiment

Two sgRNAs targeting the *ATP2C1* P1 promoter region were designed. sgRNA\_1 and sgRNA\_2 were located 90 and 391 bp upstream the P1-associated TSS, respectively. The sgRNAs were individually cloned in the CROPseq-Guide-Puro vector (Addgene #86708) as previously described (Datlinger et al., 2017). Briefly, sgRNAs were synthesized by Eurofins genomics as 74 base oligonucleotides with 18 and 35 bases of homology to the hU6 promoter and guide RNA backbone, respectively. Oligonucleotides were diluted to 100  $\mu$ M and posteriorly 1:1000 (final concentration 100 nM). Then, Cropseq-Guide-Puro was digested with BsmB1 (NEB), an 8,333 bp fragment was purified using Purelink Quick Gel extraction kit (Invitrogen). sgRNA oligonucleotides (200 fmoles) were cloned into CROPseq-Guide-Puro (11 fmoles) using NEBuilder HiFi DNA assembly master mix (NEB cat. no. E2621L) into the Lucigen Endura E. coli cells (Lucigen). Electroporation was realized in the BioRad GenePulser machine. Lentiviruses were produced by the platform de Vectorologie Lyon. A CRISPRi competent CCRF-CEM cell line was generated by lentiviral transduction with the dCas9-KRAB-MeCP2 vector (Addgene #122205) and selected for high-efficiency silencing using a GFP-targeting sgRNA control (Adamson et al., 2016). dCas9-KRAB-MeCP2 CCRF-CEM cells at  $0.3 \times 10^6$  cells/ml were independently infected with the two sgRNA lentiviruses (20 MOI) in 2 ml of complete culture medium (RPMI + 5% FBS). After 3 days of infection, cells were selected with Puromycin (2.5  $\mu$ g/ml) for 7 days.

## Classification of T-ALL samples based on P1-isoforms expression

A series of 120 T-ALL RNA-seq samples were classified according to the expression level of P1 transcripts into high P1 expression and low P1 expression. Briefly, normalized RNA-seq values from 15 *ATP2C1* transcripts were used, 4 corresponding to P1 transcripts and 11 to P2 transcripts (**Supplementary Figure 1**). For each sample, we calculated the log<sub>2</sub> ratio of the average P1- and P2-transcript levels. Samples were organized from highest to lowest ratio value (**Supplemental Figure 2**). The ranked samples were divided into four quartiles; the first quartile represented the T-ALL subset with high P1-isoforms expression and the fourth quartile the subset with low P1-isoforms expression.

## GSEA Enrichment analysis

GSEA was performed to analyze the enrichment of datasets between high and low expression of P1 *ATP2C1* groups using the software GSEA from the Broad Institute (<http://software.broadinstitute.org/GSEA/>) (Subramanian et al., 2005; Mootha et al., 2003). Briefly, the gene expression values for the two T-ALL subsets with P1 high and P1 low were formatted accordingly to the Broad Institute instructions. GSEA was performed by comparing the P1-high over P1-low samples using default parameters and hallmarks gene sets enrichment (h.all.v7.4.symbols.gmt).

## TFs and functional enrichment analysis

The TFs binding to the *ATP2C1* P1 and P2 promoters were obtained from the ENCODE database (ENCODE Project Consortium 2012) and using the UCSC human genome browser track GrCH38/hg38, the complete list can be found as Supplemental table "ENCODE", and for the enrichment analysis both TFs lists were submitted to g: Profiler

(<https://biit.cs.ut.ee/gprofiler/gost>). The top 10 most significant KEGG pathways were selected. For the TFs binding analysis of *ATP2C1* P1 and P2 promoters in the T-ALL cell lines, the database of Remap 2020 (Chèneby et al., 2020) was used.

### **Heat shock**

Jurkat cell line was transfected with *ATP2C1* P1, P2 or *HSPA1A*, *PGL3* constructs as described in the luciferase assay section. After transfection cells were cultured in 2mL of media in 12 well plates. After 20 hours, half of the cell population was treated with heat shock conditions: 42°C-1h follow by a 2h recovery at 37°C, the other half of the cell population was used as control of 37°C-1h. Cells were harvested and luciferase reporter assay was performed using the Dual-Luciferase® Reporter Assay System (Promega, E1910) following manufacture instructions. Luciferase activity was measured using a Multimode Plate Reader VICTOR Nivo™ (Perkin Elmer). Data were normalized to Renilla values and represented as the fold change of relative light units over the basic *pGL3* vector from control cells. Experiments were performed in triplicate.

### **Cell density**

Jurkat, CCRF-CEM, and Sil-ALL cell lines were seeded per triplicate in two density conditions (low and high) according to the optimal conditions of each cell line, that were as follows, low-density conditions for Jurkat, CCRF-CEM, and Sil-ALL were 0.1 M/mL, 0.2 M/mL, and 0.5 M/mL respectively and high-density conditions for Jurkat, CCRF-CEM, and Sil-ALL were 3 M/mL, 2M/mL, and 1.5M/mL respectively. After 24h, samples of 5M cells were taken for RNA extraction, cDNA synthesis, and RT-qPCR experiments as described in the gene expression analysis section.



**Supplemental Table S4. Primers**

Name	Sequence 5'-3'	Purpose
<i>P1_F ATP2C1</i>	TTTATCGACGCTGAGAAACCA	RT-qPCR
<i>P1_R ATP2C1</i>	TTGATGTCAATACAGGAATCATTGTCT	RT-qPCR
<i>P2_F ATP2C1</i>	CCATCAACCCGAGTGCAGTT	RT-qPCR
<i>P2_R ATP2C1</i>	TTTGATGTCAATACAGGAATCATTGT	RT-qPCR
<i>Total_F ATP2C1</i>	ACTTCCCAGTCAGTCTCAGGA	RT-qPCR
<i>Total_R ATP2C1</i>	GCCTTGGGCTAGCTCTGTAA	RT-qPCR
<i>B2M_F</i>	AGGACTGGTCTTTCTATCTCTTG	RT-qPCR
<i>B2M_R</i>	CGGCATCTTCAAACCTCCAT	RT-qPCR
<i>gRNA_ups ATP2C1</i>	GTTTACCCATATGTTACTGT	CRISPR-Cas9
<i>gRNA_down ATP2C1</i>	AAGTCTGAACAGCCACAAGC	CRISPR-Cas9
<i>ATP2C1P1KO_F</i>	AATCCCCTTCGCCTCTGTTT	Deletion check
<i>ATP2C1P1KO_R</i>	GGCTTCATTACAGAACTCCGC	Deletion check
<i>ATP2C1_P1_STAT3_F</i>	AATTTTTTTGtccGTCCCCTATAACAGTCTTC	Mutagenesis
<i>ATP2C1_P1_STAT3_R</i>	TAACAAATTACCTCACTCTTC	Mutagenesis
<i>ATP2C1_P1_NFATC4_F</i>	CTGTGGGTATaacAAGTCACCAGC	Mutagenesis
<i>ATP2C1_P1_NFATC4_R</i>	ACTCTCGTGAGGCTTAATTTTAATTC	Mutagenesis
<i>sgRNA1_ATP2C1</i>	AATGTAAAGTATCTGGTTAC	CRISPRi
<i>sgRNA2_ATP2C1</i>	TAAGCCTCACGAGAGTCTGT	CRISPRi

# RESULTS PART II

## 5.6 Results part II Functional study of *ATP2C1* in a T-ALL model

### 5.6.1 Goal

We previously showed in the first part of this work that *ATP2C1* is deregulated in a subset of T-ALL samples and T-ALL cell lines, we demonstrated by expression and luciferase reporter assay analysis that *ATP2C1* is up-regulated by hypoxia and PMA/Ionomycin stimulation, revealing the importance of *ATP2C1* in T cell function and T-cell activation. To assess the role of *ATP2C1*, in T-cell function, we performed the complete inactivation of *ATP2C1* in two CCRF-CEM cell line clones and performed a series of functional assays including intracellular  $\text{Ca}^{2+}$  measurement, cell proliferation assay, apoptosis, and necrosis quantification, metabolism, cell adhesion assay, and T-cell activation phenotype assessment.

### 5.6.2 Contributions

The inactivation of the *ATP2C1* gene was carried out by the Center of Immunophenomics (CIPHE) platform located in Marseille, France.

In this part of the project I participate in the experimental execution of the following:

- Validation of the *ATP2C1* knock-out clones by PCR
- Handling and proliferation assays with and without trypsin-EDTA
- I performed the apoptosis and necrosis assay before FACS, the FACS and FACS analysis were achieved with the help of Iris Manosalva, engineer of our team. Iris Manosalva was also involved in the performance and analysis of metabolism, cell adhesion, and CD69 assays.
- I contribute to the editing of the figures, and with my supervisor to the writing of the results.

### 5.6.3 Results II

#### Generation of *ATP2C1* knock-out in the CCRF-CEM cell line

For the study of the *ATP2C1* function, we inserted a stop codon with a polyadenylation site in the second exon of *ATP2C1*, which is shared by all the *ATP2C1* transcripts (Fig. 1A, top panel). We obtained two homozygous *ATP2C1* knock-outs (KO) clones from the CCRF-CEM cell line (Fig. 1A, bottom panel; hereafter KO1 and KO2).

#### ***ATP2C1* is required for maintaining homeostatic intracellular $\text{Ca}^{2+}$**

As *ATP2C1* protein is an ATPase  $\text{Ca}^{2+}$  pump localized in the Golgi apparatus, we assessed the cytosolic  $\text{Ca}^{2+}$  concentration in WT and the two KO CCRF-CEM cells (Fig. 1B). Inactivation of *ATP2C1* resulted in a reduced intracellular  $\text{Ca}^{2+}$  concentration. This is reflected by both a reduction of  $\text{Ca}^{2+}$  positive cells, as well as a lower level of  $\text{Ca}^{2+}$

concentration in positive cells (Fig. 1B left). This suggests that the ATP2C1 pump contributes to the homeostatic level of intracellular  $\text{Ca}^{2+}$ , consistent with its role in  $\text{Ca}^{2+}$  transport from the cytosol into the Golgi apparatus (Brini et al., 2012).

Stimulation with PMA/Ionomycin increases the intracellular  $\text{Ca}^{2+}$  with a maximum peak between 75min and 90min as observed in the FACS histograms (Fig. 1B left) and the quantification of median fluorescence at different time points after PMA/Ionomycin stimulation (Fig. 1B right top) in all cell lines tested WT and KO. The overall intracellular  $\text{Ca}^{2+}$  levels remain lower in the WT cell line as compared to the KO cell lines in stimulated and non-stimulated cell lines (Fig. 1B, top and bottom).

### **ATP2C1 is required for normal cell proliferation and survival**

To assess differences in cell survival between wild-type and knock-out CCRF-CEM cell lines we initially performed a standard proliferation assay. We observed a drastic impairment in the proliferation rate in the KO cells (Fig. 2A). The original CCRF-CEM cell line is semi-adherent and therefore trypsinization was used to dissociate cells from the culture plate. We noticed that KO cells stop proliferating just after the Trypsin treatment (Fig. 2A). The reason for this effect is currently unknown. However, we can hypothesize that the chelating property of the EDTA used during the trypsinization process could result in an osmotic shock of the KO cells. To avoid the detrimental effect of the Trypsin, we performed a new proliferation assay without trypsinization (see Methods section) (Fig. 2B). We still observed a significant decrease in the proliferation rate of the two KO clones. This is also consistent with a decreased oxygen consumption rate during basal respiration (Fig. 2C), observed in the two KO clones, as well as, an increased percentage of necrotic and apoptotic cells (Fig. 2D).

### **ATP2C1 is required for normal cell adhesion**

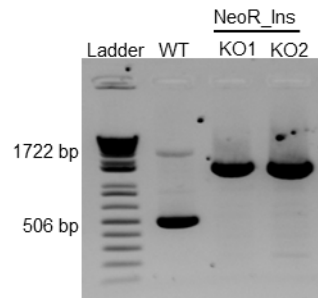
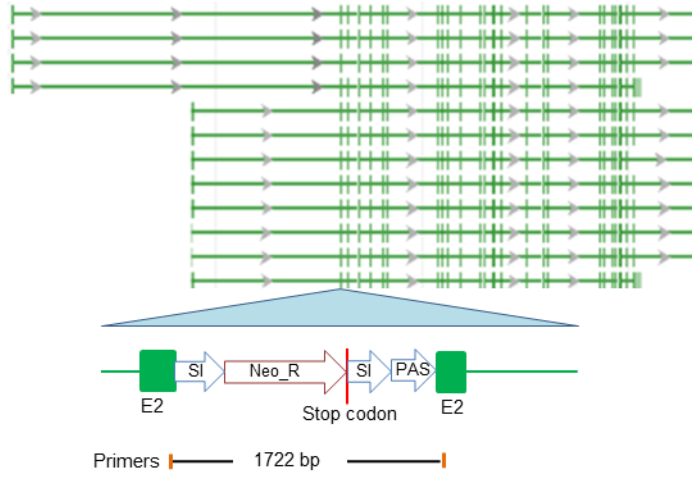
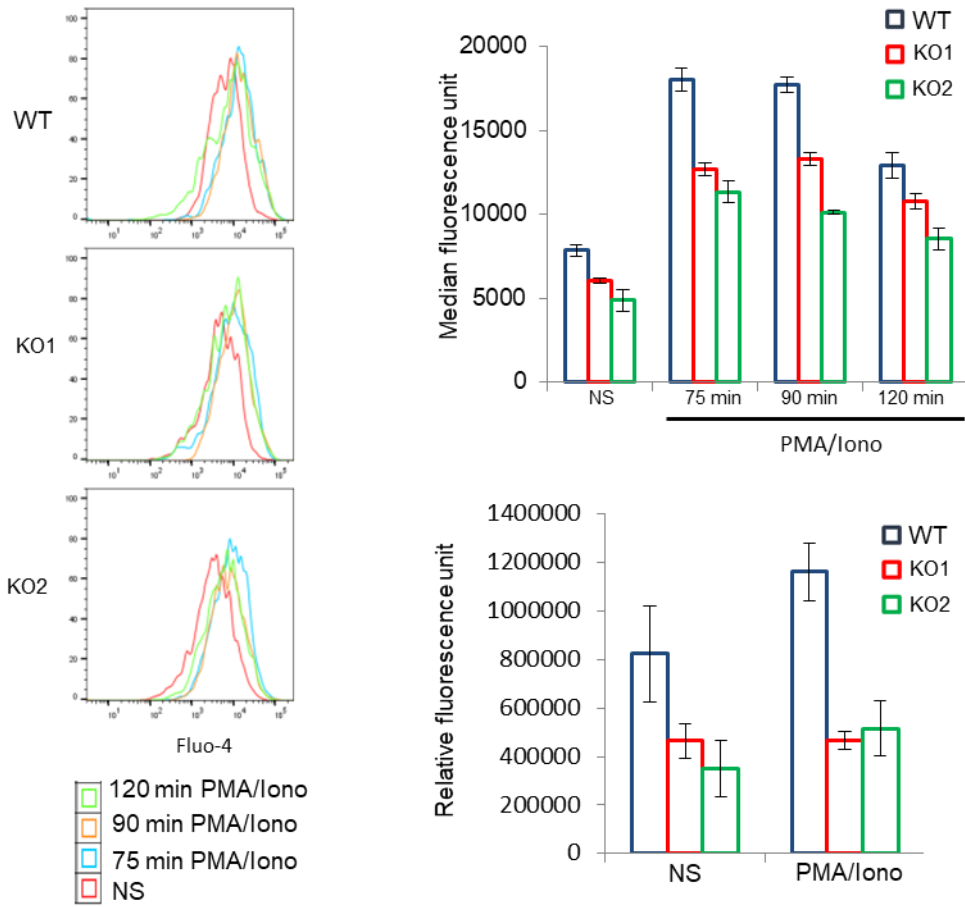
Differences in the time of disassociation between WT and KO clones were observed during harvesting procedures. To quantitatively assess cell adhesion, we quantified the time required for complete cell detachment after washing the cells with PBS (Fig. 2E). We observed that the time of detachment was significantly reduced in the two KO clones. Our results suggested that ATP2C1's control of intracellular  $\text{Ca}^{2+}$  is important for cell adhesion, consistent with the known role of  $\text{Ca}^{2+}$  in cell adhesion (Sheng et al., 2013).

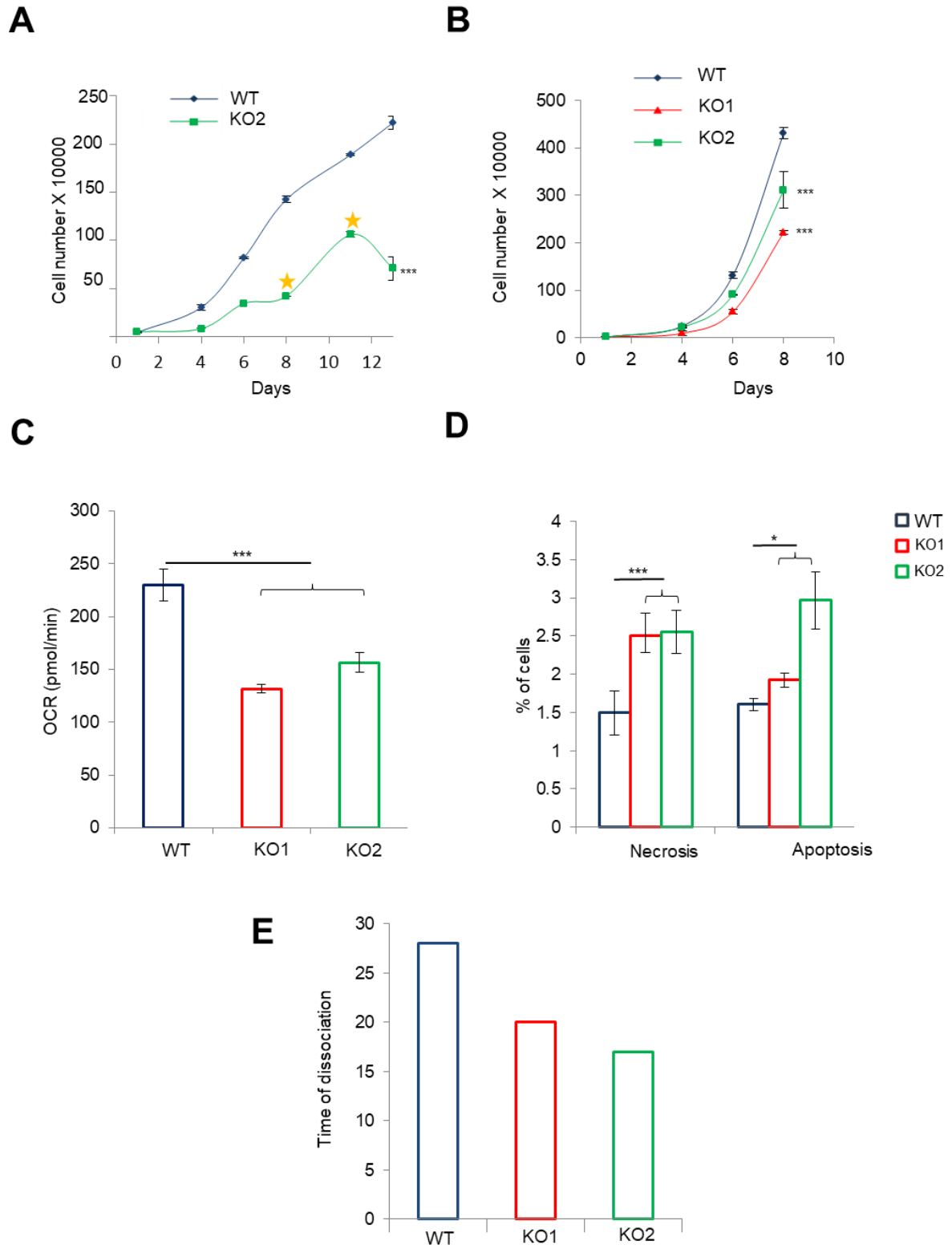
### **ATP2C1 is a negative regulator of T cell activation**

As we previously observed that ATP2C1 is induced by T cell activation (Results I), we explored the phenotypic consequences of ATP2C1 inactivation in T cell activation. WT and KO CCRF-CEM cells were stimulated during different times with PMA and Ionomycin and analyzed the percentage of CD69+ cells by FACS (Fig. 3A). We observed that after 90min of stimulation 43.5% of WT cells were CD69+, while more than half of the KO clones (84.9% and 76% of KO1 and KO2, respectively) were CD69+. The percentage of CD69+ cells reached a similar plateau after 2 hours of stimulation for both WT and KO cells (Fig. 3A). Therefore, the absence of *ATP2C1* resulted in premature activation of T cells after treatment

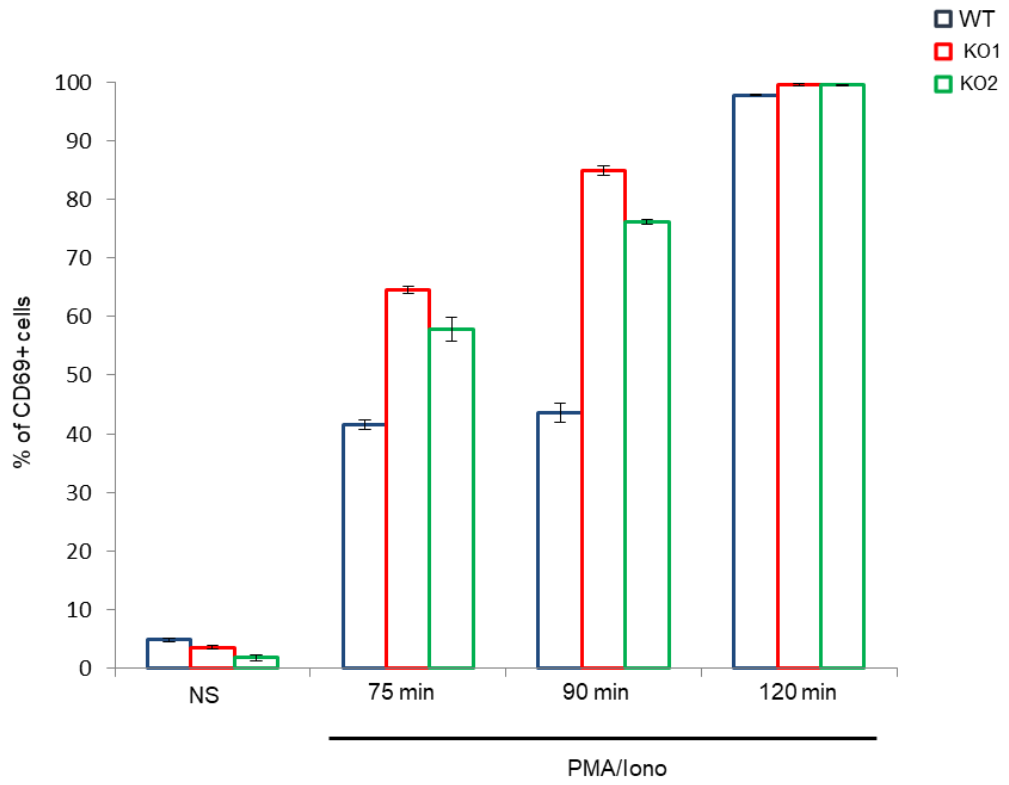
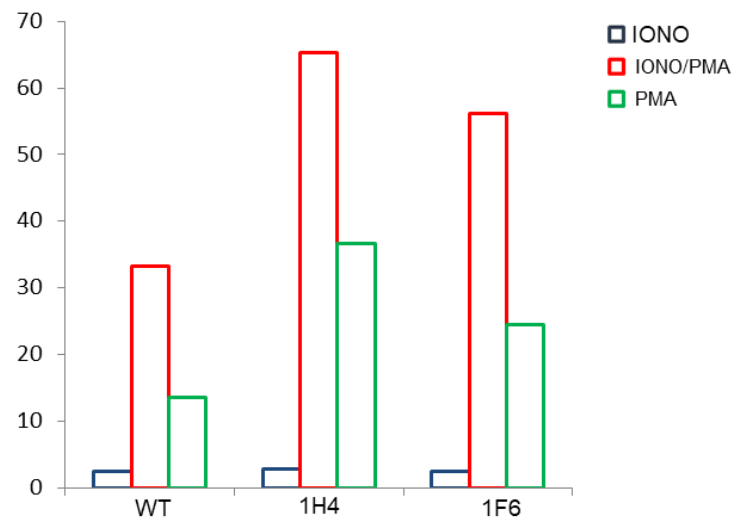
with PMA and Ionomycin, suggesting that *ATP2C1* functions as a negative regulator of T cell activation.

Since the T-cell activation is triggered by the co-occurrence of two types of signals, one dependent on  $\text{Ca}^{2+}$  and the other dependent on kinases, we aimed to assess the independent effect of PMA (Kinase dependent) and Ionomycin ( $\text{Ca}^{2+}$  dependent) in T-cell activation. We observed that after 75 min of stimulation with PMA, 13% of WT cells were CD69+, while 36% and 24 % of KO1 and KO2 were CD69+ respectively. No effect of stimulation was observed with ionomycin (Fig. 3B). Interestingly, PMA stimulation of KO clones was sufficient to achieve the same percentage of CD69 positive cells observed in WT cells stimulated with both PMA and Ionomycin, suggesting that *ATP2C1* inactivation is sufficient to compensate the Calcium-dependent activation mediated by the Ionomycin.

**A****B****Figure 1**



**Figure 2**

**A****B****Figure 3**



## Legends Figures

### Figure 1 Knock-out of the *ATP2C1* gene and intracellular $\text{Ca}^{2+}$ measurement in *ATP2C1* KO clones

**1A.** Schematic representation of the Neomycin resistance cassette (Neo\_R) with a stop codon insertion in the exon 2 shared by all the *ATP2C1* transcripts and the primers used for the detection of CCRF-CEM clones harboring a homozygous insertion of Neo\_R (KO) (upper panel). Agarose gel electrophoresis of two clones (KO1-KO2) harboring a homozygous insertion of Neo\_R (KO) is shown in the lower panel. PCR products were amplified from the genomic DNA of each clone with a set of primers designed for insertion detection. The expected amplicon size is 1722 bp for homozygous KO and 506 bp for the WT genomes.

**1B.** Representative intracellular Fluo4- $\text{Ca}^{2+}$  FACS fluorescence distributions (left) in CCRF-CEM WT (top) KO (middle) and KO2 (bottom), the different color lines represent the time after PMA/Ionomycin stimulation in which the intracellular Fluo4- $\text{Ca}^{2+}$  was quantified and the non-stimulated control (NS). The bar graphs show the median overall quantification of intracellular Fluo4- $\text{Ca}^{2+}$  fluorescence for the WT and KO cells in non-stimulated conditions and after PMA/Ionomycin stimulation at the indicated time-points measured by FACS (right top) and after 75 min of PMA/Ionomycin stimulation measured by a fluorescence plate reader (right bottom).

### Figure 2 Cell proliferation, basal respiration, apoptosis, and cell adhesion in *ATP2C1* KO clones

**2A.** Cell proliferation assay using trypsin-EDTA. CCRF-CEM WT and KO2 cells were grown for 12 days and count every 2-3 days; trypsin-EDTA was added whenever indicated by a golden star for detachment purposes. Significance was assessed by a two-way ANOVA test. Error bars represent the s.d. of three independent replicates. \*\*\*  $P \leq 0.001$ .

**2B.** Cell proliferation assay using PBS. CCRF-CEM WT, KO1, and KO2 cells were grown for 10 days and count every 2-3 days without the use of trypsin-EDTA, PBS was used instead. Significance was assessed by a two-way ANOVA test. Error bars represent the s.d. of three independent replicates. \*\*\*  $P \leq 0.001$ .

**2C.** The bar graph shows quantified Oxygen consumption rates (OCRs) (basal mitochondrial respiration) of CCRF-CEM WT and KO cell lines. Significance was assessed by a two-sided Student t-test. Error bars represent the s.d. of three independent replicates. \*\*\*  $P \leq 0.001$ .

**2D.** Percentage of apoptotic and necrotic cells evaluated by FACS after staining with annexin and propidium iodide (PI) in the CCRF-CEM WT, KO1, and KO2 cells. Significance was assessed by a two-sided Student t-test. Error bars represent the s.d. of three independent replicates. \*\*\*  $P \leq 0.001$ , \*  $P \leq 0.05$ .

**2E.** The bar graph shows the quantified time of detachment of CCRF-CEM WT and KO cell lines

### Figure 3

#### T-cell activation is exacerbated in *ATP2C1* KO clones

**1A.** Percentage of cells positive to CD69 in CCRF-CEM WT and KO cell lines after PMA/Ionomycin stimulation at the indicated time points.

**1B.** Percentage of cells positive to CD69 in CCRF-CEM WT and KO cell lines after 75 minutes of stimulation with PMA, Ionomycin, or PMA/Ionomycin.

## 5.6.4 Materials and Methods II

### CRISPR neomycin cassette insertion

*ATP2C1* knock-out in the CCRF-CEM cell line was performed by introducing a stop codon with a polyadenylation site along with a neomycin resistance cassette (NeoR), for clone selection, in the second exon shared by all the *ATP2C1* transcript isoforms. Two homozygous CCRF-CEM *ATP2C1* knock-out clones were obtained. PCR for the confirmation of NeoR cassette insertion was performed using the kit Phire Tissue Direct PCR Master Mix (Thermo Fisher Scientific, F170L), using the following set of primers: forward-5' GAG AAG TAG CAT GTG GTT 3' and reverse-5' GCC ACA TTT CAA GTC ATA TA 3'.

### Cell proliferation with trypsinization

Cells were seeded at 20,000 cells/well on day 1 and let them proliferate for 12 days; dilutions to maintain optimal cellular concentration were made at days 6 and 11 using trypsin-EDTA to dissociate the adherent cells. Cell number was assessed at the days indicated using the automated cell counter Countess 3 (Thermo Fisher) following the manufacturer's instructions.

### Cell proliferation without trypsinization

Cells were seeded at 20,000 cells/well on day 1 per triplicate and per day point, and let them grow for 8 days, culture media was added to keep optimal cellular concentration, and PBS was used to dissociate the adherent cells. Cell number was assessed at the points indicated using the automated cell counter Countess 3 (Thermo Fisher) following the manufacturer's instructions.

### Cell adhesion

Cells were cultured at  $2 \times 10^5$  cells/mL. After 48h, non-adherent cells were removed and adherent cells were detached using PBS. Determination of the time lap needed for detachment was performed by the observation on the microscope for any attached cell at the bottom of the plate.

### Metabolism

Cells were processed according to the Mito Stress Test kit and measured using a seahorse machine following previous publication (Xia et al 2021). Briefly,  $0.4 \times 10^6$  cells/well were seeded in seahorse coated cell-tak plate. One day after, the RPMI medium was removed and replaced by Medium MEM Seahorse. Cells were kept at 37 °C for 1 h and the seahorse cartridge was upload with the following final concentrations of each inhibitor: Oligomycine (4 uM), FCCP (0.25 uM), Antimycine/Rotenone (0.5 uM). Then the stress response was measured using the WAVE software from Seahorse.

### Intracellular Ca<sup>2+</sup>

The presence of Ca<sup>2+</sup> was measured using the Fluo-4 Direct Calcium Assay Kits (Invitrogen, F10471) following the manufacturer's instructions. Briefly,  $1 \times 10^6$  cells were incubated with

Fluo-4 without Probenecid at 37 °C for 30 min. Cells were centrifuged and fluorescence was observed using FORTESA X-20 cytometer or Plate Reader VICTOR Nivo™ (Perkin Elmer).

#### **PMA/Ionomycin stimulation**

Cells were stimulated with PMA (20 ng/ml) and Ionomycin (2.5 μM). Cells at different time points of stimulation were washed with PBS to stop the induction. CD69 (clone FN50, Biolegend) antibody was used to determine the number of CD69 positive cells. Cells were stained with CD69 (1:100) antibody and quantified using the Cytometer (Fortessa X-20 BD). FlowJo 10 software was used for analysis.

# **DISCUSSION AND PERSPECTIVES**

# Chapter 6 Discussion and perspectives

## 6.1 Summary of main results

During my Ph.D., I contributed to systematically study the alternative promoter usage profiles in T-ALL samples compared to both healthy thymocytes and immortalized cell lines, using H3K4me3 ChIP-seq data. By looking to profiles specific to leukemic cells, we found an alternative promoter of the *ATP2C1* gene as being preferentially active in T-ALL samples. The *ATP2C1* gene has two promoters that give rise to a complex array of transcripts, however, the most distal promoter (P1) was preferentially active in a subset of T-ALL samples. We used RNA-seq data from a series of T-ALL samples and the main thymic subpopulations to check the expression of the transcripts derived from P1 and P2 promoters. We found that only the isoforms derived from P1 were significantly up-regulated in T-ALL as compared with normal thymocytes.

We performed a GSEA to discover pathways that were preferentially enriched in the samples with the highest expression of the P1 transcripts. We found enrichment of pathways involved in the inflammatory response and protein secretion like unfolded protein response, interferon-alpha response, interferon-gamma response, and hypoxia, while in the lowest P1 expression transcripts we found an enrichment of some pathways involved in cell cycle regulation. To gain insights into the *ATP2C1* P1 function, we looked for the transcription factors bound to each promoter and performed a Gene Ontology analysis with those factors; we found that transcription factors bound to P1 are associated with pathways involved in immune response and cancer while those bound to P2 are associated with different pathways, like cell cycle regulation.

We next choose as a working model three immortalized T-ALL cell lines CCRF-CEM, JURKAT, and SIL-ALL which displays different P1 activities; high, middle, and low respectively, and corroborate by qPCR the P1 profiles observed by H3K4me3 ChIP-seq data, where the expression of the isoforms derived from P1 is higher in CCRF-CEM followed by JURKAT and a complete lack of P1 expression is observed in SIL-ALL. We also investigate the transcription factors bound to P1 and P2 in some immortalized cell lines with active or inactive P1 using ReMap database, we found some oncogenic factors like TAL1, RUNX, MYB in the cell lines that have an active P1.

We further investigate the effect of hypoxia and PMA/Ionomycin stimulation on the overall expression of *ATP2C1* and the P1- and P2-specific isoforms, we found that both treatments up-regulate the expression of *ATP2C1*, especially from the P1 isoform. To corroborate the results observed by RT-qPCR, we performed a luciferase assay to assess the promoter activity of the P1 and P2 regions. We observed an increase in the P1 activity after PMA/Ionomycin stimulation. Then, we mutated two sites for the transcription factors STAT3 and NFATC4. After the mutation of both sites, the induction of P1 activity was completely abolished. Importantly, we were able to confirm the requirement of the P1 promoter for the induction of the *ATP2C1* genes. This was achieved by performing two alternative CRISPR-based strategies: genetic deletion and epigenetic silencing of the P1 promoter by CRISPR-Cas9 and CRISPRi technologies, respectively.

To study the function of *ATP2C1* we inactivated the gene in the CCRF-CEM cell line. We performed several tests including intracellular calcium, cell proliferation, apoptosis, metabolism, cell adhesion, and CD69 measurement after stimulation. In the CCRF-CEM knock-out clones, we observed a decrease in cell proliferation and increased cell death, decreased intracellular calcium, basal mitochondrial respiration, and cell adhesion as compared to wt cells. Moreover, we observed acceleration of the kinetics of T cell activation, suggesting that  $\text{Ca}^{2+}$  accumulation in the Golgi apparatus might be required for proper control of T cell signaling.

In summary, by integrating epigenomic data from primary thymocytes and T-ALL samples we have discovered an alternative promoter specifically deregulated in T-ALL and identified a  $\text{Ca}^{2+}$  pump as a novel protein involved in the modulation of T-cell activation. Below I discussed more in detail the implications of our finding concerning the potential role(s) of *ATP2C1* in leukemic transformation and T cell activation.

## 6.2. Alternative promoter identification

In this study, we used ChIP-seq data of H3K4me3 mark to identify genes with more than one active promoter and differentially expressed between normal thymocyte subpopulations and T-ALL samples, and we were able to identify a subset of genes with alternative promoters frequently de-regulated in T-ALL samples, although other approaches can be used such as CAGE tag data (Kaczowski et al., 2016; Muratani et al., 2014) and RNA-seq (Demircioglu et al., 2019), we favored the ChIP-seq approach because it provided direct evidence for epigenetic regulation. However, for large-scale projects, it might be more useful to use RNA-seq data as much more samples are available in the literature. This is illustrated by the recent work from Demircioglu and col. as discussed below.

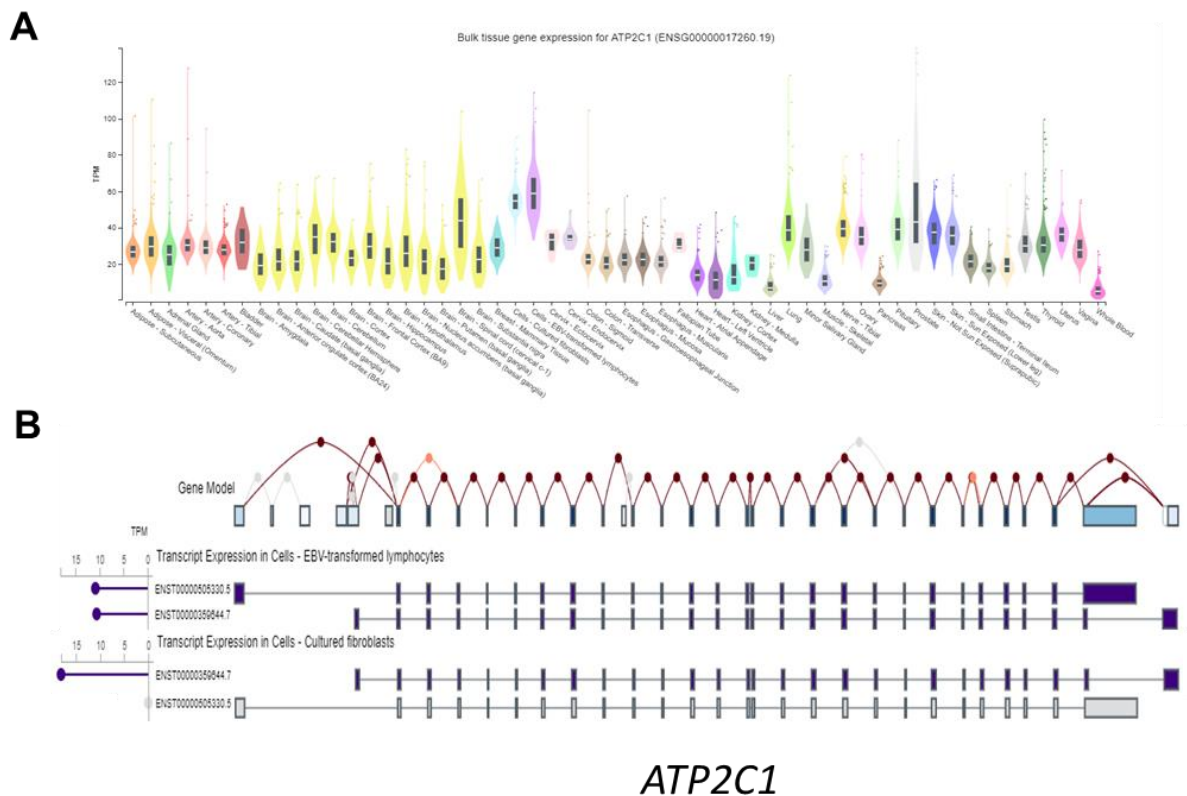
Demircioglu and col. analyzed 18,468 RNA-seq samples, providing the largest survey of active promoters in human tissues and cancers, identifying novel alternative promoters associated with cancer. The scale of the data generated permits for the first-time analysis of patient-to-patient promoter usage, suggesting that promoter selection is firmly regulated, influences the cancer transcriptome and that difference in alternative promoter usage possibly contribute to cellular transformation in cancer (Demircioglu et al., 2019) (Chapter 2 Figure 2.4). Highlighting the extensive role of alternative promoters in isoform expression throughout specific contexts and isoform diversity regulation and underscores how patient-to-patient variation in promoter usage is a link to pathological traits in cancer. They construct a comprehensive catalog of active promoters over 42 cancer types and tissues that will be a helpful resource to the comprehension of gene regulation networks and non-coding mutations in cancer (Demircioglu et al., 2019).

Our integrative analysis allows us to identify 8 genes associated with an AP deregulated in T-ALL. Although in my thesis I focused on the study of the *ATP2C1* gene, it would be interesting to study the other genes as well. Besides, our laboratory has now generated additional ChIP-seq data for an additional set of 60 primary T-ALLs. It will be worth performing a novel bioinformatic analysis with this dataset to identify additional alternative

promoters that might have escaped our initial analyses due to the relatively low number of samples. Finally, it will be of interest to investigate whether ATP2C1 alternative promoter usage is also deregulated in other hematopoietic malignancies using available RNA-seq or ChIP-seq data.

### 6.3 ATP2C1 regulation in leukemia

ATP2C1 is ubiquitously expressed. From our results, we can conclude that ATP2C1 P1 is an inducible promoter that responds to T-cell activation signaling, and to a less extent, to other stressing factors like hypoxia. Even though the ATP2C1 pump is constitutively expressed, it might be that the P1-derived isoforms have more specificity and might be important in cells that respond to stimulus through calcium signaling pathways like immune cells and neurons. Analyses of gene expression based on the GTEx database (<https://gtexportal.org/home/gene/ATP2C1>) show that ATP2C1 is expressed in most of the studied human cell types (Figure 6.1A). However, a higher expression was observed in fibroblasts and EBV-transformed lymphoblastoid cells. To assess which transcripts were expressed in these two cell types, I interrogated the GTEx database for the relative expression of the ATP2C1 transcripts. As shown in (Figure 6.1B), the fibroblast cells expressed primarily the P2-associated transcripts, while the immortalized EBV lymphoblastoid cells similarly expressed both P1- and P2-associated transcripts. These observations also support the findings that the P1 promoter is activated in transformed cells.

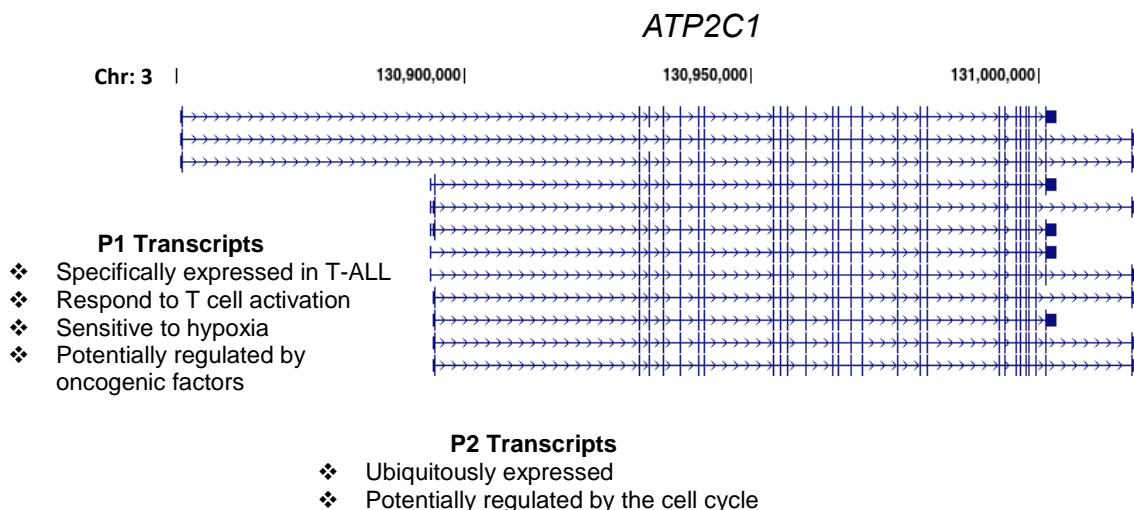




**Figure 6.1** (A) *ATP2C1* expression of most studied cell types. High expression of *ATP2C1* is observed for cultured fibroblasts and EBV-transformed lymphocytes. (B) Relative expression of two *ATP2C1* transcripts: ENST00000505330.5 (P1 derived) and ENST00000359644.7 (P2 derived) of cultured fibroblasts and EBV-transformed lymphocytes. EBV-transformed lymphocytes show similar expression of P1 and P2 transcripts, while fibroblasts mainly express P2 transcripts (adapted from (<https://gtexportal.org/home/gene/ATP2C1>)).

Based on our findings we can propose a model in which T cell precursors express the P2 transcripts that are normally regulated through the cell cycle. In a subset of T-ALL with activated immune signaling, the P1 promoter will be induced and results in a scape of *ATP2C1* expression from the cell cycle control (Figure 6.2). To explore this possibility, the expression of *ATP2C1* isoforms in the different cell cycle populations should be studied in normal and cancer cells.

The P1-associated transcripts of *ATP2C1* are specifically expressed in a subset of T-ALL samples which have an immune active phenotype. Together with the finding that P1 is an inducible promoter responding to the T cell signaling, we can hypothesize that ectopic activation of P1 might be related to the onco-genotype of the T-ALL samples. Many factors involved in T cell signaling, such as *STAT5B* and the JAK family of kinases are frequently mutated in T-ALL. To gain insight into the genotype of the T-ALL expressing the P1-transcripts, our collaborators from the Necker Hospital will compare the genotype of the T-ALL samples in the function of the expression of the P1-transcripts based on a series of 120 T-ALL samples that have been analyzed by RNA-seq in a collaborative effort between the two labs. The results will allow us to determine whether the activation of the p1 promoter is directly linked to the genotype of a T-ALL subset.



**Figure 6.2** *ATP2C1* Alternative promoters differentially regulated.

What can be the mechanism of P1 activation in a subset of T-ALLs? First of all, our epigenetics analyses have shown that the P1 promoter is found in poised states mainly associated with H3K4me1, but low levels of the silencing marks such as H3K27me3 or H3K9me3. This might explain the rapid induction of P1 promoters in cell lines where the

promoter is inactive, as observed for the Sil-ALL cell line. This can also provide a rationale for the ectopic activation of the P1 promoter in a subset of T-ALL. Moreover, we have also found that some oncogenic transcription factors like TAL1, MYB, and RUNX3 are bound to P1. However, a most extensive study in T-ALL samples must be done to corroborate the potential implication of these oncogenic factors in the regulation of P1 in T-ALL. For instance, knockdown of these factors could be performed in relevant cell lines to assess their impact on the regulation of the p1 promoter.

#### 6.4 Impact of ATP2C1 P1 isoforms in T-ALL

Jenkins and col. (Jenkins et al., 2016) studied the regulation of ATP2C1 and ATP2C2 in HCT116, a gastric cancer cell line. They showed that their expression depends on both cell density and cell deoxygenation. In our study, we also found an up-regulation of *ATP2C1* in hypoxic conditions, but not in high cell density conditions. The Jenkins study also found that changes in ROS production in cancer cells caused by increased cell density and hypoxia can auto-regulate the expression of ATP2C2, which in turn is involved in the potential removal of ROS. Moreover, the authors observed that *ATP2C2* was regulated by the cell cycle and its overexpression partially inhibits the cell cycle in the S phase. However, in this study, the authors focused more on the role of *ATP2C2*, but not *ATP2C1*. Therefore, it is plausible that *ATP2C1* could be involved in similar mechanisms. Notably, our knock-out model of *ATP2C1* showed that *ATP2C1* is required for cell proliferation and cell survival. It will be interesting to explore whether *ATP2C1* is also required for the regulation of ROS levels and cell cycle control using the *ATP2C1* knock-out cells. Moreover, given the role of *ATP2C1* in Ca<sup>2+</sup> and Mn<sup>2+</sup> homeostasis and participation in the processing and transport of proteins in the Golgi apparatus, it is plausible that the deregulation of *ATP2C1* in T-ALL might contribute to the oncogenic transformation. In particular, the ATP2C1 pump is crucial for Mn<sup>2+</sup> detoxification by transporting Mn<sup>2+</sup> from the cytoplasm into the Golgi. Cytosolic overload and depletion of Golgi-hosted Mn<sup>2+</sup> may lead to the loss of cell cycle control, genetic instability, and multinucleation. For instance, deficient Mn<sup>2+</sup> homeostasis caused by ATP2C1 mutations may be the initial event in tumorigenesis in some HHD patients and ATP2C1<sup>-/-</sup> mouse models (Garcia-Rodriguez et al., 2012).

In addition to the potential role of *ATP2C1* in cancer, we might ask whether the expression of P1-isoforms has a specific impact on the oncogenic transformation, especially in T-ALL. It is noteworthy that leukemic cells with an active P1 promoter do not have a significant increase in the overall level of ATP2C1. This indicated that P1 activation does not result in ATP2C1 overexpression. Therefore, alternative promoter usage of ATP2C1 might be required for either the expression of isoforms with distinct functions or fine-tune regulation of the gene. We did not favor the first hypothesis, because the P1 and P2 derived isoforms are predicted to encode for proteins with little differences (Point to supplemental figure), while all the functional domains of the ATPase Ca<sup>2+</sup> pump appear to be conserved between the different isoforms. Rather, we favor a model in which the activation of the P1 promoter might impact the relative expression of the ATP2C1 gene. On the one hand, given that P2 harbor binding sites for transcription factors involved in the regulation of cell cycle, we could envision that, in normal cells, ATP2C1 is regulated throughout the cell cycle, while the activation of P1 in leukemic cells will render the expression of ATP2C1 independent of the cell cycle (Figure 6.2). On the other hand, as P1 is sensitive to the “activation” status of the T-ALL, it is

plausible that its activation could help the cells to better adapt or deal with the inflammatory environment. For instance, the potential role of *ATP2C1* in limiting the T cell signaling, might prevent the leukemic cells to become over-activated and therefore increase their survival rate.

Finally, it will be essential to determine whether the expression of P1-isoforms is associated with the clinical outcome of T-ALL patients. To gain initial insight into the potential association of P1 with the patient's survival, I analyzed the dataset set generated by the work (Dermircioglu et al. 2019). As previously mentioned, in this study the authors assess the clinical relevance of alternative promoter usage in different types of solid tumors. Interestingly, I found that the *ATP2C1* P1 promoter was significantly associated with a worst outcome for glioma cancer patients. More strikingly, *ATP2C1* P1 ranked at position 16<sup>th</sup> from 92 alternative promoters significantly associated with patient survival among all the studied cancers, strongly suggesting that deregulation of *ATP2C1* alternative promoter usage might be highly clinically relevant (Figure 6.3).

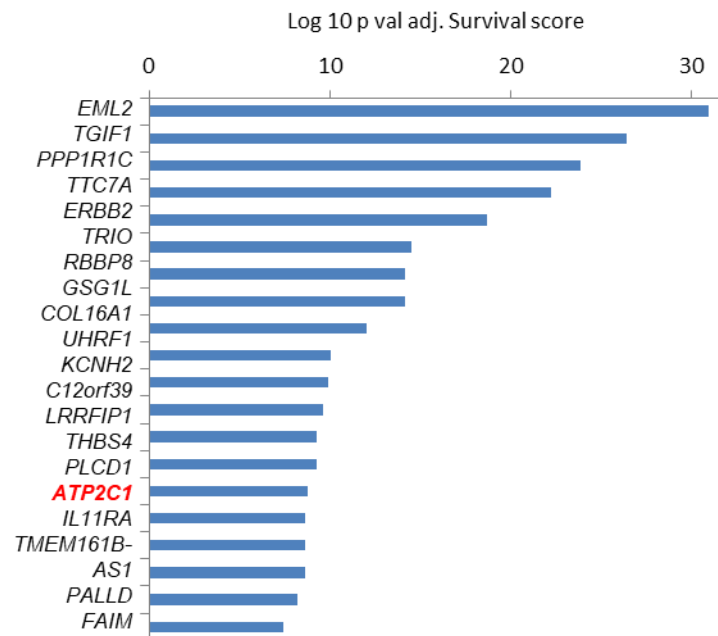


Figure 6.3 Top 20 genes with an alternative promoter specifically associated with patient survival (data from Dermircioglu et al., 2019).

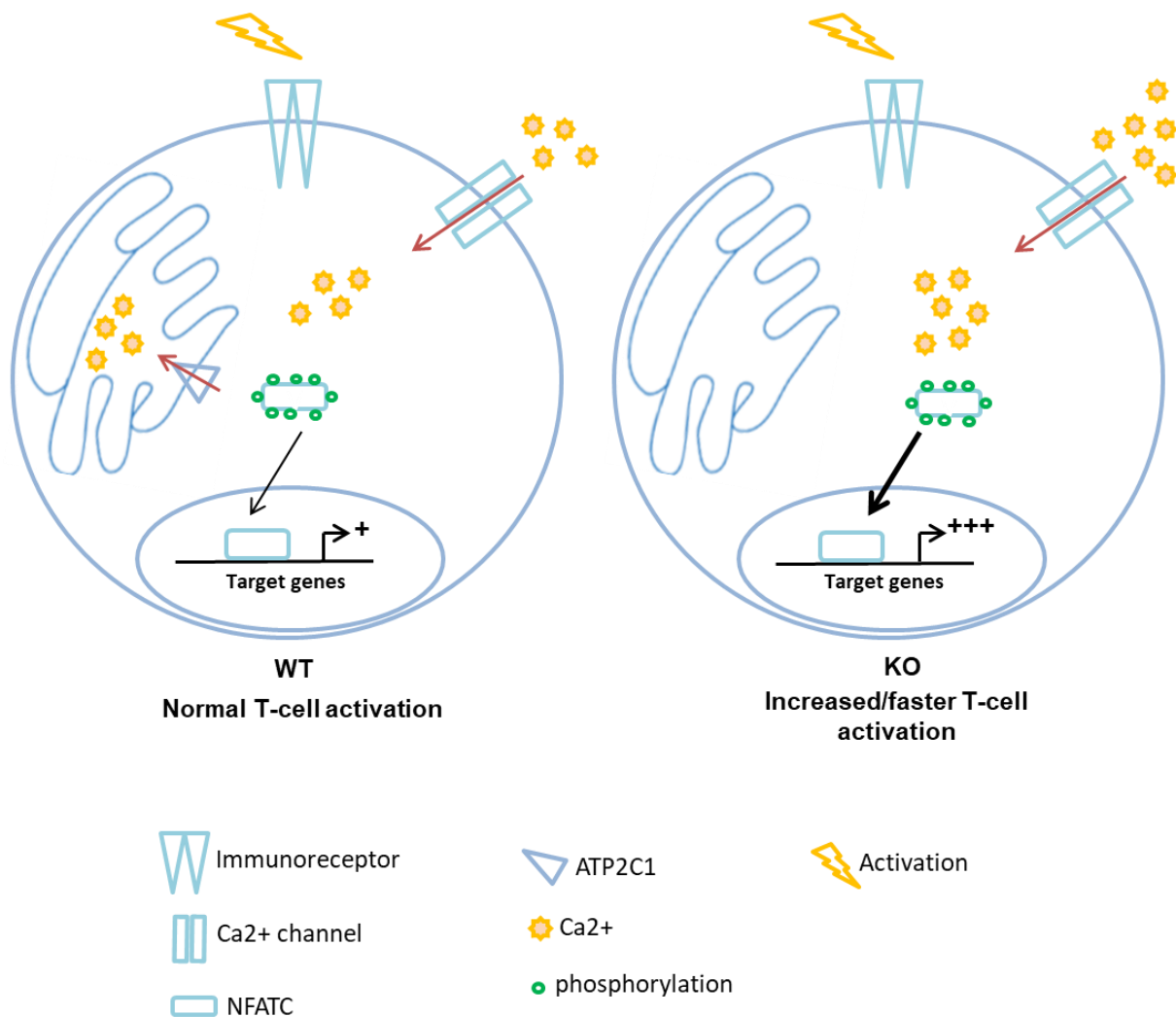
To assess whether *ATP2C1* alternative promoter usage is also significantly associated with patient's survival in T-ALL we are collaborating with the laboratory of Hematology from Necker's Hospital. In a partnership between the Necker and TAGC laboratories a series of clinically annotated 120 T-ALL samples have been analyzed by RNA-seq. I have classified these samples depending on P1-transcripts expression and provided the list to our collaborators. Using this list, they will be able to determine whether there is a statistical association between the P1 activity and the clinical outcomes.

## 6.5 Function of *ATP2C1* as a modulator of T cell activation

We initially observed that *ATP2C1* was induced by PMA/Ionomycin stimulation, while the P1 promoter appears to be induced by the T cell signaling. In addition, we also observed that P1 activation in T-ALL was associated with an immune-active phenotype. Therefore, we wondered whether there was a link between the function of *ATP2C1* and T cell activation. To this end, we assessed the efficiency of T cell activation in WT and *ATP2C1* KO cells using the CCRF-CEM as a T cell model. Surprisingly, we found that in the absence of *ATP2C1* there was a premature activation of T cells after T cell stimulation. Based on these results, we suggest that *ATP2C1* might function as a negative regulator of T cell activation.

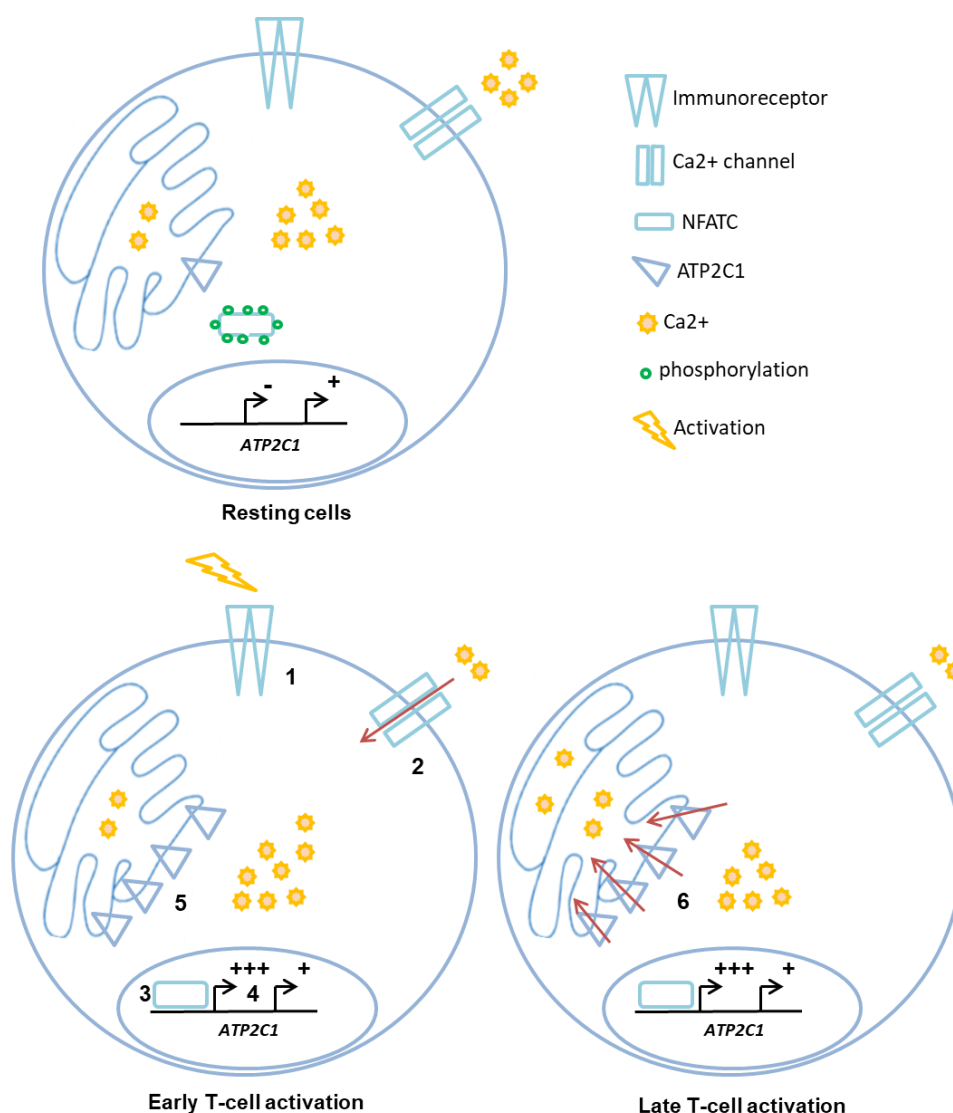
*ATP2C1* is a  $\text{Ca}^{2+}$  pump located in the Golgi apparatus (see Chapter 4). Calcium plays an essential role in T cell activation (Gwack et al., 2007; Park et al 2020). During an immune response, the engagement of T-cell receptors induces a decrease in the intracellular  $\text{Ca}^{2+}$  store and then activates store-operated  $\text{Ca}^{2+}$  entry (SOCE) to raise the intracellular  $\text{Ca}^{2+}$  concentration, which is mediated by the  $\text{Ca}^{2+}$  release-activated  $\text{Ca}^{2+}$  (CRAC) channels. A repetitive or prolonged increase in intracellular  $\text{Ca}^{2+}$  is required for the calcineurin-mediated dephosphorylation of NFAT, which is then able to translocate to the nucleus and activate the target cells (Chapter 3 Figure 3.7). It is well established that the Endoplasmic Reticulum plays an important role both as a  $\text{Ca}^{2+}$  store in the cells and for the rapid release of  $\text{Ca}^{2+}$  upon T cell stimulation (Gwack et al., 2007; Park et al 2020). In this respect, the ATPase  $\text{Ca}^{2+}$  pumps associated with the ER (SERCA) play an important role in maintaining the calcium homeostasis in the cells.

Could be possible that GA plays a similar role as the RE in T cell activation? Our results might support this hypothesis. We have observed that *ATP2C1*-deficient cells have an overall decrease of intracellular  $\text{Ca}^{2+}$ . However, this could be compatible with an increased level of cytosolic  $\text{Ca}^{2+}$ . Indeed, as illustrated in (Figure 6.4), in the absence of the *ATP2C1* pump there should be a lack of  $\text{Ca}^{2+}$  storage in the GA. This, in turn, might result in a relative increase of cytosolic  $\text{Ca}^{2+}$ . Therefore, in *ATP2C1* KO cells, the NFAT will be more efficiently dephosphorylated and translocated into the nucleus, thus resulting in a faster activation of the cells. However, to ascertain this hypothesis a series of new experiments will need to be performed. These include: (i) a more precise measurement of  $\text{Ca}^{2+}$  in the different compartment of the cell; (ii) use a more physiological activation of T cells (e.g. CD3/CD28 antibodies instead of PMA/Ionomycin); (iii) test additional readouts of T-cell activation (IL2, transcriptome; reporter assay); (iv) inactivation and functional studies of *ATP2C1* in primary T cells. If validated, these findings will shed light on a new function of the GA as a regulator of T cell signaling.



**Figure 6.4 Effect of ATP2C1 KO in T-cell activation** Schematic representation of the effect of ATP2C1 KO in T-cell activation. During T-cell activation, the stimulation of immunoreceptors, allows the entry of Ca<sup>2+</sup> into the cytosol, in normal conditions (WT cells) (left panel), the cytosolic Ca<sup>2+</sup> can be regulated by the expression of ATP2C1 pump in the Golgi apparatus (GA). Increased cytosolic Ca<sup>2+</sup> levels mediate the calmodulin-calcineurin-dependent de-phosphorylation of cytosolic NFAT, followed by the translocation of un-phosphorylated NFAT to the nucleus; NFAT binds to the promoters of T-cell activation related genes, promoting its transcription and an activated T-cell state. In the *ATP2C1* KO cells (left panel), depletion of ATP2C1 pump in the GA, impairs the up-take of cytosolic Ca<sup>2+</sup> into GA, generating higher levels and/or slower depuration of Cytosolic Ca<sup>2+</sup> in KO cells, which in turn can increase the rate of NFAT de-phosphorylation and translocation to the nucleus, where it can up-regulate the transcription of T-cell activation related genes as compared to the WT cells, inducing a faster T-cell activation response.

As I previously mentioned, *ATP2C1* is induced by T cell activation. This suggests that an increased level of *ATP2C1* might be required to reduce the level of cytosolic  $\text{Ca}^{2+}$  by pumping it into the GA lumen, and consequently constraint the calcium-mediated T cell signaling (Figure 6.5). We also showed that the P1 promoter is required for the induction of *ATP2C1*, consistent with the fact that P1 is a direct target of NFAT. Therefore, we can hypothesize that P1 is required for the *ATP2C1*-mediated modulation of T cell signaling. To test this hypothesis, T-cell activation experiments should be performed in the P1-deleted clones. In this context, the amount of *ATP2C1* pump will be maintained at a basal level without increase after T cell activation. The prediction will be that the deletion of the P1 promoter might result in “premature” T cell activation as observed with the inactivation of the full *ATP2C1* gene.



**Figure 6.5** Scheme representing the implications of *ATP2C1* in the regulation of T-cell activation. We focused our attention on the  $\text{Ca}^{2+}$  homeostasis mediated by *ATP2C1* since is the subject of this work. In the upper panel, a resting T-cell is shown, basal expression of

ATP2C1 pump located in the membrane of Golgi apparatus (GA) mediates the cytosolic  $\text{Ca}^{2+}$  homeostasis by the generation of intra-Golgi  $\text{Ca}^{2+}$  storage. During early T-cell activation (lower panel left), the activation signals are sensed by immune-receptors (1) which in turn allows the entry of  $\text{Ca}^{2+}$  to the cytosol (2) the increase of cytosolic  $\text{Ca}^{2+}$  mediates the calmodulin-calcineurin-dependent de-phosphorylation of cytosolic NFAT, followed by the translocation of un-phosphorylated NFAT to the nucleus (3), NFAT binds to the alternative promoter (P1) *ATP2C1*, promoting the transcription of the gene (4). ATP2C1 pump is over-expressed in the membrane of the Golgi apparatus (5). In the late T-cell activation (lower panel right), the overexpressed ATP2C1 pump, allows the mobilization of  $\text{Ca}^{2+}$  from the cytosol to the GA, reducing the cytosolic  $\text{Ca}^{2+}$  levels (6), restraining the T-cell activation signal. Overexpression of *ATP2C1* acts as a brake of T-cell activation.

Overall, by starting from the study of alternative promoter regulation in T-ALL, I could identify a novel regulator of T cell activation and highlight a potential new function of the GA as a Calcium store involved in the modulation of T cell signaling.

## 6.6. Long term perspectives

### 6.6.1 The link between Hailey-Hailey disease and immunodeficiencies.

Hailey- Hailey disease (HDD) is an inherited defect caused by loss-of-function mutations in *ATP2C1* producing acantholytic dermatosis. *ATP2C1* dysfunction in HDD results in an inefficient increase of external calcium in the granular layer, failing to stabilize desmosome integrity and activate  $\text{Ca}^{2+}$  sensing receptors; these receptors are responsible for triggering cell-to-cell adhesion, cell differentiation in the granular layer, and reconstituting the  $\text{Ca}^{2+}$  gradient (Brini and Carafoli 2009). In the skin of HDD patients, the Golgi  $\text{Ca}^{2+}$  uptake rate of keratinocytes slowed, and the  $\text{Ca}^{2+}$  level in the Golgi was notably lower (Behne et al., 2003). Reduced  $\text{Ca}^{2+}$  concentration in the Golgi may impair the glycosylation of desmosomes (Van Baelen et al., 2004). Desmosome formation was delayed in *ATP2C1*-deficient keratinocytes, but its assembly may be re-established by being cultured in an elevated  $\text{Ca}^{2+}$  concentration solution (Raiko et al., 2012).

In our study, we found that the CCRF-CEM semi-adherent cell line presented a distinctive phenotype between knock-out and WT conditions. The two *ATP2C1* knock-out clones were less adherent than the wild-type cells (Fig. 2E from the second part of the results), these observations are therefore consistent with the loss of cell adhesion found in the epithelial cells of the HDD patients. More generally, given our finding that *ATP2C1* is required to modulate the immune response, we could speculate that HDD patients could also have immune-related pathologies. Therefore, it could be interesting to assess whether the HDD patients have any comorbidity related to the immune system, such as inflammatory or autoimmune diseases.

### 6.6.2 Inhibitors of ATP2C1 as immunomodulators

It could be possible that inhibitors of the ATP2C1 pump can increase T-cell activation, which could be used to treat some immunodeficiencies and, if its oncogenic role is established, the inhibitors could also reduce the proliferation of cancer cells. Some volatile anesthetics have

been described as inhibitors of the plasma membrane calcium ATPase such as halothane, isoflurane, enflurane, and desflurane (Kosk-Kosicka and Roszczynska 1993). So far, no specific inhibitor for ATP2C1 has been described, therefore it remains an exciting field of research.



# Bibliography

# Bibliography

- Andersson, R., & Sandelin, A. (2020). Determinants of enhancer and promoter activities of regulatory elements. In *Nature Reviews Genetics* (Vol. 21, Issue 2). <https://doi.org/10.1038/s41576-019-0173-8>
- Aramburu, J., Rao, A., & Klee, C. B. (2001). Calcineurin: From structure to function. *Current Topics in Cellular Regulation*, 36(C). [https://doi.org/10.1016/S0070-2137\(01\)80011-X](https://doi.org/10.1016/S0070-2137(01)80011-X)
- Audia, J. E., & Campbell, R. M. (2016). Histone modifications and cancer. *Cold Spring Harbor Perspectives in Biology*, 8(4). <https://doi.org/10.1101/cshperspect.a019521>
- Aung, C. S., Kruger, W. A., Poronnik, P., Roberts-Thomson, S. J., & Monteith, G. R. (2007). Plasma membrane Ca<sup>2+</sup>-ATPase expression during colon cancer cell line differentiation. *Biochemical and Biophysical Research Communications*, 355(4). <https://doi.org/10.1016/j.bbrc.2007.02.050>
- Axelsen, K. B., & Palmgren, M. G. (1998). Evolution of substrate specificities in the P-type ATPase superfamily. *Journal of Molecular Evolution*, 46(1). <https://doi.org/10.1007/PL00006286>
- Bannister, A. J., & Kouzarides, T. (2011). Regulation of chromatin by histone modifications. In *Cell Research* (Vol. 21, Issue 3). <https://doi.org/10.1038/cr.2011.22>
- Belhocine, M., Simonin, M., Abad Flores, J. D., Cieslak, A., Manosalva, I., Pradel, L., Smith, C., Mathieu, E.-L., Charbonnier, G., Martens, J. H. A., Stunnenberg, H. G., Maqbool, M. A., Mikulasova, A., Russell, L. J., Rico, D., Puthier, D., Ferrier, P., Asnafi, V., & Spicuglia, S. (2021). Dynamic of broad H3K4me3 domains uncover an epigenetic switch between cell identity and cancer-related genes. *Genome Research*. <https://doi.org/10.1101/gr.266924.120>
- Bell, A. C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, 405(6785). <https://doi.org/10.1038/35013100>
- Bell, A. C., West, A. G., & Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, 98(3). [https://doi.org/10.1016/S0092-8674\(00\)81967-4](https://doi.org/10.1016/S0092-8674(00)81967-4)
- Belver, L., & Ferrando, A. (2016). The genetics and mechanisms of T cell acute lymphoblastic leukaemia. *Nature Reviews Cancer*, 16(8). <https://doi.org/10.1038/nrc.2016.63>
- Benayoun, B. A., Pollina, E. A., Ucar, D., Mahmoudi, S., Karra, K., Wong, E. D., Devarajan, K., Daugherty, A. C., Kundaje, A. B., Mancini, E., Hitz, B. C., Gupta, R., Rando, T. A., Baker, J. C., Snyder, M. P., Cherry, J. M., & Brunet, A. (2015). Erratum: H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency (Cell (2014) 158:673-688). In *Cell* (Vol. 163, Issue 5). <https://doi.org/10.1016/j.cell.2015.10.051>
- Bernard, D., Quatannens, B., Begue, A., Vandenbunder, B., & Abbadie, C. (2001). Antiproliferative and antiapoptotic effects of cRel may occur within the same cells via the up-regulation of manganese superoxide dismutase. *Cancer Research*, 61(6).

- Berridge, M. J. (1997). The AM and FM of calcium signalling. *Nature*, 386(6627).  
<https://doi.org/10.1038/386759a0>
- Berridge, M. J., Bootman, M. D., & Roderick, H. L. (2003). Calcium signalling: dynamics, homeostasis and remodelling. *Nature Reviews Molecular Cell Biology*, 4(7). <https://doi.org/10.1038/nrm1155>
- Bertipaglia, I., & Carafoli, E. (2007). Calpains and human disease. *Subcellular Biochemistry*, 45.  
[https://doi.org/10.1007/978-1-4020-6191-2\\_2](https://doi.org/10.1007/978-1-4020-6191-2_2)
- Boeger, H., Griesenbeck, J., Strattan, J. S., & Kornberg, R. D. (2004). Removal of promoter nucleosomes by disassembly rather than sliding in vivo. *Molecular Cell*, 14(5).  
<https://doi.org/10.1016/j.molcel.2004.05.013>
- Boyadjiev, S., & Jabs, E. (2000). Online Mendelian Inheritance in Man (OMIM) as a knowledgebase for human developmental disorders. In *Clinical Genetics* (Vol. 57, Issue 4).  
<https://doi.org/10.1034/j.1399-0004.2000.570403.x>
- Bradner, J. E., Hnisz, D., & Young, R. A. (2017). Transcriptional Addiction in Cancer. In *Cell* (Vol. 168, Issue 4). <https://doi.org/10.1016/j.cell.2016.12.013>
- Brini, M., Cali, T., Ottolini, D., & Carafoli, E. (2012). Calcium pumps: Why So many? *Comprehensive Physiology*, 2(2). <https://doi.org/10.1002/cphy.c110034>
- Brini, M., & Carafoli, E. (2009). Calcium Pumps in Health and Disease. *Physiological Reviews*, 89(4).  
<https://doi.org/10.1152/physrev.00032.2008>
- BRODIN, P., FALCHETTO, R., VORHERR, T., & CARAFOLI, E. (1992). Identification of two domains which mediate the binding of activating phospholipids to the plasma-membrane Ca<sup>2+</sup> pump. *European Journal of Biochemistry*, 204(2). <https://doi.org/10.1111/j.1432-1033.1992.tb16715.x>
- Brouland, J. P., Gélébart, P., Kovács, T., Enouf, J., Grossmann, J., & Papp, B. (2005). The loss of sarco/endoplasmic reticulum calcium transport ATPase 3 expression is an early event during the multistep process of colon carcinogenesis. *American Journal of Pathology*, 167(1).  
[https://doi.org/10.1016/S0002-9440\(10\)62968-9](https://doi.org/10.1016/S0002-9440(10)62968-9)
- Brown, C. Y., Mize, G. J., Pineda, M., George, D. L., & Morris, D. R. (1999). Role of two upstream open reading frames in the translational control of oncogene mdm2. *Oncogene*, 18(41).  
<https://doi.org/10.1038/sj.onc.1202949>
- Burge, S. M., & Wilkinson, J. D. (1992). Darier-White disease: A review of the clinical features in 163 patients. *Journal of the American Academy of Dermatology*, 27(1).  
[https://doi.org/10.1016/0190-9622\(92\)70154-8](https://doi.org/10.1016/0190-9622(92)70154-8)
- Callewaert, G., Parys, J. B., de Smedt, H., Raeymaekers, L., Wuytack, F., Vanoevelen, J., van Baelen, K., Simoni, A., Rizzuto, R., & Missiaen, L. (2003). Similar Ca<sup>2+</sup>-signaling properties in keratinocytes and in COS-1 cells overexpressing the secretory-pathway Ca<sup>2+</sup>-ATPase SPCA1. *Cell Calcium*, 34(2). [https://doi.org/10.1016/S0143-4160\(03\)00070-8](https://doi.org/10.1016/S0143-4160(03)00070-8)
- Calo, E., & Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? In *Molecular Cell* (Vol. 49, Issue 5). <https://doi.org/10.1016/j.molcel.2013.01.038>

- Capelson, M., & Corces, V. G. (2004). Boundary elements and nuclear organization. In *Biology of the Cell* (Vol. 96, Issue 8). <https://doi.org/10.1016/j.biolcel.2004.06.004>
- Carrión, A. M., Link, W. A., Ledo, F., Mellström, B., & Naranjo, J. R. (1999). DREAM is a Ca<sup>2+</sup>-regulated transcriptional repressor. *Nature*, *398*(6722). <https://doi.org/10.1038/18044>
- Catterall, W. A. (2000). Structure and regulation of voltage-gated Ca<sup>2+</sup> channels. In *Annual Review of Cell and Developmental Biology* (Vol. 16). <https://doi.org/10.1146/annurev.cellbio.16.1.521>
- Chadwick, L. H. (2012). The NIH Roadmap Epigenomics Program data resource. In *Epigenomics* (Vol. 4, Issue 3). <https://doi.org/10.2217/epi.12.18>
- Chen, K., Chen, Z., Wu, D., Zhang, L., Lin, X., Su, J., Rodriguez, B., Xi, Y., Xia, Z., Chen, X., Shi, X., Wang, Q., & Li, W. (2015). Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nature Genetics*, *47*(10). <https://doi.org/10.1038/ng.3385>
- Chen, L., & Widom, J. (2005). Mechanism of transcriptional silencing in yeast. *Cell*, *120*(1). <https://doi.org/10.1016/j.cell.2004.11.030>
- Choquette, D., Hakim, G., Filoteo, A. G., Plishker, G. A., Bostwick, J. R., & Penniston, J. T. (1984). Regulation of plasma membrane Ca<sup>2+</sup> ATPases by lipids of the phosphatidylinositol cycle. *Biochemical and Biophysical Research Communications*, *125*(3). [https://doi.org/10.1016/0006-291X\(84\)91369-X](https://doi.org/10.1016/0006-291X(84)91369-X)
- Cialfi, S., le Pera, L., de Blasio, C., Mariano, G., Palermo, R., Zonfrilli, A., Uccelletti, D., Palleschi, C., Biolcati, G., Barbieri, L., Screpanti, I., & Talora, C. (2016). The loss of ATP2C1 impairs the DNA damage response and induces altered skin homeostasis: Consequences for epidermal biology in Hailey-Hailey disease. *Scientific Reports*, *6*. <https://doi.org/10.1038/srep31567>
- Cieslak, A., Charbonnier, G., Tesio, M., Mathieu, E. L., Belhocine, M., Touzart, A., Smith, C., Hypolite, G., Andrieu, G. P., Martens, J. H. A., Janssen-Megens, E., Gut, M., Gut, I., Boissel, N., Petit, A., Puthier, D., Macintyre, E., Stunnenberg, H. G., Spicuglia, S., & Asnafi, V. (2020). Blueprint of human thymopoiesis reveals molecular mechanisms of stage-specific TCR enhancer activation. *Journal of Experimental Medicine*, *217*(9). <https://doi.org/10.1084/jem.20192360>
- Clapier, C. R., & Cairns, B. R. (2009). The biology of chromatin remodeling complexes. In *Annual Review of Biochemistry* (Vol. 78). <https://doi.org/10.1146/annurev.biochem.77.062706.153223>
- Clarke, D. M., Loo, T. W., Inesi, G., & MacLennan, D. H. (1989). Location of high affinity Ca<sup>2+</sup>-binding sites within the predicted transmembrane domain of the sarco-plasmic reticulum Ca<sup>2+</sup>-ATPase. *Nature*, *339*(6224). <https://doi.org/10.1038/339476a0>
- Cloos, P. A. C., Christensen, J., Agger, K., & Helin, K. (2008). Erasing the methyl mark: Histone demethylases at the center of cellular differentiation and disease. In *Genes and Development* (Vol. 22, Issue 9). <https://doi.org/10.1101/gad.1652908>
- Coassin, M., Ursini, F., & Bindoli, A. (1992). Antioxidant effect of manganese. *Archives of Biochemistry and Biophysics*, *299*(2). [https://doi.org/10.1016/0003-9861\(92\)90282-2](https://doi.org/10.1016/0003-9861(92)90282-2)

- Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A., & Jaenisch, R. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50). <https://doi.org/10.1073/pnas.1016071107>
- Cui, C., Merritt, R., Fu, L., & Pan, Z. (2017). Targeting calcium signaling in cancer therapy. *Acta Pharmaceutica Sinica B*, *7*(1). <https://doi.org/10.1016/j.apsb.2016.11.001>
- Cullen, P. J. (2003). Calcium signalling: The ups and downs of protein kinase C. In *Current Biology* (Vol. 13, Issue 18). <https://doi.org/10.1016/j.cub.2003.08.041>
- Culotta, V. C., Yang, M., & Hall, M. D. (2005). Manganese transport and trafficking: Lessons learned from *Saccharomyces cerevisiae*. In *Eukaryotic Cell* (Vol. 4, Issue 7). <https://doi.org/10.1128/EC.4.7.1159-1165.2005>
- Cuthbert, G. L., Daujat, S., Snowden, A. W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P. D., Tempst, P., Bannister, A. J., & Kouzarides, T. (2004). Histone deimination antagonizes arginine methylation. *Cell*, *118*(5). <https://doi.org/10.1016/j.cell.2004.08.020>
- D'Acquisto, F., & Crompton, T. (2011). CD3+CD4-CD8- (double negative) T cells: Saviours or villains of the immune response? *Biochemical Pharmacology*, *82*(4). <https://doi.org/10.1016/j.bcp.2011.05.019>
- Dang, D., & Rao, R. (2016). Calcium-ATPases: Gene disorders and dysregulation in cancer. In *Biochimica et Biophysica Acta - Molecular Cell Research* (Vol. 1863, Issue 6). <https://doi.org/10.1016/j.bbamcr.2015.11.016>
- Davies, R., Liu, L., Taotao, S., Tuano, N., Chaturvedi, R., Huang, K. K., Itman, C., Mandoli, A., Qamra, A., Hu, C., Powell, D., Daly, R. J., Tan, P., & Rosenbluh, J. (2021). CRISPRi enables isoform-specific loss-of-function screens and identification of gastric cancer-specific isoform dependencies. *Genome Biology*, *22*(1). <https://doi.org/10.1186/s13059-021-02266-6>
- Davuluri, R. v., Suzuki, Y., Sugano, S., Plass, C., & Huang, T. H. M. (2008). The functional consequences of alternative promoter use in mammalian genomes. In *Trends in Genetics* (Vol. 24, Issue 4). <https://doi.org/10.1016/j.tig.2008.01.008>
- Dawson, M. A., Bannister, A. J., Göttgens, B., Foster, S. D., Bartke, T., Green, A. R., & Kouzarides, T. (2009). JAK2 phosphorylates histone H3Y41 and excludes HP1 $\alpha$  from chromatin. *Nature*, *461*(7265). <https://doi.org/10.1038/nature08448>
- de Klerk, E., & 't Hoen, P. A. C. (2015). Alternative mRNA transcription, processing, and translation: insights from RNA sequencing. *Trends in Genetics*, *31*(3). <https://doi.org/10.1016/j.tig.2015.01.001>
- de la Serna, I. L., Ohkawa, Y., Berkes, C. A., Bergstrom, D. A., Dacwag, C. S., Tapscott, S. J., & Imbalzano, A. N. (2005). MyoD Targets Chromatin Remodeling Complexes to the Myogenin

- Locus Prior to Forming a Stable DNA-Bound Complex. *Molecular and Cellular Biology*, 25(10). <https://doi.org/10.1128/mcb.25.10.3997-4009.2005>
- de Laat, W., & Dekker, J. (2012). 3C-based technologies to study the shape of the genome. In *Methods* (Vol. 58, Issue 3). <https://doi.org/10.1016/j.ymeth.2012.11.005>
- de Meis, L., & Vianna, A. L. (1979). Energy interconversion by the Ca<sup>2+</sup>-dependent ATPase of the sarcoplasmic reticulum. In *Annual review of biochemistry* (Vol. 48). <https://doi.org/10.1146/annurev.bi.48.070179.001423>
- Defossez, P. A., Kelly, K. F., Filion, G. J. P., Pérez-Torrado, R., Magdinier, F., Menoni, H., Nordgaard, C. L., Daniel, J. M., & Gilson, E. (2005). The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. *Journal of Biological Chemistry*, 280(52). <https://doi.org/10.1074/jbc.M510802200>
- Dellis, O., Arbabian, A., Brouland, J. P., Kovács, T., Rowe, M., Chomienne, C., Joab, I., & Papp, B. (2009). Modulation of b-cell endoplasmic reticulum calcium homeostasis by epstein-barr virus latent membrane protein-1. *Molecular Cancer*, 8. <https://doi.org/10.1186/1476-4598-8-59>
- Demircioğlu, D., Cukuroglu, E., Kindermans, M., Nandi, T., Calabrese, C., Fonseca, N. A., Kahles, A., Lehmann, K. van, Stegle, O., Brazma, A., Brooks, A. N., Rättsch, G., Tan, P., & Göke, J. (2019). A Pan-cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative Promoters. *Cell*, 178(6). <https://doi.org/10.1016/j.cell.2019.08.018>
- Denes, H., Brian J, A., Tong lhn, L., Ashley, L., Violaine, S.-A., Alla A, S., Heather A, H., & Richard A, Y. (2013). Super-enhancers in the control of cell identity and disease. *Cell*.
- Denmeade, S. R., Mhaka, A. M., Rosen, D. M., Brennen, W. N., Dalrymple, S., Dach, I., Olesen, C., Gurel, B., DeMarzo, A. M., Wilding, G., Carducci, M. A., Dionne, C. A., Møller, J. v., Nissen, P., Christensen, S. B., & Isaacs, J. T. (2012). Engineering a prostate-specific membrane antigen-activated tumor endothelial cell prodrug for cancer therapy. *Science Translational Medicine*, 4(140). <https://doi.org/10.1126/scitranslmed.3003886>
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D. G., Lagarde, J., Veeravalli, L., Ruan, X., Ruan, Y., Lassmann, T., Carninci, P., Brown, J. B., Lipovich, L., Gonzalez, J. M., ... Guigó, R. (2012). The GENCODE v7 catalog of human long noncoding RNAs: Analysis of their gene structure, evolution, and expression. *Genome Research*, 22(9). <https://doi.org/10.1101/gr.132159.111>
- Diederichs, S. (2014). The four dimensions of noncoding RNA conservation. In *Trends in Genetics* (Vol. 30, Issue 4). <https://doi.org/10.1016/j.tig.2014.01.004>
- Dolmetsch, R. E., Xu, K., & Lewis, R. S. (1998). Calcium oscillations increase the efficiency and specificity of gene expression. *Nature*, 392(6679). <https://doi.org/10.1038/31960>
- Dovey, O. M., Foster, C. T., & Cowley, S. M. (2010). Histone deacetylase 1 (HDAC1), but not HDAC2, controls embryonic stem cell differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, 107(18). <https://doi.org/10.1073/pnas.1000478107>

- Du, Q., Luu, P. L., Stirzaker, C., & Clark, S. J. (2015). Methyl-CpG-binding domain proteins: Readers of the epigenome. In *Epigenomics* (Vol. 7, Issue 6). <https://doi.org/10.2217/epi.15.39>
- Duncan, E. M., Muratore-Schroeder, T. L., Cook, R. G., Garcia, B. A., Shabanowitz, J., Hunt, D. F., & Allis, C. D. (2008). Cathepsin L Proteolytically Processes Histone H3 During Mouse Embryonic Stem Cell Differentiation. *Cell*, *135*(2). <https://doi.org/10.1016/j.cell.2008.09.055>
- Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., Epstein, C. B., Frietze, S., Harrow, J., Kaul, R., Khatun, J., Lajoie, B. R., Landt, S. G., Lee, B. K., Pauli, F., Rosenbloom, K. R., Sabo, P., Safi, A., Sanyal, A., ... Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414). <https://doi.org/10.1038/nature11247>
- Elwess, N. L., Filoteo, A. G., Enyedi, A., & Penniston, J. T. (1997). Plasma membrane Ca<sup>2+</sup> pump isoforms 2a and 2b are unusually responsive to calmodulin and Ca<sup>2+</sup>. *Journal of Biological Chemistry*, *272*(29). <https://doi.org/10.1074/jbc.272.29.17981>
- Falchetto, R., Vorherr, T., Brunner, J., & Carafoli, E. (1991). The plasma membrane Ca<sup>2+</sup> pump contains a site that interacts with its calmodulin-binding domain. *Journal of Biological Chemistry*, *266*(5). [https://doi.org/10.1016/s0021-9258\(18\)49937-1](https://doi.org/10.1016/s0021-9258(18)49937-1)
- Falchetto, R., Vorherr, T., & Carafoli, E. (1992). The calmodulin-binding site of the plasma membrane Ca<sup>2+</sup> pump interacts with the transduction domain of the enzyme. *Protein Science*, *1*(12). <https://doi.org/10.1002/pro.5560011209>
- Fan, L., Li, A., Li, W., Cai, P., Yang, B., Zhang, M., Gu, Y., Shu, Y., Sun, Y., Shen, Y., Wu, X., Hu, G., Wu, X., & Xu, Q. (2014). Novel role of Sarco/endoplasmic reticulum calcium ATPase 2 in development of colorectal cancer and its regulation by F36, a curcumin analog. *Biomedicine and Pharmacotherapy*, *68*(8). <https://doi.org/10.1016/j.biopha.2014.10.014>
- Farber, J. L. (1981). The role of calcium in cell death. *Life Sciences*, *29*(13). [https://doi.org/10.1016/0024-3205\(81\)90670-6](https://doi.org/10.1016/0024-3205(81)90670-6)
- Feng, M., Grice, D. M., Faddy, H. M., Nguyen, N., Leitch, S., Wang, Y., Muend, S., Kenny, P. A., Sukumar, S., Roberts-Thomson, S. J., Monteith, G. R., & Rao, R. (2010). Store-independent activation of orai1 by SPCA2 in mammary tumors. *Cell*, *143*(1). <https://doi.org/10.1016/j.cell.2010.08.040>
- Foggia, L., & Hovnanian, A. (2004). Calcium pump disorders of the skin. In *American Journal of Medical Genetics - Seminars in Medical Genetics: Vol. 131 C* (Issue 1). <https://doi.org/10.1002/ajmg.c.30031>
- Fourrel, G., Magdinier, F., & Gilson, É. (2004). Insulator dynamics and the setting of chromatin domains. In *BioEssays* (Vol. 26, Issue 5). <https://doi.org/10.1002/bies.20028>
- Fujisawa, T., & Filippakopoulos, P. (2017). Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. In *Nature Reviews Molecular Cell Biology* (Vol. 18, Issue 4). <https://doi.org/10.1038/nrm.2016.143>

- Fulton, D. L., Sundararajan, S., Badis, G., Hughes, T. R., Wasserman, W. W., Roach, J. C., & Sladek, R. (2009). TFCat: The curated catalog of mouse and human transcription factors. *Genome Biology*, *10*(3). <https://doi.org/10.1186/gb-2009-10-3-r29>
- Furney, S. J., Higgins, D. G., Ouzounis, C. A., & López-Bigas, N. (2006). Structural and functional properties of genes involved in human cancer. *BMC Genomics*, *7*. <https://doi.org/10.1186/1471-2164-7-3>
- Galupa, R., & Heard, E. (2015). X-chromosome inactivation: New insights into cis and trans regulation. In *Current Opinion in Genetics and Development* (Vol. 31). <https://doi.org/10.1016/j.gde.2015.04.002>
- Gheorghe, M., Sandve, G. K., Khan, A., Chèneby, J., Ballester, B., & Mathelier, A. (2019). A map of direct TF-DNA interactions in the human genome. *Nucleic Acids Research*, *47*(4). <https://doi.org/10.1093/nar/gky1210>
- Gómez-del Arco, P., Kashiwagi, M., Jackson, A. F., Naito, T., Zhang, J., Liu, F., Kee, B., Vooijs, M., Radtke, F., Redondo, J. M., & Georgopoulos, K. (2010). Alternative promoter usage at the Notch1 locus supports ligand-independent signaling in T cell development and leukemogenesis. *Immunity*, *33*(5). <https://doi.org/10.1016/j.immuni.2010.11.008>
- Gomez-Villafuertes, R., Torres, B., Barrio, J., Savignac, M., Gabellini, N., Rizzato, F., Pintado, B., Gutierrez-Adan, A., Mellström, B., Carafoli, E., & Naranjo, J. R. (2005). Downstream regulatory element antagonist modulator regulates Ca<sup>2+</sup> homeostasis and viability in cerebellar neurons. *Journal of Neuroscience*, *25*(47). <https://doi.org/10.1523/JNEUROSCI.3912-05.2005>
- Gräff, J., & Mansuy, I. M. (2008). Epigenetic codes in cognition and behaviour. In *Behavioural Brain Research* (Vol. 192, Issue 1). <https://doi.org/10.1016/j.bbr.2008.01.021>
- Grice, D. M., Vetter, I., Faddy, H. M., Kenny, P. A., Roberts-Thomson, S. J., & Monteith, G. R. (2010). Golgi calcium pump secretory pathway calcium ATPase 1 (SPCA1) is a key regulator of Insulin-like Growth Factor Receptor (IGF1R) processing in the basal-like breast cancer cell line MDA-MB-231. *Journal of Biological Chemistry*, *285*(48). <https://doi.org/10.1074/jbc.M110.163329>
- Gross, D. S., & Garrard, W. T. (1988). Nuclease hypersensitive sites in chromatin. In *Annual Review of Biochemistry* (Vol. 57). <https://doi.org/10.1146/annurev.bi.57.070188.001111>
- Gwack, Y., Feske, S., Srikanth, S., Hogan, P. G., & Rao, A. (2007). Signalling to transcription: Store-operated Ca<sup>2+</sup> entry and NFAT activation in lymphocytes. *Cell Calcium*, *42*(2). <https://doi.org/10.1016/j.ceca.2007.03.007>
- Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. In *Nature Reviews Molecular Cell Biology* (Vol. 19, Issue 10). <https://doi.org/10.1038/s41580-018-0028-8>
- Harris, M. B., Mostecky, J., & Rothman, P. B. (2005). Repression of an interleukin-4-responsive promoter requires cooperative BCL-6 function. *Journal of Biological Chemistry*, *280*(13). <https://doi.org/10.1074/jbc.M412649200>



- HASSELBACH, W., & MAKINOSE, M. (1961). The calcium pump of the “relaxing granules” of muscle and its dependence on ATP-splitting. *Biochemische Zeitschrift*, 333.
- Hegde, M. R., & Crowley, M. R. (2019). Genome and Gene Structure. In *Emery and Rimoin’s Principles and Practice of Medical Genetics and Genomics*. Elsevier. <https://doi.org/10.1016/B978-0-12-812537-3.00004-4>
- Heilker, R., Manning-Krieg, U., Zuber, J. F., & Spiess, M. (1996). In vitro binding of clathrin adaptors to sorting signals correlates with endocytosis and basolateral sorting. *EMBO Journal*, 15(11). <https://doi.org/10.1002/j.1460-2075.1996.tb00650.x>
- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., & Ren, B. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*, 39(3). <https://doi.org/10.1038/ng1966>
- Hinnebusch, A. G., Ivanov, I. P., & Sonenberg, N. (2016a). Translational control by 5'-untranslated regions of eukaryotic mRNAs. In *Science* (Vol. 352, Issue 6292). <https://doi.org/10.1126/science.aad9868>
- Hinnebusch, A. G., Ivanov, I. P., & Sonenberg, N. (2016b). Translational control by 5'-untranslated regions of eukaryotic mRNAs. In *Science* (Vol. 352, Issue 6292). <https://doi.org/10.1126/science.aad9868>
- Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K., & Sharp, P. A. (2017). A Phase Separation Model for Transcriptional Control. In *Cell* (Vol. 169, Issue 1). <https://doi.org/10.1016/j.cell.2017.02.007>
- Hu, Q., Deshpande, S., Irani, K., & Ziegelstein, R. C. (1999). [Ca<sup>2+</sup>]<sub>i</sub> Oscillation frequency regulates agonist-stimulated NF-κB transcriptional activity. *Journal of Biological Chemistry*, 274(48). <https://doi.org/10.1074/jbc.274.48.33995>
- Hu, S., Xie, Z., Onishi, A., Yu, X., Jiang, L., Lin, J., Rho, H. sool, Woodard, C., Wang, H., Jeong, J. S., Long, S., He, X., Wade, H., Blackshaw, S., Qian, J., & Zhu, H. (2009). Profiling the Human Protein-DNA Interactome Reveals ERK2 as a Transcriptional Repressor of Interferon Signaling. *Cell*, 139(3). <https://doi.org/10.1016/j.cell.2009.08.037>
- Hu, Z., Bonifas, J. M., Beech, J., Bench, G., Shigihara, T., Ogawa, H., Ikeda, S., Mauro, T., & Epstein, E. H. (2000). Mutations in ATP2C1, encoding a calcium pump, cause Hailey-Hailey disease. *Nature Genetics*, 24(1). <https://doi.org/10.1038/71701>
- Hughes, T. A. (2006). Regulation of gene expression by alternative untranslated regions. *Trends in Genetics*, 22(3). <https://doi.org/10.1016/j.tig.2006.01.001>
- Hull, E. E., Montgomery, M. R., & Leyva, K. J. (2016). HDAC Inhibitors as Epigenetic Regulators of the Immune System: Impacts on Cancer Therapy and Inflammatory Diseases. In *BioMed Research International* (Vol. 2016). <https://doi.org/10.1155/2016/8797206>

- James, P., Vorherr, T., Krebs, J., Morelli, A., Castello, G., McCormick, D. J., Penniston, J. T., de Flora, A., & Carafoli, E. (1989). Modulation of erythrocyte Ca<sup>2+</sup>-ATPase by selective calpain cleavage of the calmodulin-binding domain. *Journal of Biological Chemistry*, 264(14).  
[https://doi.org/10.1016/s0021-9258\(18\)83181-7](https://doi.org/10.1016/s0021-9258(18)83181-7)
- Jelinic, P., & Shaw, P. (2007). Loss of imprinting and cancer. In *Journal of Pathology* (Vol. 211, Issue 3).  
<https://doi.org/10.1002/path.2116>
- Jenkins, J., Dmitriev, R., & Papkovsky, D. (2016). The Secretory Pathway Ca<sup>2+</sup>/Mn<sup>2+</sup>-Atpase SPCA2 Regulates Mn<sup>2+</sup>-Dependent Cell Cycle Progression in 3D Culture of Colon Cancer Cells. *Biophysical Journal*, 110(3). <https://doi.org/10.1016/j.bpj.2015.11.1806>
- Jenuwein, T., & Allis, C. D. (2001). Translating the histone code. In *Science* (Vol. 293, Issue 5532).  
<https://doi.org/10.1126/science.1063127>
- Jia, Q., Chen, S., Tan, Y., Li, Y., & Tang, F. (2020). Oncogenic super-enhancer formation in tumorigenesis and its molecular mechanisms. In *Experimental and Molecular Medicine* (Vol. 52, Issue 5). <https://doi.org/10.1038/s12276-020-0428-7>
- Jones, P. A., & Baylin, S. B. (2007). The Epigenomics of Cancer. In *Cell* (Vol. 128, Issue 4).  
<https://doi.org/10.1016/j.cell.2007.01.029>
- Kagey, M. H., Newman, J. J., Bilodeau, S., Zhan, Y., Orlando, D. A., van Berkum, N. L., Ebmeier, C. C., Goossens, J., Rahl, P. B., Levine, S. S., Taatjes, D. J., Dekker, J., & Young, R. A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 467(7314).  
<https://doi.org/10.1038/nature09380>
- Kalender Atak, Z., Gianfelici, V., Hulselmans, G., de Keersmaecker, K., Devasia, A. G., Geerdens, E., Mentens, N., Chiaretti, S., Durinck, K., Uyttebroeck, A., Vandenberghe, P., Wlodarska, I., Cloos, J., Foà, R., Speleman, F., Cools, J., & Aerts, S. (2013). Comprehensive Analysis of Transcriptome Variation Uncovers Known and Novel Driver Events in T-Cell Acute Lymphoblastic Leukemia. *PLoS Genetics*, 9(12). <https://doi.org/10.1371/journal.pgen.1003997>
- Khare, S. P., Shetty, A., Biradar, R., Patta, I., Chen, Z. J., Sathe, A. v., Reddy, P. C., Lahesmaa, R., & Galande, S. (2019). NF-κB Signaling and IL-4 Signaling Regulate SATB1 Expression via Alternative Promoter Usage During Th2 Differentiation. *Frontiers in Immunology*, 10.  
<https://doi.org/10.3389/fimmu.2019.00667>
- Khurana, E., Fu, Y., Chakravarty, D., Demichelis, F., Rubin, M. A., & Gerstein, M. (2016). Role of non-coding sequence variants in cancer. In *Nature Reviews Genetics* (Vol. 17, Issue 2).  
<https://doi.org/10.1038/nrg.2015.17>
- Kimura, S., & Mullighan, C. G. (2020). Molecular markers in ALL: Clinical implications. In *Best Practice and Research: Clinical Haematology* (Vol. 33, Issue 3).  
<https://doi.org/10.1016/j.beha.2020.101193>
- Klein-Hessling, S., Rudolf, R., Muhammad, K., Knobloch, K. P., Maqbool, M. A., Cauchy, P., Andrau, J. C., Avots, A., Talora, C., Ellenrieder, V., Screpanti, I., Serfling, E., & Patra, A. K. (2016). A threshold level of NFATc1 activity facilitates thymocyte differentiation and opposes notch-

- driven leukaemia development. *Nature Communications*, 7.  
<https://doi.org/10.1038/ncomms11841>
- Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T. K., Zacarias-Cabeza, J., Spicuglia, S., de La Chapelle, A. L., Heidemann, M., Hintermair, C., Eick, D., Gut, I., Ferrier, P., & Andrau, J. C. (2011). Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Structural and Molecular Biology*, 18(8). <https://doi.org/10.1038/nsmb.2085>
- Kondratskyi, A., Yassine, M., Kondratska, K., Skryma, R., Slomianny, C., & Prevarskaya, N. (2013). Calcium-permeable ion channels in control of autophagy and cancer. *Frontiers in Physiology*, 4 OCT. <https://doi.org/10.3389/fphys.2013.00272>
- Kouzarides, T. (2007). Chromatin Modifications and Their Function. In *Cell* (Vol. 128, Issue 4). <https://doi.org/10.1016/j.cell.2007.02.005>
- Lakshmaiah, K. C., Jacob, L. A., Aparna, S., Lokanatha, D., & Saldanha, S. C. (2014). Epigenetic therapy of cancer with histone deacetylase inhibitors. In *Journal of Cancer Research and Therapeutics* (Vol. 10, Issue 3). <https://doi.org/10.4103/0973-1482.137937>
- Lee, W. J., Roberts-Thomson, S. J., & Monteith, G. R. (2005). Plasma membrane calcium-ATPase 2 and 4 in human breast cancer cell lines. *Biochemical and Biophysical Research Communications*, 337(3). <https://doi.org/10.1016/j.bbrc.2005.09.119>
- Lemon, B., & Tjian, R. (2000). Orchestrated response: A symphony of transcription factors for gene control. In *Genes and Development* (Vol. 14, Issue 20). <https://doi.org/10.1101/gad.831000>
- Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E., & de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, 12(14). <https://doi.org/10.1093/hmg/ddg180>
- Levine, M., Cattoglio, C., & Tjian, R. (2014). Looping back to leap forward: Transcription enters a new era. In *Cell* (Vol. 157, Issue 1). <https://doi.org/10.1016/j.cell.2014.02.009>
- Li, H., Sun, X. K., & Zhu, X. J. (2003). Four novel mutations in ATP2C1 found in Chinese patients with Hailey-Hailey disease. *British Journal of Dermatology*, 149(3). <https://doi.org/10.1046/j.1365-2133.2003.05495.x>
- Li, L., He, S., Sun, J. M., & Davie, J. R. (2004). Gene regulation by Sp1 and Sp3. *Biochemistry and Cell Biology*, 82(4). <https://doi.org/10.1139/o04-045>
- Lopez-Bigas, N., De, S., & Teichmann, S. A. (2008). Functional protein divergence in the evolution of Homo sapiens. *Genome Biology*, 9(2). <https://doi.org/10.1186/gb-2008-9-2-r33>
- Lusser, A., Urwin, D. L., & Kadonaga, J. T. (2005). Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. *Nature Structural and Molecular Biology*, 12(2). <https://doi.org/10.1038/nsmb884>

- MacLennan, D. H. (1970). Purification and properties of an adenosine triphosphatase from sarcoplasmic reticulum. *Journal of Biological Chemistry*, 245(17).  
[https://doi.org/10.1016/s0021-9258\(19\)63820-2](https://doi.org/10.1016/s0021-9258(19)63820-2)
- Mandal, D. M., Wilson, A. F., Elston, R. C., Weissbecker, K., Keats, B. J., & Bailey-Wilson, J. E. (2000). Effects of misspecification of allele frequencies on the type I error rate of model-free linkage analysis. *Human Heredity*, 50(2). <https://doi.org/10.1159/000022900>
- Mandal, D., Woolf, T. B., & Rao, R. (2000). Manganese selectivity of Pmr1, the yeast secretory pathway ion pump, is defined by residue Gln783 in transmembrane segment 6: Residue Asp778 is essential for cation transport. *Journal of Biological Chemistry*, 275(31).  
<https://doi.org/10.1074/jbc.M002619200>
- Maqbool, M. A., Pioger, L., el Aabidine, A. Z., Karasu, N., Molitor, A. M., Dao, L. T. M., Charbonnier, G., van Laethem, F., Fenouil, R., Koch, F., Lacaud, G., Gut, I., Gut, M., Amigorena, S., Joffre, O., Sexton, T., Spicuglia, S., & Andrau, J. C. (2020). Alternative Enhancer Usage and Targeted Polycomb Marking Hallmark Promoter Choice during T Cell Differentiation. *Cell Reports*, 32(7).  
<https://doi.org/10.1016/j.celrep.2020.108048>
- Margueron, R., Trojer, P., & Reinberg, D. (2005). The key to development: Interpreting the histone code? In *Current Opinion in Genetics and Development* (Vol. 15, Issue 2).  
<https://doi.org/10.1016/j.gde.2005.01.005>
- Martens, J. H. A., & Stunnenberg, H. G. (2013). BLUEPRINT: Mapping human blood cell epigenomes. In *Haematologica* (Vol. 98, Issue 10). <https://doi.org/10.3324/haematol.2013.094243>
- Maston, G. A., Evans, S. K., & Green, M. R. (2006). Transcriptional regulatory elements in the human genome. In *Annual Review of Genomics and Human Genetics* (Vol. 7).  
<https://doi.org/10.1146/annurev.genom.7.080505.115623>
- Maunakea, A. K., Nagarajan, R. P., Bilenky, M., Ballinger, T. J., Dsouza, C., Fouse, S. D., Johnson, B. E., Hong, C., Nielsen, C., Zhao, Y., Turecki, G., Delaney, A., Varhol, R., Thiessen, N., Shchors, K., Heine, V. M., Rowitch, D. H., Xing, X., Fiore, C., ... Costello, J. F. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, 466(7303).  
<https://doi.org/10.1038/nature09165>
- McConkey, D. J., & Orrenius, S. (1997). The role of calcium in the regulation of apoptosis. In *Biochemical and Biophysical Research Communications* (Vol. 239, Issue 2).  
<https://doi.org/10.1006/bbrc.1997.7409>
- McGarvey, K. M., Greene, E., Fahrner, J. A., Jenuwein, T., & Baylin, S. B. (2007). DNA methylation and complete transcriptional silencing of cancer genes persist after depletion of EZH2. *Cancer Research*, 67(11). <https://doi.org/10.1158/0008-5472.CAN-06-2029>
- Metze, D., Hamm, H., Schorat, A., & Luger, T. (1996). Involvement of the adherens junction - actin filament system in acantholytic dyskeratosis of Hailey-Hailey disease. A histological, ultrastructural, and histochemical study of lesional and non-lesional skin. *Journal of Cutaneous Pathology*, 23(3). <https://doi.org/10.1111/j.1600-0560.1996.tb01469.x>

- Micaroni, M., Giacchetti, G., Plebani, R., Xiao, G. G., & Federici, L. (2016). ATP2C1 gene mutations in Hailey-Hailey disease and possible roles of SPCA1 isoforms in membrane trafficking. In *Cell Death and Disease* (Vol. 7, Issue 6). <https://doi.org/10.1038/cddis.2016.147>
- Miranda, T. B., & Jones, P. A. (2007). DNA methylation: The nuts and bolts of repression. In *Journal of Cellular Physiology* (Vol. 213, Issue 2). <https://doi.org/10.1002/jcp.21224>
- Monteith, G. R., Davis, F. M., & Roberts-Thomson, S. J. (2012). Calcium channels and pumps in cancer: Changes and consequences. In *Journal of Biological Chemistry* (Vol. 287, Issue 38). <https://doi.org/10.1074/jbc.R112.343061>
- Muscella, A., Calabriso, N., Vetrugno, C., Fanizzi, F. P., de Pascali, S. A., Storelli, C., & Marsigliante, S. (2011). The platinum (II) complex [Pt(O,O'-acac)( $\gamma$ -acac)(DMS)] alters the intracellular calcium homeostasis in MCF-7 breast cancer cells. *Biochemical Pharmacology*, 81(1). <https://doi.org/10.1016/j.bcp.2010.09.012>
- Naito, T., Tanaka, H., Naoe, Y., & Taniuchi, I. (2011). Transcriptional control of T-cell development. *International Immunology*, 23(11). <https://doi.org/10.1093/intimm/dxr078>
- Neph, S., Vierstra, J., Stergachis, A. B., Reynolds, A. P., Haugen, E., Vernot, B., Thurman, R. E., John, S., Sandstrom, R., Johnson, A. K., Maurano, M. T., Humbert, R., Rynes, E., Wang, H., Vong, S., Lee, K., Bates, D., Diegel, M., Roach, V., ... Stamatoyannopoulos, J. A. (2012). An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature*, 489(7414). <https://doi.org/10.1038/nature11212>
- Ng, S. Y., Johnson, R., & Stanton, L. W. (2012). Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO Journal*, 31(3). <https://doi.org/10.1038/emboj.2011.459>
- Niggli, V., Adunyah, E. S., Penniston, J. T., & Carafoli, E. (1981). Purified (Ca<sup>2+</sup>-Mg<sup>2+</sup>)-ATPase of the erythrocyte membrane. Reconstitution and effect of calmodulin and phospholipids. *Journal of Biological Chemistry*, 256(1). [https://doi.org/10.1016/s0021-9258\(19\)70149-5](https://doi.org/10.1016/s0021-9258(19)70149-5)
- Niggli, V., Penniston, J. T., & Carafoli, E. (1979). Purification of the (Ca<sup>2+</sup>-Mg<sup>2+</sup>)-ATPase from human erythrocyte membranes using a calmodulin affinity column. *Journal of Biological Chemistry*, 254(20). [https://doi.org/10.1016/s0021-9258\(19\)86652-8](https://doi.org/10.1016/s0021-9258(19)86652-8)
- Oda, K. (1992). Calcium depletion blocks proteolytic cleavages of plasma protein precursors which occur at the Golgi and/or trans-Golgi network. Possible involvement of Ca<sup>2+</sup>-dependent Golgi endoproteases. *Journal of Biological Chemistry*, 267(24). [https://doi.org/10.1016/s0021-9258\(18\)41949-7](https://doi.org/10.1016/s0021-9258(18)41949-7)
- Odermatt, A., Taschner, P. E. M., Khanna, V. K., Busch, H. F. M., Karpati, G., Jablecki, C. K., Breuning, M. H., & MacLennan, D. H. (1996). Mutations in the gene-encoding SERCA1, the fast-twitch skeletal muscle sarcoplasmic reticulum Ca<sup>2+</sup> ATPase, are associated with Brody disease. *Nature Genetics*, 14(2). <https://doi.org/10.1038/ng1096-191>

- Ogbourne, S., & Antalis, T. M. (1998). Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. In *Biochemical Journal* (Vol. 331, Issue 1). <https://doi.org/10.1042/bj3310001>
- Okano, M., Bell, D. W., Haber, D. A., & Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, *99*(3). [https://doi.org/10.1016/S0092-8674\(00\)81656-6](https://doi.org/10.1016/S0092-8674(00)81656-6)
- Oki, M., Aihara, H., & Ito, T. (2007). Role of histone phosphorylation in chromatin dynamics and its implications in diseases. *Sub-Cellular Biochemistry*, *41*. [https://doi.org/10.1007/1-4020-5466-1\\_14](https://doi.org/10.1007/1-4020-5466-1_14)
- Okunade, G. W., Miller, M. L., Azhar, M., Andringa, A., Sanford, L. P., Doetschman, T., Prasad, V., & Shull, G. E. (2007). Loss of the Atp2c1 Secretory Pathway Ca<sup>2+</sup>-ATPase (SPCA1) in mice causes Golgi stress, apoptosis, and midgestational death in homozygous embryos and squamous cell tumors in adult heterozygotes. *Journal of Biological Chemistry*, *282*(36). <https://doi.org/10.1074/jbc.M703029200>
- Oldershaw, K. A., & Taylor, C. W. (1990). 2,5-Di-(tert-butyl)-1,4-benzohydroquinone mobilizes inositol 1,4,5-trisphosphate-sensitive and -insensitive Ca<sup>2+</sup> stores. *FEBS Letters*, *274*(1–2). [https://doi.org/10.1016/0014-5793\(90\)81366-V](https://doi.org/10.1016/0014-5793(90)81366-V)
- Orphanides, G., Lagrange, T., & Reinberg, D. (1996). The general transcription factors of RNA polymerase II. In *Genes and Development* (Vol. 10, Issue 21). <https://doi.org/10.1101/gad.10.21.2657>
- Parekh, A. B. (2011). Decoding cytosolic Ca<sup>2+</sup> oscillations. *Trends in Biochemical Sciences*, *36*(2). <https://doi.org/10.1016/j.tibs.2010.07.013>
- Parekh, A. B., & Putney, J. W. (2005). Store-Operated Calcium Channels. *Physiological Reviews*, *85*(2). <https://doi.org/10.1152/physrev.00057.2003>
- Pedersen, P. L., & Carafoli, E. (1987). Ion motive ATPases. I. Ubiquity, properties, and significance to cell function. *Trends in Biochemical Sciences*, *12*. [https://doi.org/10.1016/0968-0004\(87\)90071-5](https://doi.org/10.1016/0968-0004(87)90071-5)
- Pekowska, A., Benoukraf, T., Ferrier, P., & Spicuglia, S. (2010). A unique H3K4me2 profile marks tissue-specific gene regulation. *Genome Research*, *20*(11). <https://doi.org/10.1101/gr.109389.110>
- Perissi, V., Aggarwal, A., Glass, C. K., Rose, D. W., & Rosenfeld, M. G. (2004). A Corepressor/Coactivator Exchange Complex Required for Transcriptional Activation by Nuclear Receptors and Other Regulated Transcription Factors. *Cell*, *116*(4). [https://doi.org/10.1016/S0092-8674\(04\)00133-3](https://doi.org/10.1016/S0092-8674(04)00133-3)
- Privalsky, M. L. (2004). The role of corepressors in transcriptional regulation by nuclear hormone receptors. In *Annual Review of Physiology* (Vol. 66). <https://doi.org/10.1146/annurev.physiol.66.032802.155556>

- Ptashne, M., & Gann, A. (1997). Transcriptional activation by recruitment. In *Nature* (Vol. 386, Issue 6625). <https://doi.org/10.1038/386569a0>
- Pundhir, S., Poirazi, P., & Gorodkin, J. (2015). Emerging applications of read profiles towards the functional annotation of the genome. In *Frontiers in Genetics* (Vol. 6, Issue MAY). <https://doi.org/10.3389/fgene.2015.00188>
- Qiao, Y., Shiue, C. N., Zhu, J., Zhuang, T., Jonsson, P., Wright, A. P. H., Zhao, C., & Dahlman-Wright, K. (2015). AP-1-mediated chromatin looping regulates ZEB2 transcription: New insights into TNF $\alpha$ -induced epithelial-mesenchymal transition in triple-negative breast cancer. *Oncotarget*, 6(10). <https://doi.org/10.18632/oncotarget.3158>
- Rahman, S., Magnussen, M., León, T. E., Farah, N., Li, Z., Abraham, B. J., Alapi, K. Z., Mitchell, R. J., Naughton, T., Fielding, A. K., Pizzey, A., Bustraan, S., Allen, C., Popa, T., Pike-Overzet, K., Garcia-Perez, L., Gale, R. E., Linch, D. C., Staal, F. J. T., ... Mansour, M. R. (2017). Activation of the LMO2 oncogene through a somatically acquired neomorphic promoter in T-cell acute lymphoblastic leukemia. *Blood*, 129(24). <https://doi.org/10.1182/blood-2016-09-742148>
- Rangarajan, A., Talora, C., Okuyama, R., Nicolas, M., Mammucari, C., Oh, H., Aster, J. C., Krishna, S., Metzger, D., Chambon, P., Miele, L., Aguet, M., Radtke, F., & Dotto, G. P. (2001). Notch signaling is a direct determinant of keratinocyte growth arrest and entry into differentiation. *EMBO Journal*, 20(13). <https://doi.org/10.1093/emboj/20.13.3427>
- Recillas-Targa, F., Pikaart, M. J., Burgess-Beusse, B., Bell, A. C., Litt, M. D., West, A. G., Gaszner, M., & Felsenfeld, G. (2002). Position-effect protection and enhancer blocking by the chicken  $\beta$ -globin insulator are separable activities. *Proceedings of the National Academy of Sciences of the United States of America*, 99(10). <https://doi.org/10.1073/pnas.102179399>
- Reményi, A., Schöler, H. R., & Wilmanns, M. (2004). Combinatorial control of gene expression. In *Nature Structural and Molecular Biology* (Vol. 11, Issue 9). <https://doi.org/10.1038/nsmb820>
- Rippe, K., Schrader, A., Riede, P., Strohner, R., Lehmann, E., & Längst, G. (2007). DNA sequence- and conformation-directed positioning of nucleosomes by chromatin-remodeling complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 104(40). <https://doi.org/10.1073/pnas.0702430104>
- Robertson, K. D., & Wolffe, A. P. (2000). DNA methylation in health and disease. In *Nature Reviews Genetics* (Vol. 1, Issue 1). <https://doi.org/10.1038/35049533>
- Rudolph, H. K., Antebi, A., Fink, G. R., Buckley, C. M., Dorman, T. E., LeVitre, J. A., Davidow, L. S., Mao, J. i., & Moir, D. T. (1989). The yeast secretory pathway is perturbed by mutations in PMR1, a member of a Ca<sup>2+</sup> ATPase family. *Cell*, 58(1). [https://doi.org/10.1016/0092-8674\(89\)90410-8](https://doi.org/10.1016/0092-8674(89)90410-8)
- Saibil, S. D., & Ohashi, P. S. (2020). Targeting T cell activation in immuno-oncology. In *Current Oncology* (Vol. 27, Issue S2). <https://doi.org/10.3747/co.27.5285>
- Saito, K., Uzawa, K., Endo, Y., Kato, Y., Nakashima, D., Ogawara, K., Shiba, M., Bukawa, H., Yokoe, H., & Tanzawa, H. (2006). Plasma membrane Ca<sup>2+</sup> ATPase isoform 1 down-regulated in human oral cancer. *Oncology Reports*, 15(1). <https://doi.org/10.3892/or.15.1.49>

- Sakabe, K., Wang, Z., & Hart, G. W. (2010).  $\beta$ -N-acetylglucosamine (O-GlcNAc) is part of the histone code. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(46). <https://doi.org/10.1073/pnas.1009023107>
- Salehi, S., Taheri, M. N., Azarpira, N., Zare, A., & Behzad-Behbahani, A. (2017). State of the art technologies to explore long non-coding RNAs in cancer. In *Journal of Cellular and Molecular Medicine* (Vol. 21, Issue 12). <https://doi.org/10.1111/jcmm.13238>
- Salz, T., Deng, C., Pampo, C., Siemann, D., Qiu, Y., Brown, K., & Huang, S. (2015). Histone methyltransferase hSETD1A is a novel regulator of metastasis in breast cancer. *Molecular Cancer Research*, *13*(3). <https://doi.org/10.1158/1541-7786.MCR-14-0389>
- Santella, L., & Carafoli, E. (1997). Calcium signaling in the cell nucleus. *The FASEB Journal*, *11*(13). <https://doi.org/10.1096/fasebj.11.13.9367344>
- Santos-Rosa, H., Kirmizis, A., Nelson, C., Bartke, T., Saksouk, N., Cote, J., & Kouzarides, T. (2009). Histone H3 tail clipping regulates gene expression. *Nature Structural and Molecular Biology*, *16*(1). <https://doi.org/10.1038/nsmb.1534>
- Schatzmann, H. J. (1966). ATP-dependent Ca<sup>+++</sup>-Extrusion from human red cells. *Experientia*, *22*(6). <https://doi.org/10.1007/BF01901136>
- Schmitt, S., Prestel, M., & Paro, R. (2005). Intergenic transcription through a Polycomb group response element counteracts silencing. *Genes and Development*, *19*(6). <https://doi.org/10.1101/gad.326205>
- Schneider, L., Fumagalli, M., & di Fagagna, F. D. A. (2012). Terminally differentiated astrocytes lack DNA damage response signaling and are radioresistant but retain DNA repair proficiency. *Cell Death and Differentiation*, *19*(4). <https://doi.org/10.1038/cdd.2011.129>
- Seet, B. T., Dikic, I., Zhou, M. M., & Pawson, T. (2006). Reading protein modifications with interaction domains. In *Nature Reviews Molecular Cell Biology* (Vol. 7, Issue 7). <https://doi.org/10.1038/nrm1960>
- Seidler, N. W., Jona, I., Vegh, M., & Martonosi, A. (1989). Cyclopiazonic acid is a specific inhibitor of the Ca<sup>2+</sup>-ATPase of sarcoplasmic reticulum. *Journal of Biological Chemistry*, *264*(30). [https://doi.org/10.1016/s0021-9258\(19\)84646-x](https://doi.org/10.1016/s0021-9258(19)84646-x)
- Sertil, O., Kapoor, R., Cohen, B. D., Abramova, N., & Lowry, C. v. (2003). Synergistic repression of anaerobic genes by Mot3 and Rox1 in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, *31*(20). <https://doi.org/10.1093/nar/gkg792>
- Shiio, Y., & Eisenman, R. N. (2003). Histone sumoylation is associated with transcriptional repression. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(23). <https://doi.org/10.1073/pnas.1735528100>
- Shogren-Knaak, M., Ishii, H., Sun, J. M., Pazin, M. J., Davie, J. R., & Peterson, C. L. (2006). Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science*, *311*(5762). <https://doi.org/10.1126/science.1124000>



- Shull, G. E., & Greeb, J. (1988). Molecular cloning of two isoforms of the plasma membrane Ca<sup>2+</sup>-transporting ATPase from rat brain. Structural and functional domains exhibit similarity to Na<sup>+</sup>,K<sup>+</sup>- and other cation transport ATPases. *Journal of Biological Chemistry*, 263(18). [https://doi.org/10.1016/s0021-9258\(18\)68354-1](https://doi.org/10.1016/s0021-9258(18)68354-1)
- Smedler, E., & Uhlén, P. (2014). Frequency decoding of calcium oscillations. In *Biochimica et Biophysica Acta - General Subjects* (Vol. 1840, Issue 3). <https://doi.org/10.1016/j.bbagen.2013.11.015>
- Smith, E., Lin, C., & Shilatifard, A. (2011). The super elongation complex (SEC) and MLL in development and disease. In *Genes and Development* (Vol. 25, Issue 7). <https://doi.org/10.1101/gad.2015411>
- Sobczak, K., & Krzyzosiak, W. J. (2002). Structural determinants of BRCA1 translational regulation. *Journal of Biological Chemistry*, 277(19). <https://doi.org/10.1074/jbc.M109162200>
- Sparber, P., Filatova, A., Khantemirova, M., & Skoblov, M. (2019). The role of long non-coding RNAs in the pathogenesis of hereditary diseases. In *BMC Medical Genomics* (Vol. 12). <https://doi.org/10.1186/s12920-019-0487-6>
- Srinivasan, L., & Atchison, M. L. (2004). YY1 DNA binding and PcG recruitment requires CtBP. *Genes and Development*, 18(21). <https://doi.org/10.1101/gad.1228204>
- Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. In *Nature* (Vol. 403, Issue 6765). <https://doi.org/10.1038/47412>
- Strehler, E. E., & Zacharias, D. A. (2001). Role of alternative splicing in generating isoform diversity among plasma membrane calcium pumps. In *Physiological Reviews* (Vol. 81, Issue 1). <https://doi.org/10.1152/physrev.2001.81.1.21>
- Strigrow, F., & Ehrlich, B. E. (1996). Ligand-gated calcium channels inside and out. *Current Opinion in Cell Biology*, 8(4). [https://doi.org/10.1016/S0955-0674\(96\)80025-1](https://doi.org/10.1016/S0955-0674(96)80025-1)
- Sugiyama, K., Sugiura, K., Hara, T., Sugimoto, K., Shima, H., Honda, K., Furukawa, K., Yamashita, S., & Urano, T. (2002). Aurora-B associated protein phosphatases as negative regulators of kinase activation. *Oncogene*, 21(20). <https://doi.org/10.1038/sj.onc.1205432>
- Suzuki, H. I., Young, R. A., & Sharp, P. A. (2017). Super-Enhancer-Mediated RNA Processing Revealed by Integrative MicroRNA Network Analysis. *Cell*, 168(6). <https://doi.org/10.1016/j.cell.2017.02.015>
- Takai, D., & Jones, P. A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proceedings of the National Academy of Sciences of the United States of America*, 99(6). <https://doi.org/10.1073/pnas.052410099>
- Thastrup, O., Cullen, P. J., Drobak, B. K., Hanley, M. R., & Dawson, A. P. (1990). Thapsigargin, a tumor promoter, discharges intracellular Ca<sup>2+</sup> stores by specific inhibition of the endoplasmic reticulum Ca<sup>2+</sup>-ATPase. *Proceedings of the National Academy of Sciences of the United States of America*, 87(7). <https://doi.org/10.1073/pnas.87.7.2466>

- The GTEx Consortium. (2013). The Genotype-Tissue Expression (GTEx) project The GTEx Consortium\* Abstract. *Database: National Center for Biomedical Information*, 45(6).
- Toyoshima, C. (2008). Structural aspects of ion pumping by Ca<sup>2+</sup>-ATPase of sarcoplasmic reticulum. In *Archives of Biochemistry and Biophysics* (Vol. 476, Issue 1). <https://doi.org/10.1016/j.abb.2008.04.017>
- Toyoshima, C., Nakasako, M., Nomura, H., & Ogawa, H. (2000). Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature*, 405(6787). <https://doi.org/10.1038/35015017>
- Toyoshima, C., & Nomura, H. (2002). Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature*, 418(6898). <https://doi.org/10.1038/nature00944>
- Trimarchi, T., Bilal, E., Ntziachristos, P., Fabbri, G., Dalla-Favera, R., Tsigos, A., & Aifantis, I. (2014). Genome-wide Mapping and Characterization of Notch-Regulated Long Noncoding RNAs in Acute Leukemia. *Cell*, 158(3). <https://doi.org/10.1016/j.cell.2014.05.049>
- Tyagi, M., Imam, N., Verma, K., & Patel, A. K. (2016). Chromatin remodelers: We are the drivers!! In *Nucleus* (Vol. 7, Issue 4). <https://doi.org/10.1080/19491034.2016.1211217>
- van Baelen, K., Dode, L., Vanoevelen, J., Callewaert, G., de Smedt, H., Missiaen, L., Parys, J. B., Raeymaekers, L., & Wuytack, F. (2004). The Ca<sup>2+</sup>/Mn<sup>2+</sup> pumps in the Golgi apparatus. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1742(1–3). <https://doi.org/10.1016/j.bbamcr.2004.08.018>
- van Baelen, K., Vanoevelen, J., Missiaen, L., Raeymaekers, L., & Wuytack, F. (2001). The Golgi PMR1 P-type ATPase of *Caenorhabditis elegans*: Identification of the gene and demonstration of calcium and manganese transport. *Journal of Biological Chemistry*, 276(14). <https://doi.org/10.1074/jbc.M010553200>
- Vanoevelen, J., Dode, L., van Baelen, K., Fairclough, R. J., Missiaen, L., Raeymaekers, L., & Wuytack, F. (2005). The secretory pathway Ca<sup>2+</sup>/Mn<sup>2+</sup>-ATPase 2 is a Golgi-localized pump with high affinity for Ca<sup>2+</sup> ions. *Journal of Biological Chemistry*, 280(24). <https://doi.org/10.1074/jbc.M501026200>
- Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., & Luscombe, N. M. (2009). A census of human transcription factors: Function, expression and evolution. In *Nature Reviews Genetics* (Vol. 10, Issue 4). <https://doi.org/10.1038/nrg2538>
- Venkatachalam, K., & Montell, C. (2007). TRP Channels. *Annual Review of Biochemistry*, 76(1). <https://doi.org/10.1146/annurev.biochem.75.103004.142819>
- Verma, A. K., Filoteo, A. G., Stanford, D. R., Wieben, E. D., Penniston, J. T., Strehler, E. E., Fischer, R., Heim, R., Vogel, G., Mathews, S., Strehler-Page, M. A., James, P., Vorherr, T., Krebs, J., & Carafoli, E. (1988). Complete primary structure of a human plasma membrane Ca<sup>2+</sup> pump. *Journal of Biological Chemistry*, 263(28). [https://doi.org/10.1016/s0021-9258\(18\)68198-0](https://doi.org/10.1016/s0021-9258(18)68198-0)

- Vermezovic, J., Adamowicz, M., Santarpia, L., Rustighi, A., Forcato, M., Lucano, C., Massimiliano, L., Costanzo, V., Bicciato, S., del Sal, G., & D'Adda Di Fagagna, F. (2015). Notch is a direct negative regulator of the DNA-damage response. *Nature Structural and Molecular Biology*, 22(5). <https://doi.org/10.1038/nsmb.3013>
- Vilar, J. M. G., & Saiz, L. (2005). DNA looping in gene regulation: From the assembly of macromolecular complexes to the control of transcriptional noise. In *Current Opinion in Genetics and Development* (Vol. 15, Issue 2). <https://doi.org/10.1016/j.gde.2005.02.005>
- Wasserman, W. W., & Sandelin, A. (2004). Applied bioinformatics for the identification of regulatory elements. In *Nature Reviews Genetics* (Vol. 5, Issue 4). <https://doi.org/10.1038/nrg1315>
- Wei, Y., Chen, J., Rosas, G., Tompkins, D. A., Holt, P. A., & Rao, R. (2000). Phenotypic screening of mutations in Pmr1, the yeast secretory pathway Ca<sup>2+</sup>/Mn<sup>2+</sup>-ATPase, reveals residues critical for ion selectivity and transport. *Journal of Biological Chemistry*, 275(31). <https://doi.org/10.1074/jbc.M002618200>
- West, A. G., & Fraser, P. (2005). Remote control of gene transcription. In *Human Molecular Genetics* (Vol. 14, Issue SPEC. ISS. 1). <https://doi.org/10.1093/hmg/ddi104>
- West, D. W., & Clegg, R. A. (1984). Casein kinase activity in rat mammary gland Golgi vesicles. Demonstration of latency and requirement for a transmembrane ATP carrier. *Biochemical Journal*, 219(1). <https://doi.org/10.1042/bj2190181>
- Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., & Young, R. A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, 153(2). <https://doi.org/10.1016/j.cell.2013.03.035>
- Wilson, A. S., Power, B. E., & Molloy, P. L. (2007). DNA hypomethylation and human diseases. In *Biochimica et Biophysica Acta - Reviews on Cancer* (Vol. 1775, Issue 1). <https://doi.org/10.1016/j.bbcan.2006.08.007>
- Xiang, M., Mohamalawari, D., & Rao, R. (2005). A novel isoform of the secretory pathway Ca<sup>2+</sup>, Mn<sup>2+</sup>-ATPase, hSPCA2, has unusual properties and is expressed in the brain. *Journal of Biological Chemistry*, 280(12). <https://doi.org/10.1074/jbc.M413116200>
- Yan, F., Mo, X., Liu, J., Ye, S., Zeng, X., & Chen, D. (2017). Thymic function in the regulation of T cells, and molecular mechanisms underlying the modulation of cytokines and stress signaling (Review). In *Molecular Medicine Reports* (Vol. 16, Issue 5). <https://doi.org/10.3892/mmr.2017.7525>
- Yan, J., Enge, M., Whittington, T., Dave, K., Liu, J., Sur, I., Schmierer, B., Jolma, A., Kivioja, T., Taipale, M., & Taipale, J. (2013). XTranscription factor binding in human cells occurs in dense clusters formed around cohesin anchor sites. *Cell*, 154(4). <https://doi.org/10.1016/j.cell.2013.07.034>
- Yang, X. J., & Seto, E. (2007). HATs and HDACs: From structure, function and regulation to novel strategies for therapy and prevention. In *Oncogene* (Vol. 26, Issue 37). <https://doi.org/10.1038/sj.onc.1210599>

- Yang, X. J., & Seto, E. (2008). The Rpd3/Hda1 family of lysine deacetylases: From bacteria and yeast to mice and men. In *Nature Reviews Molecular Cell Biology* (Vol. 9, Issue 3). <https://doi.org/10.1038/nrm2346>
- Zacarías-Cabeza, J., Belhocine, M., Vanhille, L., Cauchy, P., Koch, F., Pekowska, A., Fenouil, R., Bergon, A., Gut, M., Gut, I., Eick, D., Imbert, J., Ferrier, P., Andrau, J.-C., & Spicuglia, S. (2015). Transcription-Dependent Generation of a Specialized Chromatin Structure at the TCR $\beta$  Locus. *The Journal of Immunology*, 194(7). <https://doi.org/10.4049/jimmunol.1400789>
- Zhao, P., Zhong, W., Ying, X., Yuan, Z., Fu, J., & Zhou, Z. (2008). Manganese chloride-induced G0/G1 and S phase arrest in A549 cells. *Toxicology*, 250(1). <https://doi.org/10.1016/j.tox.2008.05.016>
- Zhao, Z., & Shilatifard, A. (2019). Epigenetic modifications of histones in cancer. In *Genome Biology* (Vol. 20, Issue 1). <https://doi.org/10.1186/s13059-019-1870-5>
- Zhu, H., Wang, G., & Qian, J. (2016). Transcription factors as readers and effectors of DNA methylation. *Nature Reviews Genetics*, 17(9). <https://doi.org/10.1038/nrg.2016.83>
- Zvaritch, E., Modyanov, N., Carafoli, E., James, P., Vorherr, T., & Falchetto, R. (1990). Mapping of Functional Domains in the Plasma Membrane Ca<sup>2+</sup> Pump Using Trypsin Proteolysis. *Biochemistry*, 29(35). <https://doi.org/10.1021/bi00487a012>

# ANNEX I



## Dynamic of broad H3K4me3 domains uncover an epigenetic switch between cell identity and cancer-related genes

Mohamed Belhocine, Mathieu Simonin, José David Abad Flores, et al.

*Genome Res.* published online June 23, 2021

Access the most recent version at doi:[10.1101/gr.266924.120](https://doi.org/10.1101/gr.266924.120)

---

<b>P&lt;P</b>	Published online June 23, 2021 in advance of the print journal.
<b>Accepted Manuscript</b>	Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.
<b>Creative Commons License</b>	This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <a href="https://genome.cshlp.org/site/misc/terms.xhtml">https://genome.cshlp.org/site/misc/terms.xhtml</a> ). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <a href="http://creativecommons.org/licenses/by-nc/4.0/">http://creativecommons.org/licenses/by-nc/4.0/</a> .
<b>Email Alerting Service</b>	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or <a href="#">click here</a> .

---

---

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Published by Cold Spring Harbor Laboratory Press

# ANNEX II

Symbol	No TSS	Summary of gene expression pattern	Reference
<i>IGF2</i>	4	The human insulin-like growth factor-II ( <i>IGF2</i> ) gene, which is transcribed from four promoters, P1-P4, is imprinted in the fetal liver but bi-allelic expression occurs in the adult liver. The fetal liver uses primarily promoters P3 and, P4 however adult liver transcribes <i>IGF2</i> from promoter P1.	Vu and Hoffman 1994
<i>HTR3B</i>	2	The <i>HTR3B</i> gene codes for the subunit B of the serotonin receptor type 3. Two alternative promoters control the expression of different <i>HTR3B</i> transcripts in the peripheral and central nervous systems. The transcription start site of P1 has been found to control the gene expression in the brain and the transcription start sites of promoter P2 have been observed in the intestine.	Tzvetkov <i>et al.</i> , 2007
<i>SLC7A7</i>	2	The human <i>SLC7A7</i> gene is mainly expressed at the basolateral membrane of the polarized epithelial cells in the renal tubules and the small intestine. The alternative promoter of <i>SLC7A7</i> gene is differentially expressed in the brain (P1).	Puomila <i>et al.</i> , 2007
<i>HRH1</i>	3	Three separate promoters lead to the human histamine H1 receptor ( <i>HRH1</i> ). <i>HRH1</i> is differentially expressed in primary cultured human airway smooth muscle (HASM) cells (P1), primary cultured human bronchial epithelial cells and bronchial epithelial cell line, and other tissues (brain) are known to express histamine H1 receptors (P1, and P3).	Swan <i>et al.</i> , 2006
<i>RUNX1</i>	3	Alternative promoters transcribe many mRNA isoforms that are differentially expressed. P1 transcribes two isoforms (b and c), leading to the production of two distinct proteins with a variety of biological functions. The switching of the promoters controls the expression of <i>RUNX1</i> during embryonic hematopoiesis. Disruption of P2 by the 12:21 chromosomal translocation results in the most	Levanon <i>et al.</i> , 1996; Pozner <i>et al.</i> , 2007; Pui <i>et al.</i> , 2001



		common subtype of childhood acute lymphoblastic leukemia.	
<i>RUNX2</i>	2	Predominantly expressed in bone, colon, heart, tonsil, head/neck, lung, and ovary. Expression was also reported in B and T cells of various developmental stages. P1 transcribes two isoforms (a and b), leading to the production of two distinct proteins with a variety of biological functions. The isoform (c) transcribed by P2 encodes a protein with a shorter and distinct N-terminus when it is compared to isoform a. Aberrant expression was also reported in some tumors (e.g. adenocarcinoma, colon tumor).	Okumura <i>et al.</i> , 2007
<i>RUNX3</i>	2	<i>RUNX3</i> , predominantly expressed in hematopoietic cells, is a tumor suppressor gene that is frequently deleted or transcriptionally silenced in cancer. Multiple transcript variants driven by two distinct promoters encoding different isoforms have been found for this gene.	Okumura <i>et al.</i> , 2007
<i>RGS4</i>	4	<i>RGS4</i> expression in the human brain is spatially and temporally regulated in the dorsolateral prefrontal and visual cortex, through differential transcription of five different isoforms from four alternative promoters.	Ding <i>et al.</i> , 2007
<i>FMO1</i>	3	The use of three alternative promoters regulates the tissue-specific transcription of <i>FMO1</i> in humans. <i>FMO1</i> is silenced postnatally in the liver, but not in the kidney. The transcription of the gene in the fetal human liver is exclusively from the P1 promoter, whereas in extra-hepatic tissues, P2 and P3 are active.	Shephard <i>et al.</i> , 2007
<i>PDE4B</i>	7	<i>PDE4B</i> , linked to schizophrenia in humans, transcribes many isoforms driven by at least seven distinct promoters.	Cheung <i>et al.</i> , 2007
<i>PRL</i>	2	The pituitary hormone prolactin (PRL), best known for its role in the regulation of lactation, is transcribed by two different promoters that regulate pituitary versus extra pituitary expression of prolactin in primates.	Gerlo <i>et al.</i> , 2006

<i>IKBK</i>	2	Two alternative first exons, one is housekeeping required for proper expression and the other is active in cells of hepatic origin at a tissue-specific site.	Fusco <i>et al.</i> , 2006
<i>ART3</i>	2	<i>ART3</i> expression in human macrophages, testis, semen, tonsil, heart, and skeletal muscle appears to be governed by a combination of differential splicing and tissue-preferential use of two alternative promoters.	Friedrich <i>et al.</i> , 2006
<i>CDC2</i>	2	In humans two transcripts exist for <i>CDC2</i> , one including and one excluding the untranslated first exon, both result in the same protein.	Veerla and Hoglund 2006
<i>PTHrP</i>	2	Three alternative promoters have been found in this gene. P3- initiated transcripts were detectable in most tumors, whereas transcripts initiated by either P1 or P2 were present in only a subset of tumors.	Richard <i>et al.</i> , 2003
<i>NAT1</i>	2	Most mRNAs of <i>NAT1</i> gene originates at the promoter, P1, an alternative <i>NAT1</i> promoter designated P3, which is the most active in some tissues, like the kidney, liver, lung, and trachea.	Husain A <i>et al.</i> , 2007
<i>CD36</i>	5	<i>CD36</i> gene has 5 alternative first exons. The alternative transcripts are all expressed in more than one human tissue and their expression patterns vary highly in skeletal muscle, heart, liver, adipose tissue, placenta, spinal cord, cerebrum, and monocytes.	Andersen <i>et al.</i> , 2006
<i>AFP</i>	2	Like the traditional <i>AFP</i> promoter (P1), the alternative promoter (P2) is active in the yolk sac and fetal liver and contributes to the early expression of the <i>AFP</i> gene.	Schoy <i>et al.</i> , 2000
<i>AQP4</i>	2	The aquaporin-4 ( <i>AQP4</i> ) gene encodes two proteins isoforms. Both protein isoforms are expressed in the brain, whereas mainly the smaller isoform is found in other tissues. However differential transcriptional regulation and tissue-specific factors regulate their relative expression by using alternative promoters.	Umenishi and Verkman 1998

<i>MITF</i>	5	<i>MITF</i> consists of 4 widely spaced multiple promoters, which generate not only the diversity in the transcriptional regulation of these promoters but also the structurally different isoforms. The 5'-flanking regions of these isoform-specific exons are termed A, H, B, and M promoters, respectively. Among these promoters, the M promoter has received particular attention, because it is functional only in melanocyte-lineage cells and is upregulated by Wnt signaling via the functional LEF-1-binding site. In contrast to <i>MITF-M</i> , other <i>MITF</i> isoforms are widely expressed in many cell types.	Shibahara <i>et al.</i> , 2001
<i>AC133</i>	5	Transcription of <i>AC133</i> (human stem cell surface protein) isoforms is controlled by 5 different alternative promoters in a tissue-dependent manner, where exon 1A-containing <i>AC133</i> transcript was specifically expressed in human CD34+ cord blood cells.	Shmelkov <i>et al.</i> , 2004
<i>NOTCH1</i>	5	Stage-specific activation of the Notch1 promoters may be fundamental for modulating levels of Notch signaling during development and leukemogenesis. The induction of promoters expressing isoforms with differential ligand requirement may support a feed-forward mechanism that augments Notch signaling required for expansion of immature thymocytes.	Gómez del Arco <i>et al.</i> , 2010
<i>BBOX1</i>	3	The transcription initiation of the human <i>BBOX1</i> gene might occur at 3 different exons, and the expression level of each type of transcript is organ-specific.	Rigault <i>et al.</i> , 2006
<i>KLK11</i>	3	Tissue-specific use of multiple promoters regulates the expression and intracellular trafficking of <i>KLK11</i> /hippostatin isoforms.	Mitsui <i>et al.</i> , 2006
<i>GR</i>	7	Alternative first exons each under the control of specific transcription factors control both the tissue-specific glucocorticoid receptor ( <i>GR</i> ) gene expression and are involved in the tissue-specific <i>GR</i> transcriptional response to stimulation.	Turner <i>et al.</i> , 2006

<i>CYP19</i>	5	Promoters have different tissue specificity and one human promoter is of retroviral origin.	Golovine et al., 2003
<i>GATA-2</i>	2	Differential expression; one promoter is predominantly erythroid; the second is housekeeping.	Pan et al., 2000
<i>NOS1</i>	9	Different first exons have different tissue specificity and translation efficiency. Promoter usage influences alternative splicing. The N-truncated form is from an intronic promoter and is also expressed in human testis.	Wang Y et al., 1997
<i>P21</i>	2	Different functions; p21B induces apoptosis whereas p21 induces cell cycle arrest.	Nozell and Chen 2002
<i>HNF4<math>\alpha</math></i>	2	Different activity and expression; one isoform is predominantly expressed in the islet of Langerhans whereas the other is expressed in the liver and kidney; one isoform has reduced activity.	Nakhei et al., 1998
<i>PPAR<math>\gamma</math></i>	4	The use of four promoters results in two isoforms with different tissue specificity.	Fajas et al., 1997
<i>TP73</i>	2	P73 gene has two independent promoters, which produce isoforms, TAp73 and DeltaTAp73, where TAp73 is a tumor suppressor, DeltaTAp73 is oncogenic.	Muller et al., 2006
<i>P63</i>	2	P63 has TAp63 and deltaNp63 isoforms derived from alternative promoters. Different usage of the promoters is observed in various cancers.	Muller et al., 2006
<i>ALK</i>	2	An alternative isoform of the anaplastic lymphoma kinase (ALK) is expressed in ~ 11% of melanomas and sporadically in other human cancer types, but not in normal tissues.	Wiesner et al., 2015

<i>MEST</i>	2	<i>MEST</i> , an imprinted gene, is only transcribed from the paternal allele. However, an isoform from a different promoter of a distinct first exon is expressed from both paternal and maternal alleles in some cancer tissues.	Pedersen <i>et al.</i> , 2002
<i>MET</i>	2	Gastric cancer exhibited tumor-specific expression of the MET receptor via an internal cryptic promoter, producing a truncated isoform lacking the amino-terminal Sema domain, which regulates receptor dimerization and signaling.	Muratani et al., 2013