Doctorate program
Milan
EXPERIMENTAL
MEDICINE

Università degli Studi di Milano

PhD Course in
Experimental Medicine

CYCLE XXXVII

PhD thesis

CHARACTERIZATION OF THE EPIGENOMIC PROFILE SHAPING HUMAN FOXP3^{+} REGULATORY T CELL IDENTITY IN THE TUMOR MICROENVIRONMENT

Candidate: **Dr. Carolina Dossena**

Matr. R13365

Tutor: **Dr. Michaela Fakiola**

Director: **Prof. Massimiliano Pagani**

Academic Year 2023-2024

SUMMARY

ABSTRACT

One of the hallmarks of cancer is the evasion of the anti-tumor immune response. Cancer cells orchestrate several mechanisms to grow unchecked while promoting a suppressive tumor microenvironment (TME). In this context, $CD4^+$ $FOXP3^+$ regulatory T cells (Tregs) have emerged as a major immunosuppressive population sustaining cancer progression. In physiological conditions they mediate immune tolerance and maintain balance in immune responses, controlling their intensity and timing. Moreover, Tregs can acquire specific phenotypes to modulate distinct immunological contexts and regulate tissue homeostasis through mechanisms that extend beyond their canonical immune functions.

Tumor-infiltrating Treg cells (TI-Tregs) are frequently found in solid tumors, and their ratio compared to $CD8^+$ effector T cells often correlates with poor patient prognosis. This makes the selective depletion or reprogramming of TI-Tregs a promising strategy in cancer immunotherapy, aiming to enhance anti-tumor immunity without disrupting Treg-dependent homeostasis in healthy tissues. Previous studies have revealed specific transcriptional programs of TI-Tregs compared to other immune subsets, including Tregs in the blood (PB-Tregs) and normal tissues adjacent to tumors (NAT-Tregs). These findings pave the way for discovering novel TI-Treg-specific targets. However, a deeper molecular profiling of this Treg cell state, including their regulatory landscape shaped by TME cues, is required to develop novel immunotherapy approaches.

Here, we focus on epigenomic rewiring as a key process that orchestrates Treg adaptation and the acquisition of a tumor-promoting phenotype. The integration of external signals dynamically alters various layers of epigenomic regulation, mediating the establishment of tailored gene expression programs. In this regard, enhancers and DNA-binding transcription factors (TFs) are pivotal in forming specific regulatory networks.

To map the TI-Treg epigenomic profile, $CD4^+$ $CD25^+$ $CD127^-$ Tregs were isolated from the blood (PB), tumors, and normal adjacent tissue (NAT) of colorectal cancer (CRC) and non-small cell lung cancer (NSCLC) patients. Chromatin accessibility and different histone modifications were profiled using ATAC-seq and ChIPmentation-seq, respectively, to reconstruct the genome-wide active and repressed chromatin states and to precisely map open chromatin regions with enhancer features. Additionally,

single-cell RNA-seq data were leveraged to discover bidirectionally transcribed noncoding enhancer RNAs (eRNAs), another feature marking active enhancers and delineating the portion of the enhancer region most likely mediating enhancer function and TF binding. TF motif analysis and *in silico* TF footprinting were employed to assess TF binding dynamics and differential regulatory networks across Treg cell states.

We map the comprehensive enhancerome of human Tregs through the integration of multiple epigenomic features assessed in independent omics datasets. This atlas includes 14,158 *bona fide* active enhancers in TI-Tregs and in normal tissue-Tregs. Notably, our combined multi-omics approach highlights that enhancers comprise a minor fraction of open chromatin regions in Tregs, underscoring the advantage of our approach in robustly annotating active enhancers compared to other accessible regions.

We reveal distinct open chromatin landscapes across PB-, NAT-, and TI-Treg cell states, with altered accessibility in cis-regulatory elements including enhancers. Focusing on TI-Treg chromatin rewiring, we find that epigenomic changes at enhancer level reflect transcriptional differences with respect to NAT-Tregs. Moreover, the existing specificities between these two Treg populations, confirmed by our findings, are driven at a much greater extent by enhancers as opposed to promoters.

Integration of epigenomic and transcriptomic data uncover different regulatory programs involving coordinated changes in enhancer activity and target gene expression across Treg cell states. We find diverse axes of Treg epigenomic regulation within the tumor. On one hand, the activation program, shared with NAT-Tregs, is further sustained in the highly suppressive TI-Tregs. On the other hand, a TI-Treg private regulatory program results from the adaptation to TME cues and involves potentially specific metabolic alterations.

Interestingly, distinct Treg regulatory circuits are driven by both shared and specific TF families. We confirm BATF as a crucial regulator of the activation program, shared between TI- and NAT-Tregs. Additionally, we uncover potentially relevant TFs for TI-Treg specific adaptation, with factors from non-canonical NFkB pathway among the top drivers.

Overall, comprehensive understanding and dissection of multilayer epigenomic regulation can inform on novel targets for specific modulation of Tregs in the TME, with potential clinical relevance for the design of novel immunotherapies.

DISCLOSURE OF RESEARCH INTEGRITY

This research was conducted in strict adherence to the principles and values outlined in the European Code of Conduct for Research Integrity, including reliability, rigor, honesty, respect, and transparency. All experimental procedures involving human cells derived from patients were performed with the highest ethical standards. Consent was obtained from all patients, and their confidentiality was maintained throughout the study. The design, execution, analysis, and reporting of this research were carried out with accuracy, ensuring the reproducibility and reliability of the results. At every stage of the project, data collection and analysis were performed with transparency, and any potential limitations were clearly acknowledged. This statement is made in consultation with my Supervisor, confirming that the research conducted is in full compliance with ethical guidelines and research integrity policies.

ABBREVIATIONS

ATAC-seq	Assay for Transposase-Accessible Chromatin with high-throughput sequencing
BAM	Binary Alignment Map
BtcEnh	Bidirectionally transcribed Enhancer
ChIPmentation-seq	Chromatin Immunoprecipitation followed by Tn5 transposase-mediated tagmentation sequencing
CRC	Colorectal Cancer
DAR	Differentially Accessible Region
DEG	Differentially Expressed Gene
eRNA	Enhancer RNA
FACS	Fluorescence-Activated Cell Sorting
FDR	False Discovery Rate
FLD	Fragment Length Distribution
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
FRiP	Fraction of Reads in Peaks
GSEA	Gene Set Enrichment Analysis
H3K27ac	Histone 3 lysine 27 acetylation
H3K27me3	Histone 3 lysine 27 tri-methylation
H3K36me3	Histone 3 lysine 36 tri-methylation
H3K4me1	Histone 3 lysine 4 mono-methylation
H3K4me3	Histone 3 lysine 4 tri-methylation
HiChIP	High-throughput Chromatin Conformation Capture with Immunoprecipitation
HVG	Highly Variable Gene

NAT	Normal-Adjacent Tissue
NAT-Treg	Normal-Adjacent Tissue Regulatory T cell
NES	Normalized Enrichment Score
NFR	Nucleosome-Free Region
NSCLC	Non-Small Cell Lung Cancer
OCR	Open Chromatin Region
PB	Peripheral Blood
PB-Treg	Peripheral Blood Regulatory T cell
PC	Principal Component
PCA	Principal Component Analysis
scRNA-seq	Single-cell RNA sequencing
Tconv	$CD4^+$ Conventional T cell
TES	Transcription End Site
Tex	$CD8^+$ Exhausted T cell
TF	Transcription Factor
TFBS	Transcription Factor Binding Site
TI	Tumor-Infiltrating
TI-Treg	Tumor-infiltrating Regulatory T cell
TME	Tumor Microenvironment
Tmem	$CD8^+$ Memory T cell
Tr1	$CD4^+$ $FOXP3^-$ Type 1 Regulatory T cell
Treg	Regulatory T cell
TSS	Transcription Start Site
UMAP	Uniform Manifold Approximation and Projection

1 Introduction

1.1 Introduction to the immune system

The immune system is a complex network of cells and molecules that defend the body against pathogens(1). It consists of two main categories: the innate immune system, which provides the first line of defense through physical barriers and immune cells like macrophages, neutrophils, and natural killer (NK) cells, and the adaptive immune system, which develops specialized responses to specific pathogens and retains memory of past infections. Central to the adaptive immune response are T cells, which originate in the bone marrow and mature in the thymus. T cells are classified into several distinct lineages, each with specialized functions. $CD4^+$ T helper cells orchestrate immune responses by releasing cytokines that activate other immune cells, including B cells and macrophages. This lineage includes various subtypes, such as Th1, Th2, and Th17 cells, each tailored to different pathogens and inflammatory contexts. $CD8^+$ cytotoxic T cells directly kill infected or cancerous cells by recognizing foreign antigens presented via MHC class I molecules. Regulatory T cells (Tregs) is a critical adaptive immune subset, which maintain immune tolerance and prevent excessive responses that could harm host tissues. Additionally, memory T cells are formed after initial immune encounters and persist long-term, enabling a rapid and effective response to subsequent infections. Together, these adaptive T cell populations are vital for establishing long-lasting immunity.

1.2 The crucial role of the immune system in cancer progression

1.2.1 Definition of cancer

Cancer is a significant global health issue and a leading cause of illness and death among both children and adults(2). The lethality of malignant tumors arises from their uncontrolled growth and ability to spread within normal tissues, leading to damage and loss of function. This behavior is driven by faulty regulation of cell proliferation, resistance to apoptosis, and the capacity of tumor cells to invade surrounding tissues and metastasize to distant locations. Cancer is a complex disease resulting from the crosstalk between genetic, epigenetic, and environmental determinants. Indeed, the acquisition of genetic mutations conferring survival advantage to normal cells can be

mediated and/or potentiated by many different intrinsic and extrinsic factors. The development of solid tumors is a continuum process evolving from the formation of pre-cancerous lesions during tumor initiation, to tumor invasion and metastatisation. In each of these key steps, the host immune system has profound effects and ultimately determines the disease's outcome.

1.2.2 Historical basis for the recognition of immune system in cancer control

The first observations linking cancer to immunity date back to the late nineteenth century, first with Rudolf Virchow's hypothesis that cancer could be caused by tissue inflammation, followed by the injection of a mix of bacterial products by the surgeon William B. Coley – the so-called Coley's Toxins – to cancer patients to stimulate their immune system against the tumor(3). In 1909, Paul Ehrlich hypothesized that the immune system could prevent the formation of tumors from neoplastic lesions(3). Nonetheless, it took several decades of research and progress in both fields of oncology and immunology, including the discovery of T lymphocytes as mediators of the adaptive immune response, to understand the importance of the immune system in cancer development.

According to the concept of "immunosurveillance" originally proven around 1960(3), the immune system can recognize and destroy transformed cells before tumors are formed and eradicate them after they are manifested. However, because of the continuous evolution of cancer cells, a phase of immune *equilibrium* can be perpetuated without the tumor being clinically detectable for a variable amount of time that can even be many years, until eventually less antigenic tumor clones are selected – the so-called "immunoediting"(4). Those cells that resist the immune attack are the ones that can overcome the immune response and induce a tolerant and immunosuppressive environment. Nowadays, this ability of cancer to escape the immune system is recognized as a cancer hallmark(5). Moreover, in parallel to a "branched evolution model" of cancer progression, that focuses solely on tumor cell features and the co-existence of multiple malignant clones deriving from a common ancestor, a "parallel immune selection model" has been proposed, which puts emphasis on the driving role of the immune system in cancer fate determination(6).

1.2.3 Introduction to immunotherapies

Starting from the late nineties, the new vision of cancer as a disease profoundly related to immunity paved the way for the first immunotherapies. The goal of cancer immunotherapy is to optimally tune the cancer-immunity cycle, enabling correct presentation of cancer antigens, T cell priming and activation, trafficking of T cells to the tumor, optimal recognition and killing of tumor cells, but at the same time avoiding uncontrolled autoimmune inflammatory responses(7).

Encouraged by the promising results obtained with hematopoietic stem cell transplantation in leukemia patients, many different approaches based on soluble immune mediators and immune cells were tested, not only employing T cells, but also other cell types, like dendritic cells (DCs) and natural-killer (NK) cells. Some of them have been subsequently approved by the U.S. FDA (Food and Drug Administration), like chimeric antigen receptor (CAR) T cells for hematologic malignances. A big breakthrough, made possible by the advancements in molecular immunology, has been the development of inhibitory antibodies targeting CTLA-4 and PD-1, prototypes of the so-called immune checkpoint proteins, which are crucial negative regulators of T cell activation. Nowadays, improving the effectiveness of immune checkpoint blockade therapies, expanding the molecular targets, and increasing the number of patients that can benefit from it represent some of the most important challenges in translational cancer research.

1.2.4 Roles of Innate and Adaptive immunity in promoting tumor growth

Clinically relevant immune responses are mainly mediated by $CD8^+$ cytotoxic T lymphocytes (CTLs). Indeed, T cells can be frequently found at the invasive margins, infiltrating the tumor and in lymph nodes draining the tumor site. A higher number of $CD8^+$ CTLs, memory T cells and $CD4^+$ Th1 correlates with better prognosis in different cancer types including colon cancer(8).

Tumor-specific immune responses are stimulated by the expression of several types of molecules by cancer cells. These can be antigens encoded by genes with passenger mutations that frequently occur in cancer because of genomic instability, products of oncogenic viruses – if the tumor is a consequence of a viral infection – or proteins expressed in excessive amount, or in not appropriate time or location. It is important to mention that the neoantigenicity of tumors varies a lot between distinct types of tumors. For instance, colorectal cancers (CRC) with microsatellite instability

(MSI), resulting from defective DNA repair mechanisms, express a larger amount of neoantigens than microsatellite-stable CRCs(9).

Both innate and adaptive responses to tumors can be elicited, with different mechanisms observed for each. Tumor-specific $CD8^+$ CTLs represent the main contributors to anti-tumor immunity by killing malignant cells exposing peptides deriving from tumor antigens. Notably, to mount an effective response, the antigens must be presented by dendritic cells in the lymph nodes, in the presence of the appropriate costimulatory signals. Specific $CD4^+$ helper T cell subsets can contribute to the host immune defense against cancer by enhancing $CD8^+$ T cell activation and creating a favorable environment for the immune response, primarily through cytokine secretion. Concerning the innate compartments, Natural Killer (NK) cells and M1 macrophages have been implicated in the immune surveillance against cancer, with similar mechanisms also used to kill infected cells or microorganisms.

It is now clear that the immune system plays a double role in cancer progression, since there is evidence that it can both prevent and promote the growth of solid tumors, and the regulation of this complex relationship determines the prognosis of the patient. In this regard, chronic inflammation is considered a highly significant risk factor for tumor development, and most external tumorigenic agents promote an inflammatory condition that may persist and become chronic(9). Some subsets of the innate immunity may have a direct role in malignant transformation, like myeloid cells for their ability to generate free radicals that can lead to DNA damage, or by secreting soluble mediators that promote tissue remodeling, angiogenesis, and cell growth. The adaptive compartment has been recognized as responsible for enhancing tumor progression with several mechanisms. In this regard, a key role is played by $CD4^+$ T_H2 helper cells or Treg cells, which suppress the T_H1 and CTL responses against tumors and instead promote the expansion of other cell types favoring tumor growth.

1.2.5 The heterogeneous tumor immune landscape

The cross-talk between tumor and immune cells takes place in the context of the tumor microenvironment (TME), which is the intricate ecosystem resulting from the interaction of cancer cells with other cell types and non-cellular component(10,11). Both innate and adaptive immune cells can be recruited and hijacked by cancer for the maintenance of immunosuppression. However, several immune cell types have been demonstrated to play both pro- and anti-tumorigenic roles. One reason for these

contrasting evidence lies in the fact that the cancer immune landscape is profoundly heterogeneous across cancer types and even across patients with the same subtype, concerning the type, frequency, activation state and location of innate and adaptive immune cells(12). In this regard, the importance of the “immune contexture”, defined by type, density, functional orientation, and location of the immune infiltrate, has been recognized across all stages of tumor progression, with relevant prognostic and therapeutic implications(13). For example, an immunoscore based on density and location of $CD45RO^+$ memory T cells and CTLs has revealed higher accuracy in predicting the outcome of CRC patients than TNM classification system based on tumor features(9).

In general, while CTLs, T_H1 and memory T cells are associated with prolonged survival, the varying effects of other T subsets on tumor prognosis may be attributed to their phenotypic plasticity, depending on TME signals and tissue-specific features(13).

Histological studies on tumor biopsies collected from tumor patients before treatment identified at least three distinct immune profiles(14). The immune-inflamed phenotype is characterized by the presence of $CD4^+$ and $CD8^+$ T cells in the tumor parenchyma, with or without myeloid and monocytic cells. These features of the tumor microenvironment are indicative of a pre-existing antitumor response that has been inhibited by immunosuppression mediated by tumor cells. In the immune-excluded phenotype, immune cells are present but restricted to the stroma compartment surrounding or penetrating the tumor, but they are never close to tumor cells. This indicates that cancer cells might have inhibited the migration of potentially effective tumor-specific T cells within the tumor. Lastly, the immune-desert profile is characterized by poor presence of T cells either in the tumor bed or at the margins and might reflect the absence of a pre-existing antitumor response.

The three immune profiles differentially correlate with response to checkpoint inhibitors, particularly those targeting PD-1/PD-L1. Indeed, clinical response to this therapy is less likely to occur in patients with non-inflamed tumors, corresponding to the second and third profiles.

Many factors eventually contribute to the specific immune profile of a cancer patient, including tumor-intrinsic properties and extrinsic factors. Among the former, the overall mutational burden and tumor epigenetic aberrations are crucially involved in

determining the tumor microenvironment and immunogenicity. On the other hand, host genetics – concerning genetic variation in immune-response genes – as well as the gut microbiome and several environmental cues can be crucial factors shaping the immune profile. The combination of all these forces defines the so-called "cancer-immune set point," which reflects the equilibrium between promotion and inhibition of anticancer immunity. The set point also corresponds to the threshold that must be surpassed to elicit effective control of cancer by the immune system when treating the patient with immunotherapy.

1.2.6 Evasion of immune response by tumors

A major focus of tumor immunology is to clarify the mechanisms by which cancer cells evade the immune attack, in order to target them with therapeutic approaches and improve the effectiveness of the host response. Most escape mechanisms can be divided into immune-mediated (e.g., the downregulation of the antigen presentation machinery or the expression of inhibitory immune molecules) or related to tumor cell-intrinsic properties that directly or indirectly shape the immune response in favor of cancer progression. These features include genetic and epigenetic aberrations and altered cancer metabolism(12).

According to the immunoediting phenomenon, less immunogenic tumor clones are more likely to survive as a consequence of selective pressure by the immune system. Possible underlying mechanisms are the loss of tumor antigen expression and the downregulation of class I major histocompatibility complex (MHC) molecules, or proteins involved in the assembly and transport of MHC on the cell surface, possibly due to mutations in genes encoding these molecules.

Active inhibition of antitumor responses can be mediated by the engagement of T cell immune checkpoint proteins, which normally inhibit excessive immune responses and prevent autoimmunity. Tumor cells can upregulate the expression of PD-1 and CTLA-4 on T cells that are chronically exposed to tumor antigens in an immunosuppressive context. Indeed, tumor-infiltrating (TI) T cells often show a dysfunctional phenotype, often referred to as exhausted, characterized by an impairment of the effector functions. At the same time, tumor cells can amplify the expression of ligands of the immune checkpoints, such as PD-L1, to engage the receptors on the T cells. Moreover, cancer cells produce immunosuppressive cytokines (e.g., TGF-beta), which can inhibit the activation and function of innate and adaptive immune cells.

Other mechanisms involve the generation of an immunosuppressive tumor microenvironment by inducing the differentiation and activation of immune subsets that inhibit T cell activation and differentiation into antitumoral Th1 and CTLs. Among others, myeloid-derived suppressor cells (MDSCs) and regulatory T cells (Treg) have major roles in the immune microenvironment. MDSCs are immature myeloid precursors able to exploit various suppressive mechanisms, including cytokine secretion and the promotion of Treg differentiation. The latter T cell subset, frequently found in the immune infiltrate of solid tumors, has received considerable attention in recent years as a novel potential target in cancer immunotherapy and represents the focus of this work.

1.3 Regulatory T cells

Immunological tolerance can be defined as the unresponsiveness to an antigen, mediated by exposure to that antigen. Tolerance to self-antigens is a physiological property of the immune system, and the failure of mechanisms mediating self-tolerance gives rise to autoimmune diseases(1). Given the fact that the specificities of T cell receptors (TCR) are provided by random gene recombination and not influenced by what is foreign or self for each individual, it is not unlikely that some developing T cells can recognize self-antigens.

The first experimental evidence pointing to the existence of a specialized T cell subset able to prevent autoimmunity and to restrain pathogenic immune responses came from neonatal thymectomy and from studies of transplantation tolerance in chicken-quail chimeras and in mice(15). Then, a seminal discovery by Sakaguchi and colleagues in 1995 described regulatory T cells (Tregs) as an immunosuppressive subpopulation of $CD4^+$ T cells expressing high levels of the IL-2 receptor alpha chain (CD25)(16). Nonetheless, since this marker is also upregulated in all activated T cells, the isolation of the tolerance-inducing subpopulation remained a challenging task. Crucial insights into the biology of Tregs were provided by studies on mice and humans with a loss-of-function mutations in the forkhead box P3 (FOXP3) gene(17,18). Indeed, patients with FOXP3 mutations are affected by a T-cell dependent rare autoimmune systemic syndrome called IPEX (immune dysregulation, polyendocrinopathy, enteropathy, X-linked)(18). The assessment of FOXP3 expression in $CD4^+$ $CD25^+$ Treg cells and other functional studies revealed that the FOXP3 mutations are associated with Treg deficiency highlighting the central role of this transcription factor in the differentiation, phenotypic stability and regulatory function of Treg cells(19–23). These observations have established a dominant and indispensable role for Tregs in maintaining self-tolerance and in regulating immune homeostasis.

Treg cells can be distinguished also by the low expression of IL-7 receptor (CD127)(24) and by the constitutively high levels of CTLA4(25), together with other established Treg cell surface molecules, including the glucocorticoid-induced tumor necrosis factor (TNFR) family-related protein (GITR)(26) and the inducible T-cell costimulator (ICOS)(27). Notably, human activated conventional T cells can transiently upregulate at lower levels the aforementioned markers along with FOXP3 following

TCR stimulation. Nevertheless, gating on $CD4^+$ $CD25^{hi}CD127^{low}$ T cells in human peripheral blood identifies more than 90% of Treg cells, as confirmed by high FOXP3 expression(28).

1.3.1 Generation of Tregs

Under normal conditions, the number of Treg cells is low, accounting for 5-10% of the total $CD4^+$ T lymphocytes. One of their main roles is to protect from autoimmunity and pathogenic immune responses that can damage tissues. They develop in three settings: i) in the thymus, from $CD4^+$ T cells that recognize primarily self-antigens; ii) in the peripheral tissues, from naïve $CD4^+$ T cells able to recognize both self and foreign antigens in the absence of an activating environment generated by the innate immunity; and iii) after inflammatory reactions, to regulate excessive immune responses(1).

During T lymphocyte maturation within the thymus, those able to recognize self-antigens presented by thymic cells encounter negative selection; however, some self-reactive $CD4^+$ T cells differentiate into Tregs specific for these antigens and leave the thymus to mediate self-tolerance in the periphery. These are called natural or thymic Tregs (nTregs or tTregs) and contribute to the mechanisms of central tolerance. The choice between the two fates is dictated by the strength and duration of TCR signal. In particular, Treg selection is instructed by TCR with intermediate affinity, which is below that causing deletion and above that mediating positive selection(15). Additional signals that are essential for Treg differentiation in the thymus are IL-2 and, to a lesser degree, IL-7 and IL-15(15). A two-step model of tTreg cell differentiation proposes that a high-affinity TCR signal upregulates CD25 with a consequent increase in the responsiveness of Treg precursor cells to IL-2 signals, conferring survival advantage and inducing FOXP3 expression via STAT5 regulation(29).

Peripherally induced Tregs (pTregs or iTregs) are generated extrathymically in selected niches, mainly in response to foreign antigens including environmental allergens, food and commensal metabolites, but also to tissue-specific antigens non abundant in the thymus(15,30). Indeed, the TCR repertoires of tTregs and pTregs are largely non-overlapping, supporting complementary antigen coverage and non-redundant functions for the two subsets(31). The requirements for pTreg differentiation are a high-affinity TCR signal together with suboptimal costimulation, meaning that the chronical exposure to antigens in absence of inflammatory conditions will favor FOXP3

induction as a mechanism of peripheral tolerance(1,15). This process is further facilitated by high levels of TGF-beta, which directly regulates Foxp3 expression and mediates survival advantage to pTreg precursors(32). It remains to be established which markers can reliably distinguish tTregs from pTregs in humans, although many have been proposed: one example is Helios, a transcription factor (TF) member of the Ikaros family, highly expressed in tTregs and not in pTregs(30,33,34).

1.3.2 Treg maintenance and stability

Foxp3 acts as a molecular orchestrator of the Treg cell program. It is true that Foxp3-deficient Tregs exhibit some phenotypic and transcriptional overlap with $Foxp3^+$ cells, including the inability to proliferate in response to TCR stimulation (i.e., anergy), the dependence on paracrine IL2, the low expression of IL7R and the high amount of CD25, CTLA4 and GITR(35,36). In this regard, important Treg signature genes are already expressed before FOXP3 induction at thymic Treg precursor stage and some of them play crucial roles in FOXP3 upregulation itself(37). Moreover, it has been shown that a large part of the transcriptional landscape of Tregs is not transactivated by Foxp3(38). However, without Foxp3 the suppressive function of Treg cells is completely abrogated(35). In fact, Foxp3 is critical for amplifying and stabilizing the expression of signature genes while repressing immune response-promoting genes such as those for proinflammatory cytokines' production(35,36). Therefore, Foxp3 is critical for Treg-mediated suppression but independent of pre-existing features, required for Treg lineage commitment, that indeed precede and promote its induction. In this regard, a significant role of Treg-specific epigenomic features has been established for Treg cell development. As will be discussed later, epigenetics represents a fundamental level of transcriptional regulation. Canonical epigenetic changes include histone modification and DNA methylation of CpG residues; these events modify the accessibility of TFs and the transcriptional machinery to regulatory regions of the genome, thus impacting on the transcription of cell-type specific gene programs.

During Treg differentiation, specific CpG demethylation takes place at regulatory regions of Treg signature genes including Foxp3, which correlates with gene activation and is Foxp3-independent(39). These regions act as specific enhancers for the constitutive expression of Treg genes, since they exhibit open chromatin and H3K27ac modification in naïve Tregs(40). These epigenetic mechanisms may allow the Treg

phenotypes to be inherited along their proliferation in the periphery and can be a reliable marker to be used together with Foxp3 expression for identifying stably differentiated Tregs.

Interestingly, Foxp3 regulation and DNA hypomethylation act complementary: following TCR stimulation, Foxp3 works as a transcriptional repressor at the promoter of many genes such as IL2 and IFNG(39); on the other hand, demethylated regions are present at genes upregulated in steady state condition, such as *IL2RA (CD25)* and *CTLA4* contributing to the maintenance of Treg stability.

Notably, differently from tTreg cells, where DNA demethylation represents one of the first events in Treg lineage determination prior to Foxp3 induction, the generation of pTregs can be initiated by a transient Foxp3 upregulation, that gradually evolves in stable Foxp3 expression and in the Treg-type hypomethylation pattern(40).

For the maintenance of the Treg cell functional and transcriptional program, a continuous and elevated expression of Foxp3 is required(22,23). Multiple conserved non-coding sequences (CNS) proximal to Foxp3 locus are involved in stable Foxp3 gene expression. CNS1, which contains binding sites for TFs like NFAT, RAR/RXR and TFG-beta-activated SMAD3, is primarily involved in the induction of Foxp3 in peripheral naïve $CD4^+$ T cells(41,42). CNS2 is the regulatory region presenting DNA hypomethylation, which mediates binding of protein complexes including Runx1, CREB/ATF, NF-kB and Foxp3 itself, as an autoregulatory loop. This region is required for a stable heritable active state of the Foxp3 locus in the progeny of dividing Tregs and, therefore, for Treg lineage stability(41). CNS3 represents a regulatory element with enhancer epigenomic features, which can be bound by the NF-kB family member c-Rel and acts early in Foxp3 induction at precursor stages(41).

Treg functional stability has been debated over the past years, particularly because there are some studies reporting loss of Foxp3 expression in Tregs under certain inflammatory and/or autoimmune conditions(15). Furthermore, loss of suppressive activity by Tregs can be followed by a gain in effector features, which may contribute to autoimmunity(43,44). However, *in vivo* tracking of Foxp3-expressing Tregs over time demonstrated that Tregs are highly stable under basal and inflammatory conditions(45). One explanation, which takes also in consideration the key role of epigenetics, is that $Foxp3^+$ T cells constitute a heterogenous population in physiological conditions and a small fraction possesses incomplete Treg-type

epigenome(39,46); under certain circumstances, these Foxp3$^+$ cells might convert to Foxp3- T conventional cells. In contrast, the majority of Foxp3$^+$ cells have fully established Treg epigenome and thus constitute a stable cell lineage.

1.3.3 Suppressive mechanisms

Treg cells use multiple immunomodulatory mechanisms to suppress immune responses that target both innate and adaptive immune cells. Their suppressive functions are highly flexible and organized into modules that can be activated specifically depending on the tissue and inflammatory context, including contact-dependent mechanisms, secretion of cytokines and metabolic perturbation of target cells. This adaptable nature ensures an effective and appropriate control of immune homeostasis in diverse types of environments.

The surface molecule CTLA-4 mediates an important example of contact-dependent suppressive mechanism in Tregs. It is a member of the CD28 receptor family and has a stronger affinity for B7 ligands compared to CD28 expressed on the membrane of other T cells. Through CTLA-4, Treg cells compete with other T cells for the engagement of B7 ligands on dendritic cells, thereby limiting the initial costimulation-dependent activation of these T cells in secondary lymphoid organs. Moreover, CTLA-4 can down-modulate the expression of B7 molecules on antigen presenting cells (APCs) and also capture them by trans-endocytosis, thus reducing the number of available ligands for CD28-mediated costimulation(47,48). In addition, through CTLA-4, Tregs can stimulate the expression of indoleamine 2,3-dioxygenase (IDO) in dendritic cells, an enzyme that catabolizes and depletes the essential amino acid tryptophan from the environment, thus inducing T cell starvation(49).

Other contact-dependent mechanisms that act on dendritic cells (DC) involve the interaction of LAG-3, a CD4 homolog expressed on Treg cells, with MHC class II molecules on DC, which inhibits their maturation and costimulatory capacity(50), and may also cause physical removal of the MHC-II complex from their surface(51). Moreover, the Ig family member TIGIT, highly expressed on Treg cells, upon interaction with dendritic cells can induce the upregulation of immunosuppressive IL-10 and TGF-beta by the latter(52).

Among the other cell-surface molecules that mediate immunosuppression, the constitutive expression of CD25 on Tregs, which is indispensable for differentiation, acts also as a mechanism for depriving effector T cells of IL-2 and, therefore, inhibiting

their proliferation(53). In addition, two ectoenzymes, ATP apyrase (CD39) and ecto-5'-AMP-nucleotidase (CD73), are responsible for the generation of adenosine, which in turn mediates impairment of both effector T cells and dendritic cells(54).

Several proteins secreted by Tregs play immunomodulatory roles. Among them, IL-10 and TGF-beta represent two key effector molecules. IL-10 is crucial for the control of immune homeostasis in barrier tissues like the colon, skin and lung(55). TGF-beta is produced by Tregs to limit Th17 and Th1 responses at the mucosal interfaces; it has been proposed as a central cytokine regulating autoimmune and allergic processes, with important consequences related to its gene dosage and interaction with the gut microbiota(56). Similarly, IL-35 derived from Tregs has been suggested to play a role in immune tolerance in the gut(57).

A different mechanism by which Treg directly suppress effector cells is the cell-mediated cytolysis, originally suggested by the finding that Treg express granzyme B. Through the release of this molecule, Treg cells can kill both responder T cells and APCs(58).

1.3.4 Treg functions and plasticity

While originally Tregs were interpreted as a universal suppressive population, more recent advances in the field of Treg biology led to the identification of large developmental and phenotypic diversity in Tregs, with the existence of heterogeneous subsets. Indeed, not only do they maintain immune homeostasis by dominant tolerance, but have also been widely recognized as key players in tissue homeostasis, by participating in tissue inflammation resolution and healing processes(59,60). In addition to the division based on the site of induction, a functional classification distinguishes a 'central' Treg population, with circulating characteristics resembling those of naïve conventional T cells, and several 'effector' populations, with an activated profile reflecting recent antigen encounter, similarly to other effector T cells(61).

These two populations differ importantly in their migration capacity and the requirements for their maintenance. Most of the Tregs present in the circulation and within lymphoid tissues express lymphoid homing molecules, in particular CC-chemokine receptor 7 (CCR7) and the adhesion receptor CD62L, encoded by the *SELL* gene, also known as L-selectin(62,63). Notably, the homeostasis of this 'central' population of Tregs, also called 'naïve' or 'resting' Tregs, mainly depends on paracrine

IL2 secretion in T cell zones of secondary lymphoid tissues(63). Naïve/resting Tregs are not quiescent and possess a baseline suppressive function. However, the subsequent antigen encounter via direct TCR recognition, and possible additional differentiation stimuli, determine the acquisition of the activated phenotype, with varying expression of markers such as CD69, CD103, CD38, CD44, KLRG1, ICOS, GITR and granzyme B(64,65). Effector Tregs make up a small fraction of Tregs in the circulation and in lymphoid organs. Whilst naïve/resting Tregs typically downregulate CCR7 and enable migration through non-lymphoid tissues, mediated by other trafficking receptors including CCR4 and CCR8(62). Interestingly, their survival seems less dependent on IL-2 and mediated instead by sustained ICOS signalling(63). Central and effector Treg populations can be described also as $FOXP3^{low}CD45RA^{hi}CD25^{low}$ and $FOXP3^{hi}CD45RA^{low}CD25^{hi}$, respectively(66).

A large body of research has focused on the functional plasticity of effector Tregs that allows an efficient control of the immune response in distinct immunological contexts. Indeed, based on the original trigger and the underlying cytokine milieu, $CD4^+$ effector T cells are induced to polarize to specific subsets that elaborate distinct effector mechanisms for a precise control of the specific inflammatory context. For example, TGF-beta and IL-6 drive the differentiation of IL-17-producing T_H17 cells against extracellular bacteria or fungi, while IFN-gamma and IL-12 induce T_H1 cells during infection with intracellular pathogens(1). Treg cells are fundamental to carefully regulate all kinds of immune responses, to ensure their efficient resolution without excessive tissue damage. In this regard, a shared use of transcription factors and trafficking mediators between effector Tregs and various T_H cells has been observed, that allows an efficient curtail of the corresponding polarized T_H immune response by Treg cells. This mirroring in transcription factor requirement, whose advantage may be the colocalization and/or co-survival of regulatory and effector cells in precise settings, drives a further functional specialization of Tregs, that acquire the molecular machinery needed to restrain specific responses. For example, STAT3, a master TF for T_H17 subsets, is upregulated in Tregs for optimal suppression of T_H17-skewed immune response and mediates the acquisition of CXCR6 to migrate to the intestine and the production of IL-10(67). Instead, Tregs that suppress T_H1 cell subsets upregulate the TF T-bet, which is also required for T_H1 differentiation. In turn, T-bet mediates the upregulation of the chemokine receptor CXCR3, which facilitates the

accumulation of Tregs in T_H1 inflammatory contexts(68). Parallel specifications of effector Tregs have also been described for the regulation of T_H2- and T_{FH}-mediated reactions, that critically rely on IRF4(69) together with GATA3(70) and BCL6(71), respectively. All these environmental response factors can work together with Foxp3 to induce either a temporary or permanent cell state, allowing them to engage distinct suppression mechanisms matched to the specific type of response being controlled.

Another important source of diversity among Tregs, providing additional evidence for their high adaptability to different environmental stimuli, is the heterogeneous pool of tissue-resident Treg populations with unique phenotypes and specialized functions present in most non-lymphoid tissues, including skin, intestinal mucosa, adipose tissue, lung, skeletal muscle, brain, and placenta(72). They normally localize in healthy tissues, although in small number, in tissues with different grades of inflammation (e.g., infected tissues, autoimmune target sites, grafts, atherosclerotic plaques and tumors) and even immunoprivileged sites(73). There are several hypotheses proposed for their accumulation in the tissues, which are not mutually excluded(73). Chemokine-based migration (e.g., through CCR4 expression)(74) and the TCR specificity for self-antigens(75) are two main mechanisms suggested, even for Tregs recruitment in healthy and non-challenged tissues. Additionally, the conversion of local or circulating conventional $CD4^+$ T cells could explain their accumulation in non-lymphoid tissues, particularly in the gut lamina propria(76).

1.4 Tissue resident Tregs in non-lymphoid tissues

Tissue-resident Tregs are phenotypically and functionally different from lymphoid tissue Tregs for certain properties, such as the expression of transcription factors and effector molecules, the chemokine receptor profile and the mechanisms of action.

One of the best characterized examples of tissue-resident Treg population is that present in visceral adipose tissue (VAT). VAT-resident Tregs differentially express many genes compared to their lymphoid organ counterparts(77). The master regulator of adipocyte differentiation and function, peroxisome proliferator-activated receptor-gamma PPAR-gamma, is also upregulated in these cells and confers to VAT-Tregs unique adaptation and properties, also related to lipid uptake and metabolism(78).

There is growing evidence that the functions of Tregs within non-lymphoid tissues extend far beyond the regulation of local immune responses. For instance, VAT Tregs

have been reported to affect local and systemic metabolism, whereas tissue repair and regeneration effects have been reported for Tregs in multiple tissues including skeletal muscles, skin, heart, and lungs, after acute or chronic injuries(72). In particular, skin Tregs have been found to exert a protective and homeostatic role in healthy tissues, but also to promote tissue healing after various insults, favoring wound closure and dampening fibrosis(59). This regenerative property has been commonly associated with the upregulation of the growth factor amphiregulin (AREG), which, in fact, is part of a pan-tissue signature upregulated in several tissue-resident compartments (72,79). Tissue-resident Treg activity can also stimulate the proliferation and differentiation of non-lymphoid cell precursors. For instance, Treg cells in the gut can control intestinal stem cell renewal through IL-10, bone marrow Tregs regulate and are frequently colocalized with hematopoietic stem cells, and skin Tregs control the activities of the stem cells present at the level of the hair follicle, where they concentrate. Notably, these activities can be mediated by direct interaction between Tregs and the target parenchymal or stromal cells, or by indirect effect via other immune cells, such as macrophages(72,80).

Importantly, most of the evidence on tissue-resident Tregs derives from studies performed on mice. While some of them have been recapitulated also in human studies, less information is known about the phenotype and function of Tregs resident in human tissues and, as a general consideration, there is greater heterogeneity in the environmental exposures between human and laboratory mice, which might reveal a more complex landscape of tissue-Treg specifications(72).

Recently, a combination of epigenetic and transcriptomic studies also based on single-cell methods lead to the definition of a common signature for tissue-resident Tregs that is conserved among mice and humans. Multiple genes encoding for transcription factors (e.g., BATF, IRF4, NFIL3, ID2, RORA, and HIF1A), different chemokine receptors and effector molecules are included in this signature, which reveals signs of tissue adaptation(60,72,81–83).

Tissue-resident Tregs present in mucosal tissues deserve particular attention, due to the unique barrier function exerted by mucosal sites, which requires highly coordinated regulation to ensure tolerance to non-harmful external agents while simultaneously protecting the organism from potential dangers. In this regard, the resident Tregs in

these tissues play a crucial role, because they keep the balance between pathogen clearance and tissue preservation.

The respiratory tract is constantly exposed to external antigens and pathogens and tissue-resident Tregs cover several key roles in the protection of this barrier. At the homeostasis, 10-20% of total lung Tregs in specific pathogen-free (SPF) mice display a tissue-specific profile according to the aforementioned tissue-specific signature. Although lung tissue Tregs share “core” genes with Tregs from other non-lymphoid tissues, lung tissue-specific cues might confer additional influence to their phenotype. Amphiregulin, IL-18 and IL-33 signaling drive tissue repair functions of this population(84). Additionally, they have been implicated in the restraining of Th2-mediated allergic responses to innocuous antigens through an IL-10-dependent mechanism(84).

Tregs are the most important cell population for the maintenance of oral tolerance and the containment of the gut microbiota, and they constitute 10-15% and 25-35% of total $CD4^+$ cells in the small intestine and colon lamina propria, respectively(80). While small intestine Tregs are predominantly developed against dietary antigens, colon Tregs are largely originated by crosstalk with the microbial commensals. The majority of intestinal Tregs are peripherally induced, with TGF-beta and retinoic acid (RA) potentially facilitating this process, but the thymic origin is not excluded(80). At least three functionally distinct subsets of intestinal Tregs have been identified, and the expression of TFs and surface molecules guides their specification. The $GATA3^+$ $Helios^+$ ($Nrp1^+$) subset constitute around one third of intestinal Tregs and has been suggested to derive from the thymus, at least in mice(85). Consistently with the expression of GATA3, the IL-33 receptor ST2, and AREG, this subset has a role in the regulation of T_H2 responses, recognizes the alarmin IL-33 during inflammatory events and directly contributes to tissue repair(80,84). On the other hand, $ROR\gamma t^+$ $Helios^-$ constitute about 50% of total colonic Tregs, have mainly extra-thymic origin, and are involved in colon tolerance to both commensal and pathobionts(86). They require the expression of c-Maf for their development and function as highly activated effector Tregs, as the high expression of CTLA-4, ICOS and CD44 suggests(84). The secretion of IL-10 by these Tregs is thought to contribute to their suppressive function, which is particularly directed against T_H17 inflammatory responses(84). Lastly, a $ROR\gamma t^-$ subset has been identified as a supposedly pTreg population, which constitutes about

15% of colonic Tregs, but around 50% of the small intestine compartment, suggesting a predominant role for this population in dietary tolerance and restraint of food allergy(87).

Very recent and interesting insights on the biology and dynamics of tissue-resident Tregs have been provided in a study from Burton et al(88). The authors explore tissue-resident Tregs of various origins, revealing a shared molecular signature and short-term residency patterns. Through transcriptomics, genetic testing, and migration models, the study shows that Treg cells exhibit pan-tissue migration, with only slight preferences for specific tissues like the gut. Indeed, gut-resident Treg cells have distinctive characteristics compared to Tregs from other tissues: they express higher levels of Ccr9 and Ccr5, which are associated with gut-homing, lower levels of Itgb1 and Sell, and exhibit a more restricted clonal migration, with limited TCR sharing with other non-gut tissues. Functionally, they demonstrate a stronger preference for re-entry into the gut during migration studies, suggesting a specialized gut-residency model compared to more flexible non-gut Tregs. Overall, the findings highlight the transient nature of tissue Treg cells and propose a reversible model of tissue residency, where Treg cells seed multiple tissues and adapt to local environments.

1.5 Tregs within the tumor microenvironment

1.5.1 Prognostic value of Tregs within solid tumors

Nowadays it is broadly recognized that Tregs play a pivotal role in tumor progression as an immunosuppressive population and contribute to the determination of the clinical outcome of cancer patients.

Several aspects regarding the mechanisms leading to Treg infiltration and activity in tumors need still to be clarified. Tregs can be recruited into tumors in response to chemokines secreted by tumor cells and innate immune cells; by Curiel and colleagues it was observed that tumor-derived macrophages from patients with ovarian cancer secrete CCL22, a chemokine that mediates trafficking of Tregs to tumors(89). Moreover, Tregs can be expanded *in situ*, and proliferate efficiently in response to tumor-derived factors (TGF-β, IL-10) within the TME(1). Lastly, the generation of suppressive Tregs from non-suppressive CD25− conventional T cells driven by tumor-derived TGF-β and adenosine has mainly been studied in murine models, but their contribution remains to be confirmed in human cancers(1).

Treg cells infiltrate almost all tumors, *e.g.* head and neck, lung, liver, gastrointestinal tract, breast, pancreas, and ovary(90,91). For most cancer types, a high density of Treg cells in the TME is associated with a poor prognosis for patients. In this regard, $CD8^+$ /$FoxP3^+$ T cell ratio is positively correlated with patient. In fact, the abundance of the Treg compartment shifts the equilibrium in favor of the suppression of effector T cells, thus allowing malignant cell proliferation and expansion. Nevertheless, while effector cells are clearly involved in directly killing tumor cells, the role of Treg cells in cancer immunity has been more controversial. Although Treg cells are known for their immunosuppressive functions, which can synergize with cancer cells to inhibit antitumor responses, paradoxically, their presence within tumors has been associated with better prognosis in certain cancers, including colon, esophageal, and bladder cancers. This seemingly contradictory evidence may be explained by the heterogeneity of the Treg population across different tissues. In colorectal cancer (CRC), for example, it has been shown that CRCs can be classified based on the infiltration of suppression-competent $FOXP3^{hi}$ Treg cells versus $FOXP3^{lo}$ nonsuppressive T cells(92). The latter, which lack the naive T cell marker CD45RA and exhibit unstable FOXP3 expression, secrete inflammatory cytokines. Notably, CRCs with a higher presence of $FOXP3^{lo}$ T cells show better prognosis compared to those dominated by $FOXP3^{hi}$ Treg cells. This study highlights the functional diversity of $FOXP3^+$ T cells in CRC and suggests that targeting $FOXP3^{hi}$ Treg cells could enhance antitumor immunity, while promoting $FOXP3^{lo}$ non-Treg cells may help prevent tumor formation.

1.5.2 Metabolic profile of tumor-infiltrating Treg cells

The unique metabolic profile of Tregs allows them to survive and exert their immunosuppressive activity with high efficiency in the TME, which is characterized by scarcity of fundamental nutrients, including glucose and glutamine, by abundance of suppressive metabolites, like lactate, and prohibitive conditions, exemplified by hypoxia and low ambient pH. While such an environment constitutes an important barrier for an appropriate anti-tumor T cell activity, Tregs are favored under several aspects. For instance, Tregs are resistant to a low-glucose environment, since they rely on oxidative phosphorylation for their sustainment, and the presence of lactate can enhance their suppressive activity(93,94). At the same time, CTLA-4-mediated

blocking of the costimulatory signals establishes their independency from glycolysis, thus conferring functional stability to Tregs(95).

1.5.3 Suppressive mechanisms in the TME

Tregs in the tumor microenvironment utilize similar immunosuppressive mechanisms to those described in their physiological functions. For example, the consumption by Tregs of IL-2, which is needed for the activation and survival of other T cells, results in hampered activity of antitumor effector T cells(96). Tumor-reactive $CD4^+$ T cells, in addition to their helper functionality, require IL-2 signal to activate a Blimp-1-dependent transcriptional program that leads to the expression of granzyme B, that promotes rejection of established tumors. Therefore, IL-2 deprivation by Tregs limits the differentiation program of cytotoxic $CD4^+$ T cells in the TME(97). Also, the Treg-derived CD73- and CD39-mediated production of adenosine, which binds to the adenosine receptor A2A on effector T cells, has been implicated in the inhibition of T cell antitumor activity in mouse models(96). Direct killing of effector T cells in the TME can be mediated by Tregs in a granzyme B- and perforin-dependent mechanism, whereas APC activity is counteracted by Tregs through the production of immunosuppressive cytokines and through CTLA4-dependent downregulation of costimulatory signals on tumor-associated DCs(96). Moreover, this mechanism mediates an unstable interaction between DCs and Tregs, which is sufficient for the activation of their suppressive function, but inadequate to induce an excessive proliferation that would be detrimental for Tregs(98). Therefore, a CTLA-4-dependent feedback loop has been proposed as a mechanism to regulate Treg suppression function in the TME.

1.5.4 Molecular characterization of tumor-infiltrating Tregs

Given their crucial role in dampening the anti-tumor immunity and their negative correlation with survival in several cancer types, Treg cells are emerging as primary targets of immunotherapy.

Initial investigations have revealed that *in vivo* depletion of Tregs using monoclonal antibodies against CD25 enhances anti-tumor T cell responses and induces regression of experimental tumors(99).

Different attempts in modulation of human Treg cells have been made by targeting their surface molecules. TI-Tregs express high levels of the Treg-associated

molecules CTLA-4 and GITR. Treatment with a soluble form of the natural ligand of GITR (GITRL), or with blocking antibodies to CTLA-4, was proved to reduce the suppression mediated by human liver tumor-infiltrating $CD4^+$ $Foxp3^+$ Tregs, thus restoring the functions of effector T cells(99).

The importance of CTLA-4 for Treg cells was first discovered after the generation of $Ctla4^{-/-}$ mice that displayed early and fatal impairment of self-tolerance, similarly to what happens in $foxp3^{-/-}$ mice(99). Importantly, it was recently shown that anti-CTLA-4 mAb efficacy depends on FcγR-mediated depletion of $CD4^+$ Tregs within the TME(100). Treg cell depletion after anti-CTLA-4 therapy has been reported to increase anti-tumor specific immune responses and to reduce tumor burden in clinical experiments. However, some efforts still must be put to address important issues related to the safety and specificity of these therapies. One crucial consideration is that the lack of specificity for tumor infiltrating lymphocytes, that share most of their molecules targeted for therapy with effector lymphocytes, can cause severe autoimmunity following systemic Treg cell depletion. These adverse effects could be avoided if selective depletion of TI-Treg cells was feasible. Therefore, the molecular characterization of TI-Tregs should aid towards the development of highly specific therapies that effectively target the regulatory compartment without affecting the overall immune homeostasis. To this purpose, our laboratory performed a comprehensive transcriptome analysis of human $CD4^+$ Treg cells infiltrating two different tumor types, non-small cell lung cancer (NSCLC) and CRC(101). TI-Treg cells in these tumors were highly suppressive and displayed a distinct transcriptional profile compared not only to other $CD4^+$ T cell subsets but also to Treg cells infiltrating normal tissues. Notably, similar changes were described in TI-Tregs derived from different cancer types, highlighting a common layer of adaptation in response to the TME stimuli. The study resulted in the identification of a molecular signature specific for TI-Tregs. Among the genes preferentially expressed, some regulatory checkpoints were found, as GITR, OX40 and CTLA-4, which are also known to be associated with increased suppressor activity. Other well characterized genes as BATF, TIGIT, CCR8, LAG-3, TIM-3, PD-1 and its ligand PD-L1 were also found to be upregulated in TI-Treg cells. Remarkably, high expression in whole-tumor samples of key TI-Treg signature genes, such as LAYN, MAGEH1, or CCR8, correlated with poor patient prognosis. All these data indicate that the TME drives acquisition of specific features of TI-Treg cells.

Importantly, a deeper molecular characterization of this unique phenotype is needed, in order to identify exclusive markers that could be exploited as targets of novel immunotherapies.

1.6 Colorectal cancer and non-small cell lung cancer

CRC is the third most common cancer globally and the second leading cause of cancer-related deaths, with a 5-year survival rate of 63%(102). Its incidence is rising, particularly in developing countries adopting Western lifestyles, with risk factors including obesity, sedentary behavior, red meat consumption, alcohol, and tobacco use. The Consensus Molecular Subtype (CMS) classification categorizes CRC into four subtypes: CMS1, which is immunogenic with high microsatellite instability (MSI-H) and rich immune infiltration; CMS2, characterized by Wnt/Myc pathway activation; CMS3, notable for KRAS mutations and metabolic dysregulation; and CMS4, the mesenchymal and most aggressive subtype, marked by epithelial–mesenchymal transition (EMT), angiogenesis, and stemness pathways. CMS1 has a favorable prognosis due to high immune infiltration, while CMS4 is associated with poor prognosis despite the presence of immune cells, due to immunosuppressive factors. In contrast, CMS2 and CMS3 are immunologically “cold,” with little lymphoid or myeloid infiltration, making them potential candidates for immunotherapeutic strategies to boost anti-tumor immunity.

NSCLC, which accounts for 85% of all lung cancers, is a leading cause of cancer-related deaths(103). It includes adenocarcinoma, squamous-cell carcinoma, and large-cell carcinoma, with adenocarcinoma being the most prevalent subtype. NSCLC is driven by mutations in genes such as EGFR, KRAS, and ALK, which have become targets for therapeutic interventions. NSCLC also exhibits a high mutational burden, which can generate tumor-specific neoantigens that make it susceptible to immunotherapies like PD-1 inhibitors. However, resistance to these therapies often arises, due to mechanisms like downregulation of MHC, defects in IFN-γ signaling, and altered IDO expression. The TME of NSCLC contains diverse immune cell populations, including T cells ($CD4^+$, $CD8^+$, and double-negative T cells), macrophages, and dendritic cells. Prognostically, the presence of Tregs and M2 macrophages correlates with poor outcomes, whereas $CD8^+$ T cells and tertiary lymphoid structures are linked to better survival.

1.7 Epigenomic regulation of gene expression

How the genome integrates environmental signals for the regulation of gene expression? The answer to this question can be found in the world of epigenetics.

Historically, the word “epigenetics” was used to describe events that could not be explained by genetic principles. Conrad Waddington, who is given credit for coining the term, defined epigenetics as “the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being” (Waddington, 1942).

Epigenetics, in a broad sense, is a bridge between genotype and phenotype, a phenomenon that changes the final outcome of a locus or chromosome without changing the underlying DNA sequence. Thus, cellular differentiation may be considered an epigenetic phenomenon, largely governed by changes in what Waddington described as the “epigenetic landscape” rather than alterations in genetic inheritance. More specifically, epigenetics may be defined as the study of any potentially stable and, ideally, heritable change in gene expression or cellular phenotype that occurs without changes in Watson-Crick base-pairing of DNA.

Epigenetic mechanisms include DNA-methylation, histone modifications, non-coding RNAs, chromatin accessibility, and nucleosomes remodeling. The crosstalk between all these mechanisms enables prompt cell reaction to environmental changes.

1.8 Chromatin Structure and organization

Chromatin is a dynamic and structured DNA-protein complex that compacts eukaryotic chromosomal DNA within the nucleus. The primary structural unit of chromatin is the nucleosome, consisting of 145-147 base pairs of DNA wrapped around a histone octamer made of two each of H2A, H2B, H3, and H4. The addition of histone H1 further stabilizes the structure, creating the chromatosome. Nucleosomes are connected by linker DNA, which is more susceptible to nuclease digestion. Histone proteins, particularly their N-terminal tails, are highly conserved and subject to post-translational modifications like acetylation and methylation, crucial for gene expression regulation. These modifications influence chromatin structure and the interaction with regulatory factors, facilitating or restricting DNA accessibility. Nucleosomes compact the DNA, protect it from damage, and serve as a scaffold for epigenetic signals, allowing

chromatin to switch between repressive and accessible states depending on cellular needs.

1.9 Chromatin accessibility and ATAC-seq

Chromatin accessibility refers to the extent to which nuclear macromolecules can interact with genomic DNA, and it is influenced by nucleosome occupancy, modifications, and the binding of non-histone proteins. Nucleosome composition and post-translational modifications, such as histone acetylation, play a key role in regulating chromatin states, affecting transcription factor binding and chromatin remodeling. Open chromatin regions (OCRs), typically found at regulatory elements like promoters and enhancers, are essential for gene regulation and transcriptional activation.

Several techniques have been developed to measure chromatin accessibility, including Mnase-seq, Dnase-seq, FAIRE-seq, and ATAC-seq. Among them, ATAC-seq (Assay for Transposase-Accessible Chromatin with sequencing)(104) is the most efficient and widely used method due to its ability to simultaneously assess open chromatin, nucleosome positioning, and transcription factor binding, while requiring minimal starting material. ATAC-seq employs an engineered Tn5 transposase, which simultaneously fragments the DNA and integrates sequencing adapters into regions of the genome that are accessible and less tightly bound by nucleosomes, enabling high-throughput identification of open chromatin regions (OCRs).

The computational analysis of ATAC-seq data involves multiple steps: quality control, alignment to a reference genome, peak calling to identify OCRs, and annotation to associate peaks with genes. Further downstream analyses include differential accessibility, motif enrichment, and footprinting to explore the regulatory landscape of chromatin. ATAC-seq has become a powerful tool for genome-wide characterization of chromatin accessibility and gene regulation.

1.10 Enhancer biology

Enhancers are pivotal regulatory DNA elements that modulate gene expression by interacting with promoters, often over large genomic distances. They are characterized by specific histone modifications, such as histone H3 lysine 4 monomethylation (H3K4me1) and histone H3 lysine 27 acetylation (H3K27ac), which help distinguish

them from promoters and mark them as active regions(105). These modifications also contribute to maintain an open chromatin state, which is necessary for the binding of TFs and other regulatory proteins. The dynamic nature of these histone marks allows enhancers to rapidly respond to cellular signals, thus modulating gene expression in a context-dependent manner.

Enhancers contain binding sites for TFs and co-factors, which are crucial for the recruitment and assembly of the transcriptional machinery. The binding of TFs to enhancers in turn facilitates the recruitment of chromatin remodelers and histone-modifying enzymes that alter the chromatin landscape, thereby enhancing the accessibility of the enhancer region.

One of the key mechanisms by which enhancers exert their function is through physical interactions with target promoters, a process known as enhancer-promoter looping. This looping is also facilitated by the binding of TFs and co-factors that bring enhancers and promoters into close proximity, allowing for the direct transfer of regulatory signals. This spatial organization is crucial for the precise regulation of gene expression, enabling enhancers to activate transcription over long distances.

Active enhancers are often transcribed by RNA polymerase II into non-coding RNA molecules known as enhancer RNAs (eRNAs)(106,107). These eRNAs are typically short, non-polyadenylated, and bidirectionally transcribed. Although the exact functions of eRNAs are still under investigation, they are believed to play a role in stabilizing enhancer-promoter interactions. eRNAs may also interact with TFs to enhance transcription efficiency. Additionally, eRNAs have been implicated in chromatin remodeling and the recruitment of histone modification enzymes. Different studies highlight that eRNA transcription is highly cell type-specific, and positively correlated with enhancer activity and target gene expression.

In summary, enhancers are complex regulatory elements that integrate multiple molecular mechanisms to fine-tune gene expression. Through TF binding, chromatin remodeling, enhancer-promoter looping, and the production of eRNAs, enhancers play a critical role in orchestrating cell type-specific transcriptional programs.

1.11 Current knowledge on the epigenomic profile of human tumor-infiltrating Tregs

Recent studies have significantly advanced our understanding of the molecular profile of human TI-Tregs through comprehensive investigations of their transcriptome, employing also single cell-based methodologies. These efforts have shed light on potential regulatory networks and specific gene programs associated with TI-Tregs. Among the different contributions, Zheng and colleagues presented a pan-cancer single-cell landscape of tumor-infiltrating T cells, utilizing scRNA-seq on T cells derived from 21 different cancer types. They identified a highly abundant and tumor-enriched activated Treg population characterized by TNFRSF9^{+} expression. This study also highlighted several signature genes for this tumor-reactive subset, with HIVEP1 emerging as a key TF regulator. Furthermore, they identified 18 TFs that are shared between terminally exhausted T cells and TI-Tregs. Additional studies have molecularly characterized Tregs across various malignancies, including hepatocellular carcinoma(108), colorectal cancer (CRC) and non-small cell lung cancer (NSCLC)(109) intrahepatic cholangiocarcinoma(110) and head and neck squamous cell carcinoma(111). Collectively, these works have elucidated the transcriptional specificities of TI-Tregs and proposed key TFs that drive their phenotype.

In parallel, recent studies integrating CRISPR screens with single-cell transcriptome profiling have aimed to dissect the crucial regulatory networks governing human Tregs. Notable among these is the work by Marson et al., which explores the regulatory roles of several known and unknown TFs within human Tregs, including FOXP3, PRDM1, and IRF4(112). Additionally, a study led by Obradovic et al.(113) identified 17 candidate master regulators (MRs) that are critical for the transcriptional state and recruitment of TI-Tregs, confirming the essentiality of eight MRs in regulating TI-Treg behavior through pooled CRISPR-Cas9 screening. Moreover, emerging research has begun to focus on the epigenomic profiles of both tumor and non-tumor Tregs. These studies indicate that the chromatin landscape of Tregs is dynamically shaped upon interaction with the TME, consistently with their transcriptional specifications. Various epigenomic layers are being explored to uncover chromatin rewiring in TI-Tregs. In this context, insights into chromatin accessibility changes have been reported in studies involving Tregs from hepatocellular carcinoma(114) and lung cancer(115).

The chromatin landscape and the transcription factors (TFs) act synergistically to drive the adaptation of Tregs within the TME. Different TFs have been already described as key regulators of TI-Treg activation (or repression) and proposed as potential novel targets for cancer immunotherapy.

One of the most studied TFs in the context of Treg differentiation and activation both in healthy tissues and in tumors is BATF, that belongs to the AP-1/ATF factor superfamily and dimerizes with members of the Jun family of proteins. This factor has shown important roles in many T cell lineages, including $CD8^+$ T cells(116), IL-17–producing helper T cells(117) and follicular helper T cells(118). BATF has been reported to play an essential role in reprogramming Tregs in healthy human tissues. In a study from Hayatsu et al. published in 2017, the authors identified BATF as a crucial regulator in tissue-specific regulatory T cells, with mutations in Foxp3 altering its interaction with BATF and affecting Treg cell function, leading to tissue-restricted inflammation(119). In a second paper, Delacher et al. reported BATF as a key TF driving the molecular program for the differentiation of precursor cells into nonlymphoid-tissue regulatory T cells, shaping their chromatin accessibility and guiding their functional specialization in tissue homeostasis and regeneration(82). Recently, the role of BATF has been clarified in the context of TI-Tregs. One study highlighted BATF as a key co-regulator with IRF4 in controlling the immunosuppressive molecular program of intratumoral Tregs(120). Subsequently, Itahashi et al. identified BATF as essential for the differentiation of naïve Treg cells into effector Tregs in the TME of NSCLC patients, with increased chromatin accessibility at BATF-binding sites mediating this process. In BATF-deficient mice, Treg cell numbers were significantly reduced in tumors, while $CD8^+$ T cell infiltration remained unchanged. High BATF expression in Tregs was also associated with poor prognosis in cancer patients, highlighting its role in promoting tumor tolerance(115). Another paper indicated BATF as a key *trans*-regulator of a highly suppressive $TNFR^+$Treg subpopulation enriched in the TME of head and neck squamous cell carcinomas (HNSCC), and in multiple other tumor types, including melanoma and lung cancer(111). Experiments on *ex vivo* activated human primary Tregs with BATF deletion suggest that this factor regulates the survival, trafficking and suppressive function of human activated Tregs, and limits their excessive activation. Moreover,

consistently with previous studies, a BATF integrated network was found to control key signature genes involved in the activation and differentiation of the TNFR$^+$ TI-Tregs, including CXCR6, TNFRSF4, TNFRSF9, RELB, ICOS, and CTLA4, supporting the notion that BATF regulates the identity of these TI-Tregs(111).

As mentioned above, IRF4 is a member of the IFN regulatory factor family that cooperates with BATF/Jun family proteins. In the past, its role in the differentiation of effector Tregs under T_H1- and T_H2-type inflammatory conditions and its synergistic action together with Blimp-1 for their functional maturation has been established, leveraging ChIP-qPCR for the investigation of relevant TF binding sites in mice(121). Moreover, IRF4, along with its partner BATF, orchestrates a transcriptional program that drives the differentiation and suppressive function of a subset of effector Tregs within the TME. These IRF4$^+$Tregs express a range of immunosuppressive molecules, are more prevalent in tumors, and correlate with exhausted T cells and poor prognosis across multiple cancers. IRF4 largely relies on BATF to enhance its DNA binding at AP-1/IRF composite elements (AICEs), driving the transcription of genes necessary for Treg activation(120).

A member of the Forkhead box family of TFs, FOXO1, is known to act in cooperation with FOXO3 to control Treg differentiation. Foxo proteins are shown to be essential for Treg cell development and function, regulating Foxp3 expression and preventing Treg cells from adopting effector T cell characteristics(122). Moreover, ChIP-seq analysis has revealed that Foxo1 binds to around 300 target genes, including *Ifng*, regulating a unique genetic program crucial for repression of pro-inflammatory genes in Tregs, that is independent of Foxp3(122). Studies on mice reveled that Foxo1 in TI-Tregs is downregulated, leading to enhanced immune suppression by facilitating Treg migration to non-lymphoid tissues and inhibiting CD8$^+$ T cell responses. In fact, manipulation of Foxo1 showed potential for selectively breaking tumor immune tolerance without triggering autoimmunity(123).

Helios (encoded by IKZF2) and Eos (encoded by IKZF4) are two important Ikaros family members with established crucial roles for Treg development, stability, and function. Treg-specific Helios deficiency in mice leads to impaired suppressive function *in vivo*, linked to reduced FoxP3 expression, impaired STAT5 activation, and consequently increased effector cytokine production, ultimately causing autoimmune responses(124,125). Importantly, selective Helios deficiency in CD4 Tregs promotes

their instability within the TME, leading to their conversion into proinflammatory effector cells and enhancing antitumor immunity, presenting a potential strategy for cancer immunotherapy(126). On the other hand, Eos is thought to act as a crucial mediator of the Foxp3-dependent gene repression, by directly interacting with Foxp3 and inducing chromatin modifications that establish gene silencing. Eos is critical for multiple aspects of Treg suppressor function, as mice with Treg-specific Eos deletion develop autoimmune conditions and show reduced tumor growth accompanied by enhanced $CD8^+$ T cell-mediated production of IFN-γ/TNF-α, despite exhibiting normal Treg function in vitro(127).

The NF-kB pathway is known to be highly utilized by both conventional T cells and Tregs. However, a study published in 2017 highlighted that canonical NF-κB signaling regulates a unique transcriptional program in Tregs compared to Tconvs, including lineage-specific genes like FoxP3 and Ikzf2, crucial for Treg identity and function. Moreover, by performing NF-kB ChIP-sequencing and ATAC-seq, distinct NF-κB binding patterns and altered chromatin accessibility at these sites were found in Tregs compared to Tconvs, suggesting that Treg-specific chromatin structures play a role in this differential regulation(128). Both canonical subunits of NF-kB, c-Rel (encoded by REL gene) and RelA, and alternative NF-kB components, encoded by NFKB2 and RELB genes, have a central role in Treg development, maintenance and function(128,129). Inhibition of c-Rel specifically impairs the generation and function of activated Tregs, and has been linked to reduction of melanoma growth and higher efficacy of anti-PD-1 immunotherapy(130). Notably, the NF-kB signalling is tightly related to the tumor necrosis factor receptor superfamily (TNFRSF) signaling. Indeed, TNFRSF signaling activates RelA to regulate cell survival and proliferation of eTregs in both lymphoid and non-lymphoid tissues, including intestinal $ROR\gamma t^+$ Treg cells.

Recently, TFs involved in the non-canonical NF-kB pathway, including NFKB2, RELB and BCL3, have been described to regulate gene programs highly active in the aforementioned $TNFR^+$ TI-Tregs(111).

2 Aim of the project

Regulatory T cells (Tregs) are crucially involved in tumor progression, mainly for their immunosuppressive mechanisms which counteract anti-tumor immune responses keeping a favorable microenvironment for cancer cells. Therefore, tumor-infiltrating Tregs (TI-Tregs) emerge as novel attractive targets in cancer immunotherapy, as their depletion or functional reprogramming can reactivate tumor immunity. However, the strategies currently being explored focus on targets shared with non-tumor Tregs and potentially other T cell subsets important for immune homeostasis and anti-tumor response. For this reason, we need novel approaches that allow selective targeting of Tregs within the tumor. In this regard, a distinct transcriptional profile has been uncovered in TI-Tregs compared to Tregs in healthy tissues, thus creating an opportunity to find actionable vulnerabilities in the tumor-promoting Treg cell state. In this study, we aim to understand how Tregs integrate the tumor microenvironment cues at epigenomic level to acquire the observed transcriptional changes, which reflect a functional adaptation within the TME. Indeed, epigenomic features are dynamic in response to external stimuli and enable functional adaptation by mediating specific regulatory networks. In detail, the project will pursue the following Specific Aims:

1) Build a comprehensive atlas of regulatory elements with *bona fide* enhancer features in human tumor and non-tumor Tregs
2) Reveal TI-Treg specific regulatory networks mediated by enhancers and DNA-binding transcription factors (TFs), as key *cis*- and *trans*-regulators of differential gene programs.
3) Identify key TFs shaping the TI-Treg cell phenotype, with the potential to provide novel approaches for the design of selective immunotherapies.

To address these aims, we profile and integrate different epigenomic features of *ex vivo* Tregs deriving from distinct tissue sources of cancer patients. We focus primarily on chromatin accessibility, histone modifications, and transcription of enhancer RNAs, as features allowing to identify the Treg *bona fide* “enhancerome”. We cross epigenomic and transcriptomic profiles for TI-Tregs and non-tumor Tregs seeking specific alterations, and measure TF activity at enhancer level with state-of-the-art computational methods. By combining multiple omics, we envisage to pinpoint novel candidate targets that might prove to be clinically relevant.

3 Material and methods

3.1 Isolation of T cells from human primary tissues

Regulatory T cells (Treg) and other T cells were isolated from tumor resections of Non-Small Cell Lung Cancer (NSCLC) or Colorectal Cancer (CRC) patients, non-neoplastic tissue adjacent to the tumor (NAT) and peripheral blood (PB) of the same patients, and from peripheral blood from healthy donors.

Human lung and colorectal biopsies were obtained respectively from Humanitas Research Hospital (Rozzano) and San Paolo Hospital (Milan), following ethical approval from their Institutional Review Boards. Informed consent was obtained from all patients prior to acquisition of the samples. Samples were confirmed to be tumor or normal based on pathologist assessment and were obtained prior to treatment.

Fresh tissues were processed by mechanical and enzymatic dissociation in order to obtain a single-cell suspension. The tissue dissociation was necessary for solid tissues such as NSLC tissue, CRC tissue and their non-neoplastic counterparts.

The dissociation of colon and lung tissues and the isolation of mononuclear cells have been performed by Dr. Ramona Bason and described in detail as part of her doctoral thesis.

3.2 Purification of human Treg cells by Fluorescence-Activated Cell Sorting (FACS)

Treg cells were purified by FACS using the following fluorochrome conjugated antibodies: anti-CD4 APC/Cy7 (Biolegend clone OKT4), anti- CD127 PE (Miltenyi, clone MB15-18C9) and anti-CD25 FITC (eBioscience, clone 4E3), using FACSAria II (BD biosciences). $CD4^+$ Tregs were sorted as $CD4^+$ $CD25^+$ $CD127^{dim}$ cells. In addition to the Treg cells, other four different populations of immune cells have been isolated starting from the tissue specimens, according to their specific markers: $CD4^+$ T conventional cells (Tconv) as $CD3^+$ $CD4^+$ $CD45RO^+$ cells; Type 1 regulatory T cells (Tr1) as $CD3^+$ $CD4^+$ $CD27^+$ $CD195^+$ cells; $CD8^+$ memory T cells (Tmem) were isolated as $CD3^+$ $CD8^+$ $CD45RO^+$ $TIM3^-$ $CD39^-$ cells, whereas $CD8^+$ exhausted T cells (Tex) as $CD3^+$ $CD8^+$ $CD45RO^+$ $TIM3^+$ $CD39^+$ cells.

To validate the expression of specific Treg markers, the following fluorochrome conjugated antibodies were used after sorting: anti-OX40 (Biolegend, clone Ber-

ACT35), anti-CTLA-4 PE (eBioscience, clone eBio20A (20A, A3)), and anti-FOXP3 APC (eBioscience, PCH101 clone).

3.3 Epigenomic techniques

The sample preparation, bulk ATAC-seq (Assay for Transposase-Accessible Chromatin with sequencing) protocol, and ChIPmentation-seq (Chromatin Immunoprecipitation followed by Tn5 transposase-mediated tagmentation sequencing) protocol, have been performed by Dr. Ramona Bason and described in detail in her doctoral thesis.

ATAC-seq was performed on *ex vivo* Treg cells, Tr1, Tconv, Tmem, and Tex cells, derived from 3 CRC patients and 7 NSCLC patients. For the sequencing, the number of starting cells used for each experiment was 5000. Properly amplified genomic DNA libraries were sequenced on the HiSeq4000 Illumina platform, using the paired-end sequencing mode with 125 bp reads, with an average sequencing depth of 240 million reads for each DNA library.

ChIPmentation with sequencing was performed on the same sorted cell populations derived from primary tumors of 10 NSCLC patients (of which 7 of them were pooled in two separate samples to have enough starting material) and the peripheral blood of 6 healthy donors. The number of starting cells was minimum 15,000. Data for a complete set of the most commonly used histone marks was generated consisting of H3K4me3, H3K27ac, H3K4me1, H3K36me3, and H3K27me3.

Properly amplified libraries were sequenced on the HiSeq2500 Illumina platform, using the single-end sequencing mode with 50-75 bp reads, with an average sequencing depth of 40 million reads for each DNA library.

3.4 ATAC-seq data processing and quality control

For the primary analysis of ATAC-seq data (*i.e.,* quality control of sequencing data, alignment to the reference genome, reads filtering, peak calling) a Nextflow pipeline available on the nf-core community (*https://nf-co.re/atacseq*)(130(131) for ATAC-seq data was implemented and optimized for our lab and run with the support of Singularity container technology to ensure reproducibility.

The sequence quality along all sequenced reads was estimated using FastQC v0.11.9(132) and MultiQCv1.9(133). To remove masked adapter sequences and to

cut sequences from low-quality score regions, TrimGalorev0.6.5 was used. In particular, base calls with Q score < 20 were trimmed from the 3' end of the reads and reads shorter than 36 bp were filtered out. The reads were aligned to the human hg38 reference genome (UCSC hg38 assembly) using BWA-mem(Burrows- Wheeler Alignment)(134), sorted using SAMtoolsv1.10 and directly converted into binary files (BAM). Multi-mapped reads, PCR duplicate and low-quality reads were marked and removed using Picardv2.23.1 and SAMtoolsv1.10. Sequences of mitochondrial origin as well as reads mapped to ENCODE hg38 blacklisted regions were discarded using BEDtoolsv2.29.2 and SAMtoolsv1.10. The Ataqv tool(135), developed by Parker Lab, was used to compare different metrics across all the experimental samples, including the Fragment Length Distribution and the TSS enrichment. From mapped reads of Treg samples, peaks were called with MACS2v2.2.7.1(136), using the following parameters: --nomodel --shift -100 --extsize 200 --qvalue 0.01--keep-dup all --call-summits. FeatureCounts was then utilized to count reads in all the called peaks from all the replicates.

3.5 Generation of ATAC-seq consensus peak-set and peak concordance analysis

Our Treg ATAC-seq dataset included 5 tumor-infiltrating Treg (TI-Treg) samples, 5 NAT-Treg samples and 4 PB-Treg samples. A consensus peak-set across all Treg samples was created using BEDtoolsv2.29.2 *merge*. We retained all the peaks present in at least two replicate samples out of all the 14 ATAC-seq Treg samples. Peaks found in unplaced and unlocalized scaffolds, peaks mapped to Y chromosome and to Pseudo-Autosomal Regions (PARs), were removed resulting in 97,669 Treg consensus peaks, that correspond to Treg open chromatin regions (OCRs).

To check the across-tissue concordance of called peaks, PB-, NAT-, and TI-Treg condition peaks were defined as peaks called in at least one of the Treg samples deriving from the same tissue.

3.6 ATAC-seq data normalization and transformation with DESeq2

Raw ATAC-seq counts for the consensus peak-set have been normalized using the DESeq2 method(137). All peaks were confirmed to pass a threshold for the sum of raw counts across all the samples equal to 10. The DESeqDataSet object was created

using the design formula *"~subject + tissue*", where "subject" corresponds to the patient ID variable, while "tissue" corresponds to the tissue of origin variable (PB, NAT, or TI). Indeed, given the availability of matched samples, we performed a paired analysis, to explore differences between Treg states (tissue) taking into account patient variability (subject). The '*DESeq'* function was utilized to estimate size factors and estimate dispersion to the mean peak intensity with a local fitting. To the DESeq-normalized counts a 'regularized log' transformation was applied with the '*rlog'* function, to stabilize the variance between samples for peaks with small counts. The transformation was not blind to the experimental design, as suggested by DESeq2 documentation, when the data transformation is performed for downstream analysis like Principal Component Analysis.

3.7 Global exploration of sample-to-sample distances with Principal Component Analysis (PCA)

To assess sample distances while accounting for patient-to-patient variability, rlog-transformed data were used as input for the '*RemoveBatchEffect'* function from the Limma package, setting as batch the "subject" variable. Then, 500 top consensus peaks with the highest variance across Treg samples were considered to perform PCA on the batch-corrected counts using the '*prcomp'* function, with the *center* and *scale* parameter set both as true.

3.8 Genomic annotation of the consensus peak-set

To annotate OCRs to genes, we first assigned them univocally to the gene with nearest Transcription Start Site (TSS), using HOMERv4.11 *annotatePeaks* and the annotation GTF file from UCSC hg38 assembly. Then, the consensus OCRs were intersected with genomic windows $\pm$ 500 bp from the TSS (or multiple TSSs) of known hg38 genes, to distinguish TSS-proximal and TSS-distal OCRs. An additional strategy was then used to improve the gene annotation for those TSS-distal OCRs. In fact, a public dataset of H3K27ac HiChIP (High-throughput Chromatin Conformation Capture with Immunoprecipitation) on human PB-Tregs, as reported by Mumbach et al. (2017)(138), was leveraged to get information about physicaly interacting putative enhancer-promoter regions. We used UCSC LiftOver tool to convert the published region pair coordinates from h19 to hg38. Then the regions were intersected with the

TSS genomic windows, to tag the pairs as “Enhancer-Enhancer”, “Enhancer-Promoter” or “Promoter-Promoter”, and subsequently intersected with our consensus OCRs, using *Bedtools intersect*. Only the “Enhancer-Promoter” region pairs were considered for the annotation of TSS-distal OCRs. When an overlap between our OCRs and the publicly available genomic interacting pairs could be found, priority was given to the annotation based on the enhancer-promoter looping. To select univocal gene annotation for TSS-distal OCRs involved in more than one region pair, a strategy was developed to prioritize protein coding genes over other gene biotypes, and genes with higher mean normalized expression in TI-Tregs, from our RNA-seq dataset(101), over lowly expressed genes.

3.9 Genomic Track visualization

For the visualization of ATAC-seq tracks of each sample, the GenomeCoverageBed function from bedtoolsv2.29.2. was used to generate bedGraph tracks, scaling the coverage by a constant factor calculated as 1 million divided by the number of mapped reads. The tracks were converted into the bigwig format using ENCODE bedGraphToBigWigv357. The Integrative Genomics Viewer (IGV)(139) software was used for exploring and browsing the tracks.

3.10 Differential accessibility analysis separately for NSCLC and CRC samples

A differential chromatin accessibility analysis was performed using DESeq2 on ATAC-seq counts from the 14 Treg samples, applying the design formula “~ disease + disease:subject + disease:tissue”. This analysis aimed to assess differences between TI-Tregs, NAT-Tregs, and PB-Tregs, separately for CRC and NSCLC. Specifically, comparisons between TI-Tregs vs. NAT-Tregs and TI-Tregs vs. PB-Tregs were explored within each disease context. Additionally, interaction terms (disease:tissue) were included in the design to investigate whether the effect of tissue on chromatin accessibility varied depending on the disease context.

3.11 ChIPmentation data processing and quality control

The quality control of the reads was performed using FastQC v.0.11.9 and MultiQCv1.9 (http://www.bioinformatics.babraham.ac.uk/projects/). The reads were aligned to the human UCSC hg38 reference genome using BWA- mem v.0.7.17,

sorted using SAMtoolsv1.10 and directly converted into binary files (BAM). Multi-mapped reads, PCR duplicates and low-quality reads were marked and removed using Picardv2.23.1 and SAMtools. The quality of the immunoprecipitation samples was assessed using various tools from the DeepTools suite(140), including the *plotHeatmap* tool, which was used to visualize the signal relative to the TSS, transcription end sites (TES) of genes and the gene bodies. For the visualization of ChIPmentation-seq tracks of each sample, bedGraph tracks were generated using GenomeCoverageBed function from bedtoolsv2.29.2., scaling the coverage by a constant factor calculated as 1 million divided by the number of mapped reads. The tracks were converted into bigwig using ENCODE bedGraphToBigWigv357. Samples with a low signal-to-noise ratio, as determined by i) visual inspection of the heatmap of signal enrichment along genes created with DeepTools and ii) and genomic track visualization on IGV, were excluded from downstream analyses.

3.12 De novo chromatin state characterization

De novo chromatin state characterization was performed using a multivariate Hidden Markov Model, a machine learning approach (ChromHMM v1.12) considering five histone modifications (H3K4me3, H3K27ac, H3K4me1, H3K36me3 and H3K27me3) across our ChIPmentation library. All samples were downsampled to a maximum depth of 18 million reads, based on their average sequencing depth. The read counts for all the considered samples were computed in non-overlapping 200-bp bins across the entire genome. The binarization step was performed comparing ChIPmentation-seq read counts to corresponding IgG sample read counts as control to adjust the binarization threshold locally and thus reduce the technical noise. Chromatin state annotation tracks were generated separately for each cell population and biological replicate. Several models were trained in parallel considering 8, 10, 12, 14, and 15 number of states. Of these, the 14-state model captured the key interactions between histone marks and identified an active enhancer state, thus it was used for downstream analyses.

3.13 Identification of active enhancers

To identify the highly active enhancer elements, all the “Active Enhancer” and “Genic Enhancer” regions from ChromHMM states were selected. These two states are

defined by the co-presence of high levels of H3K27ac and H3K4me1 signal in the case of active enhancers and the co-occupancy of H3K27ac, H3K36me3 and H3K4me1 for genic enhancers. Separate consensus peaksets were generated for active and genic enhancers retaining regions present in at least two biological replicates (PB or TI-Treg). The consensus peaksets were further filtered for RPKM counts > 1 based on H3K27ac signal, in order to keep highly active regions. Then, a combined consensus peakset was built for active and genic enhancers using DiffBind v2.10.0(141). This generated a consensus peakset of 8,163 regions across the seven samples.

Integration of ChromHMM states and ATAC-seq data

A spider plot was generated to evaluate the overlap of consensus OCRs within chromatin states annotated in Treg samples. The enrichment score for each ChromHMM-defined state was calculated as the ratio between the total number of OCR peaks with midpoints overlapping a given chromatin state and the total genomic length covered by that chromatin state.

To quantify overlap, Bedtools v2.29.2 was used to intersect consensus OCRs with ChromHMM-annotated enhancer regions. ATAC-seq peaks were classified as enhancers if at least 40% of their base pairs overlapped with a ChromHMM-defined enhancer region.

3.14 Integration of 5' single-cell RNA-seq libraries from human T lymphocytes

We integrated 11 publicly available 5' single-cell RNA-seq libraries from sorted CD4+/CD3+ T lymphocytes to identify Treg populations and perform enhancer RNA analysis using the ReapTEC pipeline. The dataset comprised 4 samples from tumor tissue, 1 from metastasis, 3 from NAT, and 3 from PB, derived from 4 patients with CRC or NSCLC.

Preprocessing was conducted using the python package Scanpy (v1.6.0)(142). For quality control, cells with fewer than 200 detected genes and genes detected in fewer than 0.1% of total cells were excluded. Cells with a high percentage of mitochondrial or ribosomal gene counts, abnormal total gene number or total counts were also removed based on visual inspection of the aforementioned quality metric distributions. Counts were normalized per cell using a scaling factor of 10^4 and log-transformed. Highly variable genes were identified using default Scanpy parameters and excluding

mitochondrial and ribosomal genes. Harmony(143) was employed to correct for batch effects, particularly for variations across different subjects. A K-Nearest Neighbor graph was constructed based on Euclidean distance in PCA space (*scanpy.pp.neighbors* with n_neighbors=15, n_pcs=20). Uniform manifold approximation and projection (UMAP) was used for visualization of the data.

Leiden clustering was performed at a resolution of 0.8. Three clusters were removed from further analysis due to a low number of cells, batch-specific representation, or contamination identified by marker gene expression from non-T cell lineages (e.g., innate immune or B cells). Clusters were annotated by assessing the expression of canonical Treg markers, including *FOXP3*, *CTLA4*, and *IL2RA*. The FOXP3$^+$ Treg cluster was confirmed through differential expression analysis between clusters and gene set enrichment analysis with previously published Treg gene sets(109). Cell barcodes corresponding to this cluster were selected for downstream ReapTEC analysis.

3.15 Discovery of enhancer RNAs

For the identification of enhancer RNAs (eRNAs) from single-cell transcriptomes, we utilized the ReapTEC (Read-level pre-filtering and transcribed enhancer call) pipeline(144). ReapTEC is designed to process 5′ scRNA-seq, allowing the precise identification of transcription start sites (TSSs) based on read-level features associated with capped RNA fragments.

The FASTQ files were mapped to the human genome (hg38) using STARsolo (v2.7.10b), implementing a whitelist of cell barcodes to filter for Treg cells. Uniquely mapped reads (MAPQ=255) were extracted using SAMtools (v1.16.1), retaining only the first read (R1) in each paired-end alignment. Duplicated reads were removed using UMI-tools (v1.1.2), ensuring unique transcript quantification across cell barcodes. To isolate capped RNA fragments, reads containing a mismatched guanine (G) base were extracted. These bases are tagged as soft-clipped G nucleotides by STARsolo, which captures the reverse transcription at the 5′ cap of the nascent RNA molecule. These reads representing possible TSS regions were converted to 1bp and strand-separated using SAMtools to ensure accurate identification of forward and reverse transcription events.

Bidirectionally transcribed candidate enhancers were identified based on a previously published computational workflow(145). Separate clusters of forward or reverse 1bp reads located within 10 bp of each other were initially merged. Then forward and reverse tag clusters within 400bp of each other were paired to identify candidate eRNAs. Subsequently, those that overlapped known TSS regions of protein-coding genes (±300 bp) were filtered out to exclude promoter-derived transcription. To quantify eRNAs and determine their bidirectionality, reads overlapping eRNAs were counted in a strand-specific manner using bigWigAverageOverBed. Only eRNAs with a bidirectionality score of $|D| < 0.8$ were retained for further analysis, ensuring that the identified regions reflected enhancer activity.

3.16 Integration of enhancer RNAs with ChromHMM-derived chromatin states and Open Chromatin Regions

To assess the enrichment of eRNAs within chromatin states, as defined by ChromHMM in Treg samples, we followed a similar approach to the one used for OCRs. A difference in the analysis of eRNAs was the detailed exploration of enrichment within the *Active TSS* state. For this, we subdivided regions annotated as *Active TSS* into two categories: regions overlapping the TSSs of known genes and other regions, located mainly within gene bodies or intergenic regions.

Additionally, we generated a consensus of “active” chromatin states by selecting regions annotated as *Active TSS*, *Flanking TSS*, and *Flanking TSS Downstream* that were present in at least two out of the seven biological replicates of Treg samples. The consensus of active chromatin states and enhancer regions was derived using ChromHMM annotations.

For the integration analysis, we used Bedtools v2.29.2 to intersect the eRNAs with three datasets: (1) the consensus of enhancer regions, (2) the consensus of active chromatin states derived by ChromHMM, and (3) the OCR peak set from ATAC-seq data. An eRNA was considered to overlap a given region if at least 50% of its base pairs intersected with that region.

3.17 Annotation of OCRs as enhancers or promoters

Following the integration of ATAC-seq data with ChromHMM-derived enhancers and eRNAs, we annotated our consensus of OCRs for downstream analyses. TSS-

proximal OCRs, defined as OCRs within genomic windows ± 500 bp from the TSS of genes, were annotated as “Promoter”. TSS-distal OCRs with either (1) evidence of enhancer chromatin state, thus overlapping with the ChromHMM-annotated enhancers, or (2) evidence of eRNA transcription, or (3) presence of both, were annotated as “Enhancer”. All the other OCRs were annotated as “Other”.

3.18 Integration of OCRs with Treg bidirectionally transcribed enhancers dataset

With the aim of refining enhancer OCR coordinates for downstream analyses involving transcription factor motif mapping, Enhancer OCRs were intersected with a public dataset of bidirectionally transcribed eRNAs active in human PB-Treg cells, derived from ReapTEC analysis of a scRNA-seq dataset of $CD4^+$ T lymphocytes from healthy donors(144). Henceforth, these public eRNA are referred to as btcEnhs to distinguish from the eRNAs discovered in this work. We considered 43,900 btcEnh identified in the dataset with activity in the Treg cluster. A btcEnh was considered to overlap an OCR if at least 50% of its base pairs intersected with that region.

3.19 Refinement of OCR coordinates based on ATAC-seq and enhancer RNAs

For downstream transcription factor (TF) motif analysis, the original OCR coordinates were refined by defining 500 bp windows centered on midpoints determined by their overlap with eRNAs identified in this work or public btcEnh regions for Tregs. Specifically:

- For enhancer OCRs overlapping with eRNAs, the midpoint of the eRNA core region - defined as the interval between the forward and reverse tag clusters indicating the potential TSS of enhancers - was used to center the 500 bp region.
- For enhancer OCRs overlapping a public btcEnh but not eRNAs, the btcEnh midpoint was used as the center.
- For promoter OCRs, enhancer OCRs overlapping ChromHMM active enhancer regions, and other regions, the consensus OCR summit base-pair (the peak of the highest merged ATAC-seq signal across all Treg samples) was used to center the 500 bp window.

Next, a two-step process was applied to generate a refined, non-overlapping peak set of 500 bp regions. First, for OCRs overlapping multiple eRNAs or btcEnh regions, we prioritized the window closest to the ATAC-seq summit of the OCR. In the second step, overlaps across different OCR types were resolved by prioritizing enhancer regions over promoters and promoters over other regions. Among regions of the same category, those encompassing the ATAC-seq summit and with the smallest offset were favored. If regions remained unresolved, the region with the highest mean ATAC-seq counts (baseMean) across all Treg samples was selected.
This process resulted in a consensus peak set of 97,379 fixed, non-overlapping OCRs, which was used for all subsequent transcription factor analyses.

3.20 Differential chromatin accessibility analysis across Treg cell states

A differential chromatin accessibility analysis was performed on all the 97,669 consensus OCRs with DESeq2 using ATAC-seq counts from the 14 Treg samples, applying the design formula "~ subject + tissue" as previously mentioned. Differentially accessible regions (DARs) were defined as having an adjusted p-value < 0.05.

3.21 Correlation between RNA-seq and ATAC-seq data

We used 32 Treg samples from the bulk RNA-seq dataset published by De Simone et al.(101), including 14 TI-Treg, 7 NAT-Treg, and 11 PB-Treg, from CRC and NSCLC patients. FPKM counts were calculated for the Treg RNA-seq dataset and mean FPKM counts were computed for TI-Treg, NAT-Treg and PB-Treg conditions, by aggregating counts for all Treg replicates from the same tissue. For each gene in the RNA-seq dataset, an accessibility score per Treg condition was computed by summing the mean ATAC-seq DESeq2-normalized counts for all the OCRs annotated to the gene. For genes with only one OCR annotated, the mean count of that OCR was used as the accessibility score. The mean FPKM count and the accessibility score per gene per condition were correlated to define relationships between chromatin accessibility and gene expression between the two independent datasets.

3.22 Differential gene expression analysis between TI-Tregs and NAT-Tregs

Raw RNA-seq counts for 22,295 genes from the previously mentioned dataset were analyzed with DESeq2 using "~tissue" as the design formula. The coefficient

contrasting TI-Treg vs NAT-Treg was used for testing differentially expressed genes, which were defined as genes with adjusted P-value < 0.05.

3.23 Calculation of accessibility change score for each gene

For each gene differentially expressed between TI- and NAT-Tregs, a score representing the variation in chromatin accessibility between the two conditions was calculated. In particular, the median Log2FoldChange of ATAC-seq count in TI-Treg compared to NAT-Treg was computed separately for all enhancer and promoter OCRs, thus obtaining two accessibility change scores for each gene, one at enhancer level, and one at promoter level. These scores were then correlated with the Log2FoldChange of RNA-seq count for each gene.

3.24 K-means clustering of differentially accessible enhancers

We identified a total of 5,721 differentially accessible enhancer OCRs (DAR enhancers) from pairwise comparisons of chromatin accessibility between TI-Tregs, NAT-Tregs, and PB-Tregs. For these DAR enhancers, the mean chromatin accessibility per condition was calculated using rlog-transformed ATAC-seq counts. These values were then standardized to z-scores to remove scaling differences and emphasize relative changes in accessibility between conditions.

We performed k-means clustering on these z-scores, testing both 3 and 5 cluster solutions. The optimal number of clusters was initially suggested as 3 by Gap Statistics. However, by increasing to 5 clusters, we were able to differentiate a distinct group of enhancers that exhibited preferential increases in chromatin accessibility specifically in TI-Tregs. This finer resolution was not apparent in the 3-cluster solution Given the biological relevance of this TI-Treg-specific group, the 5-cluster solution was retained for downstream analyses and functional validation, providing a more nuanced understanding of enhancer activity across Treg subtypes.

Gene Set Enrichment Analysis

For Gene Set Enrichment Analysis, we used the *gseapy* package with 1000 permutations. Differentially expressed genes between TI-Tregs and NAT-Tregs were ranked by the -log10(adjusted p-value) multiplied by the Log2FoldChange negative or positive sign. Significant enrichment or depletion was reported for FDR < 0.05.

3.25 Computing highly variable genes across Treg cell states

We identified highly variable genes (HVGs) across Treg cell states using our RNA-seq dataset following a systematic approach. First, DESeq2-normalized counts were preprocessed with a winsorization function to mitigate the influence of extreme expression values on variance and mean estimates. Specifically, the top and bottom 5% of expression values for each gene were clipped, with values outside these bounds set to the 95th and 5th percentiles, respectively. This procedure limits the disproportionate influence of outliers on variability estimates.

On the winsorized dataset, the squared coefficient of variation (CV^2) was calculated for each gene, providing a normalized measure of variance relative to the mean expression. We then fitted a generalized linear model (GLM), restricting the fit to the top 5% of highly expressed genes with a CV^2 greater than 0.3, thus avoiding bias from lowly expressed genes.

Next, the ratio of observed to expected variance was computed for each gene, reflecting its deviation from the model fit. This ratio, scaled by the degrees of freedom, was compared against a chi-square distribution to determine statistical significance. Multiple testing correction was applied using the false discovery rate (FDR), and genes with an adjusted p-value < 0.001 were considered HVGs.

3.26 Gene modules identification

For all highly variable genes (HVGs) with at least one differentially accessible (DAR) enhancer, we calculated the proportion of enhancers within each of the five clusters. The proportion was determined by dividing the number of enhancers in each cluster by the total number of enhancers associated with the gene. Using these enhancer cluster proportions, cosine similarity was computed across all HVGs, providing a quantitative measure of how similar genes are based on their enhancer cluster profiles.

The resulting cosine similarity matrix was then used to perform hierarchical clustering with Ward's method ("ward.D2"), which minimizes the total variance within clusters. Based on exploratory analysis and biological considerations, 12 gene clusters were selected. This selection aimed to separate genes with enhancer regulation predominantly restricted to Cluster 4 (representing highly private TI-Treg enhancer

regulation) from genes with combinatorial enhancer regulation, particularly those associated with enhancers from both Cluster 2 and Cluster 4. The goal was to assess whether genes regulated by enhancers from multiple clusters exhibit distinct gene expression programs compared to those with more exclusive enhancer regulation.

3.27 Pathway analysis

The R interface to the gProfiler tool was utilized to perform statistical enrichment analysis of gene ontology terms and biological pathways from multiple databases, including Reactome, KEGG and WikiPathways. A custom background gene set included all the genes associated with at least one consensus OCR. To correct p-values for multiple testing, the “g_SCS” algorithm from gProfiler was used. The explanation of the algorithm is reported on the gProfiler documentation site: https://biit.cs.ut.ee/gprofiler/page/docs. The p-value threshold used for significant terms was set as to 0.05.

3.28 Motif enrichment analysis

Transcription factor (TF) motifs were mapped onto the enhancer OCRs of each previously defined enhancer cluster separately. Motif enrichment analysis was performed using the Homer *findMotifsGenome.pl* tool, with non-overlapping 500 bp genomic regions as input, ensuring that the analysis was restricted to the core enhancer regions. For each enhancer cluster, a binomial test was conducted, using all consensus OCRs with refined coordinates as the background set.

The TF motif database was a custom set, comprising 1,179 non-redundant motifs from the TRANSFAC database, supplemented with 13 additional non-redundant motifs unique to the JASPAR database or the Human Transcription Factor Catalog (https://humantfs.ccbr.utoronto.ca/index.php). To ensure the reliability and non-redundancy of the TRANSFAC motifs, we applied a stringent selection process: for each DNA-binding factor, consensus motifs derived from experimental evidence, particularly ChIP-seq data from human samples, were prioritized. If available, we selected motifs marked as "recommended" by TRANSFAC, leading to the selection of a single consensus motif for each factor in humans. This approach included not only canonical TFs with high binding specificity but also other DNA-binding proteins such as DNMT and TET, commonly categorized as TFs due to their binding activities.

TF motifs were considered significantly enriched if they had both a p-value < 0.05 and a q-value < 0.05. To increase the robustness of the results, we filtered out TF motifs that were identified in fewer than 10 target regions and motifs of TFs non expressed in any of the Treg conditions, according to our RNA-seq dataset.

For aggregation at the TF class level, we obtained TF family annotations from both the TRANSFAC database and the Human Transcription Factor Catalog. These classifications were integrated and manually curated to harmonize TF definitions across different sources, minimizing the number of unclassified TFs. This process resulted in the identification of 56 distinct TF classes. For each TF class, the most significant TF motif within the class, represented by the maximum negative log(p-value), was used to compare the enrichment significance across enhancer clusters.

3.29 Footprinting analysis with Tobias

Transcription factor footprinting analysis was performed with TOBIAS, a python- based suite of command-line tools. To ensure reproducibility, a Nextflow- based pipeline wrapping all the main functionalities of the suite was used (https://github.molgen.mpg.de/loosolab/TOBIAS-nextflow/tree/master/TOBIAS_MAPOKS)(146). As input, we used ATAC-seq signals from 4 PB-, 4 NAT- and 4 TI-Tregs matched by patient. As a first step, ATAC-seq signals are merged by condition to generate condition *.bam* files. The ATAC-seq condition data were corrected for intrinsic Tn5 bias in two steps using the ATACorrect module: first, by defining each Tn5 cutsite as the 5' end of a read shifted by $^{+}$ 5 bp at the plus strand and -4 bp at the minus strand to center the transposition event. Then, by calculating expected Tn5 cutsites per region, using a dinucleotide weight matrix (DWM) to estimate the background probability of Tn5 insertion based on its sequence preference, and subtracting these expected cuts from the observed signals to yield corrected tracks. A continuous combined footprint score was computed using ScoreBigwig module, with a function considering both read depletion due to TF binding and accessibility signal. Once base-wise footprint scores were computed, the BINDetect module was used to scan the sequence information of original consensus OCRs for known TFs binding motifs. Position weight matrices (PWMs) of our TF motifs previously described for motif enrichment analysis were used. Once a motif was confidently matched (pval < 1e−4) to a DNA sequence within OCRs, the base-wise

score was converted by TOBIAS into a motif-wise score by taking the highest score among the ones of the bases composing the motif. A global threshold to the score was set to distinguish between bound and unbound sites, with a significance p-value set to as 0.01. For the differential footprinting analysis between multiple conditions, the individual site scores were summarized with the differential binding module by TOBIAS in order to identify differentially active TFs across conditions.

3.30 TF differential binding analysis for clusters of enhancers

To identify transcription factors (TFs) with significantly higher footprint scores within specific enhancer clusters (Cluster 2 and Cluster 4) for TI-Tregs, we used transcription factor binding sites (TFBS) derived from footprint analysis conducted with TOBIAS. The analysis focused on motifs found within non-overlapping 500 bp OCRs annotated as enhancers. A linear regression model was applied for each TF to determine whether TI-Treg footprint scores were significantly associated with enhancer clusters. For each enhancer cluster (Cluster 2 or Cluster 4), a binary variable (*is.roi*) was assigned, indicating whether the binding site was located within the target cluster or outside (i.e. in other enhancer OCRs). The model controlled for chromatin accessibility by including the log2-transformed mean chromatin accessibility score as a covariate. This score represents represents the average chromatin accessibility level for the enhancer that hosts the TFBS, specifically in TI-Tregs. The formula used was: log2(TI-Treg_footprint_score) ~ is.roi + log2(TI-Treg_accessibility_score), where TI-Treg_footprint_score represents the footprint score calculated with TOBIAS. For each TF, the coefficient and p-value for the is.roi term were extracted, indicating whether the footprint scores were significantly higher in the target enhancer cluster compared to other enhancers. P-values were adjusted using the false discovery rate (FDR) method. This analysis was conducted in parallel across TFs, with computation performed using a multi-core setup.

4 Results

4.1 Integration of multiple epigenomic features draws a comprehensive map of the enhancerome in Regulatory T cells from cancer patients

4.1.1 Experimental workflow for the identification of active enhancers

To identify active enhancers in tumor-infiltrating regulatory T cells (TI-Tregs), we adopted a multi-omics combined approach to profile key epigenomic features in highly purified human Treg cells. These features constitute interdependent layers for the identification of a chromatin region as a functional enhancer element and only their integration allows to define *bona fide* enhancers. First, the assay for transposase-accessible chromatin with sequencing (ATAC-seq) maps open chromatin regions (OCRs), which include putative *cis*-regulatory elements with binding sites for *trans*-regulators. Second, ChIPmentation with sequencing allows the genome-wide profiling of functionally distinct histone tail modifications and the *de novo* reconstruction of chromatin states with active enhancer features. Third, the precise mapping of transcription start sites (TSS) on single Treg cell transcriptomes is exploited for the discovery of bidirectionally transcribed enhancer RNAs (eRNAs), which allows for robust enhancer identification and refinement of the *core* functional enhancer region (**Figure 1**). We conduct these analyses on primary Treg cells from tumor tissue (TI-Tregs), normal adjacent tissue (NAT-Tregs) and circulating peripheral blood (PB-Tregs), collected from chemotherapy-naïve patients diagnosed with either colorectal cancer (CRC) or non-small cell lung cancer (NSCLC). These Treg populations represent distinct cell states, defined here as variations in functional and epigenetic characteristics influenced by their tissue environment. By comparing these cell states, we aimed to uncover epigenetic differences associated with Treg function in different tissue environments. In total, 173 libraries of T lymphocyte samples from 20 subjects were included in this study: 52 ATAC-seq libraries, 106 ChIPmentation-seq libraries for 5 different histone marks, and 11 single-cell RNA-seq (scRNA-seq) libraries (**Table 1**).

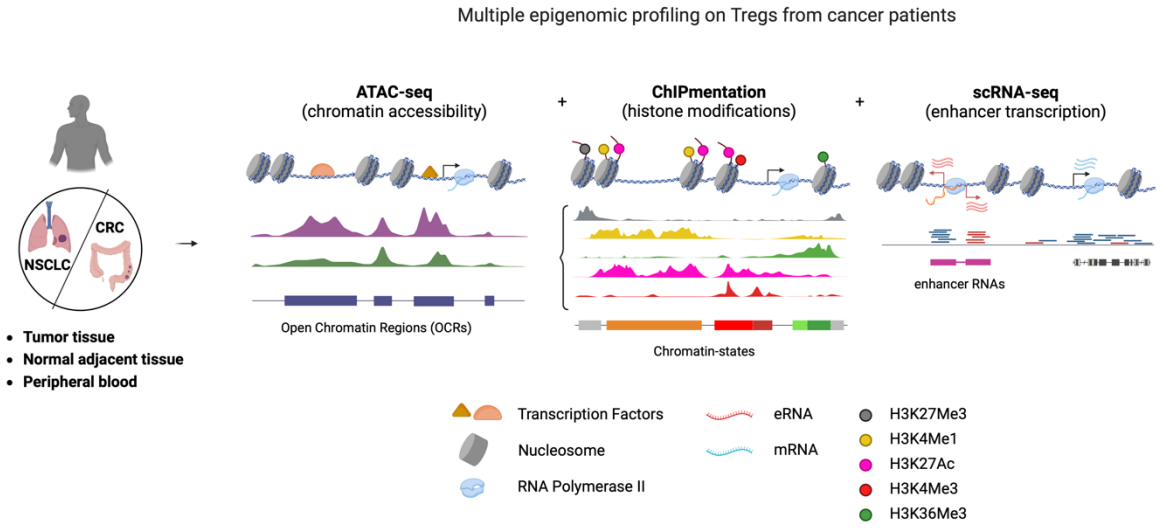

***Figure 1. Experimental workflow of the project**. Our dataset includes regulatory T cells (Tregs) isolated from either Non-Small Cell Lung Cancer (NSCLC) or Colorectal Cancer (CRC) cancer patients, deriving from three different tissue sources, the tumor, the normal tissue adjacent to the tumor (NAT) and the peripheral blood (PB). To identify high-confidence Treg enhancers, we integrated data from three distinct omics approaches: ATAC-seq, to map the open chromatin regions depleted of nucleosomes and accessible to the transcription machinery and transcription factors (TFs); ChIPmentation, to profile different histone modifications and combine them for* de novo *reconstruction of biologically meaningful chromatin states; single-cell RNA-seq for the discovery of bidirectionally transcribed enhancer RNAs (eRNAs), highly correlated with enhancer activity and target gene expression. This comprehensive integration increases our confidence that the identified regions are functional enhancers.*

Tissue	Omic profiling method						
	ATAC-seq	ChIP-mentation					5' scRNA-seq
		H3K4Me3	*H3K4Me1*	*H3K27Ac*	*H3K36Me3*	*H3K27Me3*	
Peripheral blood	15	10	9	11	9	8	3
Tumor	19	13	13	14	9	10	5
Normal adjacent to tumor	18						3
	173 sequencing libraries for T lymphocytes						
Tot. libraries	52	106					11
Tot. subjects	11	11					4

***Table 1. Summary of T lymphocyte omic libraries used in this study**.*

4.1.2 Collection of the ATAC-seq dataset of human T lymphocytes from cancer patients

All the biological material, samples, and ATAC-seq and ChIP-mentation experiments, were processed, collected, and performed, respectively, in our laboratory by dr. Ramona Bason as part of her PhD project. The scRNA-seq dataset used for

identification of eRNAs in Tregs, along with the bulk RNA-seq dataset, exploited to derive TI-Treg-specific gene signatures, have previously been published in two separate studies from our lab(101,147).

To perform ATAC-seq, primary tumors have been collected from NSCLC and CRC patients, together with matching normal adjacent tissue and peripheral blood samples. From each source, $CD4^+$ $CD25^+$ $CD127^{low}$ Treg cells have been isolated and purified by flow cytometry. In parallel with the Treg cell populations, additional T lymphocyte populations have been sorted and analyzed, including $CD4^+$ conventional T cells (Tconv, sorted as $CD3^+$ $CD4^+$ $CD45RO^+$), $CD4^+$ $FOXP3^-$ Type 1 Regulatory T cells (Tr1, sorted as $CD3^+$ $CD4^+$ $CD27^+$ $CD195^+$), $CD8^+$ memory T cells (Tmem, $CD3^+$ $CD8^+$ $CD45RO^+$ $TIM3^-$ $CD39^-$), and $CD8^+$ exhausted T cells (Tex, $CD3^+$ $CD8^+$ $CD45RO^+$ $TIM3^+$ $CD39^+$).

Using our optimized ATAC-seq protocol for low starting material, we obtained high-quality ATAC-seq libraries from a minimum of 5,000 Treg cells collected from patients. The complete ATAC-seq dataset consisted of 52 sequenced samples, derived from 10 human subjects; primary tumors were obtained from 7 NSCLC and 3 CRC patients, together with matching normal adjacent tissues and peripheral blood samples.

4.1.3 ATAC-seq libraries quality control

After sequencing, multiple quality control metrics were used to assess ATAC-seq library quality. The sequencing quality check of all the T cell libraries revealed a good read quality without any major warnings **(Figure 2, bottom panel**). The alignment of the reads to the reference genome (Human Genome 38) and the peak calling on the mapped fragments confirmed a good quality of ATAC-seq libraries (**Figure 2, top panel**). Indeed, almost all samples have more than 50 million mapped reads, the minimum cut-off suggested for open chromatin detection and differential analysis, whereas some of them have nearly 200 million reads, which is recommended for TF computational footprinting (148). Concerning the peak calling, almost all the samples have a number of peak included between 25,000 and 50,000, with some samples reaching 75,000 called peaks (**Figure 2, middle panel**).

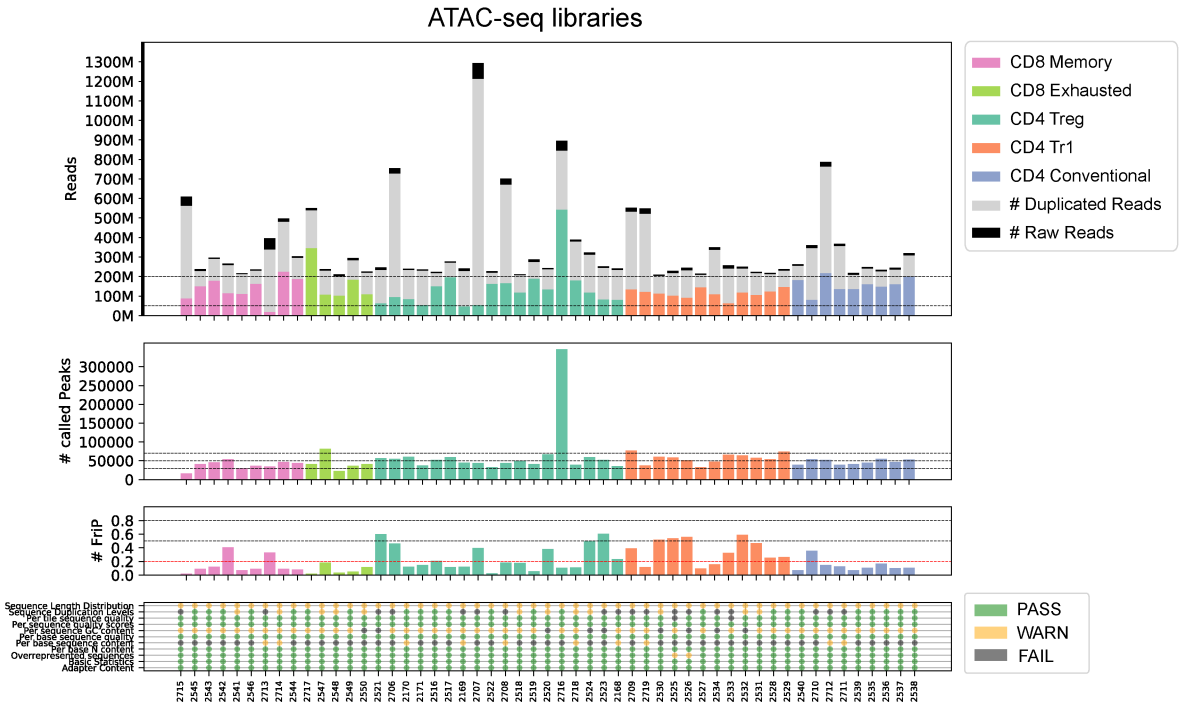

Figure 2. Summarized report of ATAC-seq data quality control. *For each sample the number of reads (top), the number of called peaks and fraction of reads in peaks (FRiP) scores (middle), and the scores relative to the read sequencing metrics with FastQC (bottom), are represented in bar plots. The black portion of each bar represents the fraction of unmapped reads, the grey one the fraction of low-quality reads (i.e., not uniquely mapped, duplicate reads and reads originated from mitochondrial DNA or repetitive blacklisted regions) that is filtered out in the post- alignment processing. The colored bars represent the final pool of reads that passed all the quality controls. The color scheme of the bars indicates different cell types: pink represents CD8⁺ memory T cells, light green represents CD8⁺ exhausted T cells, green represents CD4⁺ Treg cells, orange represents CD4⁺ FOXP3⁻ regulatory T cells, and blue represents CD4⁺ conventional T cells. The dotted lines indicate relevant thresholds for each of the metrics, as suggested by ENCODE guidelines. The dot plot color scheme (bottom) reflects the outcomes of the FastQC quality checks for each metric: green indicates values that meet optimal quality standards, orange indicates values that are flagged as warnings but still acceptable, and grey indicates values that fall below the defined quality thresholds.*

The fragment length distribution (FLD) was then compared across all the libraries and with a reference distribution that is considered reliable when the tagmentation process is correctly performed. As expected(148), the majority of samples showed decreasing and periodic peaks in the first 100 bp, at ~ 200 bp and 400 bp, corresponding to nucleosome-free regions (NFR), and mono- and di-nucleosomes, respectively. **Figure 3A** reports the FLD for all the Treg ATAC-seq samples, but a similar profile was observed for all the other samples. The enrichment of transposase Tn5 insertion events on the TSS of genes genome-wide was also evaluated as a metric for quantifying signal-to-noise ratio, with the peak of enrichment observed within a window of 200 bp from the TSS (**Figure 3B**).

However, a specific evaluation of each Treg library revealed suboptimal metrics for two TI-Treg samples, one with an unrealistic number of called peaks and an unexpected shape of the FLD, whilst the other one having relatively low signal-to-noise ratio as observed from the TSS enrichment. For this reason, these samples and their matching NAT-Treg samples were excluded from the dataset. Overall, the chromatin accessibility profiles of 5 TI-Tregs, 5 NAT-Tregs and 4 PB-Tregs samples, deriving from 3 CRC and 2 NSCLC patients were analyzed.

Collectively, these results indicate an overall good quality of our sequencing reads, aligned reads and called peaks, permitting further downstream analyses of the ATAC-seq data.

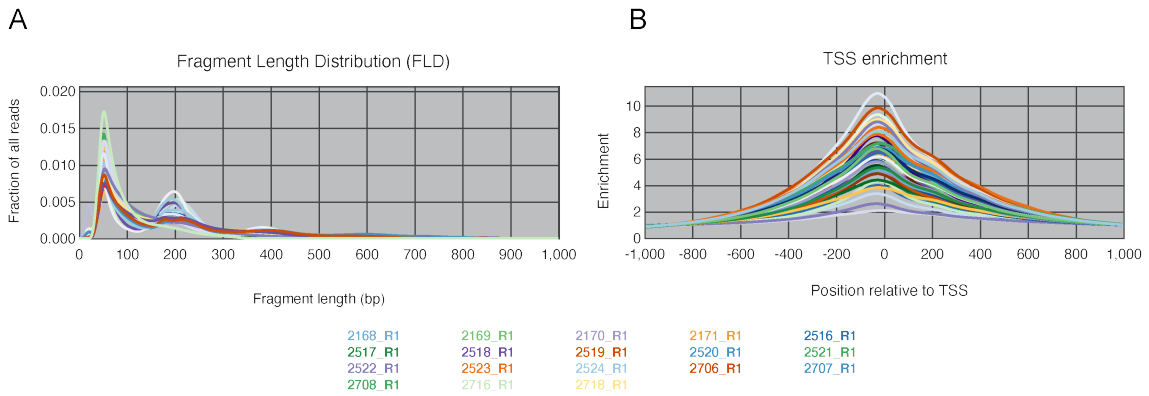

Figure 3 ***Additional quality control metrics for Treg ATAC-seq dataset.*** *Fragment Length Distribution (FLD) plot showing the distribution of fragment sizes across all Treg samples. (A) The expected peak around ~100 bp represents nucleosome-free regions, while the peaks around 200 bp and 400 bp reflect mononucleosome- and dinucleosome-bound fragments, indicating successful enrichment of accessible chromatin regions. (B) A high enrichment score at the TSS is indicative of good data quality and successful capture of open chromatin regions. The color scheme in both plots corresponds to individual samples, as indicated by the sample identifiers at the bottom.*

4.1.4 Within- and across-tissue similarities and differences in the chromatin accessibility landscape of patient-derived Tregs

By merging overlapping ATAC-seq peaks found in at least two Treg samples of the dataset, we generated a consensus peak-set of OCRs, resulting in the identification of 97,669 intervals. These OCRs represent the comprehensive collection of accessible chromatin regions for circulating, tumor-infiltrating, and normal tissue-Tregs.

To assess the similarities and differences of the chromatin accessibility landscapes across Treg samples obtained from different patients, we first counted the number of shared peaks between increasing number of Treg samples derived from the same type of tissue, irrespectively of cancer type. In PB- and TI-Tregs, the peaks conserved

in all the samples constitute the largest group, whereas highly conserved peaks represent the second largest group in NAT-Tregs, which also display a higher number of private peaks compared to TI-Tregs (**Figure 4A**).

This suggests that Tregs derived from different tumors display a more conserved chromatin accessibility landscape, whereas Tregs from corresponding normal tissues exhibit a larger number of tissue-specific private peaks reflecting different phenotypic and functional specification based on the tissue.

In addition, we found that the majority of Treg OCRs from one tissue are shared across the other two Treg cell states (**Figure 4B**). Around 62% of consensus peaks were called in at least one biological replicate from each tissue, and 22% of all OCRs were present in all the 14 Treg samples. This indicates an overall similar landscape of accessible chromatin across these three Treg types, as expected, since we are comparing three cell states of the same Treg cell lineage.

At this point, we decided to explore the sample-to-sample relationships of Tregs at epigenomic level taking into account subject-related differences. Therefore, we performed a Principal Component Analysis (PCA) on the chromatin accessibility counts after correcting for patient-derived variability. The first principal component (PC1) captured most of the variance (89%), clearly distinguishing TI-Tregs from PB-Tregs. NAT-Tregs are grouped in between and are further separated from the other samples by PC2 (**Figure 4C**). Intriguingly, the distances between the three different populations are larger in NSCLC samples compared to CRC samples. It is important to observe that TI-Treg from CRC and NSCLC grouped separately from their counterparts in NAT and PB. This is concordant with the distinct transcriptional profile of TI-Tregs(101,149) and suggests that their gene expression signature is likely to be driven by an epigenetic reprogramming in response to tumor microenvironment (TME) cues which are shared across different tumor types.

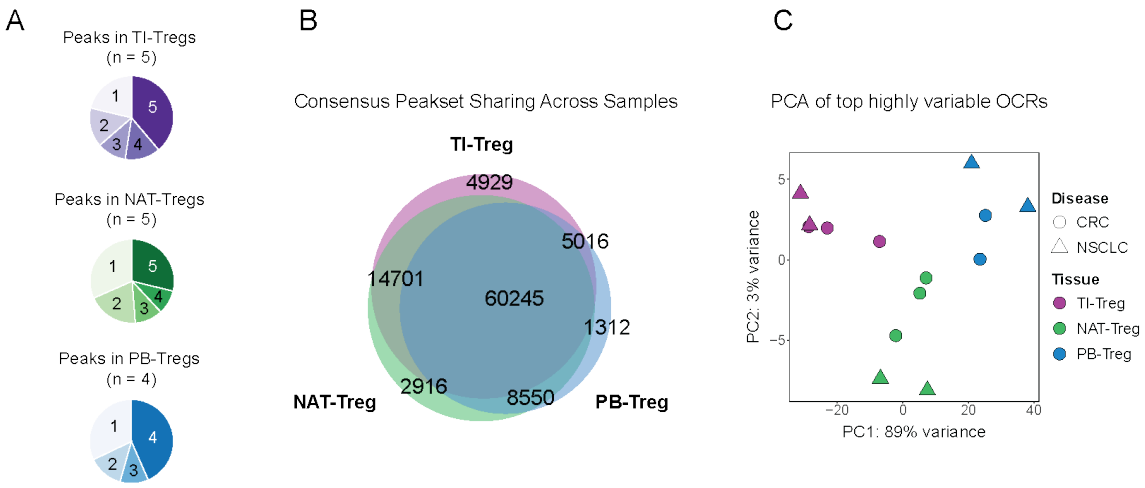

Figure 4 ATAC-seq analysis of Open Chromatin Regions in Tregs from different tissues and disease contexts*. (A) Pie charts showing the number of peaks called in individual samples or shared across increasing numbers of samples (2, 3, 4, or 5 samples) within each condition: tumor-infiltrating Treg (TI-Treg), normal adjacent tissue Treg (NAT-Treg) and peripheral blood Treg (PB-Treg). (B) Venn diagram displaying the overlap of consensus peaks between Tregs from the three different tissues, with the largest overlap observed for 60,245 peaks common to all three Treg conditions. (C) Principal Component Analysis (PCA) of top 500 highly variable OCRs across samples. Each point represents a sample, with TI-Tregs (purple), NAT-Tregs (green), and PB-Tregs (blue) from CRC (circles) and NSCLC (triangles). The separation along PC1 and PC2 illustrates the variance in OCR landscape based on tissue origin and disease context.*

For the genomic annotation of OCRs, we first assigned them univocally to the nearest TSS of genes. Additionally, a public dataset of H3K27ac High-throughput Chromatin Conformation Capture with Immunoprecipitation (HiChIP) on human PB-Tregs, as reported by Mumbach et al.(150), was leveraged to retrieve information on physically interacting putative enhancer-promoter regions. In the case of an overlap between our OCRs and the publicly available genomic regions, priority was given to the annotation based on existing enhancer-promoter looping (**Figure 5A**). For all except two of the 3,916 pairs of interacting regions present in the HiChIP dataset, at least one mate of the pair overlapped an OCR from our dataset, and for 90% of the pairs both regions overlapped with consensus OCRs (data not shown).

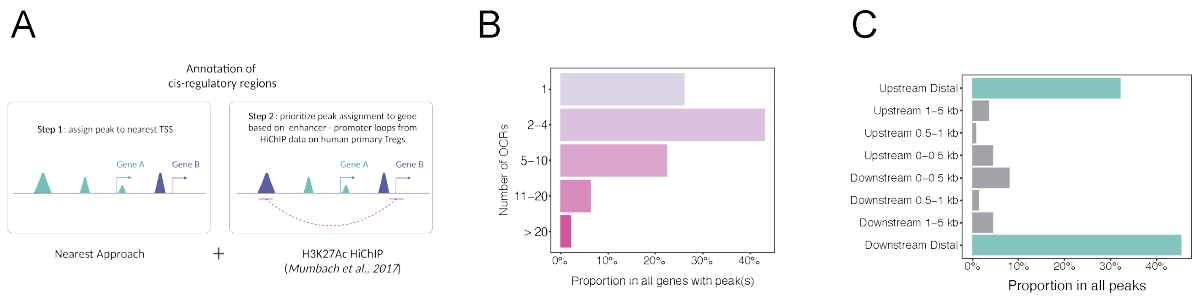

Figure 5. Annotation of Treg OCRs to genes. *(**A**) Schematic view of the strategy used to annotate OCRs to known genes. OCRs are first assigned to the gene with the nearest TSS; the annotation of TSS-distal OCRs is complemented with a public dataset of enhancer-promoter interacting pairs on PB-Tregs based on a H3K27ac HiChIP experiment(150). (**B**) Bar plot displaying the proportion of genes associated to one or to multiple OCRs. (**C**) Bar plot showing the distribution of OCRs across different genomic windows defined based on the distance from the TSS of known genes.*

In total, almost 70% of non-zero expressed genes in Tregs were associated to at least one OCR, and most genes were associated with multiple OCRs (**Figure 5B**).

Upon examining the genomic position of the OCRs in relation to their known associated genes, around 80% of them are found more than 5 kb upstream or downstream from the gene TSS (**Figure 5C**). This suggests the presence of multiple distal regulatory elements within the consensus peak-set, including enhancers. Genomic loci with accessible chromatin peaks are found at the loci of genes encoding for key Treg markers, such as *CTLA4*, *IL2RA*, *ICOS*, *TNFRSF9* (4-1BB), *TNFRSF18* (GITR), *ENTPD1* (CD39), *IL10*, and *TGFB1*, with multiple OCRs present both across their gene body and the surrounding intergenic regions (**Figure 6**).

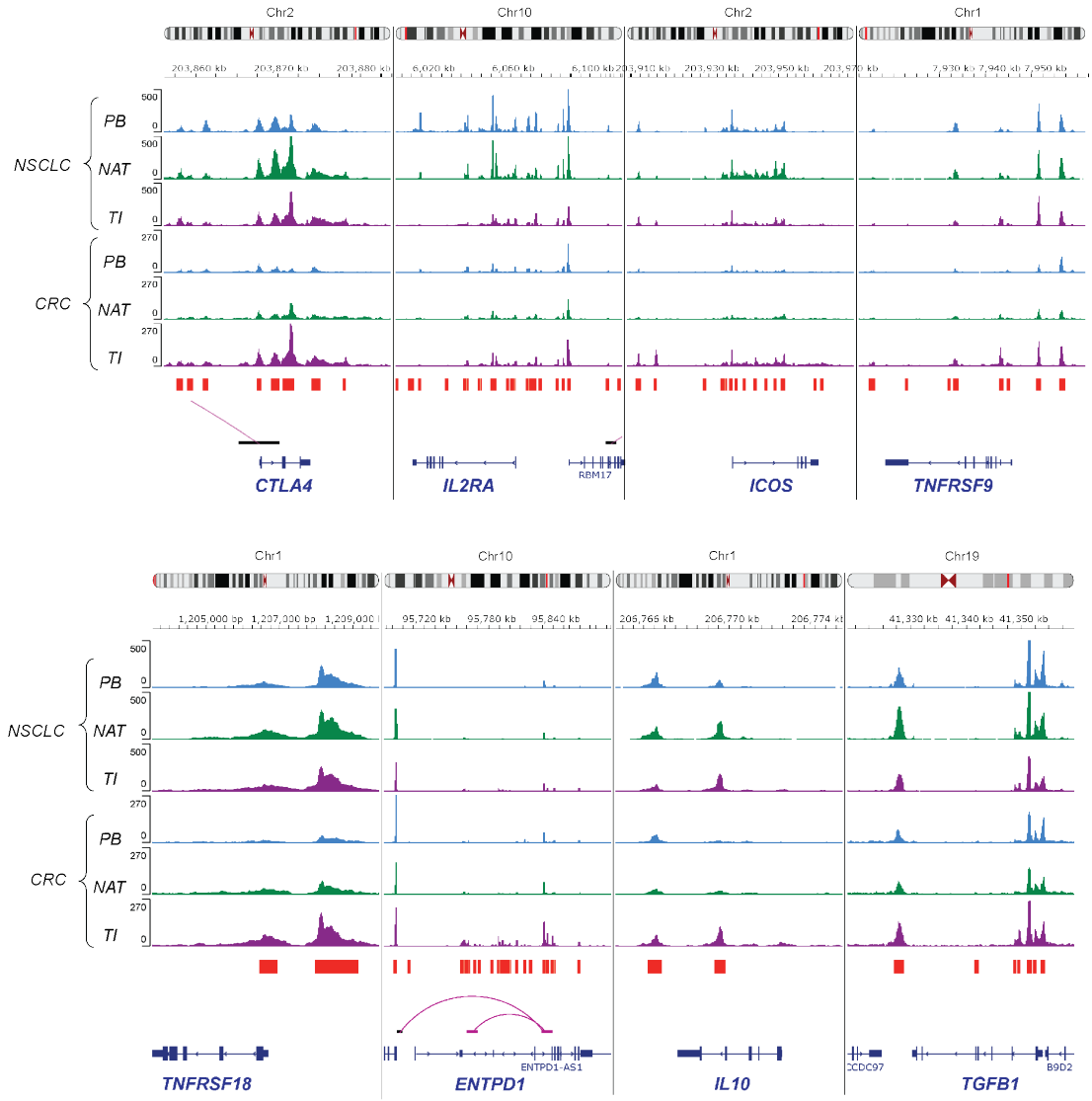

Figure 6 ***Integrated view of different gene loci for Treg marker genes****. From bottom to top: gene tracks from UCSC annotation of RefSeq genes, genomic tracks for interacting region pairs from the Mumbach et al. dataset with arcs representing the existing interaction, red tracks representing consensus OCRs, and representative ATAC-seq signals for matched Treg samples across tissue conditions for a CRC and a NSCLC patient. Purple color represents TI-Treg samples, green color marks NAT-Treg samples and blue color represents PB-Treg samples. The genomic coordinates, and the position of the gene locus, marked with a red bar, with respect to the entire chromosome are reported on top.*

4.1.5 Tumor-infiltrating Tregs from CRC and NSCLC patients display similar chromatin accessibility specificities

To further explore the observed similarities in the epigenomic profiles of TI-Tregs derived from CRC and NSCLC, as revealed by PCA, we a conducted differential accessibility analysis of TI-Tregs with respect to PB-Tregs separately for CRC and NSCLC. Both analyses revealed around 8K differential accessible regions (DARs) at *P* adjusted < 0.05, with similar number of regions with significantly increased or

decreased accessibility in CRC and NSCLC TI-Treg cells. In comparing the log2 fold-change of DARs (**Figure 7A**), we found a good concordance in the gained and lost regions identified for CRC or NSCLC TI-Treg cells. Indeed, the vast majority of DARs identified in one comparison are also significant for the other with the same directionality, whilst the remaining non-significant OCRs still exhibit the same fold change sign in the two comparisons. Overall, the number of DARs with opposite direction between the two comparisons are very few, suggesting similar changes in chromatin accessibility between CRC and NSCLC Tregs compared to PB-Tregs. Notably, contrasting TI-Tregs against NAT-Tregs separately from colon and lung gave similar results, with less shared but overall concordant DARs, despite intrinsic differences in the CRC and NSCLC NAT-Tregs (**Figure 7B**).

In addition, the interaction between disease and tissue type was evaluated in the comparison of TI-Tregs to PB-Tregs. Contrasting the interaction coefficients revealed 21 peaks with a significant increase and 30 peaks with a significant decrease in chromatin accessibility (adjusted p-value < 0.05). Given the relatively low number of peaks, these findings indicate that the additional influence of disease on the effect of tissue origin on Tregs is relatively limited. This suggests that while some peaks do show differential accessibility, the overall impact of tissue context on Tregs is not significantly altered by disease.

Overall, these findings suggest that the chromatin accessibility landscapes of TI-Treg cells derived from CRC and NSCLC undergo similar alterations with respect to circulating Tregs. In light of this, TI-Treg cells isolated from CRC and NSCLC were considered together in downstream analyses.

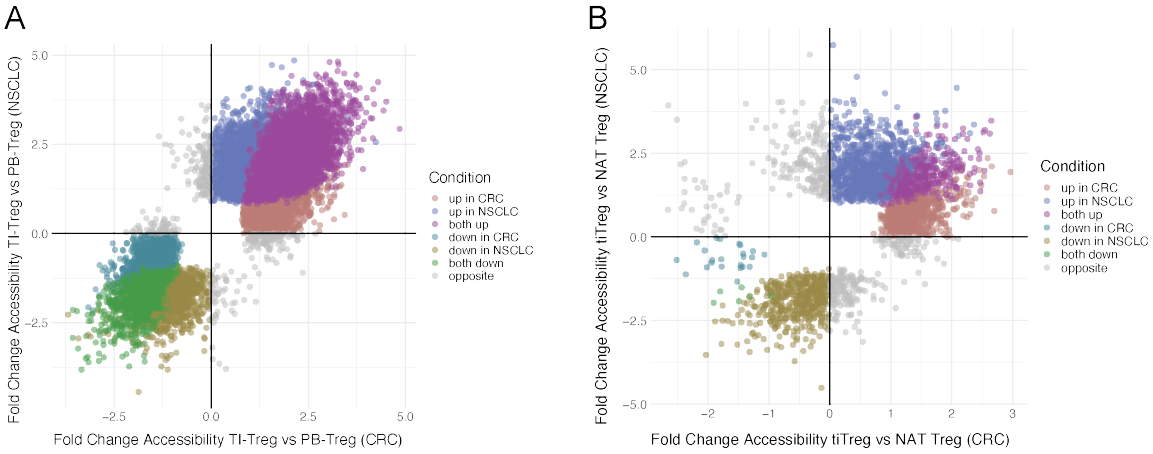

Figure 7. Similar chromatin changes in TI-Tregs derived from CRC and NSCLC patients compared to non-tumor Tregs*. Scatter plots showing the Log2 Fold Change of chromatin accessibility between (**A**) TI-Tregs and PB-Tregs, and (**B**) TI-Tregs and NAT-Tregs, across CRC (x-axis) and NSCLC (y-axis). Dots are colored based on*

significance of accessibility differences in CRC and NSCLC, as well as the sign of Log2 Fold Change. Grey dots indicate peaks with opposite Log2 Fold Change between the two diseases. P-value threshold for differentially accessible regions (DARs) is set at 0.05.

4.1.6 Assessment of genome-wide histone mark profiles in human T cell subsets

After having defined the accessible regions in different Treg states, we wanted to map genome-wide the location of active enhancers in both resting and activated Tregs, with the aim of defining the human Treg enhancerome and TI-Treg specific regulatory networks. A hallmark of active enhancers is the presence of specific histone tail modifications, in particular the acetylation on H3K27 (H3K27ac) and the mono-methylation on H3K4 (H3K4me1), which distinguishes them from active promoters, marked instead by tri-methylation of the same lysine. Therefore, we decided to explore this epigenomic layer to identify *bona fide* enhancer regions. To overcome the technical challenges posed by the systematic epigenomic characterization of primary cells, we employed a modified protocol of chromatin immunoprecipitation with sequencing optimized for low input DNA, called ChIPmentation-seq. We used this assay to generate genome-wide maps of histone modifications defining active regions, marked by H3K4me3, H3K27ac, and H3K4me1, transcribed regions, marked by H3K36me3, and repressed regions, marked by H3K27me3. Primary T lymphocytes were isolated from 10 NSCLC patients, and to obtain samples with enough number of cells for an immunoprecipitation experiment, 5 and 2 of them were pooled in two separate samples, thus obtaining in total 5 patient-derived samples. In addition, the peripheral blood of 6 healthy donors is also used to isolate circulating T cells. Our complete ChIPmentation dataset consisted of 118 sequenced libraries of four different populations of immune cells: $CD4^+$ Treg, $CD4^+$ Conventional ($CD4^+$ Tconv), $CD8^+$ Memory ($CD8^+$ Tmem) and $CD8^+$ Exhausted T cells ($CD8^+$ Tex).

4.1.7 Quality control of ChIPmentation-seq libraries

The alignment statistics for read mapping to the reference genome allowed to confirm the overall good quality of the ChIPmentation libraries. Indeed, almost all samples had between 10M and 20M of mapped reads. The bar plot in **Figure 8** shows the quality control metrics for Treg libraries.

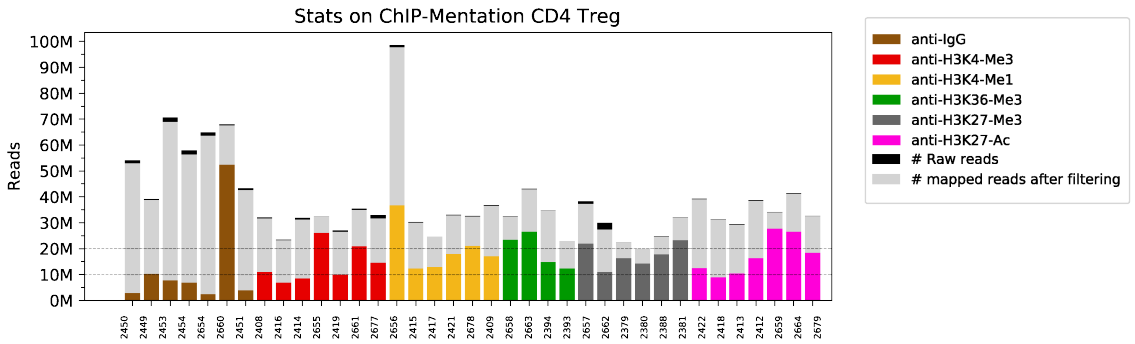

Figure 8. Summarized report of Treg ChIPmentation-seq data quality control*. For each Treg sample the number of reads is represented in a bar plot. The black portion of each bar represents the fraction of unmapped reads, the grey one the fraction of low- quality reads (i.e., not uniquely mapped, duplicate reads and reads originated from mitochondrial DNA or repetitive blacklisted regions) that is filtered out in the post- alignment processing. The colored bars represent the final pool of reads that passed all the quality controls. Red color labels H3K4me3, yellow color H3K4me1, green color H3K36me3, gray color H3K27me3, pink color H3K27ac, and brown color marks the immunoprecipitation positive control IgG.*

The global distribution of histone marks obtained with ChIPmentation is consistent with their expected localization in relation to the TSS and Transcription End Sites (TES) of genes and their gene body (**Figure 9A**). Indeed, a sharp enrichment at the TSS was observed for H3K4me3, as well as for H3K27ac, which in addition had higher enrichment outside of genes. H3K4me1 had a broader enrichment around the TSS with a depletion of signal at the TSS. Along the gene body we found the strongest enrichment for H3K36me3, whereas H3K27me3 signal is mostly enriched just before the TSS. As a control, IgG ChIP samples were included to account for non-specific background signal. As expected, the IgG tracks displayed minimal enrichment across the genome, confirming the specificity of the histone modification signals observed. This lack of notable enrichment is evident in **Figure 9B**, further validating the robustness of the ChIPmentation data and the reliability of the histone mark profiles. Indeed, integrated visualization of the histone modification tracks along known Treg genes revealed an epigenomic landscape consistent with the expected functional state of the Treg population, as exemplified with a genomic view of the *CTLA4*, *ICOS* and *CD28* genes loci from TI-Treg samples (**Figure 9B**). The libraries with suboptimal alignment metrics were carefully checked for their enrichment profile relative to gene TSS and their genomic tracks inspected, and this procedure resulted in a final sample size of 106 high-quality ChIPmentation-seq libraries.

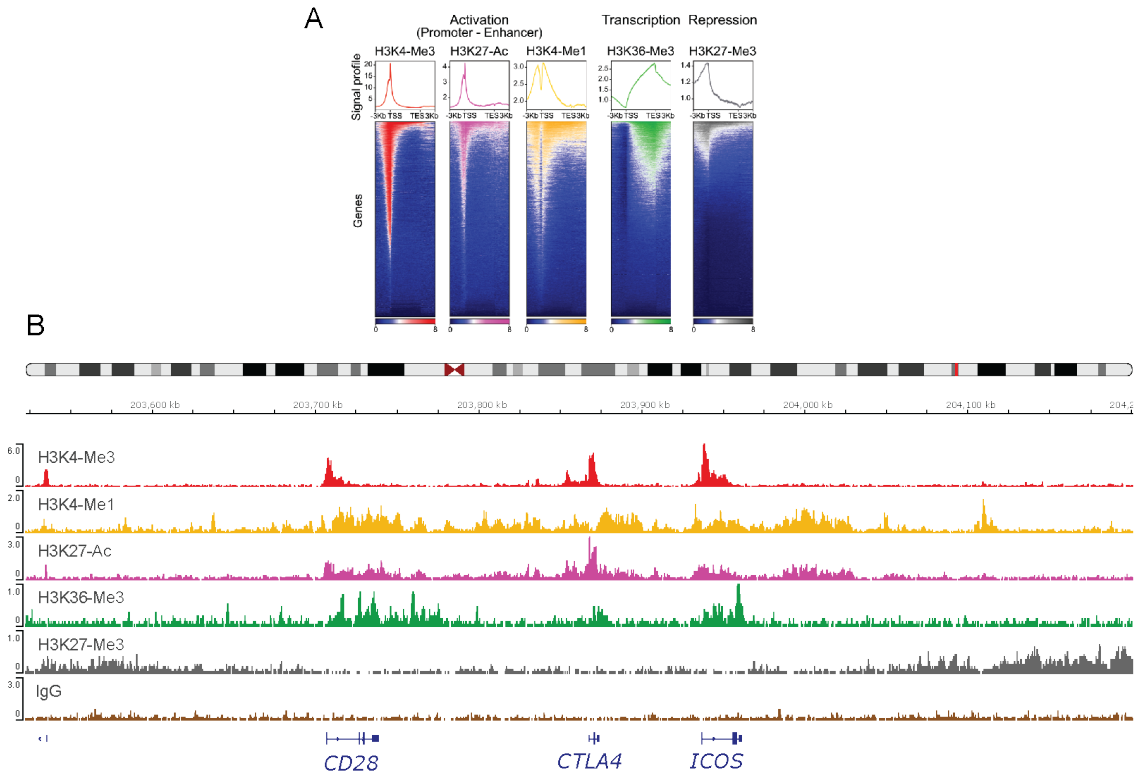

Figure 9 Genome-wide and local distribution of histone modifications. *(**A**) Representative density plots (top) of average intensity and corresponding heatmaps (bottom) displaying the relative distribution of H3K4me3, H3K27ac, H3K4me1, H3K36me3, and H3K27me3 signals at regions surrounding +/- 3kb of the gene body for all the genes present in GENCODEv25. (**B**) Representative genomic view showing the distribution of histone modification signals across a selected genomic region on chromosome number 2. In both panels, red color labels H3K4me3, yellow color H3K4me1, pink color H3K27ac, green color H3K36me3, gray color H3K27me3, and brown color marks the positive control IgG.*

4.1.8 De novo discovery of Chromatin States with ChromHMM

To capture the epigenomic landscape of TI-Tregs in a systematic manner rather than based on individual epigenomic features considered separately, we exploited a machine learning-based tool called ChromHMM(151) for *de novo* chromatin state reconstruction using the complete ChIPmentation dataset. To strengthen our discovery, we built the chromatin state model including not only Treg samples, but also ChIPmentation data for the other $CD4^+$ and $CD8^+$ T cells isolated from tumor tissues of NSCLC patients and PB of healthy donors.

Using ChromHMM, we explored the combinatorial patterns of the five histone marks in a 14-state model and predicted specific genomic features with high resolution and robustness across samples. The heatmap in **Figure 10A** shows the frequency in which different histone modifications are co-present in the same genomic region (i.e., histone

marks emission probability). Each chromatin state, based on the combinatory enrichment or depletion of the different histone marks, as well as on the enrichment of known genomic features (as shown in **Figure 10A**, left part), was assigned an annotation term in accordance with the Roadmap Epigenomics Consortium guidelines. We identified three chromatin states enriched in active histone marks (Active, Flanking, and Flanking Downstream TSS) displaying strong enrichment for H3K4me3, and either H3K27ac or H3Kme1. The state with an enrichment of H3K4me1 and H3K27ac was annotated as “Active Enhancer”, whilst the additional presence of H3K36me3 over the gene body defined “Genic Enhancers”. As “Weak Enhancer”, we defined the state characterized by the presence of H3K4me1 alone. The two transcription and two repression states were characterized by the presence of H3K36me3 and H3K27me3, respectively, with different signal intensities. The state with enrichment for both the active mark H3K4me3 and the repressed mark H3K27me3 was defined as “Bivalent”. “Quiescent” states marked regions without any significant enrichment of histone marks. To confirm the robustness of our results, we verified that the proportion of each chromatin state was comparable across all T lymphocyte samples (**Figure 10B**).

Then, using the ChromHMM-defined “Active” and “Genic” enhancer states, we identified a total of 13,050 regions with enhancer features, observed in at least two PB- and/or TI-Tregs, with a high enrichment for H3K27ac signal.

Overall, we exploited a machine learning-based approach to perform a systematic reconstruction of chromatin states in Treg cell populations. In doing so, we created a reference resource of active regulatory elements (including promoters and enhancers), as well as elongated, repressed, and quiescent regions. All in all, the localization of chromatin regions with features of active enhancers from human Treg samples provides leads on the transcriptional regulation of TI-Tregs.

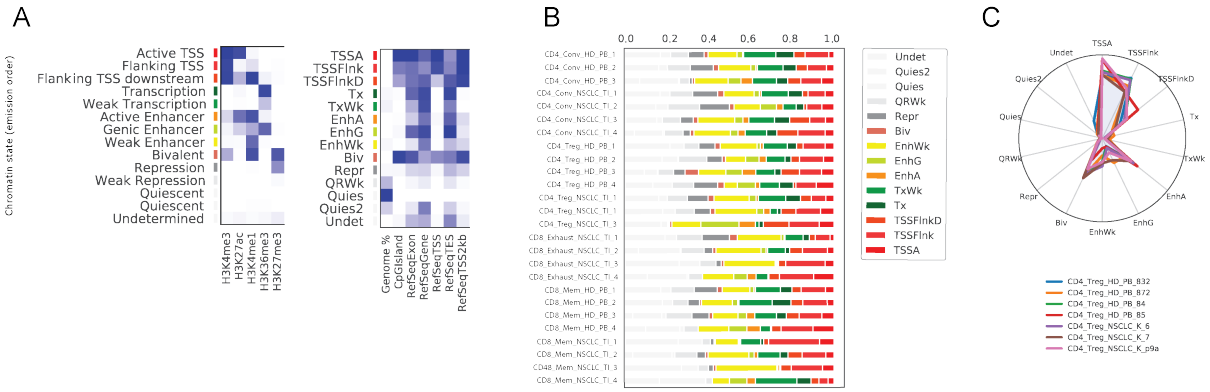

Figure 10. De novo reconstruction of chromatin states of tumor-infiltrating and peripheral blood T lymphocytes and integration with open chromatin landscape. *(**A**) Combinatorial pattern of histone marks in a 14-state model using ChromHMM. The heatmaps show the frequency of the histone modifications found in each state (Emission plot, left) and the enrichment of each chromatin state for specific types of functional elements, such TSS, Transcription End Site (TES), exons, gene bodies (right, representative heatmap for a TI-Treg replicate). A darker blue color corresponds to a greater fold enrichment. TSSA, Active TSS; TSSFlnk, Flanking TSS; TSSFlnkD, Flanking Downstream TSS; Tx, Transcription; TxWk, Weak transcription; EnhA, Active enhancer; EnhG, Genic enhancer; EnhWk, Weak enhancer; Biv, Bivalent; Repr, Repression; ReprWk, Weak repression; Quies, Quiescent; Undet, Undetermined. (**B**) Genome-wide distribution of the 14 chromatin states annotated in each T lymphocyte sample. (**C**) Spider plot depicting the enrichment of OCRs within the chromatin states annotated in TI- and PB-Treg samples. The stroke color of each polygon represents a Treg sample.*

We then compared the ChromHMM states with the chromatin accessibility data generated for PB-, NAT- and TI-Treg cells. As expected, for the ChromHMM-based genome-wide annotations of all the Treg samples, the accessible chromatin was enriched in active promoter and active enhancer states, as well as in weak enhancers and bivalent TSS states, whereas transcribed and repressive states were depleted for accessible chromatin (**Figure 10C**).

These results suggest that the ChromHMM-defined chromatin states of Treg cells constitute a robust atlas of genome-wide regulatory elements that enables a more accurate genomic annotation of ATAC-seq-defined OCRs.

4.1.9 Identification of bidirectionally transcribed enhancer RNAs in human Tregs

Enhancers identified with ChromHMM are generally longer than average ATAC-seq peaks (around 700 bp). It is known that active enhancers can be transcribed into short non-coding RNAs (eRNAs), which often correlate with enhancer activity. Notably, the portion of enhancer transcribed into eRNA is likely to harbour the DNA-bound TFs and thus may contain informative TF binding sites (TFBS). Therefore, we decided to identify the precise location of bidirectionally transcribed eRNAs for two reasons: first, to describe an additional epigenomic layer which can further strengthen the

characterization of active enhancers in Tregs; second, we aimed to potentially refine our genomic map of active enhancers by leveraging the TSS of the eRNAs, to define *core* enhancer regions, which can be used for downstream analyses. To identify eRNA transcripts in Tregs, we took advantage of a newly published pipeline named ReapTEC (standing for Read-level prefiltering and Transcribed Enhancer Call) for the precise mapping of enhancer bidirectional transcription start sites using 5' scRNA-seq data(144). To this purpose, we leveraged our in-house dataset of sorted $CD3^+$ and $CD4^+$ T lymphocytes deriving from tumor, NAT and PB samples of 1 CRC and 3 NSCLC patients. Following library integration and harmonization across patients, quality checks and contaminant cell removal, we performed unsupervised Leiden clustering identifying 14 clusters. We recognized Treg cell population based on the expression of canonical Treg marker genes, including FOXP3, IL2RA and CTLA4 (**Figure 11A-C**). We also confirmed the higher expression of TI-Treg signature genes in the Treg cluster compared to the other T lymphocyte clusters (**Figure 11D**).

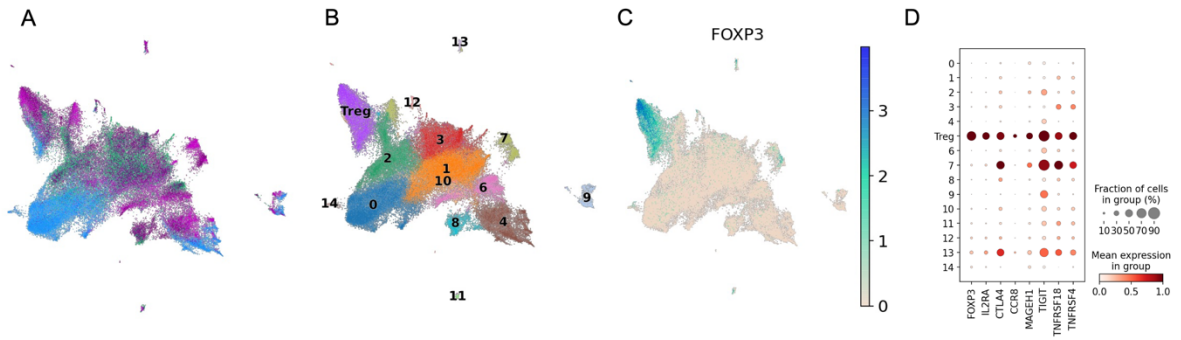

Figure 11. Identification of FOXP3+ Tregs in the single-cell RNA-seq dataset of sorted CD3+/CD4+ T lymphocytes from cancer patients*. Uniform Manifold Approximation and Projection (UMAP) visualization of the integrated scRNA-seq dataset from 1 CRC and 3 NSCLC patients, following harmonization and batch correction. Labels indicate: (**A**) tissue origin of T lymphocytes, from blood (blue), normal tissue adjacent to tumor (green), and tumor tissue (purple); (**B**) 14 clusters identified using the Leiden algorithm, including the Treg cluster identified using canonical markers; (**C**) expression levels of the FOXP3 gene, used to define Tregs. (**D**) Dot plot showing the fraction of cells expressing Treg marker genes and scaled mean expression levels across all identified clusters.*

By applying ReapTEC on the single Treg cells from the integrated dataset (**Figure 12A**), we identified in total 7,379 eRNAs. The average length of enhancer RNA based on identified bidirectional TSS was around 250 bp, which is shorter than the average ATAC-seq peak length.

As a general trend for eRNAs, we were able to observe that the portion of the enhancer defined by the bidirectional TSS is highly overlapping with the summit of the ATAC-seq peak and depleted of the H3K27ac histone mark, which is instead enriched immediately upstream and downstream of the transcribed portion. As an example, we

report the integrated view of an eRNA found within an intronic region of the *CTLA4* gene, with the chromatin accessibility signal and the enriched histone marks for genic enhancer chromatin states (**Figure 12B**). As expected, enhancers with bidirectional transcription displayed significantly higher H3K27ac signal, indicating that such enhancers are stronger activated than non-transcribed enhancers (**Figure 12C**).

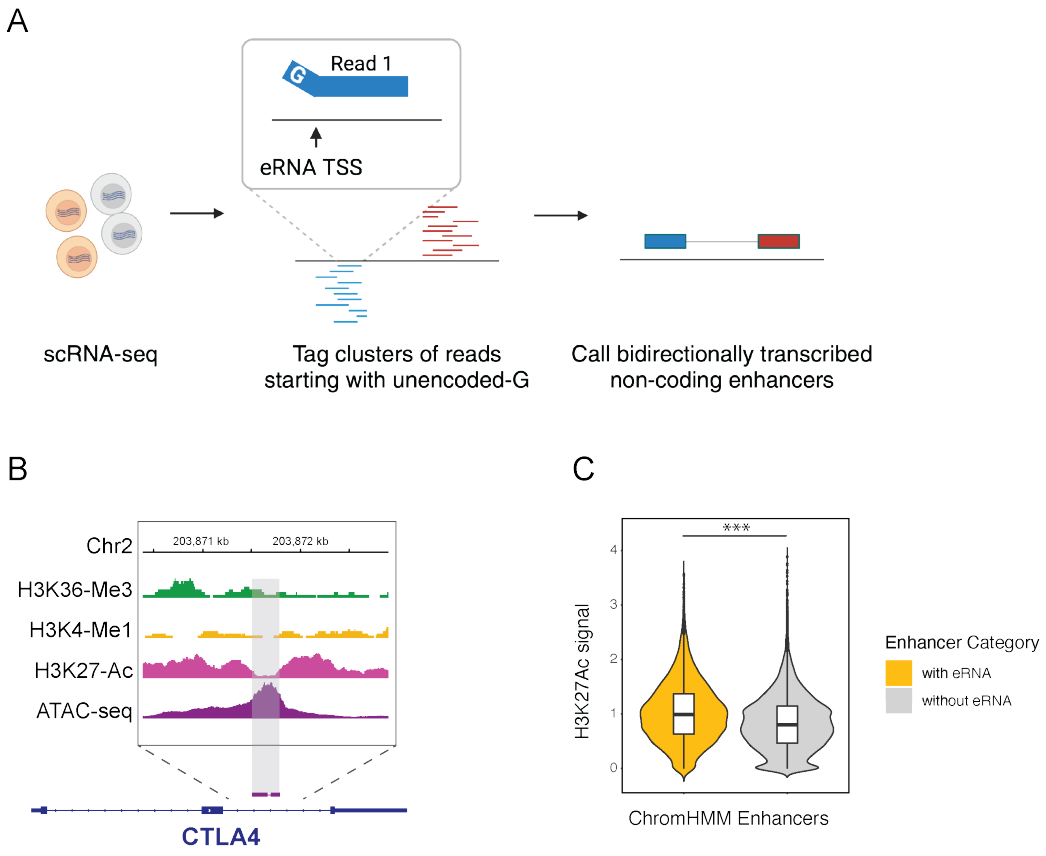

Figure 12. Identification of bidirectionally transcribed enhancers from human single-cell Treg transcriptomes. *(**A**) Schematic overview of the ReapTEC pipeline, outlining key steps for the precise mapping of enhancer RNA (eRNA) TSS from single-cell transcriptomes, enabling the identification of non-coding enhancer transcripts. (**B**) Genomic visualization of ATAC-seq signal (purple), alongside H3K27ac (pink), H3K4me1 (yellow), and H3K36me3 (green) histone mark signals, showing an eRNA identified in Treg cells within the CTLA4 gene locus. (**C**) Violin plot illustrating the distribution of H3K27ac signal, a marker of enhancer activation, across ChromHMM-annotated enhancer regions, comparing those with or without eRNA intersections. Statistical comparison was performed using a Mann-Whitney test, *** p < 0.001.*

The genomic views in **Figure 13** display additional examples of eRNA transcripts identified for key Treg genes, such as *IKZF2*, *TNFRSF18* (GITR) and *TIGIT*. Overall, we identified a consensus of bidirectionally transcribed eRNAs from single cell transcriptomes of different Treg cell states derived from cancer patients.

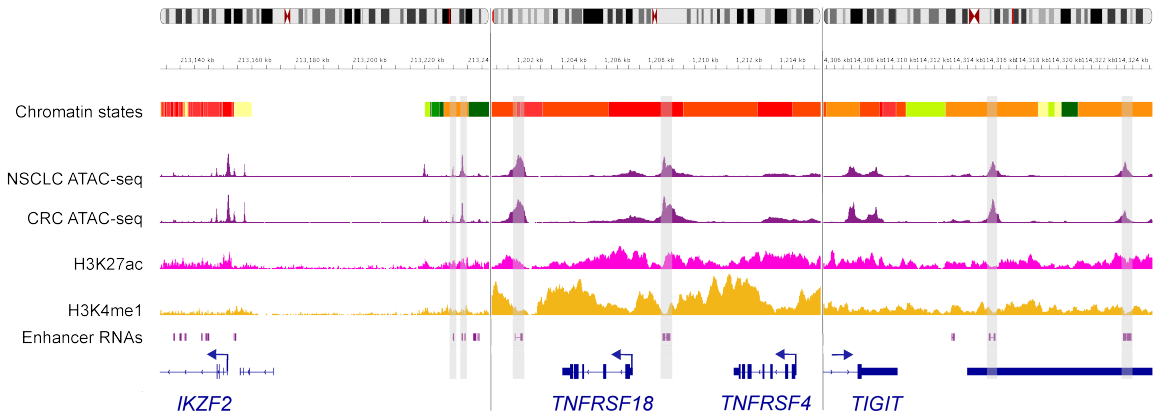

Figure 13. Integrated genomic view at Treg gene loci with transcribed enhancers. *From bottom to top: gene tracks from UCSC annotation of RefSeq genes, with arrows indicating the directon of gene transcription; eRNAs identified with the ReapTEC pipeline represented in purple; yellow and pink tracks of H3K4me1 and H3K27ac signals for representative TI-Treg ChIPmentation samples; purple tracks of ATAC-seq signal for two representative TI-Treg samples from CRC and NSCLC patients, respectively; representative chromatin state annotation track for one NSCLC TI-Treg sample. Sites where eRNAs are identified are marked with grey bars.*

4.1.10 Integration of eRNAs with the other epigenomic layers

Integration of the transcribed enhancers with chromatin accessibility showed that around 75% of all the eRNAs identified overlap with an OCR, and 67% also overlap with an active chromatin state (**Figure 14A**). As with OCRs (**Figure 10C**), we evaluated the enrichment of eRNAs in the different ChromHMM-based annotations for the Treg samples. As expected, eRNAs are mainly enriched in active chromatin states (**Figure 14B**).

Taken together, these results indicate high consistency between the three epigenomic layers of chromatin accessibility, enhancer chromatin states and enhancer transcription, although they were captured by different omics technologies performed on independent Treg samples deriving from different cancer patients.

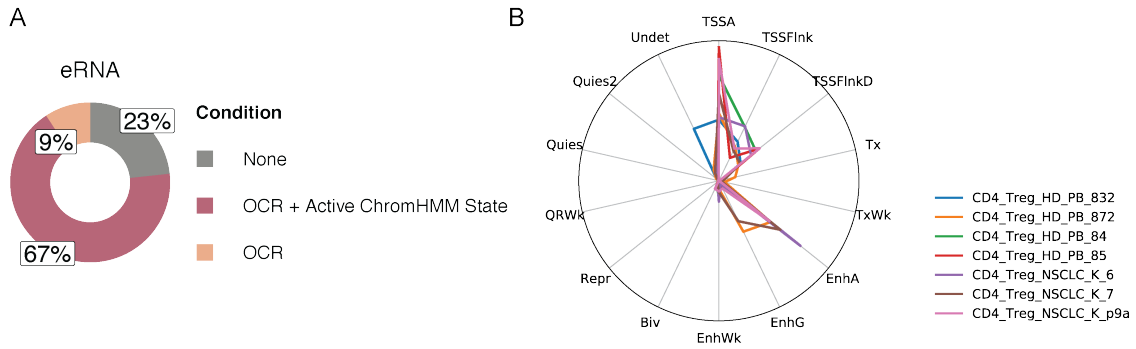

Figure 14. Crossing Chromatin Accessibility and Chromatin States with enhancer RNAs *(**A**) Donut plot displaying the overlap between eRNAs and other epigenomic features: 67% of the eRNAs identified intersect both Treg OCR and one of the "active" ChromHMM-derived chromatin states (i.e., Active Enhancers, Genic Enhancers, active TSS, Flanking TSS, Downstream Flanking TSS); 9% intersect only with OCRs, and 23% do not intersect with OCRs. (**B**) Spider plot showing the enrichment of eRNAs within the chromatin states annotated in TI- and PB-*

Treg samples. The stroke color of each polygon represents a Treg sample. The stronger enrichment is observed for Active and Genic Enhancer states and active TSS-related chromatin states.

Finally, we defined a consensus of patient-derived Treg enhancerome, as a set of accessible chromatin regions distal from the TSS, with evidence either of an active enhancer chromatin state or of bidirectional transcription, or both. The circular plot in **Figure 15A** represents the genome-wide integration between the three omics for the identification of *bona fide* Treg enhancers. In total, we annotated 14,158 OCRs as enhancers and 15,732 OCRs as promoters. Notably, enhancer regions constitute around 15% of the total number of OCRs indentified in patient-derived Tregs (**Figure 15B**), underlying the importance of investigating more precisely the identity of accessible chromatin regions for the identification of enhancers and not relying solely on the identification of open chromatin.

In conclusion, we mapped in a comprehensive manner and using multi-omics integration the active enhancers in human Treg cells from the blood, the tumor tissue of CRC and NSCLC patients and from their normal tissue adjacent to tumor.

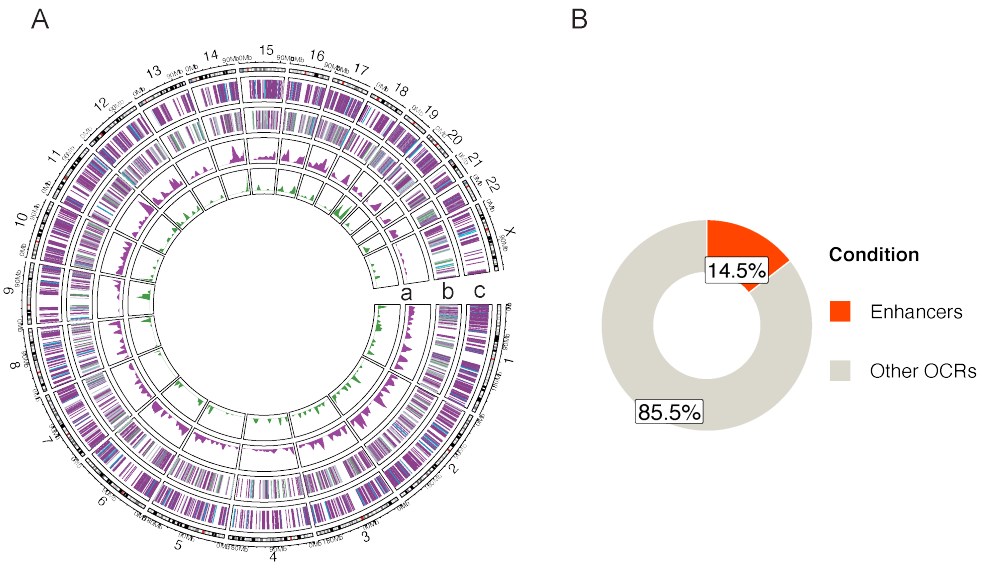

Figure 15. Multi omics integration of different epigenomic features allows to map of bona fide enhancers in the human Treg chromatin landscape*. (**A**) Circos plot depicting the cancer patient-derived Treg enhancerome identified through integration of a) chromatin accessibility measured with ATAC-seq b) eRNAs identified from single Treg cell transcriptomes, and c) ChromHMM-annotated active enhancers identified from combination of multiple histone modification profiles. For ATAC-seq the signal in DARs between TI-Treg and NAT-Treg (see Results section 4.2) are reported for two representative NAT-Treg (green color) and TI-Treg (purple color) samples. (**B**) Donut plot showing the fraction of Treg OCRs annotated as enhancers following the described integration approach.*

4.2 Enhancer accessibility changes reflect transcriptional differences between TI-Tregs and NAT-Tregs

After qualitatively assessing the enhancerome of Tregs from different tissue origins, a central question arises: can we identify differences in the open chromatin landscape between TI-Tregs and normal tissue-derived Tregs, particularly at the enhancer level? Indeed, changes in the enhancer epigenomic profile may drive distinct regulatory programs in Tregs. To explore this, we utilized ATAC-seq as a quantitative method to measure chromatin accessibility, serving as a proxy for enhancer activity. At the same time, promoter regions, identified as OCRs proximal to the TSS of known genes considering fixed genomic windows of ± 500 bp, were also included in the analysis to address potential different patterns of chromatin rewiring at the level of enhancers and promoters.

Pairwise comparison of ATAC-seq counts between the three cell states for each patient revealed statistically significant differences in OCRs. We focused on differentially accessible regions (DAR) annotated as enhancers and promoters for the three analyzed contrasts (i.e., TI-Tregs *versus* PB-Tregs, NAT-Tregs *versus* PB-Tregs, and TI-Tregs *versus* NAT-Tregs). The total number of DAR enhancer and promoters differs across the comparisons and, as expected, there are more changes between TI-Tregs and PB-Tregs than between TI-Tregs and NAT-Tregs, given that the Treg subsets in the latter comparison are both activated tissue-Tregs (**Figure 16A**). We identified in total 7,165 DARs including enhancers and promoters, according to our annotation, comparing across all the three Treg cell states. Focusing on TI-Tregs *versus* NAT-Tregs, the total number of DARs across all OCRs is 4,418, composed by 3,589 more accessible and 829 less accessible regions in TI-Tregs. Focusing *cis*-regulatory elements, 1,275 enhancers and 114 promoters show increased accessibility in TI-Treg, whereas 211 enhancers and 97 promoters have decreased accessibility, respectively (*P*-value adjusted < 0.05). The volcano plot in **Figure 16B** focuses on the DARs identified as enhancers or promoters. The analysis identified a higher percentage of regions with increased accessibility in TI-Tregs than regions with lower accessibility.

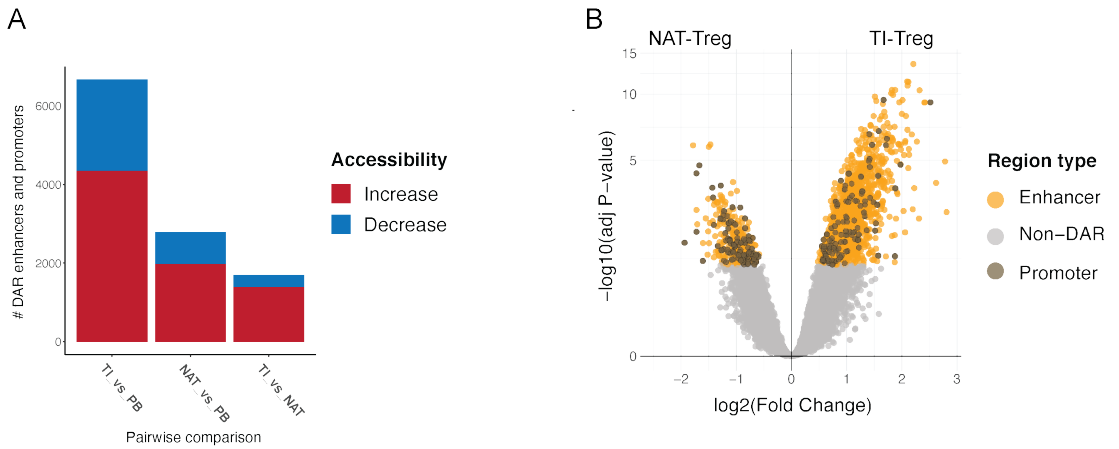

Figure 16. Cis-regulatory elements with differential accessibility across Treg cell states. *(**A**) Bar plot showing the number of OCRs annotated as promoters and enhancers with statistically significant changes in chromatin accessibility for the following pairwise comparisons: TI- versus PB-Tregs, NAT- versus PB-Tregs, and TI- versus NAT-Tregs. Promoters are defined as OCRs within ±500 bp of the TSS of known genes. OCRs with increased accessibility are shown in red, while those with decreased accessibility are in blue. (**B**) Volcano plot illustrating log2 fold changes in chromatin accessibility, in the x-axis, versus the associated -log10(adjusted p-value), in the y-axis, for all cis-regulatory elements in the paired comparison between 5 TI-Treg and 5 NAT-Treg samples. DAR enhancers and promoters are marked in orange and ochre, respectively. Differential accessibility analysis was performed using DESeq2 with the Wald test, and p-values were adjusted for multiple testing using the Benjamini-Hochberg method, with significance set at an adjusted p-value < 0.05.*

4.2.1 Defining transcriptional changes between TI-Tregs and NAT-Tregs

At this point, we wanted to exploit the bulk RNA-seq dataset published in our previous study(101) on the TI-Treg transcriptomic profile, to establish relationships between the transcriptomic and the epigenomic landscapes. For this reason, we initially assessed whether the two datasets mirror each other, since the RNA-seq and ATAC-seq samples are not paired.

By calculating an overall accessibility score per gene (see Methods 3.21), a good correlation between this score and the gene expression level is observed (**Figure 17A**). This suggests that gene locus accessibility and level of transcription are overall well integrated, the dynamic ranges of the ATAC-seq and RNA-seq counts are comparable and we can use the two datasets to evaluate consistency between the epigenomic and transcriptomic landscape.

In contrast to the TI-Treg versus PB-Treg comparison, that can primarily identify differences due to the activation and highly suppressive status of Tregs in the tumor compared to their resting state, the comparison with NAT-Tregs may reveal distinct features that are unique to TI-Tregs, shedding light into how they get activated in the TME and adapt to its specific cues. Therefore, we leveraged the bulk RNA-seq dataset to derive a specific TI-Treg signature compared to NAT-Tregs. By comparing 15 TI-

Tregs vs 8 NAT-Tregs from lung and colon, 614 and 566 genes resulted differentially up- and down-regulated, respectively, between the two cell states (adjusted p-value < 0.05) (**Figure 17B**). Therefore, we considered upregulated genes in TI-Tregs vs NAT-Tregs as a highly specific TI-Treg signature. We then compared this signature with the TI-Treg gene signature defined in our previous work(101). This gene set is also highly relevant, as it was derived by comparing TI-Tregs not only with NAT- and PB-Tregs, but also with other immune cell types both in the periphery and within the tumor. Notably, around 50% of the 309 De Simone *et al.* signature genes are also found specifically upregulated in TI-Tregs compared to NAT-Tregs.

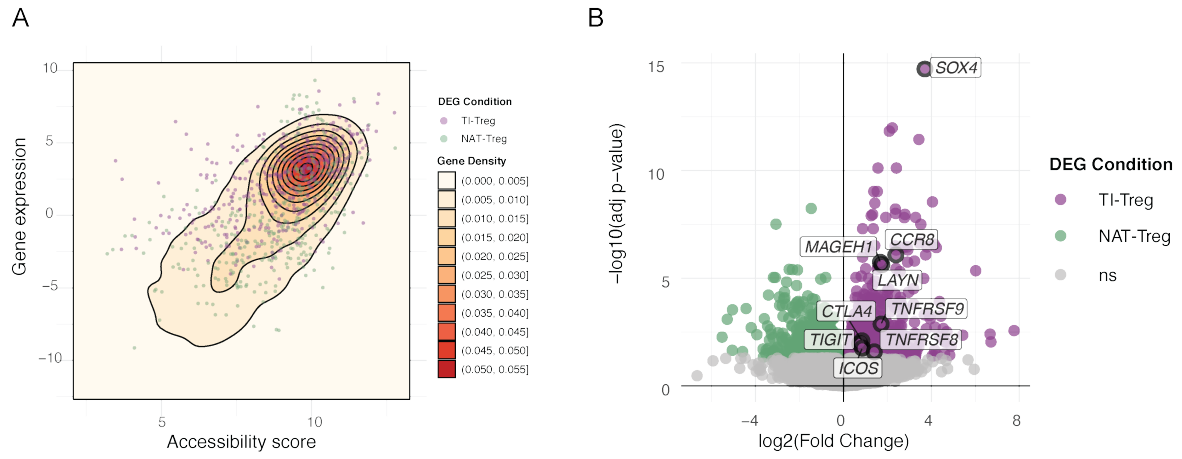

Figure 17. Differential gene expression analysis of TI-Tregs vs NAT-Tregs informed by RNA-seq and ATAC-seq consistency. *(**A**) 2D density plot correlating RNA-seq and ATAC-seq datasets. For each gene, an accessibility score was calculated as the sum of mean normalized counts for all OCRs annotated to that gene in TI-Tregs. The x-axis represents the accessibility score, while the y-axis shows gene expression values as the TI-Treg mean FPKM count per gene. Colors represent different density regions, as indicated in the color legend. Overlayed dots indicate genes with significant differential expression between TI- and NAT-Tregs (see panel **B**). (**B**) Volcano plot illustrating log2 fold changes in gene expression (x-axis) versus the associated -log10(adjusted p-value) (y-axis) for the comparison between 15 TI-Treg and 8 NAT-Treg samples. Purple and green dots represent differentially expressed genes (DEGs) upregulated in TI-Tregs and in NAT-Tregs, respectively. Differential gene expression analysis was performed using DESeq2 with the Wald test, and p-values were adjusted for multiple testing using the Benjamini-Hochberg method, with significance set at an adjusted p-value < 0.05.*

4.2.2 Integrating transcriptional and enhancer chromatin profiles of TI-Tregs and NAT-Tregs

At this point, we sought to determine whether the transcriptional changes identified between TI-Tregs and NAT-Tregs mirror their epigenomic alterations, particularly at the enhancer level. Although the genes associated with DAR enhancers represent only a fraction of the differentially expressed genes (DEGs), this fraction increases with the magnitude of gene expression change. (**Figure 18A**). To better represent the

relationship between enhancer chromatin changes and transcriptional changes at the level of each gene, an enhancer accessibility change score was first calculated for each DEG (**Figure 18B**; see *Methods Section 3.23*). This score was subsequently correlated with the log2FoldChange in expression of the corresponding DEG. The scatter plot, along with the fitted linear model, reveals a positive correlation (**Figure 18C**), suggesting that the transcriptional differences between TI-Tregs and NAT-Tregs reflect, and are likely to be driven by, corresponding changes in chromatin accessibility at enhancer level.

The same correlation was also performed with an accessibility score using only promoter regions, giving a similar result (data not shown). This is however more expected, considering the direct relationship between the gene transcription and the promoter activity. Interestingly, there is also a positive correlation between the increment in transcription level and the number of DAR enhancers of a gene (**Figure 18D**), confirming also the robust identification of enhancers with our approach.

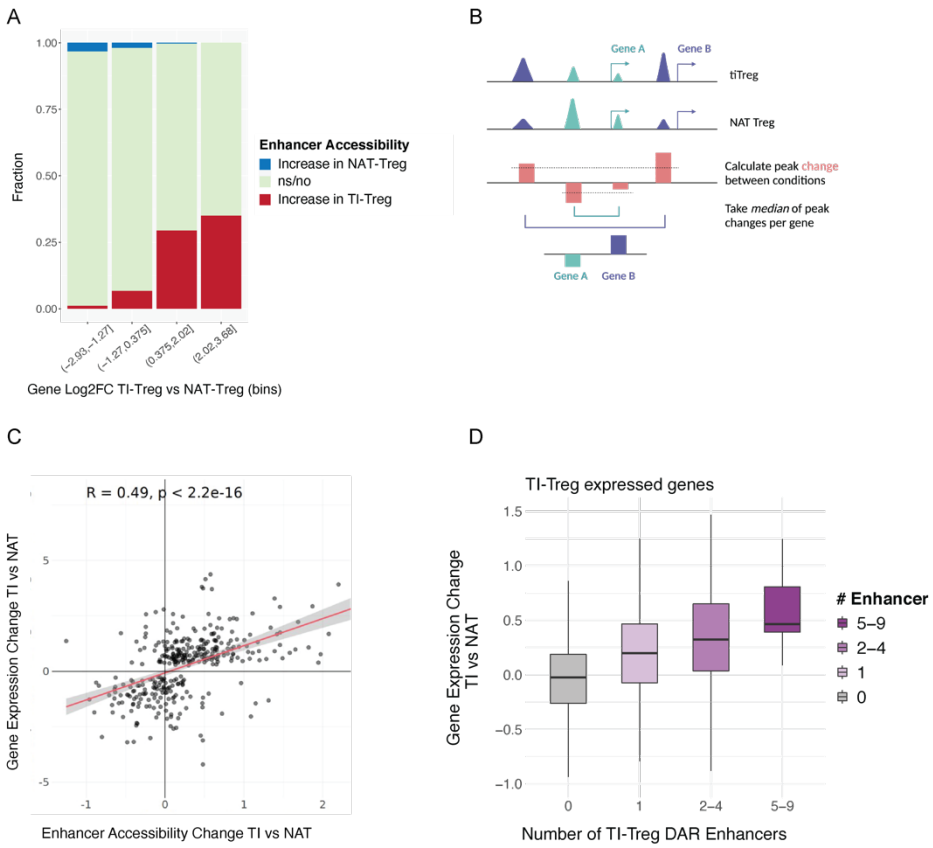

Figure 18. Concordant enhancer accessibility and target gene expression changes between TI-Tregs and NAT-Tregs. *(**A**) Bar plot displaying DEGs divided into four bins based on the log2 fold change between TI-Tregs and NAT-Tregs. The plot illustrates the fraction of genes in three categories: (1) genes with enhancers that show*

*a significant increase in accessibility in TI-Tregs (labeled in red), (2) genes with enhancers that show a significant increase in accessibility in NAT-Tregs, and (3) genes either with enhancers that exhibit non-significant chromatin changes or have no associated enhancers (shown in light green). (**B**) Schematic illustrating the method used to calculate an accessibility change score for each gene, comparing TI-Tregs to NAT-Tregs, based on the median log2 fold change in enhancer accessibility. (**C**) Scatter plot demonstrating a significant correlation between gene expression changes and enhancer accessibility changes for DEGs between TI-Tregs and NAT-Tregs (linear regression model: R = 0.49, p < 0.0001). (**D**) Box plot depicting the correlation between the number of DAR enhancers in TI-Tregs and increment of target gene expression in TI-Tregs. Genes are categorized into four bins based on the number of DAR enhancers associated.*

Focusing on enhancers, we highlight DEGs with the highest enhancer accessibility change. Upregulated genes with increased enhancer activity in TI-Tregs include genes such as *IL21R*, *LAYN,* and *ICOS*, whereas downregulated genes with decreased enhancer activity included *HIVEP2*, *TGFBR2* and *CD6* (**Figure 19**).

With these results, we showed that the transcriptional reprogramming of TI-Tregs reflects the acquisition of a distinct epigenomic landscape, that is represented by crucial chromatin accessibility alterations at enhancer level, and that the differential activity of enhancer regions establishes a regulatory network that distinguishes TI-Tregs from NAT-Tregs.

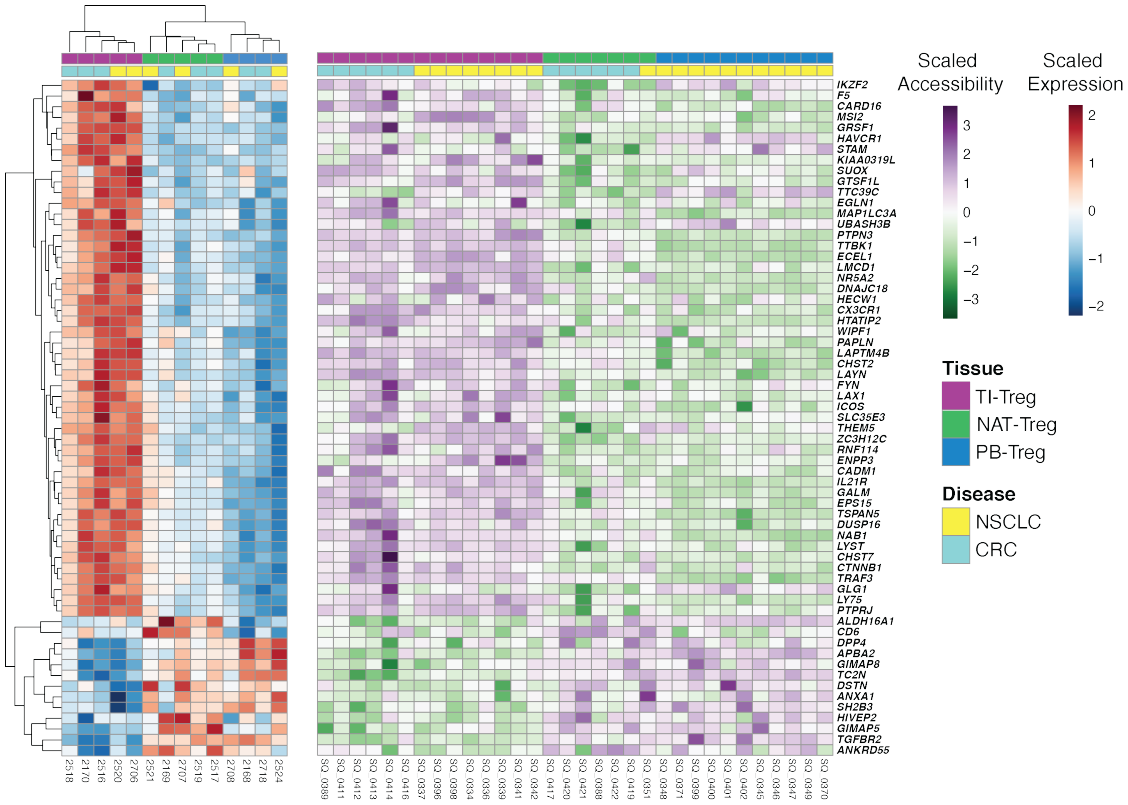

Figure 19. ***Enhancers and target genes with concordant epigenomic and transcriptomic changes across Treg cell states.*** *The heatmaps displays the scaled accessibility levels of DAR enhancers (left) associated with DEGs (right) in TI-Tregs compared to NAT-Tregs. Red shades represent higher accessibility values, blue colors represent lower accessibility. For genes with multiple DAR enhancers, only the enhancer with the lowest p-value for differential accessibility is shown. For DAR enhancers more accessible in TI-Tregs, the top 50 enhancers with the lowest p-values are reported. Hierarchical clustering of enhancers and Treg samples is performed using Euclidean distances. The right heatmap illustrates the scaled expression of target genes across Treg conditions. Purple colors represent higher expression values, green colors represent lower expression values.*